# The Markovian Binary Tree: A Model of the Macroevolutionary Process

Nectarios Kontoleon

*Thesis submitted for the degree of*

*Doctor of Philosophy*

*in*

*Applied Mathematics*

*at*

*The University of Adelaide*

*(Faculty of Engineering, Computer and Mathematical Sciences)*

Discipline of Applied Mathematics

School of Mathematical Sciences


March 14, 2006

# Contents

# List of Figures

# Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

DATE: ....27/3/06.....

# Acknowledgements

I would like to thank the Federal Government of Australia for its financial support through the Australian Postgraduate Award. Without this support I would not have been able to undertake a PhD. I would also like to thank my two supervisors, Professors Nigel Bean and Peter Taylor. They have both been exceptional supervisors always eager to help and always extremely supportive throughout my entire PhD. I do not believe a student can ask for better supervisors. I would finally like to thank my family who were also extremely supportive of my efforts during this time. Thank you.

# Dedication

This thesis is dedicated to my Dad. To teliosa Stelara mou! Thoxa to Theo!

# Abstract

One of the fundamental problems in biology is concerned with deciphering and understanding the nature of evolution. The results of evolution can be seen through the diversity of life found on earth today. The relationships between species can be ascertained using a variety of biological and statistical techniques. These relationships can be pictorially represented on a tree diagram called a phylogenetic tree. It has been found that many phylogenetic trees are imbalanced, meaning that the subtrees of phylogenetic trees differ in shape. The focus of this thesis is to develop physically-reasonable mathematically-tractable models of the speciation process. We do not wish to model the evolutionary process at the genetic level but rather, to model the process at the species level as represented by the branching structures of phylogenetic trees.

The simplest models of macroevolution generate branching structures that are either too balanced or too imbalanced. Therefore it has become increasingly fashionable to model the macroevolutionary process using continuous-time Markovian multitype branching processes (ctMMTBP). Continuous-time MMTBPs provide the flexibility needed to generate tree structures with any level of imbalance. However, the major pitfall of using ctMMTBPs is that they do not have an algorithmic approach for ascertaining measures useful in macroevolution.

The model that is proposed is called the Markovian binary tree and provides an alternative representation of the binary-split ctMMTBP. This representation is made possible by re-interpreting the transition structure of the ctMMTBP. The MBT has sufficient flexibility to account for the variation in branching structures

of phylogenetic trees and is amenable to algorithmic analyses. MBTs can also be written as level-dependent quasi-birth-and-death processes (LDQBD).

We show that many of the current models of the macroevolutionary process are subsumed by the MBT. In particular, we show that the most flexible of these models, the Multi-rate model (MR), which is also a ctMMTBP, can be subsumed by the MBT in the limit as $t \to \infty$. We do this by transforming the MR into an MBT. This model has a simpler interpretation than the MR model and now the probability that a random tree eventually evolves to some topology, $\mathcal{T}$, has an analytic solution.

Since the MBT is a LDQBD, the myriad numerical algorithms within the theory of matrix analytic methods can be modified to apply to the MBT. Indeed, we show that despite the MBT being a level-dependent QBD process, two level-independent algorithms can be modified for determining the probability of eventual extinction of the process. These algorithms are called the Depth and Order algorithms and are based on different physical interpretations of the evolution of MBTs. These algorithms can also be applied to find the extinction probability of MR model trees. Surprisingly, we show that level-independent quadratically convergent algorithms cannot be modified to the MBT and that level-dependent quadratically convergent algorithms are generally less efficient than the linearly-convergent Order algorithm. We also develop an algorithm for the MBT that determines the average imbalance.

The MBT is generalized to the Markovian tree (MT), characterized by the fact that branch points need not be binary. The MT provides an alternative framework for the ctMMTBP and bridges the gap between branching processes and matrix analytic methods. Finally, we provide the Depth and Order algorithms for the MT model.

# Chapter 1

# Introduction

## 1.1 Macroevolution and Mathematical Modelling

Earth is home to a staggering amount of diversity of life. How did such diversity arise? The difficulty in answering such a question makes it all the more enticing to attempt to solve it. One can begin to understand the mechanisms behind evolution by studying the process at the microscopic level, that is, by studying the changes that occur at the genetic level, or at the macroscopic level, that is, by studying the changes that occur at the species level. In macroevolution to be more specific, we are concerned with identifying the differences between species, quantifying these differences and then understanding just how and why these differences arose. The relationships between species can be represented pictorially in diagrams called phylogenetic trees. Phylogenetic trees give information on how related two species may be and in some cases predict the time since these species diverged from their most recent common ancestor.

There are, of course many problems associated with the biological and statistical determination of phylogenetic trees [22]. For example, one very important source of information, the paleaontological record (the fossil record) is incomplete. Therefore, in order to infer the phylogenetic tree shape from an incomplete data set requires the

use of statistics and a stochastic model of the macroevolutionary process. This can in principle then, produce a phylogenetic tree shape that has the highest probability of representing the actual tree shape [22].

Phylogenetic tree shape is important [1, 10, 11, 12, 15, 22, 30, 26, 33, 34] because it gives clues as to how the rates of macroevolution, that is, the rates of species generation and the rates of species extinction, have changed over time and in different physical locations [22]. The rate of change of the macroevolutionary process can have profound effects on the shape of the phylogenetic trees [22]. Phylogenetic trees that demonstrate significant rate variation are imbalanced. That is, different portions of the tree have different shapes. For example, some portions of the tree may be densely populated with many short branches, whereas other portions may be sparsely populated with long branches. Therefore the shape of well constructed phylogenetic trees can give clues as to the processes that may have driven macroevolution and thus generated tree shape [22].

As we have stated above, to aid in the construction of phylogenetic trees one needs to make use of stochastic modelling [1, 11, 22, 26, 30]. In order for a model to be reasonable, it must have the ability to generate useful information and to be mathematically tractable with physically reasonable assumptions.

Stochastic models are important in that they provide a probability distribution over the finite number of possible phylogenetic tree shapes that have a finite number of species. The stochastic models that have been utilized [10, 11, 30] are very simple in that they do not allow for any variation in the rate of the macroevolutionary process. One of these simple models is the well known, constant-rates birth-and-death (crBD) model and another is the proportional-to-distinguishable arrangements (PDA) model, see [22] and references therein. As expected, these models cannot account for the levels of imbalance that are found in phylogenetic trees, because they do not allow for rate variation. The crBD model predicts trees that are too balanced whereas the PDA model predicts trees that are too imbalanced [30]. Consequently, the next step in the development of physically reasonable mathematically tractable

models is to allow for rate variation.

The process of macroevolution can be thought of as a continuous-time branching process. That is, a process that begins with some particles that have the ability to generate new particles at random time intervals. This is exactly what is happening at the species level in macroevolution, a species will at some random points spawn a new species. It is generally believed that at any time point only one new species is generated [22]. This is a reasonable assumption, since it seems unlikely that two or more new species will be created simultaneously. The use of more sophisticated branching process models was originally suggested by Mooers and Heard [22] and then re-iterated by Aldous [1].

The continuous-time Markovian multi-type branching process (ctMMTBP) [2, 21] is an excellent candidate for a macroevolutionary model because it allows for variations in the rates of speciation and variations in extinction rates. Unfortunately though, the ctMMTPB is difficult to analyse and there is very little algorithmic development.

Despite this, Pinelis [26] proposed a model based on the continuous-time Markovian multi-type branching process (ctMMTBP) called the multi-rate (MR) model. It was called the multi-rate model to emphasize the fact that this model allows for significant rate variation. The MR model assigns to each species individual speciation and extinction rates. For example, some species have the capacity to generate new species more rapidly than others, whereas other species can become long-lived evolving only very slowly. The MR model encompasses all the models that do not allow for rate variation.

In this thesis we propose a model of the macroevolutionary process that is also a continuous-time Markovian multi-type branching process which we have called the Markovian binary tree model (MBT). The MBT requires us to interpret the ctMMTBP in a subtly different way. This new interpretation admits a different representation to the conventional ctMMTBP representation. Consequently, a whole new vista of modelling flexibility is opened up to the MBT because this representa-

tion provides an excellent platform from which to develop a sound algorithmic basis. Consequently, the answers to questions that a biologist may have can be potentially solved using the MBT. Thus, due to its representation and interpretation, the MBT has a significant advantage over the MR. In fact, the MR model can be shown to be encompassed by the MBT.

In the next section we discuss the layout of this thesis.

## 1.2 A guide to the thesis

We begin by giving an introduction to the world of branching processes in Chapter 2. Branching processes have a rich history of theoretical development [2, 9]. The first process that we discuss is the discrete-time Galton-Watson process, the cornerstone of branching process theory. This process is then generalized to its continuous-time counterpart, the continuous-time Markovian branching process. Following these preliminaries we then discuss the continuous-time Markovian multi-type branching process. This branching process provides the core from which the MBT is constructed and we therefore take some time in explaining it carefully.

In Chapter 3 we begin by discussing the macroevolutionary biological background. We briefly introduce phylogenetic trees and then discuss some of the important tree topological concepts. The next step we take is to discuss the most important quantitative measure of tree imbalance: Colless' index of imbalance. The remainder of Chapter 3 is devoted to introducing some of the most important macroevolutionary models. The constant-rates birth-and-death (crBD) model which has an important place in applied probability, the proportional-to-distiguishable arrangements (PDA) model, the super-PDA model and finally the multi-rate (MR) model of Pinelis [26]. We also show how the crBD model generates the PDA model. Finally, a discussion of the MR model is given.

Having introduced branching processes and the biological background we next introduce the theory of matrix analytic methods in Chapter 4. We commence by dis-

cussing the Poisson process, followed by the phase-type renewal process. The phase-type renewal process is the generalization of the Poisson process to non-exponential inter-event distributions. We next discuss the Markovian arrival process (MAP). The MAP is the generalization of the phase-type renewal process to include correlations. The MAP generates the dynamics of the MBT. The concept of the hidden and observable transitions of the MAP is used to alter the interpretations of particle transitions in the ctMMTBP and create the MBT interpretation. The level-independent quasi-birth-and-death process (LIQBD) is then introduced and we analyze the algorithm of Neuts, the algorithm $U$ and the level-independent logarithmic reduction algorithm. The algorithm of Neuts and the algorithm $U$ form the basis for analogous algorithms for the MBT that determine the probability of eventual extinction. The final process we discuss is the level-dependent quasi-birth-and-death process (LDQBD). The LDQBD process is the framework within which we represent the Markovian binary tree. The last topic we discuss is the level-dependent logarithmic reduction algorithm.

In Chapter 5 we begin by representing the Markovian binary tree (MBT) as a level-dependent quasi-birth-and-death process. We re-interpret the ctMMTBP process such that each evolving branch of an MBT has its own copy of the MAP. Since the MBT is a ctMMTBP, more specifically, a binary-branch point ctMMTBP, we also write the basic branching process equations for the MBT. From these equations we obtain the equation for the probability of eventual extinction of the process. The final sections of Chapter 5 are devoted to showing that all the models discussed in Chapter 3 are special cases of the MBT. We show, in particular, that the MR model can also be written in terms of an MBT and is thus subsumed by the MBT.

In Chapter 6 we demonstrate the power and flexibility of the MBT by developing an algorithm that calculates the mean imbalance conditional on tree size. We show that there exists a simple MBT with one parameter that has sufficient flexibility to span the entire range of theoretically allowed imbalance values for size five trees. We also demonstrate that even though this one parameter model was designed specif-

ically for size five trees, this model still generates interesting behaviour for larger size trees. It still spans most of the allowed imbalance values and therefore retains much of the flexibility seen for size 5 trees. The final section of Chapter 6 is devoted to calculating the computational complexity of the algorithm.

Chapter 7 continues the algorithmic development of the MBT, where we specifically concentrate on finding the minimal non-negative solution to the equation for the probability of eventual extinction of the MBT process. We begin by developing the Depth algorithm which is analogous to the algorithm of Neuts. We show that the difficulty in describing the sample paths in the algorithm of Neuts is removed if the sample paths are transformed into binary trees. The Order algorithm which is analogous to the algorithm $U$ is also developed. This algorithm has an interesting physical interpretation based on a concept called the order of a tree. The Order algorithm is shown to converge linearly with respect to order. A comparison of the Depth and Order algorithms is made and we show that the Order algorithm converges at a faster rate than the Depth algorithm because it considers more topologies at each iteration. We conclude Chapter 7 by analyzing the quadratically convergent logarithmic reduction algorithms. It is shown that a level-independent logarithmic reduction algorithm is not possible for the MBT and that the level-dependent logarithmic reduction algorithm will in general perform worse than the Order algorithm.

The success with which algorithms were developed in Chapters 6 and 7 leads us to the generalization of the MBT. The general Markovian tree (MT) is introduced in Chapter 8. We begin by representing the MT in a matrix analytic form, just as we did for the MBT, and then write the general ctMMTBP definition of the MT. By writing the general ctMMTBP as an MT we commence developing algorithms that may be of use in a physical modelling context. Therefore as a starting point, we develop the Depth and Order algorithms for the probability of eventual extinction of the MT. These algorithms reduce to the Depth and Order algorithms of the MBT if each branch point is forced to be binary.

# Chapter 2

# Branching Processes

## 2.1 Introduction

Evolutionary biologists face the daunting task of providing a framework with which to explain the observed diversity of life found on earth. The relationships between the species can be represented through the use of tree diagrams, called phylogenetic trees. The task then, is to decipher the shape of the phylogenetic tree of life and to determine the mechanisms that generated that particular shape. However, given the incompleteness of the biological record and the scarce knowledge of the factors that cause macroevolution, this is indeed a daunting task. At a more modest level, evolutionary biologists have studied the shapes of some of the subtrees of the tree of life by using biological and statistical techniques. As a result, there is now an emphasis on developing models of the macroevolutionary process [1, 11, 10, 22, 26, 30].

There are two possible avenues with which to pursue the development of a model of macroevolution,

- to develop a model that is based solely on physical considerations, or
- to construct a model that can account for the tree shapes that arise in nature, without attempting to provide a complete mechanistic basis for their

generation.

The first approach is currently extremely difficult to implement since it is plagued by a lack of understanding of the underlying biological mechanisms that cause macroevolution. In this thesis, we choose the second approach. Thus we shall develop a model that can account for the tree shapes that appear in nature, which in addition is also based on some reasonable physical considerations.

Qualitatively then, a species under some evolutionary constraints will continue evolving and at some point during its evolution will either become extinct or give rise to new daughter species while it then continues to evolve. Viewed in this light, a branching process seems to be a perfect candidate as a model of macroevolution. In fact, Mooers and Heard [22] stated this very succinctly, "*most biological taxa have arisen by a branching process of descent with modification*".

The remainder of this chapter is devoted to discussing some important branching processes. The aim is to describe the fundamental nature of the models currently used in macroevolutionary modelling in addition to allowing us to introduce the model that is proposed in this thesis. This chapter is organised as follows. In Section 2.2 we discuss the simplest type of branching process called the Galton-Watson process (GW). The Galton-Watson process is a discrete-time single-type branching process and is the simplest of all the branching processes. In Section 2.3 we describe the continuous-time analogue of the GW process: the single-type continuous-time Markovian branching process (ctMBP). Finally, in Section 2.4 the ctMBP is generalized to the multi-type analogue, called the continuous-time Markovian multi-type branching process (ctMMTBP).

## 2.2 The Galton-Watson Process

### 2.2.1 Definition

Excellent introductions to the Galton-Watson process can be found in [2] and [9]. Let the random variable $Z_l$ denote the number of particles that are present at time $l$ given that the process commenced with one particle at time 0. Each particle that is present evolves independently of all the others and of its preceding history. At time $l+1$ a particle can either give rise to no offspring with probability $p_0$ or with probability $p_k$ give rise to $k$ daughter particles, for $k \geq 1$.

The generating function of the offspring distribution of one particle is given by,

$$f(s) = \mathbb{E}[s^{Z_1}] = \sum_{k=0}^{\infty} p_k s^k, \ |s| \geq 1, \tag{2.2.1}$$

and the expected number of particles in the first generation spawned by a single particle is given by

$$\mathbb{E}(Z_1) = \left.\frac{df(s)}{ds}\right|_{s=1} = \sum_{k=0}^{\infty} k p_k. \tag{2.2.2}$$

The iterates of the above probability generating function are,

$$f_0(s) = s, \qquad f_1(s) = f(s), \qquad f_{n+1}(s) = f(f_n(s)). \tag{2.2.3}$$

Let $P(i,m)$ be the one-step probability that the process will have $m$ particles given that there were $i$ at the previous step, in other words,

$$P(i,m) = P[Z_{l+1} = m \big| \ Z_l = i]. \tag{2.2.4}$$

Clearly then,

$$\sum_{m=0}^{\infty} P(1,m) s^m = f(s). \tag{2.2.5}$$

Suppose that the process commences with $i$ particles and since the offspring distribution at the next generation is the sum of $i$ independent random variables, the

probability generating function for the offspring distribution is given by the convolution of the $i$ individual offspring probability generating functions. Hence,

$$\sum_{m=0}^{\infty} P(i,m)s^m = \left[f(s)\right]^i, \tag{2.2.6}$$

for $i \geq 1$.

Let $P_n(i,m)$ be the probability that there are $m$ particles at $n+1$ given that there were $i$ particles at time 1. Then, following, [2],

$$\begin{aligned}
\sum_{m=0}^{\infty} P_{n+1}(1,m)s^m &= \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} P_n(1,k)P(k,m)s^m && (2.2.7)\\
&= \sum_{k=0}^{\infty} P_n(1,k)\sum_{m=0}^{\infty} P(k,m)s^m && (2.2.8)\\
&= \sum_{k=0}^{\infty} P_n(1,k)\left[f(s)\right]^k, && (2.2.9)
\end{aligned}$$

where in the first step we have used the Chapman-Kolmogorov equations, and in the final step we have used equation (2.2.6).

Using arguments similar to equation (2.2.6) it can be shown that,

$$\sum_{m=0}^{\infty} P_n(i,m)s^m = [f_n(s)]^i. \tag{2.2.10}$$

In other words, the $n$-th step probability generating function of the *process* is given by the product of the $n$-th step probability generating functions of each of the *individual* branching processes commenced by the $i$ initial particles.

### 2.2.2 Transience of the Non-zero States and the Extinction Probability

It has been shown [9] that all the non-zero finite particle states of the process are transient, thus, with probability one, $Z_l \to 0$ or $Z_l \to \infty$ as $l \to \infty$

The probability of eventual extinction, $q$, of the process is the probability that as $l \to \infty$ there are no living particles remaining. In other words, $q = \lim_{l\to\infty} P[Z_l = 0]$.

It can be shown, [2, 9] that the extinction probability is the smallest non-negative root of the equation

$$s = f(s). \tag{2.2.11}$$

Furthermore, it can also be shown that, if $\mathbb{E}(Z_1) \leq 1$ and the variance is greater than zero when $\mathbb{E}(Z_1) = 1$ then $q = 1$ and if $\mathbb{E}(Z_1) > 1$ then $q < 1$. The process is called subcritical if $\mathbb{E}(Z_1) < 1$, critical if $\mathbb{E}(Z_1) = 1$ and supercritical if $\mathbb{E}(Z_1) > 1$. Now, since all the non-zero finite particle states of the process are transient we have,

$$P[\lim_{l\to\infty} Z_l = 0] = q = 1 - P[\lim_{l\to\infty} Z_l = \infty]. \tag{2.2.12}$$

## 2.3 The One-Dimensional Continuous-Time Markovian Branching Process

### 2.3.1 Definition

Consider the following continuous-time process: a particle that is alive at time $t$ will live for an exponentially distributed lifetime with mean $1/\alpha$, at which point it will either give rise to no offspring with probability $p_0$ or will give rise to $m \geq 1$ offsping with probability $p_m$. Each particle evolves independently of all the other particles and of its history. Such a process is called a one-dimensional continuous-time Markovian branching process (ctMBP). The probability generating function of the offspring distribution is again given by,

$$f(s) = \sum_{m=0}^{\infty} p_m s^m. \tag{2.3.1}$$

Now let $Z(t)$ denote the number of particles alive at time $t$. Let $P_{1m}(t) = P[Z(t) = m \big| Z(0) = 1]$ be the probability that at time $t$ there are $m$ particles given that the process commenced with only one particle. The probability generating function of

the number of particles at time $t$ is,

$$F(s,t) = \sum_{m=0}^{\infty} P_{1m}(t)s^m. \tag{2.3.2}$$

The probability that the process will have $m$ particles by time $t+\tau$ given that it had $i$ particles at time $\tau$ is denoted by $P_{im}(t+\tau;\tau) = P[Z(t+\tau) = m \big| Z(\tau) = i]$. Now the process is homogeneous with respect to time, so $P_{im}(t+\tau;\tau) = P_{im}(t)$. Since each particle evolves independently with respect to all other particles, the probability generating function for the number of particles given $i$ initial particles is the $i$-fold convolution of $F(s,t)$, hence,

$$\sum_{m=0}^{\infty} P_{im}(t)s^m = \left(\sum_{m=0}^{\infty} P_{1m}(t)s^m\right)^i = \big[F(s,t)\big]^i. \tag{2.3.3}$$

Let $Q_{ij}$ be the rate at which the process goes from a state with $i$ particles to a state with $j$ particles. We then have

$$Q_{ij} = i\alpha p_{j-i+1}, \tag{2.3.4}$$

for $j = i-1$ and $j > i$, $Q_{ij} = 0$ for $j < i-1$, and finally for $i = j$

$$Q_{ii} = -i\alpha p_0 - i\alpha \sum_{k=i+1}^{\infty} p_{k-i+1} = -i\alpha(1-p_1), \tag{2.3.5}$$

since $\sum_{k=0}^{\infty} p_k = 1$. The interpretation of $Q_{ij}$ is as follows. The rate at which each single particle spawns $j-i+1$ particles can be seen to be $\alpha p_{j-i+1}$ since the particle must die, which occurs with rate $\alpha$, and at its death it spawns $j-i+1$ particles with probability $p_{j-i+1}$. Since any one of the $i$ particles can do this, the total rate is therefore $i\alpha p_{j-i+1}$. Thus the total number of particles is $j$, comprised of the $i-1$ initial particles and the $j-i+1$ newly spawned particles.

We can now write down both the Kolmogorov forward and backward equations for this system. The forward equation is,

$$\frac{d}{dt}P_{ij}(t) = \sum_{k=1}^{\infty} P_{ik}(t)Q_{kj}, \tag{2.3.6}$$

Since $Q_{kj} = 0$ for $j < k - 1$ we have,

$$\frac{d}{dt}P_{ij}(t) = \sum_{k=1}^{j+1} P_{ik}(t)Q_{kj}. \tag{2.3.7}$$

Substituting equations (2.3.4) and (2.3.5) into equation (2.3.7) we obtain,

$$\frac{d}{dt}P_{ij}(t) = -j\alpha P_{ij}(t) + \sum_{k=1}^{j+1} P_{ik}(t)k\alpha p_{j-k+1}. \tag{2.3.8}$$

The backward equation can be derived from

$$\frac{d}{dt}P_{ij}(t) = \sum_{k=1}^{\infty} Q_{ik}P_{ik}(t), \tag{2.3.9}$$

and yields

$$\frac{d}{dt}P_{ij}(t) = -i\alpha P_{ij}(t) + i\alpha \sum_{k=i-1}^{\infty} p_{k-i+1}P_{kj}(t). \tag{2.3.10}$$

By multiplying equations (2.3.8) and (2.3.10) by $s^j$ and then summing from $j = 0$ to infinity we get,

$$\frac{\partial}{\partial t}F(s,t) = u(s)\frac{\partial}{\partial s}F(s,t), \quad \text{(forward equation)}, \tag{2.3.11}$$

and

$$\frac{\partial}{\partial t}F(s,t) = u\big(F(s,t)\big), \quad \text{(backward equation)}, \tag{2.3.12}$$

where

$$u(s) = \alpha\big(f(s) - s\big). \tag{2.3.13}$$

### 2.3.2 Non-Explosiveness and the Mean of the Process

In Harris, [9], it was shown that the process is non-explosive, that is, $Z(t) < \infty$ for all $t < \infty$ almost surely, if

$$\int_{1-\epsilon}^{1} \frac{ds}{f(s) - s} = \infty, \tag{2.3.14}$$

for every $\epsilon > 0$. The condition

$$\left.\frac{d}{ds}f(s)\right|_{s=1} < \infty, \tag{2.3.15}$$

is sufficient to ensure that the process is non-explosive [2]. This condition implies that the mean number of particles produced by a single particle upon its death is finite.

The mean number of particles of the process at time $t$ is defined by,

$$M(t) = \mathbb{E}[Z(t)|Z(0) = 1]. \tag{2.3.16}$$

Since, $M(t) = \frac{\partial}{\partial s}F(s,t)\big|_{s=1}$ we can differentiate the Kolmogorov backward equation with respect to $s$ to obtain,

$$\frac{d}{dt}M(t) = \lambda M(t), \tag{2.3.17}$$

where

$$\lambda = \left.\frac{d}{ds}u(s)\right|_{s=1} = \alpha\left(\left.\frac{d}{ds}f(s)\right|_{s=1} - 1\right). \tag{2.3.18}$$

Recall that $\frac{d}{ds}f(s)\big|_{s=1}$ is just the mean number of offspring generated when a particle expires. The solution to equation (2.3.17) is given by,

$$M(t) = \exp(\lambda t), \tag{2.3.19}$$

and observe that, if

- $\lambda > 0$ then $\lim_{t\to\infty} M(t) = \infty$ and the process is supercritical,
- $\lambda = 0$ then $\lim_{t\to\infty} M(t) = 1$ and the process is critical, in fact $M(t) = 1$ for all $t$, and finally if,
- $\lambda < 0$ then $\lim_{t\to\infty} M(t) = 0$ and the process is subcritical.

### 2.3.3 Transience of the Non-Zero States and the Extinction Probability

Harris [9] has shown that the Galton-Watson process is imbedded in the continuous-time Markovian branching process. Since all the non-zero finite-particle states of the

Galton-Watson process are transient, this implies that all the non-zero finite-particle states of the ctMBP are also transient and as a result,

$$P[\lim_{t\to\infty} Z(t) = 0] = 1 - P[\lim_{t\to\infty} Z(t) = \infty]. \tag{2.3.20}$$

Define the probability of extinction at time $t$ to be

$$q(t) = P[Z(t) = 0 \big| Z(0) = 1] = F(0,t). \tag{2.3.21}$$

It is not hard to see from the definition of $F(s,t)$ that $q(t)$ is a non-decreasing function of $t$. From the Kolmogorov backward equation (2.3.10) one obtains,

$$\frac{d}{dt}q(t) = u(q(t)), \tag{2.3.22}$$

with initial condition $q(0) = 0$. It is shown in [2] that $q = \lim_{t\to\infty} q(t)$ is the minimal non-negative solution of

$$u(s) = 0. \tag{2.3.23}$$

## 2.4 The Continuous-Time Markovian Multi-type Branching Process

### 2.4.1 Definition

Having discussed the Galton-Watson and the one dimensional continuous-time Markovian branching process we are now in a position to introduce the continuous-time multi-type Markovian branching process (ctMMTBP). The main point of difference between this process and the one dimensional process, is that there are now $n$ different particles *types* as opposed to only one type in the ctMBP.

We shall follow the development in Athreya and Ney [2]. Suppose we have a process with $n$-particle types, each particle of type $i \in \{1,\ldots,n\}$ has a life-span that is exponentially distributed with mean $1/a_i$, and upon its death will produce offspring of the $n$-types with distribution $p^{(i)}(j_1, j_2, \ldots, j_n)$, where $j_k \in \{0\} \cup \mathbb{Z}^+$

represents the number of particles of type $k \in \{1, \ldots, n\}$ that will be spawned. The particles upon their birth evolve independently of each other and of the past.

The offspring probability generating function given that the process begins with one particle of type $i$, for $i \in \{1, 2, \ldots, n\}$, is

$$f^{(i)}(s_1, s_2, \ldots, s_n) = \sum_{j_1, j_2, \ldots, j_n \geq 0} p^{(i)}(j_1, j_2, \ldots, j_n) s_1^{j_1} s_2^{j_2} \ldots s_n^{j_n}. \tag{2.4.1}$$

We say that the process is singular if the generating functions (2.4.1) only consist of terms that are linear in $s_i$ for all $i \in \{1, \ldots, n\}$. We call a branch point a singular branch point, if an $i$-type particle transforms into a $j$ type particle, for $i \neq j$. A binary branch point occurs when a particle of type $i$ terminates and spawns two daughter particles of types $j$ and $k$, for any $j, k = 1, 2, \ldots, n$.

Let $\boldsymbol{j} = (j_1, \ldots, j_n)$ and $\boldsymbol{i} = (i_1, \ldots, i_n)$ denote two vectors such that $j_k, i_k \in \{0\} \cup \mathbb{Z}^+$ for all $k \in \{1, \ldots, n\}$. Let $\boldsymbol{Z^i}(t) = (Z_1^{\boldsymbol{i}}(t), \ldots, Z_n^{\boldsymbol{i}}(t))$ be the number of particles of each type at time $t$ given that the process began with $\boldsymbol{i}$ particles at time 0. Let $P(\boldsymbol{i}, \boldsymbol{j}; t)$ be the probability that a process beginning with $\boldsymbol{i}$ particles at time 0 will have $\boldsymbol{j}$ particles at time $t$. The generating function is given by,

$$F(\boldsymbol{i}, \boldsymbol{s}; t) = \mathbb{E}[s_1^{Z_1^{i_1}(t)} \ldots s_n^{Z_n^{i_n}(t)}] = \sum_{\boldsymbol{j} \in (\{0\} \cup \mathbb{Z}^+)^n} P(\boldsymbol{i}, \boldsymbol{j}; t) s_1^{j_1} \ldots s_n^{j_n}, \tag{2.4.2}$$

where $\boldsymbol{s} = (s_1, \ldots, s_n)$ and $(\{0\} \cup \mathbb{Z}^+)^n$ is the $n$-fold cartesian product of $\{0\} \cup \mathbb{Z}^+$. For ease of exposition we henceforth denote $s_1^{Z_1^{i_1}(t)} \ldots s_n^{Z_n^{i_n}(t)}$ by $\boldsymbol{s}^{\boldsymbol{Z^i}(t)}$

Let $\boldsymbol{e}_i$ be the vector with one in the $i$-th component and zero in the other $n - 1$ components. Let $Z_i(t)$ be the number of particles of type $i$ present at time $t$. Due to the independence of the evolution of each of these particles, each particle initiates another multi-type branching process. Therefore, let $Z_j^{k,i}(\tau)$ be the number of particles of type $j$ that are generated by the $k$-th particle of type $i$ in a time interval of length $\tau$. The total number of particles of type $j$ that are present at time $t + \tau$ is given by the sum of the type $j$ particles generated by the $Z_i(t)$ particles at time

$t$ for all $i = 1, 2, \ldots, n$. As a result, the total number of particles of type $j$ are

$$Z_j(t+\tau) = \sum_{i=1}^{n} \sum_{k=1}^{Z_i(t)} Z_j^{k,i}(\tau). \tag{2.4.3}$$

In terms of the generating functions of the particle distribution, equation (2.4.3) can be written as

$$\boldsymbol{F}(\boldsymbol{s}; t+\tau) = \boldsymbol{F}(\boldsymbol{F}(\boldsymbol{s}; \tau); t), \tag{2.4.4}$$

where $\boldsymbol{F}(\boldsymbol{s}; t) = \big(F(\boldsymbol{e}_i, \boldsymbol{s}; t), \ldots, F(\boldsymbol{e}_n, \boldsymbol{s}; t)\big)$ and

$$F(\boldsymbol{e}_i, \boldsymbol{s}; t) = \mathbb{E}[\boldsymbol{s}^{\boldsymbol{Z}^{\boldsymbol{e}_i}(t)}].$$

The Kolmogorov differential equations play an important role in the theory of Markovian processes. For the case of the ctMMTBP both the forward and backward equations were first derived by Sevastyanov [32]. The forward equations are,

$$\frac{\partial}{\partial t} F(\boldsymbol{e}_i, \boldsymbol{s}; t) = \sum_{k=1}^{n} u^{(k)}(\boldsymbol{s}) \frac{\partial}{\partial s_k} F(\boldsymbol{e}_i, \boldsymbol{s}; t), \tag{2.4.5}$$

where

$$u^{(i)}(\boldsymbol{s}) = a_i[f^{(i)}(s_1, \ldots, s_n) - s_i], \tag{2.4.6}$$

for all $i \in \{1, 2, \ldots, n\}$. Sevastyanov [32] cleverly derived the Kolmogorov forward equations using a probability generating function approach.

The backward equations are given by,

$$\frac{\partial}{\partial t} F(\boldsymbol{e}_k, \boldsymbol{s}; t) = u^{(k)}[\boldsymbol{F}(\boldsymbol{s}; t)], \tag{2.4.7}$$

for all $k \in \{1, \ldots, n\}$. The backward equations have a simple physical interpretation, and consequently they can be derived in a more intuitive fashion than the forward equations. We shall use the argument as presented in [5]. Let the process commence with one particle of type $k$. The lifetime of this particle, $T$, is exponentially distributed with parameter, $a_k$, thus $P_k(T \leq t) = 1 - \exp(-a_k t)$. Now by conditioning on the lifetime of the particle we have,

$$F(\boldsymbol{e}_k, \boldsymbol{s}; t) = \mathbb{E}[\boldsymbol{s}^{\boldsymbol{Z}^{\boldsymbol{e}_k}(t)} | T > t] \exp(-a_k t) + \int_0^t \mathbb{E}[\boldsymbol{s}^{\boldsymbol{Z}^{\boldsymbol{e}_k}(t)} | T = x] a_k \exp(-a_k x) dx. \tag{2.4.8}$$

The first term in equation (2.4.8) represents the situation where the original particle has not yet died, as result it is the only particle in the process, and therefore

$$\mathbb{E}[\boldsymbol{s}^{\boldsymbol{Z}^{\boldsymbol{e}_k}(t)}|T > t]\exp(-a_k t) = s_k \exp(-a_k t). \tag{2.4.9}$$

The second term of equation (2.4.8) represents the situation where at time $T = x \leq t$, the initial particle dies and generates $\boldsymbol{j}$ new particles. In the remaining time $t-x$, each of the $\boldsymbol{j}$ new particles (spawned from the original $k$-type particle) generate their own ctMMTBP. Thus,

$$\begin{aligned}\mathbb{E}[\boldsymbol{s}^{\boldsymbol{Z}^{\boldsymbol{e}_k}(t)}|T = x] &= \sum_{\boldsymbol{j}} p^{(k)}(\boldsymbol{j}) F(\boldsymbol{e}_1, \boldsymbol{s}; t-x)^{j_1} \dots F(\boldsymbol{e}_n, \boldsymbol{s}; t-x)^{j_n} \\ &= f^{(k)}\big(\boldsymbol{F}(\boldsymbol{s}; t-x)\big).\end{aligned} \tag{2.4.10}$$

Substituting (2.4.9) and (2.4.10) into equation (2.4.8) we obtain

$$F(\boldsymbol{e}_k, \boldsymbol{s}; t) = s_k \exp(-a_k t) + \int_0^t f^{(k)}\big(\boldsymbol{F}(\boldsymbol{s}; t-x)\big) a_k \exp(-a_k x). \tag{2.4.11}$$

If we multiply through by $\exp(a_k t)$, we obtain,

$$F(\boldsymbol{e}_k, \boldsymbol{s}; t)\exp(a_k t) = s_k + \int_0^t f^{(k)}\big(\boldsymbol{F}(\boldsymbol{s}; t-x)\big) a_k \exp(a_k(t-x))dx. \tag{2.4.12}$$

Now changing the variable of integration from $x$ to $u = t - x$ we obtain,

$$F(\boldsymbol{e}_k, \boldsymbol{s}; t)\exp(a_k t) = s_k + \int_0^t f^{(k)}\big(\boldsymbol{F}(\boldsymbol{s}; u)\big) a_k \exp(a_k u))du. \tag{2.4.13}$$

Finally differentiating equation (2.4.13) with respect to $t$, using the Fundamental Theorem of Calculus and then multiplying through by $\exp(-a_k t)$ we obtain the Kolomogorov backward equation,

$$\frac{\partial}{\partial t} F(\boldsymbol{e}_k, \boldsymbol{s}; t) = u^{(k)}\big[\boldsymbol{F}(\boldsymbol{s}; t)\big], \tag{2.4.14}$$

for all $k \in \{1, \dots, n\}$.

### 2.4.2 Non-Explosiveness and the Mean of the Process

The process is not explosive, that is, regular, [2] if

$$\left.\frac{\partial f^{(i)}(\boldsymbol{s})}{\partial s_j}\right|_{\boldsymbol{s}=\boldsymbol{e}} < \infty, \tag{2.4.15}$$

for all $i, j = 1, 2, \ldots, n$, where $\boldsymbol{e}$ is a vector of ones of the appropriate dimension and two vectors are considered equal to each other if all their components are equal. In other words, the process is non-explosive if the expected number of particles of any type $j$ given that a birth occurs from a particle of type $i$ is finite for all $i, j = 1, 2, \ldots, n$.

The condition (2.4.15) can also be shown [2] to imply that

$$m_{ij}(t) = \mathbb{E}[Z_j(t)\big|Z(0) = \boldsymbol{e}_i] < \infty. \tag{2.4.16}$$

Let the matrix of the expected number of particle types at time $t$ be denoted by $M(t) = \{m_{ij}(t)\big|i, j = 1, \ldots, n\}$. From equation (2.4.4) it is easy to show that $M(t)$ satisfies the semi-group property [2], namely

$$M(t+u) = M(t)M(u), \tag{2.4.17}$$

for $t, u \geq 0$, and from equation (2.4.14) to show the continuity condition,

$$\lim_{t\to 0} M(t) = I, \tag{2.4.18}$$

where $I$ is the $n \times n$ identity matrix. Now (2.4.17) and (2.4.18) imply that there exists a matrix $A$ [2] which is the infinitesimal generator of the semigroup $\{M(t)\big|t \geq 0\}$ such that

$$M(t) = \exp(At). \tag{2.4.19}$$

Each element of the matrix $A$, say $A_{ij}$, can be interpreted as being the average rate at which a particle of type $i$ gives rise to particles of type $j$. In other words, $A_{ij}$ is given by the rate, $a_i$ at which a particle of type $i$ gives birth multiplied by the mean

number of particles of type $j$ that are a created by that initial type $i$ particle. Thus we write, $A_{ij} = a_i b_{ij}$ where

$$b_{ij} = \left.\frac{\partial f^{(i)}(\boldsymbol{s})}{\partial s_j}\right|_{\boldsymbol{s}=\boldsymbol{e}} - \delta_{ij}, \tag{2.4.20}$$

where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}. \tag{2.4.21}$$

The process is called positive regular if there exists some $t = t_0$, such that $m_{ij}(t_0) > 0$ for all $i, j$. From the theory of positive matrices [31] there exists a strictly positive eigenvalue $\rho(t_0)$ of $M(t_0)$, called the Perron-Frobenius eigenvalue, which has the property that any other eigenvalue $\rho$ of $M(t_0)$ is such that $|\rho| < \rho(t_0)$ and the algebraic and geometric multiplicities of the Perron Frobenius eigenvalue are both one. The eigenvalues of $M(t)$ are of the form, $\exp(\lambda_i t)$ for $i = 1, 2, \ldots, n$, where for all $i$, $\lambda_i$ are the eigenvalues of the matrix $A$. Both $M(t)$ and $A$ have the same eigenvectors. Now let $\lambda_1$ be such that $\rho(t_0) = \exp(\lambda_1 t_0)$. Consequently, $\lambda_1$ is real and $\lambda_1 > \mathrm{Re}(\lambda_k)$ for all $\lambda_k = 2, 3, \ldots, n$, [31].

### 2.4.3 Transience of the Non-Zero States and the Extinction Probability

The proof of the transience of the non-zero finite states and of the extinction probability are well known and relatively simple for the discrete-time multi-type branching process (also known as the Galton-Watson multi-type branching process) [9]. With minor modifications these proofs carry over to the continuous-time Markovian multi-type branching process [2, 9]. Thus, if the continuous-time Markovian multi-type branching process is positive regular and non-singular, all states with a finite number of particles are transient. Hence with probability one all realizations of the process will either eventually become extinct or the total number of branches will tend to infinity [2].

Let $q^{(i)}$ be the probability that the process beginning with one particle of type $i$ will eventually become extinct. Let $\boldsymbol{q} = (q^{(1)}, \ldots, q^{(n)})$. It can be shown [2] using the backward equation (2.4.14) that $\boldsymbol{q}$ is the minimal non-negative solution of

$$\boldsymbol{u}(\boldsymbol{s}) = \boldsymbol{0}, \tag{2.4.22}$$

where $\boldsymbol{u}$ is given by equation (2.4.6). This is equivalent to the condition [9]

$$\boldsymbol{q} = \boldsymbol{f}(\boldsymbol{q}), \tag{2.4.23}$$

where $\boldsymbol{f}$ is given by equation (2.4.1), for the discrete-time case. In fact, one can show, as Harris [9] did for the discrete-time case, that the solution $\boldsymbol{q}$ is either equal to $\boldsymbol{e}$, or all its components must be strictly less than one. Similar arguments may be used for the continuous-time case [2]. Consequently, if $\lambda_1 > 0$ then $\boldsymbol{q} < \boldsymbol{e}$ componentwise and if $\lambda_1 \leq 0$ then $\boldsymbol{q} = \boldsymbol{e}$. The process is then called sub-critical, critical or super-critical depending on whether $\lambda_1$ is less than, equal to, or greater than zero respectively.

In the discrete-time super-critical case, Harris [9] has shown that if the process is positive regular and non-singular, then

$$\lim_{n\to\infty} \boldsymbol{f}_n(\boldsymbol{q}_0) = \boldsymbol{q}, \tag{2.4.24}$$

where $\boldsymbol{q}_0$ is any starting vector in the unit cube of appropriate dimension and $\boldsymbol{f}_n(\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{f}_{n-1}(\boldsymbol{s}))$. This provides an algorithm for solving for the probability of eventual extinction of the process. In Chapter 8 we derive an algorithm that utilizes a similar equation to equation (2.4.24) and then develop another algorithm which converges to $\boldsymbol{q}$ in a significantly more efficient manner.

# Chapter 3

# Models of Macroevolution

## 3.1 Introduction

As already stated in Chapter 1, one of the fundamental problems facing evolutionary biologists is to explain the diversity of life found on earth [22]. Attempts at providing solutions to this problem should in principle provide some level of understanding of the factors that have influenced diversification during evolutionary history. The macroevolutionary manifestation of these factors results in changes in the rates of speciation and extinction of species [11, 22]. The consequence of this, as Mooers and Heard [22] stated in their review article, is that "most biological taxa have arisen by a branching process of descent with modification". In the context of developing a suitable model to attempt to describe the macroevolutionary process, this statement implies that a multi-type branching process provides a useful starting point, a point to which we shall return.

## 3.2 Phylogenetic Trees: Species Relationships

The relationships between species in evolutionary history are represented pictorially by a phylogenetic tree. Phylogenetic trees are constructed using information that is obtained from observational data. This data, may be genetic or paleaonto-

logical for instance. Since this data is incomplete, phylogenetic trees can only be inferred. These inferred trees may or may not represent the underlying actual tree, the structure of which cannot be known [22].

A phylogenetic tree consists of a root node, internal nodes, leaf nodes, internal branches which connect two internal nodes, or which connect the root node to an internal node, and leaf branches that connect one internal and one leaf node, or connect the root node to the leaf node if the tree has only one branch. The lengths of the branches in well constructed phylogenetic trees should in principle represent the age of the species. Figure 3.2.1 depicts a hypothetical phylogenetic tree for extant species $A, B, C, D, E, F, G$ and extinct species $a, b$ and $c$. Based on Figure 3.2.1 one could conclude that species $A$ and $B$ are more closely related to each other than say $A$ and $C$. Furthermore, one could infer that $A$ and $B$ should be less related to each other than say $F$ and $G$ since $F$ and $G$ diverged at a later time. In practice, such inferences should be made with caution, because the true tree with all extant and extinct species and correct branch lengths is not known. In fact, there is seldom enough information to be able to accurately include the extinct species in the analysis.

There are a number of statistical and practical problems associated with the reconstruction of phylogenetic trees. For a review of these consult [22]. However, despite these possible problems, phylogenetic trees can provide insight into the macroevolutionary process.

## 3.3 Tree Topology

The topology of a tree is defined [22] to be the branching pattern of that tree when the lengths of the branches and the labels of species at the leaves are ignored. Thus, any tree can be drawn in a topological fashion if all the branch lengths are made equal; this is depicted in Figure 3.3.1. Trees $A$ and $B$ are topologically identical, the only difference is that the branches of tree $A$ have varying lengths whereas the

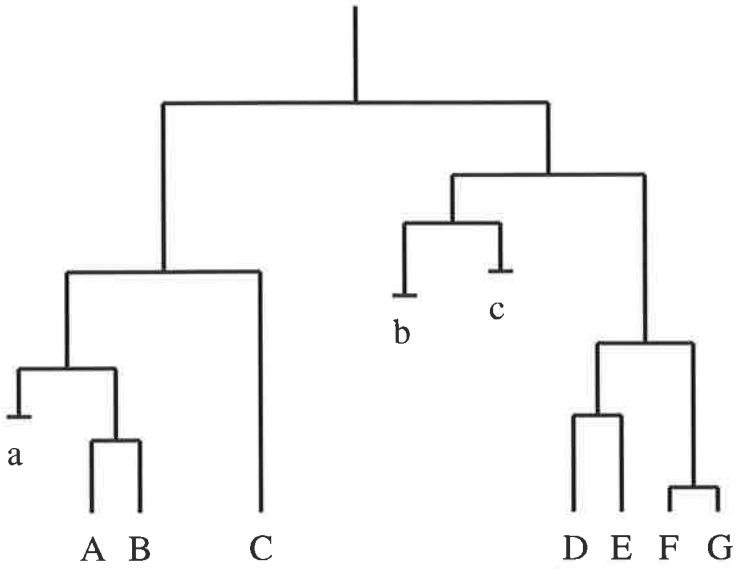


Figure 3.2.1: A hypothetical phylogenetic tree

branches of tree $B$ have identical lengths. The branch lengths of a tree represented topologically do not reflect the ages of the branches.

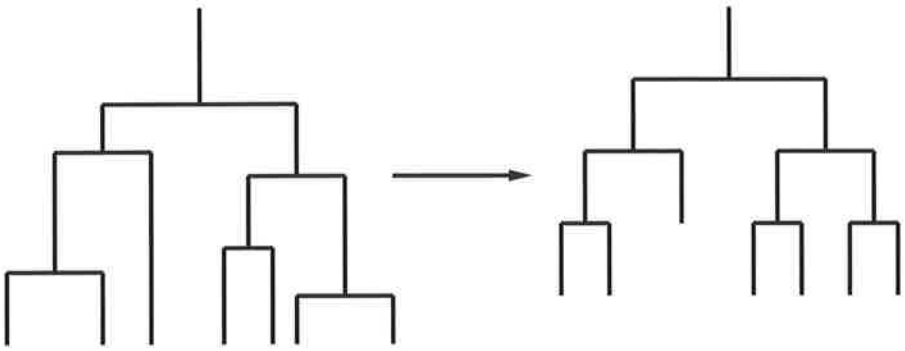


Figure 3.3.1: Two representations of the same tree

The topology or shape of a tree conveys information regarding the positional relationships between species. The topology can also provide information on the

propensity for speciation or extinction. The driving forces behind speciation or extinction are likely to be complex and varied, for example, some may be biogeographical others may be genetic [12, 22]. The consequences of these macroevolutionary driving forces is that the shape of the subtrees differ; as a result the tree is imbalanced. Imbalance can be quantified and there are a number of measures of imbalance, each one with its own interpretation [4, 15, 22]. Nonetheless the more imbalanced a phylogenetic tree is, the more varied the rates of speciation and extinction in different parts of the tree [22]. Changing conditions at different physical locations influence speciation; evolution may either "speed up", "slow down" or cease altogether in these locations.

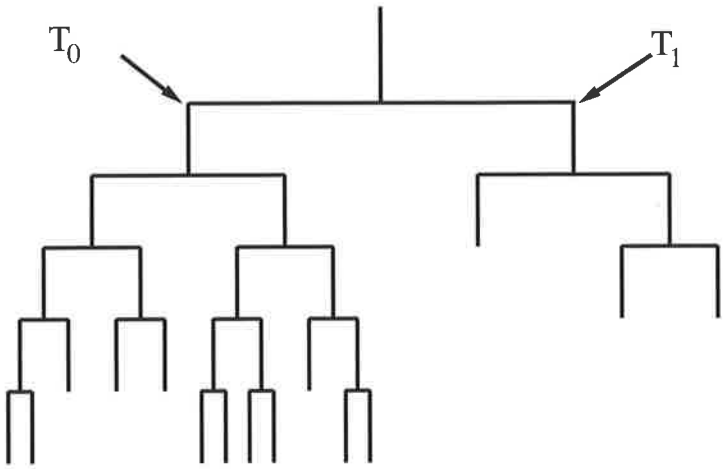


Figure 3.3.2: A tree with varying speciation rates

For example, consider the topology of a phylogenetic tree where the left subtree undergoes speciation much more rapidly than the right subtree. An example of such a tree is given in Figure 3.3.2. We have labeled the left subtree $\mathcal{T}_0$ and the right subtree $\mathcal{T}_1$. The subtree $\mathcal{T}_0$ has undergone many more speciation events than subtree $\mathcal{T}_1$, illustrating that varying speciation and extinction rates can have dramatic effects on the topology and hence the imbalance of a tree. The greater the variation in the rates of speciation and extinction within different parts of the tree, the more imbalanced the tree will be. The topology of the tree thus conveys information

about the historical macroevolutionary process.

There has been considerable interest in studying the imbalance generated by different probability models of tree generation [1, 4, 8, 10, 11, 12, 15, 22, 26, 30]. A model of macroevolution provides a probability measure on the space of tree shapes. As a result some macroevolutionary models may predict more balanced topologies, whereas others might predict less balanced topologies.

To conclude this section, we define the concept of topological isomorphism. Two trees are topologically isomorphic if by a suitable interchange of the left and right branches at each node those two trees can be made identical. The trees depicted in Figure 3.3.3 are topologically isomorphic since the first tree can be transformed into the second tree by interchanging the left and right branches at nodes 0, 1, 2, 3, and 4.

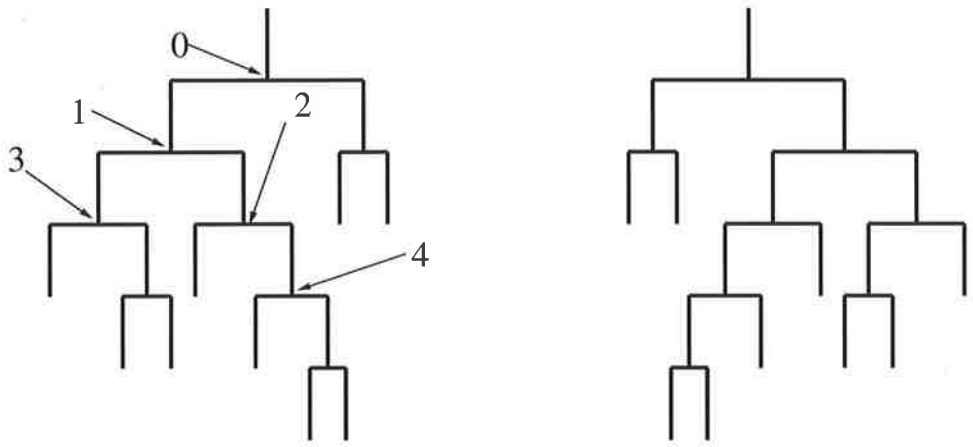


Figure 3.3.3: Two topologically isomorphic trees

Note that in this thesis any two topologically isomorphic trees are considered as having *distinct* topologies. Whereas in [22] any two topologically isomorphic trees are considered as having the *same* topology. Furthermore, there is a subtle difference in the use of terminology, what we call a topology is called a tree in [22] and [30].

## 3.4 A Labelling System for Binary Trees

We define three types of nodes for a binary tree (or for a general tree):

- internal nodes, also called branch points,
- extinct leaf nodes, and
- unstable leaf nodes.

The reason why we have chosen to use these three nodes types is due to the fact that a significant portion of this thesis is concerned with demonstrating that the models of Pinelis [26], who utilized this particular choice of node types, is encompassed by the macroevolutionary model that we propose in Chapter 5. As a result, it is necessary to go through and define branch types and nodes types in more detail.

An internal branch is defined to be a branch that has completed its evolution and is not a leaf branch. An extinct branch is a leaf branch that has completed its evolution. An unstable branch is defined to be a branch that has not completed its evolution. Thus an unstable branch will either generate a new daughter and become internal or it will become extinct.

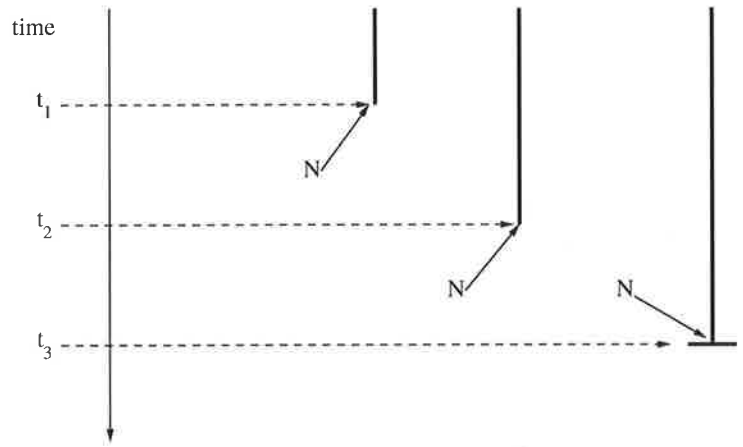


Figure 3.4.1: An example of the evolution of an unstable leaf node.

Internal nodes, also known as branch points, have a fixed position in the tree and do not change as the tree evolves and their name suggests they are neither leaf nodes nor root nodes. An extinct leaf node is the node that is at the end of an extinct branch. The position of an extinct leaf node is fixed it does not change as the tree evolves. The best way to explain an unstable leaf node is to consider Figure 3.4.1. In this figure we depict a single branch. This branch is depicted at three different times, $t_1$, $t_2$ and $t_3$. Since the branch is evolving, the position of leaf node $N$ alters from $t_1$ to $t_2$ and finally to $t_3$. At time $t_3$ the branch becomes extinct and the position of the leaf node is finally fixed and it becomes an extinct leaf node. Thus an unstable leaf node is not fixed whilst a branch continues to evolve but it becomes fixed if the branch undergoes a branch point and becomes an internal node or if the branch becomes extinct the node becomes an extinct leaf node.

From a topological viewpoint however, the single branches in Figure 3.4.1 are all identical, it makes no difference that the unstable node, $N$, was not fixed up until time $t_3$.

Here and throughout, we encase the labels of nodes in square brackets to ensure that each node can be recognized without ambiguity. We begin labelling from the [0] node. This node is either the unstable leaf node of the root branch, the extinct leaf node of a single branch tree, or the first internal node of a tree that has at least two branches. We later give a label to the root node, which is the parent node to [0]. Suppose that $[\psi] = [0, i_1, \ldots, i_m]$, where $i_1, \ldots, i_m \in \{0, 1\}$, is a node of a binary tree. Let $|\psi| = m + 1$ be defined as the depth of the node $[\psi]$. The node that is connected to the left of $[\psi]$, called the daughter node, is labelled,

$$[\psi, 0] = [0, i_1, \ldots, i_m, 0],$$

whereas the node that is connected to the right of $[\psi]$, called the parental subnode of $[\psi]$, is labelled by

$$[\psi, 1] = [0, i_1, \ldots, i_m, 1].$$

As a matter of correct terminology, the parent node of $[\psi] = [0, i_1, \ldots, i_m]$ is always that node that is at a depth of $m$ with label $[0, i_1, \ldots, i_{m-1}]$ and all nodes, $[\psi, \boldsymbol{e}_k]$, where $\boldsymbol{e}_k$ is an $1 \times k$ row vector of ones, are the parental sub-nodes of $[\psi]$ provided that these nodes exist. Figure 3.4.2 depicts a binary tree topology with all its nodes labelled.

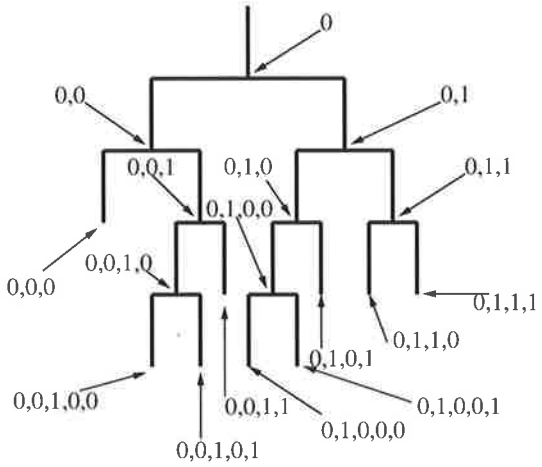


Figure 3.4.2: An example of the labelling of a binary tree.

Once again, let $[\psi] = [0, i_1, \ldots, i_m]$ be any node except for the root node of a binary tree. The function $\alpha$ has the following action on $[\psi]$,

$$\alpha(\psi) = [0, i_1, \ldots, i_{m-1}],$$

that is, $\alpha(\psi)$ is the parent node of $[\psi]$. We hence label the root node of a binary tree by $\alpha(0)$. The function $\theta$, on the other hand, acts on any internal node, $[\psi]$, or the root node $\alpha(0)$ so that,

$$\theta(\psi) = [\psi, 1],$$

and

$$\theta(\alpha(0)) = [0].$$

Suppose that $[\psi] = [0, i_1, \ldots, i_m]$ is a node of a binary tree. The branch segment between the nodes $[\psi]$ and $[\psi, i_{m+1}]$ is represented by,

$$([\psi], [\psi, i_{m+1}]),$$

where $i_{m+1} \in \{0, 1\}$. Consider the branch $([\alpha(\psi)], [\psi])$, then,

- if the branch is extinct, we write $([\alpha(\psi)], [\psi])^{(e)}$, and also write $[\psi]^{(e)}$ to denote the extinct leaf node of that branch,
- if the branch is internal, we write $([\alpha(\psi)], [\psi])^{(i)}$, and also write $[\psi]^{(i)}$ to denote the internal node of that branch, and
- if the branch is unstable we write, $([\alpha(\psi)], [\psi])^{(u)}$ and also write $[\psi]^{(u)}$ to denote the unstable leaf node of that branch.

If a branch type is unimportant we do not specify a superscript.

Let $\mathcal{T}$ denote the topology of a binary tree. If $[\psi]$ is any internal node of a tree of this topology, $\mathcal{T}$, then the tree of topology, $\mathcal{T}_{[\psi]}$, based on node $[\psi]$ can be written as the ordered set,

$$\mathcal{T}_{[\psi]} = \left\{([\alpha(\psi)], [\psi])^{(i)}, \mathcal{T}_{[\psi,0]}, \mathcal{T}_{[\psi,1]}\right\},$$

where $\mathcal{T}_{[\psi,0]}$ and $\mathcal{T}_{[\psi,1]}$ are the topologies of the daughter and parental subtrees whose first internal branch points occur at nodes $[\psi, 0]$ and $[\psi, 1]$ respectively, and for a single branch topology that is extinct we have,

$$\mathcal{T}_{[\psi]} = \{([\alpha(\psi)], [\psi])^{(e)}\},$$

or for a single branch topology that is unstable we have,

$$\mathcal{T}_{[\psi]} = \{([\alpha(\psi)], [\psi])^{(u)}\}.$$

We say that $\mathcal{T}_{[\psi]}$ is the parent tree of the daughter, $\mathcal{T}_{[\psi,0]}$, and parental, $\mathcal{T}_{[\psi,1]}$ subtrees. Consequently, at a branch point, say node $[\psi]$, the branch $([\psi], [\psi, 0])$ is refered to as the daughter branch and the branch $([\psi], [\psi, 1])$ is refered to as the parental branch.

Having discussed internal, unstable and extinct nodes we wish to introduce one more node and branch type, called a quasi-stable node and a quasi-stable branch [26]. A quasi-stable node can be thought of as being similar to an unstable node, except

for one important difference, a quasi-stable node can never become an internal node or an extinct leaf node, therefore once a quasi-stable node is formed, that portion of the branch becomes a non-extinct leaf branch so the node is never fixed. We denote a quasi-stable node by $[\psi]^q$ and a quasi-stable leaf branch as $([\alpha(\psi)], [\psi])^q$.

## 3.5 Macroevolutionary Models

One of the aims of the biologist is to decipher the possible causes of rate variation and to understand how each leaves its "footprint" on macroevolution. The development of stochastic models of the macroevolutionary process may shed some light on the manner in which such complex systems have evolved over time. Models that act as good starting points must have the flexibility to account for any of the myriad possible tree shapes that have been inferred from biological and/or paleontological evidence. However, due to the complexity of the macroevolutionary process a balance needs to be found between the need to provide a sound underlying biological basis and the need to provide algorithmic tractability. To attempt, *a priori*, to include all the known macroevolutionary factors into one model would prove to be intractable.

Branching processes and in particular multi-type branching processes have been utilised in biological applications for some time [14]. Mooers and Heard [22] and then Aldous [1] proposed the potential use of a ctMMTBP in a macroevolutionary context while Pinelis [26] developed a model called the multi-rate model (MR) which was based on the ctMMTBP. One of the major drawbacks to using the ctMMTBP in a modelling context follows from the fact that there seems to be an insufficient number of numerical algorithms from which to calculate the useful measures of the model [5]. In fact the major problem with using the MR approach of Pinelis lies in the fact that there are no reasonable algorithmic approaches except for the simplest of model types. Dorman, Sinsheimer and Lange [5] have identified this problem and provided a step in the right direction. They considered a ctMMTBP with Poisso-

nian immigration and numerically integrated the Kolmogorov backward differential equations. They then applied a finite Fourier transform to obtain the marginal distributions for the generating functions of the probabilities of particle numbers and immigrant particle numbers. As a result, they were able to calculate the mean and variance of particle numbers, and determined numerically the probability of extinction at time $t$. Dorman, Sinsheimer and Lange [5] found that in the supercritical case, as time gets large, the algorithm for determining the probability of extinction failed [5].

The remainder of this chapter is devoted to discussing some general probability concepts for tree topologies in Section 3.6, followed by a more in depth look at tree topological properties in Section 3.7, and then an introduction to Colless's measure of imbalance in Section 3.8. Section 3.9 reviews one of the simplest and most studied branching models, the birth-and-death model. In Section 3.10 the proportional-to-distinguishable arrangements model (PDA) is also reviewed, and we show that the subcritical birth-and-death model generates the PDA model. Finally, in Section 3.11 the most complex model to date is discussed, the multi-rate model, which is a continuous-time Markovian multi-type branching process [26].

## 3.6 Probability Measures and Tree Topology

The state space in which most branching processes are studied is the non-negative integers for one-dimensional branching processes, or the space of $n$-dimensional vectors with non-negative integer components for $n$-dimensional branching processes [2, 9]. However, since we wish to model the macroevolutionary process, such state spaces are not the most useful or insightful. As emphasized previously, the topology of phylogenetic trees reveals much about the underlying macroevolutionary processes. Consequently, having a process on the space of particle numbers is not nearly enough, we need to be able to keep track of the history (lineage) of all the particles in a branching process if we wish to use it as a model of macroevolution.

Knowledge of the history of the particles allows us to map the realization of the process to the space of tree topologies. Thus, instead of analyzing branching processes on the space of positive integers we shall analyze branching processes in their more natural format: on the state space of tree topologies which we denote by $\mathbb{T}$. The classical branching process framework can always be recovered by counting the number of leaf branches of a topology.

To be more precise, in this alternative framework, the evolution (ageing) of a particle traces out a branch of the tree whose branch length is the age of the particle since birth. In any realization of the process, a particle may die or give rise to new particles. If this particle gives birth to new particles we consider this parental particle as still remaining alive. As a result, each realization of the branching process for all times generates a tree whose branch lengths are dependent on the ages of all the particles. However, we are not interested in all this information, but rather the topology of the tree as it evolves. We recover the topology of the tree by applying a mapping from the space of trees to the space of tree topologies. Denote this mapping by $\mathcal{M}$. This mapping is a many to one mapping since there are an uncountably infinite number of trees that all have the same topology but differ only in their branch lengths. Note that as time evolves the space of topologies, $\mathbb{T}$, is exactly the same. What changes is the set of trees that map to each topology in $\mathbb{T}$.

### 3.6.1 Branch Types and Associated Probability Measures

There are three important generic branch types that play an important role in what follows. These three generic branch types are

1. extinct branches
2. unstable branches, and
3. quasi-stable branches.

Let us explain each one in turn. An extinct branch is generated when a particle dies, that is, it ceases to evolve. An unstable branch is generated whilst a particle is still actively evolving and capable of either giving birth to new particles or becoming extinct. A quasi-stable branch is generated when a particle is not extinct but is unable to give birth to new particles. The use of the terms *unstable* and *quasi-stable* is borrowed from Pinelis [26]. This concept of a quasi-stable branch type (particle type) is crucial to the modeling in Pinelis's paper [26] on the Multi-rate model which we discuss later. From here and throughout we shall refer to branches and particles interchangeably.

We can define different forms of the mapping $\mathcal{M}$ depending on the branch types that interest us. Recall, that the mapping $\mathcal{M}$ disregards branch length; this is still true of the variant mappings that we discuss here. The mapping that gives us the topology of a complete tree for any $t$ is denoted by $\mathcal{M}^a = \mathcal{M}$. The mapping that gives the topology of the extinct portion of a tree for any $t$ by pruning all branches except for extinct ones is denoted by $\mathcal{M}^e$. The mapping that gives us the topology of the unstable portion of the tree at any time $t$ by pruning all branches except for unstable ones is denoted by $\mathcal{M}^u$. The mapping that gives us the topology of the quasi-stable portion of a tree at any time $t$ by pruning all branches except for quasi-stable ones is denoted by $\mathcal{M}^q$. It is important to note that all these mappings map to the same space $\mathbb{T}$ independent of branch type. However, so that no ambiguity arises when discussing certain models or the trees that are generated by those models,

- if $\mathcal{T}$ is the topology of the entire tree, we write $\mathcal{T}^a$,
- if $\mathcal{T}$ is the topology of the extinct portion of a tree, we write $\mathcal{T}^e$,
- if $\mathcal{T}$ is the topology of the unstable portion of a tree, we write $\mathcal{T}^u$, and
- if $\mathcal{T}$ is the topology of the quasi-stable portion of a tree, we write $\mathcal{T}^q$.

The branching process models that we analyze are continuous time models and so the topology to which a tree is mapped will change in time. To illustrate this point,

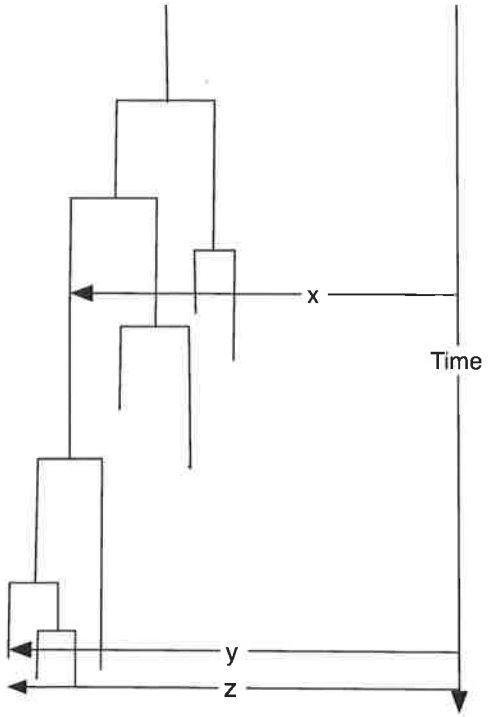


Figure 3.6.1: Same number of branches at different times does not mean topology is the same

consider the tree depicted in Figure 3.6.1 . This tree has evolved until time $t = z$. This tree consists of unstable and extinct branches but no quasi-stable branches. Suppose we wanted to know the topology of the tree at $t = x$. At $t = x$ there are four unstable branches and no extinct branches. The topology of this tree is shown in Figure 3.6.2. At $t = y$ there are also four unstable branches, but between $t = x$ and $t = y$ four branches have become extinct. The topology of the unstable portion of the tree at $t = y$ is also depicted in Figure 3.6.2. Note that the topology of the unstable portion of the tree at $t = y$ is not the same as the topology of the unstable portion of the tree at $t = x$, not because of differing branch lengths, but because the actual shape of the tree at those two times is different.

We shall finish this section with a short discussion of probability measures. As stated above, we think of the process as being a mapping from the space of realiza-

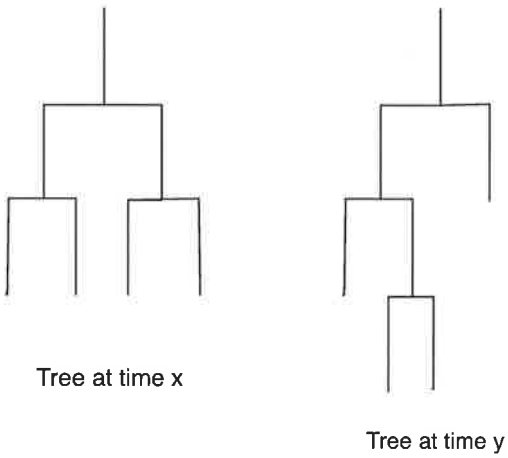


Figure 3.6.2: Same number of branches at different times does not mean topology is the same

tions, or tree histories at each $t \geq 0$, to the space of fixed topologies and that for each $t \geq 0$ there are uncountably infinitely many trees that map to a single topology. Hence, the probability that a realization of the process has topology $\mathcal{T}$ at time $t$ is given by the measure of the space of all trees of differing branch lengths that are mapped to $\mathcal{T}$ at time $t$. In the remainder of the thesis we denote this probability by $p(\mathcal{T}, t) = p(\mathcal{T}^a, t)$ and we say that this is the probability that a tree has topology $\mathcal{T}$ at time $t$. The probability that the unstable portion of a tree has topology $\mathcal{T}$ at time $t$ is denoted by $p(\mathcal{T}^u, t)$, and similar definitions apply for $p(\mathcal{T}^e, t)$ and $p(\mathcal{T}^q, t)$. In other words, because there are four different mappings from the space of tree histories of the process to the space of tree topologies there are also four different probability measures for the process. These measures are determined by what portion of the tree interests us, whether it be the entire tree, the extinct portion, the unstable portion or the quasi-stable portion. Thus the mapping from the space of tree histories to the space of topologies is determined by the measure that is being used. For example, we are interested in the measure for the unstable portion of the tree and its associated mapping, $\mathcal{M}^u$, then the probability that a tree will have topology $\mathcal{T}$ at time $t$ is $p(\mathcal{T}^u, t)$ and this is clearly not necessarily the same as the

probability that that same topology has under one of the different measures. It is important to also note that although the mapping removes all other branch types except for the ones of interest in calculating the actual measure itself, knowledge of the entire tree history remains important, and this will be evident when we discuss the multi-rate model in later sections.

The size of a topology, $\mathcal{T} \in \mathbb{T}$, is given by the number of leaf branches and is denoted by $\left|\mathcal{T}\right|$. So if $\left|\mathcal{T}\right| = s$ we say that $\mathcal{T}$ is of size $s$. Similarly,

- $\left|\mathcal{T}^a\right|$ is the size of the topology of the entire tree,
- $\left|\mathcal{T}^e\right|$ is the size of the topology of the extinct portion of a tree,
- $\left|\mathcal{T}^u\right|$ is the size of the topology of the unstable portion of a tree, and
- $\left|\mathcal{T}^q\right|$ is the size of the topology of the quasi-stable portion of a tree.

Denote the subset of $\mathbb{T}$ such that all $\mathcal{T}$ in this subset have $\left|\mathcal{T}\right| = s$ by $\mathbb{T}_s$. As before, this space remains invariant to the mapping used, so that exactly the same topologies are in $\mathbb{T}^s$ regardless of whether we are interested in $\mathcal{T}^a$, $\mathcal{T}^e$, $\mathcal{T}^u$ or $\mathcal{T}^q$.

For the remainder of this chapter let $z \in \{a, e, u, q\}$. Let $\mathbb{S}^z_{s,t}$ be the space of trees that at time $t$ are mapped by $\mathcal{M}^z$ to $\mathbb{T}_s$. The probability that a tree is in $\mathbb{T}_s$ at time $t$ is just given by the measure of the set $\mathbb{S}^z_{s,t}$, that is, the measure of the set of trees that are mapped by $\mathcal{M}^z$ to $\mathbb{T}_s$, which is denoted by $p(\left|\mathcal{T}^z\right| = s, t)$ for simplicity. Furthermore, $\lim_{t\to\infty} p(\left|\mathcal{T}^z\right| = s, t) = p(\left|\mathcal{T}^z\right| = s)$. We will also denote this probability by $p(\mathbb{T}^z_s, t)$, where we place the superscript $z$ on $\mathbb{T}_s$ to denote the fact that we calculate this under the measure that generates trees of type $z$.

At this point it is worthy of note that most of the analysis that is to be performed in this thesis is not a transient analysis, but instead the analysis is performed as $t \to \infty$. In this regime, if the branching process model that generates the trees is subcritical, then the tree consists, almost surely, of a finite number of extinct branches, or of a finite number of extinct and quasi-stable branches. The branch types that are present as $t \to \infty$ should be clear from the context.

## 3.7 Some Topological Concepts

Denote the set of trees that are topologically isomorphic to $\mathcal{T} \in \mathbb{T}$ by $\mathbb{F}(\mathcal{T})$. If $\mathcal{J} \in \mathbb{F}(\mathcal{T})$ then $\mathbb{F}(\mathcal{J}) = \mathbb{F}(\mathcal{T})$. If the daughter and parent subtrees at node $[\psi]$ are not from the same topologically isomorphic class, we say that node $[\psi]$ is an uneven node. Let the number of uneven nodes in a tree of topology $\mathcal{T}$ be denoted by $\epsilon_{\mathcal{T}}$. It is clear that $\epsilon_{\mathcal{T}}$ can be calculated [26] by

$$\epsilon_{\mathcal{T}} = I\{\mathbb{F}(\mathcal{T}_{[0,0]}) \neq \mathbb{F}(\mathcal{T}_{[0,1]})\} + \epsilon_{\mathcal{T}_{[0,0]}} + \epsilon_{\mathcal{T}_{[0,1]}}, \tag{3.7.1}$$

where $I\{A\}$ is the indicator function of the condition $A$, and if $\left|\mathcal{T}\right| = 1$ then, $\epsilon_{\mathcal{T}} = 0$.

If a tree of topology, $\mathcal{T}$, has $\epsilon_{\mathcal{T}}$ uneven nodes, then there are $2^{\epsilon_{\mathcal{T}}}$ trees that are topologically isomorphic to $\mathcal{T}$, that is, the set $\mathbb{F}(\mathcal{T})$ contains $2^{\epsilon_{\mathcal{T}}}$ trees. Figure 3.7.1 depicts the two distinct topologically isomorphic classes of size 4. Class 1 consists of one tree because there are no uneven nodes, whereas, there are four trees in the second class because there are two uneven nodes.

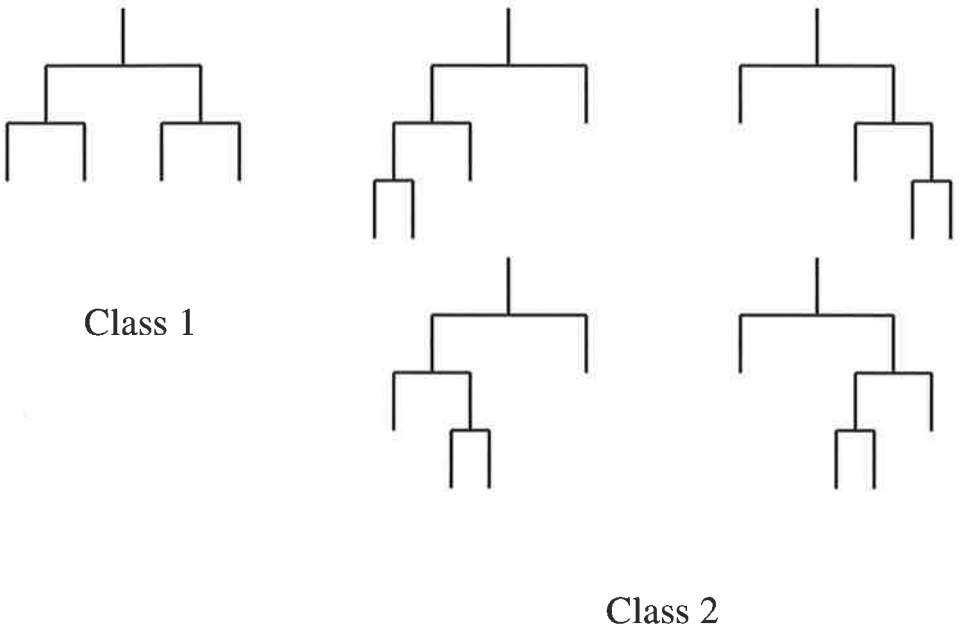


Figure 3.7.1: The two topologically isomorphic classes of size four.

The set of topologies of size $s$, $\mathbb{T}_s$, can be partitioned into its topologically

isomorphic classes. These classes are disjoint, and span all of $\mathbb{T}_s$. Thus, if we let $\mathbb{F}_{i,s}$ denote the $i$-th topologically isomorphic class, then $\mathbb{T}_s = \{\mathbb{F}_{1,s}, \ldots, \mathbb{F}_{T_s,s}\}$, where $T_s$ is the number of topologically isomorphic classes, and is given, for $s > 1$, by

$$T_s = \begin{cases} \frac{1}{2}\left(T_{s/2} + \sum_{i=1}^{s-1} T_i T_{s-i}\right) & \text{if } s \text{ is even,} \\ \frac{1}{2}\sum_{i=1}^{s-1} T_i T_{s-i} & \text{if } s \text{ is odd,} \end{cases} \tag{3.7.2}$$

with $T_1 = 1$; see [30] and references therein.

The topologically isomorphic set, $\mathbb{F}_{i,s}$ can be constructed by combining two topologically isomorphic sets, $\mathbb{F}_{j,l}$ and $\mathbb{F}_{k,s-l}$,

$$\mathbb{F}_{i,s} = \begin{cases} \left\{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{j,l}, \mathbb{F}_{k,s-l}\right\} \bigcup \left\{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{k,s-l}, \mathbb{F}_{j,l}\right\} & \text{if } \mathbb{F}_{j,l} \neq \mathbb{F}_{k,s-l}, \\ \left\{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{j,l}, \mathbb{F}_{k,s-l}\right\} & \text{if } \mathbb{F}_{j,l} = \mathbb{F}_{k,s-l}, \end{cases} \tag{3.7.3}$$

for some $j \in \{1, 2, \ldots, T_l\}$, and $k \in \{1, 2, \ldots, T_{s-l}\}$.

This concept of topologically isomorphic class is of course a purely topological concept. When the actual probability measure along with its associated mapping is important, we will label topologically isomorphic classes to reflect this. In other words, if the current measure and mapping generates trees of branch type $z$, we label the topologically isomorphic class, $\mathbb{F}_{i,s}$ by $\mathbb{F}^z_{i,s}$.

## 3.8 Colless's Index of Imbalance

As we have stated previously, there has been considerable research performed on gaining some understanding of the imbalance of phylogenetic trees through the process of macroevolution. Measures of imbalance have therefore gained a prominent place in the study of macroevolution [1, 11, 15, 22, 26] and the degree of imbalance of a topology can be quantified using a variety of indices, see for example, [15]. The one that has been most utilised is Colless's index of imbalance $I_c$ [4, 22]. Consider a topology, $\mathcal{T}$ of size $|\mathcal{T}|$. Colless's index, $I_c(\mathcal{T})$, for $\mathcal{T}$ is the total of the absolute value of the difference between the number of leaves of the daughter and parent

subtree at each and every node. Let the space of internal nodes of, $\mathcal{T}$, be denoted by $\mathbb{B}_{\mathcal{T}}$. If $[\psi] \in \mathbb{B}_{\mathcal{T}}$, then, as before, the daughter subtree of $[\psi]$ is denoted by $\mathcal{T}_{[\psi,0]}$ and the parental subtree of $[\psi]$ is denoted by $\mathcal{T}_{[\psi,1]}$. Colless's index of imbalance is defined to be

$$I_c(\mathcal{T}) = \sum_{\psi \in \mathbb{B}_{\mathcal{T}}} \left| |\mathcal{T}_{[\psi,0]}| - |\mathcal{T}_{[\psi,1]}| \right|. \tag{3.8.1}$$

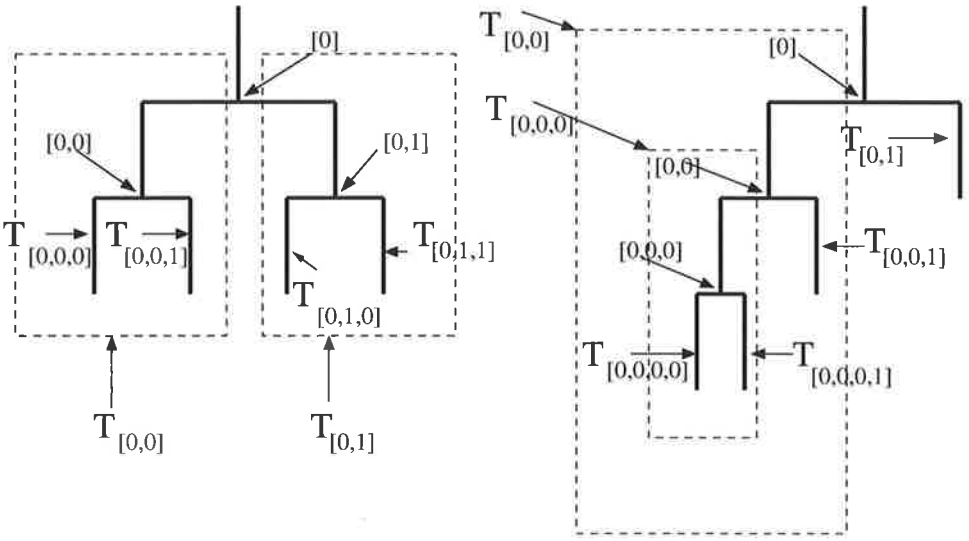


Figure 3.8.1: Colless's index of imbalance for two trees

We now calculate $I_c(\mathcal{T})$ for the two topologies depicted in Figure 3.8.1. The daughter and parent subtrees at each node are labelled for both trees. Colless' index of imbalance for the first tree is,

$$\begin{aligned} I_c(\mathcal{T}) &= \sum_{\psi \in \mathbb{B}_{\mathcal{T}}} \left| |\mathcal{T}_{[\psi,0]}| - |\mathcal{T}_{[\psi,1]}| \right| \\ &= \left| |\mathcal{T}_{[0,0]}| - |\mathcal{T}_{[0,1]}| \right| + \left| |\mathcal{T}_{[0,0,0]}| - |\mathcal{T}_{[0,0,1]}| \right| + \left| |\mathcal{T}_{[0,1,0]}| - |\mathcal{T}_{[0,1,1]}| \right| \\ &= \left| 2 - 2 \right| + \left| 1 - 1 \right| + \left| 1 - 1 \right| = 0, \end{aligned}$$

and for the second tree $I_{\mathcal{T}}$ is,

$$\begin{aligned} I_c(\mathcal{T}) &= \sum_{\psi \in \mathbb{B}_{\mathcal{T}}} \left| |\mathcal{T}_{[\psi,0]}| - |\mathcal{T}_{[\psi,1]}| \right| \\ &= \left| |\mathcal{T}_{[0,0]}| - |\mathcal{T}_{[0,1]}| \right| + \left| |\mathcal{T}_{[0,0,0]}| - |\mathcal{T}_{[0,0,1]}| \right| + \left| |\mathcal{T}_{[0,0,0,0]}| - |\mathcal{T}_{[0,0,0,1]}| \right| \\ &= \left|3-1\right| + \left|2-1\right| + \left|1-1\right| = 3. \end{aligned}$$

In fact, Colless' index of imbalance ranges from zero for the most balanced tree to $\big(|\mathcal{T}|-1\big)\big(|\mathcal{T}|-2\big)/2$ for the most imbalanced tree and so it can be normalised if one wishes to do so.

Colless's index of imbalance for trees can be calculated recursively from the values of the two lower order tree shapes, namely the parent and daughter subtrees at [0], thus,

$$I_c(\mathcal{T}) = I_c(\mathcal{T}_{[0,0]}) + I_c(\mathcal{T}_{[0,1]}) + |s-2l|, \tag{3.8.2}$$

where $I_c(\mathcal{T}_{[0,0]})$ and $I_c(\mathcal{T}_{[0,1]})$ are Colless's indices of imbalance for the daughter and parental subtrees at [0], with $|\mathcal{T}_{[0,0]}| = l$ and $|\mathcal{T}_{[0,1]}| = s-l$ leaf branches respectively.

Let $p(\mathcal{T}^z, t : |\mathcal{T}^z| = s) = p_s(\mathcal{T}^z, t)$ be the probability that a random tree has a topology $\mathcal{T}^z$ at time $t$ conditioned on it having size $s$. We write,

$$p_s(\mathcal{T}^z) = \lim_{t\to\infty} p_s(\mathcal{T}^z, t).$$

In addition, we denote the expected value of $I_c$ for trees of size $s$ by, $\mathbb{E}_s^z[I_c, t] = \mathbb{E}^z[I_c, t|s]$, and define it by,

$$\mathbb{E}_s^z[I_c, t] = \sum_{\mathcal{T}^z \in \mathbb{T}_s} I_c(\mathcal{T}^z) p_s(\mathcal{T}^z, t). \tag{3.8.3}$$

The value of $\mathbb{E}_s^z[I_c, t]$ is dependent on the stochastic model of macroevolution. Since we are mainly interested in the behaviour of the models as $t \to \infty$ let,

$$\mathbb{E}_s^z[I_c] = \lim_{t\to\infty} \mathbb{E}_s^z[I_c, t],$$

and so the expected value of Colless's index of imbalance becomes

$$\mathbb{E}_s^z[I_c] = \sum_{\mathcal{T}^z \in \mathbb{T}_s} I_c(\mathcal{T}^z) p_s(\mathcal{T}^z). \tag{3.8.4}$$

From a computational point of view, these expressions require a summation over all the possible topologies of a given size. We can simplify this expression by noting that $I_c(\mathcal{T}^z)$ is the same for all trees within a topologically isomorphic class. This is not difficult to see, when one considers that topologically isomorphic classes are related by rotations at uneven nodes, these rotations only interchange the daughter and parent subtrees and so do not affect $I_c(\mathcal{T}^z)$. Therefore the mean imbalance can be re-written in terms of topologically isomorphic classes. Let $I_c(\mathbb{F}_{i,s})$ denote the imbalance of each of the trees in $\mathbb{F}_{i,s}$, then we can re-write the mean imbalance as,

$$\mathbb{E}_s^z[I_c] = \sum_{i=1}^{T_s} I_c(\mathbb{F}_{i,s}^z) p_s(\mathbb{F}_{i,s}^z), \tag{3.8.5}$$

where, $p_s(\mathbb{F}_{i,s}^z)$ is the conditional probability as $t \to \infty$ that a tree is mapped to $\mathcal{T}^z \in \mathbb{F}_{i,s}^z$ given it has size $\left|\mathcal{T}^z\right| = s$. Written in this way, the sum is not over all topologies but over all topologically isomorphic classes of a given size and so is computationally more efficient.

## 3.9 Birth and Death Model

One of the simplest models of macroevolution is the birth-and-death model (BD). This model has been studied extensively in the probability literature in a wide range of contexts. In the context of macroevolution, it is a common assumption [1, 22, 26, 30] that the rates of speciation and extinction for all the branches are the same. In this case the model is called the equal rates Markov model [1, 8, 10, 26, 30]. However, we shall always refer to it as the constant-rates birth-and-death model (crBD) since we believe this to be a better description of the process. Under this assumption the qualitative evolution is such that any branch of the tree has probability $\lambda dt$ of giving birth to a daughter branch in a time interval $dt$ and probability $\mu dt$ of becoming extinct in a time interval $dt$, independently of the rest of the tree.

The crBD model has only two types of branch states, a branch is either unstable or extinct. We partially follow the analysis of this model as is given in [26, Appendix

A]. In [26, Appendix A] the topology of a tree is given by only the unstable portion of that tree; all the tree's extinct branches are therefore pruned. Let $p9\emptyset)$ be the probability that a time $t$ there are no unstable branches. Note the use of the empty set symbol to denote that such a tree has no topology since there are no unstable branches. The probability that a tree will be extinct by time $t$ is given by,

$$p(\emptyset,t)=\int_0^t \mu\exp\big(-(\lambda+\mu)x\big)dx+\int_0^t \lambda\exp\big(-(\lambda+\mu)x\big)p(\emptyset,t-x)p(\emptyset,t-x)dx. \tag{3.9.1}$$

This equation can easily be understood by noticing that a tree will become extinct if the parental branch, $([\alpha(0)],[0])$ becomes extinct within $(0,t]$ before undergoing any births, or that there is a birth at node $[0]$ at time $x$ and the daughter and parental subtrees both subsequently become extinct by time $t$.

Now the probability that no events occur by time $t$, and so the tree consists of only a single unstable branch $([\alpha(0)],[0])^{(u)}$, is given by, $\exp\big(-(\lambda+\mu)t\big)$. However, there is an extra term which is due to the application of the mapping $\mathcal{M}^u$, so that if any branch point occurs where one branch becomes extinct by time $t$ then it is pruned and therefore the tree retains a one unstable branch topology. Therefore the correct probability that a tree will have a topology consisting of only one unstable branch by time $t$ is given by,

$$p(\left|\mathcal{T}^u\right|=1,t)=\exp\big(-(\lambda+\mu)t\big)+\int_0^t \lambda\exp\big(-(\lambda+\mu)x\big)2p(\left|\mathcal{T}^u\right|=1,t-x)p(\emptyset,t-x)dx.$$

More generally, the probability that a tree commencing from $[\alpha(0)]$ will be mapped to a topology $\mathcal{T}^u$ that has $\left|\mathcal{T}^u\right|\geq 2$ at time $t$ is

$$\begin{aligned} p(\mathcal{T}^u,t) &= \int_0^t \lambda\exp\big(-(\lambda+\mu)x\big)\Big(2p(\mathcal{T}^u,t-x)p(\emptyset,t-x) \\ &\quad +p(\mathcal{T}^u_{[0,0]},t-x)p(\mathcal{T}^u_{[0,1]},t-x)\Big)dx, \end{aligned} \tag{3.9.2}$$

where $\mathcal{T}^u_{[0,0]}$ and $\mathcal{T}^u_{[0,1]}$ are the topologies of the daughter and parental subtrees commencing from node $[0]$. Equation (3.9.2) has the following interpretation: at time 0 the tree begins at, $[\alpha(0)]$, and in the interval $(x,x+dx)$ the root branch under-

goes a birth with probability $\lambda e^{-(\lambda+\mu)x}dx$. From this branch point there are two possibilities:

1. in the time interval $t-x$ one of the daughter or the parental branch can become extinct while the other evolves into a tree with topology $\mathcal{T}^u$, or

2. in the time interval $t-x$ the daughter branch eventually evolves into a subtree that has topology $\mathcal{T}^u_{[0,0]}$ and the parental branch evolves into a subtree that has topology $\mathcal{T}^u_{[0,1]}$. The daughter and parental subtrees evolve independently.

Due to the independence of the evolution of the parent and daughter branches in both cases, the probability of each of these two scenarios is just the product of the probability of each of the individual topologies, that is, $2p(\mathcal{T}^u, t-x)p(\emptyset, t-x)$ and $p(\mathcal{T}^u_{[0,0]}, t-x)p(\mathcal{T}^u_{[0,1]}, t-x)$. We then integrate $x$ from 0 to $t$ as the original branch point can occur at any time in that interval. Thus the probability of a tree having topology $\mathcal{T}^u$ at time $t$ can be determined recursively from the lower order tree probabilities using equation (3.9.2).

Harding [8] studied the case of the pure birth process (where $\mu = 0$) by considering the embedded process only at birth points. He showed the probability that a tree of size $s$ belongs to the topologically isomorphic class $\mathbb{F}(\mathcal{T}^u)$ is given by

$$p\big(\mathbb{F}(\mathcal{T}^u)\big) = \frac{2^{I\{\mathbb{F}(\mathcal{T}^u_{[0,0]})\neq\mathbb{F}(\mathcal{T}^u_{[0,1]})\}}}{s-1} p\big(\mathbb{F}(\mathcal{T}^u_{[0,0]})\big)p\big(\mathbb{F}(\mathcal{T}^u_{[0,1]})\big), \tag{3.9.3}$$

where $\mathbb{F}(\mathcal{T}^u_{[0,0]})$ and $\mathbb{F}(\mathcal{T}^u_{[0,1]})$ are the corresponding topologically isomorphic classes for the daughter and parent subtrees at node [0], respectively.

Pinelis [26, Appendix A] tried to prove that this relation still holds in the transient crBD model. To show that equation (3.9.3) was valid in this regime, he first wished to show by induction that a random tree, commencing with one branch, evolves into a topology $\mathcal{T}^u$ of size $\left|\mathcal{T}^u\right| = r$ by time $t$, with probability

$$p(\mathcal{T}^u, t) = \kappa(\mathcal{T}^u)p(\mathbb{T}^u_r, t). \tag{3.9.4}$$

Pinelis made the assumption that the factor $\kappa(\mathcal{T}^u)$ obeyed the following equation,

$$\kappa(\mathcal{T}^u) = \frac{\kappa(\mathcal{T}^u_{[0,0]})\kappa(\mathcal{T}^u_{[0,1]})}{\left|\mathcal{T}^u\right| - 1}, \tag{3.9.5}$$

with $\kappa(\mathcal{T}^u) = 1$ if $\left|\mathcal{T}^u\right| = 0$ or 1. Pinelis then assumed that equation (3.9.4) is valid for all topologies with sizes less than $s$. To show that equation (3.9.4) was valid for $\mathcal{T}^u$ of size $s$, Pinelis substituted $p(\mathcal{T}^u, t)$ from equation (3.9.4) into equation (3.9.2) to obtain,

$$\begin{aligned}
\kappa(\mathcal{T}^u)p(\mathbb{T}^u_s, t) &= \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big)\Big(2p(\mathcal{T}^u, t-x)p(\emptyset, t-x) \\
&\quad + p(\mathcal{T}^u_{[0,0]}, t-x)p(\mathcal{T}^u_{[0,1]}, t-x)\Big)dx \\
&= \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big)\Big(2\kappa(\mathcal{T}^u)p(\mathbb{T}^u_s, t-x)p(\emptyset, t-x) \\
&\quad + \kappa(\mathcal{T}^u_{[0,0]})p(\mathbb{T}^u_k, t-x)\kappa(\mathcal{T}^u_{[0,1]})p(\mathbb{T}^u_{s-k}, t-x)\Big)dx,
\end{aligned} \tag{3.9.6}$$

where the daughter, $\mathcal{T}^u_{[0,0]}$, and parent, $\mathcal{T}^u_{[0,1]}$, subtrees at node [0] are of sizes $k$ and $s-k$ respectively. The flaw in going from equation (3.9.2) to equation (3.9.6) is that in equation (3.9.6) Pinelis used

$$p(\mathcal{T}^u, t) = \kappa(\mathcal{T}^u)p(\mathbb{T}^u_s, t), \tag{3.9.7}$$

on the right hand side, which was exactly what he was trying to prove. He assumed it to be true on the right hand side in order to eventually show that it is true on the left hand side. It is our purpose to show rigorously that equation (3.9.4) is indeed the solution to equation (3.9.2).

**Theorem 1** *If, $p(\mathcal{T}^u, t)$, given by equation (3.9.4) is a solution to equation (3.9.2) then $\kappa(\mathcal{T}^u)$ must satisfy,*

$$\kappa(\mathcal{T}^u) = \frac{\kappa(\mathcal{T}^u_{[0,0]})\kappa(\mathcal{T}^u_{[0,1]})}{\left|\mathcal{T}^u\right| - 1}, \tag{3.9.8}$$

*with $\kappa(\mathcal{T}^u) = 1$ if $\left|\mathcal{T}^u\right| = 0, 1$.*

**Proof :** We begin the proof by assuming that $p(\mathcal{T}^u, t)$, given in equation (3.9.4) is a solution to equation (3.9.2) and then showing that $\kappa(\mathcal{T}^u)$ must satisfy equation (3.9.8). Consequently, substituting equation (3.9.4) into equation (3.9.2) gives,

$$\begin{aligned}\kappa(\mathcal{T}^u)p(\mathbb{T}^u_s,t) &= \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big)\Big(2\kappa(\mathcal{T}^u)p(\mathbb{T}^u_s,t-x)p(\emptyset,t-x)\\ &\quad +\kappa(\mathcal{T}^u_{[0,0]})p(\mathbb{T}^u_k,t-x)\kappa(\mathcal{T}^u_{[0,1]})p(\mathbb{T}^u_{s-k},t-x)\Big)dx.\end{aligned} \tag{3.9.9}$$

At this point we cannot simplify the right hand side any further.

As has been stated earlier, the probability that a random tree will have size $s$ at time $t$ is $p(\mathbb{T}^u_s, t)$. Hence by the same argument as that used to derive equation (3.9.2), we have

$$\begin{aligned}p(\mathbb{T}^u_s,t) &= \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big)\Big(2p(0,t-x)p(\mathbb{T}^u_s,t-x)\\ &\quad +\sum_{l=1}^{s-1} p(\mathbb{T}^u_l,t-x)p(\mathbb{T}^u_{s-l},t-x)\Big)dx.\end{aligned} \tag{3.9.10}$$

One can find the exact expressions for $p(\mathbb{T}^u_s, t)$ in [26, Appendix A], where the original derivation of these equations was given in [13]. Using these exact expressions in [26, Appendix A], one can easily deduce that the product of $p(\mathbb{T}^u_l, t-x)p(\mathbb{T}^u_{s-l}, t-x)$ is the same for all $l \in \{1, 2\ldots, s-1\}$. Consequently, we choose some $k \in \{1, 2\ldots, s-1\}$ and equation (3.9.10) becomes

$$\begin{aligned}p(\mathbb{T}^u_s,t) &= \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big)\Big(2p(0,t-x)p(\mathbb{T}^u_s,t-x)\\ &\quad +(s-1)p(\mathbb{T}^u_k t,-x)p(\mathbb{T}^u_{s-k},t-x)\Big)dx.\end{aligned} \tag{3.9.11}$$

Multiplying the above equation by $\kappa(\mathcal{T}^u)$ gives,

$$\begin{aligned}\kappa(\mathcal{T}^u)p(\mathbb{T}^u_s,t) &= \kappa(\mathcal{T}^u)\int_0^t \lambda\exp(-(\lambda+\mu)x)\Big(2p(0,t-x)p(\mathbb{T}^u_s,t-x)\\ &+ (s-1)p(\mathbb{T}^u_k,t-x)p(\mathbb{T}^u_{s-k},t-x)\Big)dx.\end{aligned}$$

If we subtract the above equation from both sides of equation (3.9.9) we obtain,

$$\int_0^t \lambda\exp\big(-(\lambda+\mu)x\big)p(\mathbb{T}^u_k,t-x)p(\mathbb{T}^u_{s-k},t-x)\Big(\kappa(\mathcal{T}_{[0,0]}u)\kappa(\mathcal{T}_{[0,1]}u)-\kappa(\mathcal{T}^u)(s-1)\Big)dx = 0. \tag{3.9.12}$$

Since the exponential function is always greater than zero, and $p(\mathbb{T}_k^u, t-x)p(\mathbb{T}_{s-k}^u, t-x)$ is non-zero, (otherwise only extinct trees would be possible), we have that

$$\kappa(\mathcal{T}_{[0,0]}^u)\kappa(\mathcal{T}_{[0,1]}^u) - \kappa(\mathcal{T}^u)(s-1) = 0. \tag{3.9.13}$$

Re-arranging the above equation then gives us equation (3.9.8). Furthermore, $\kappa(\mathcal{T}^u)$ is well defined, since

$$\mathcal{T}^u = \{([\alpha(0)], [0])^{(i)}, \mathcal{T}_{[0,0]}^u, \mathcal{T}_{[0,1]}^u\}, \tag{3.9.14}$$

is the unique representation of $\mathcal{T}^u$.

Finally, equation (3.9.2) is the integral form of the Kolmogorov backward equation, and therefore has a unique solution, so the solution to equation (3.9.2) given by equation (3.9.4) subject to equation (3.9.8) is the unique solution. ■

**Remark 1** *Suppose that $\mathcal{T}_1^u$ and $\mathcal{T}_2^u$ are in the same topologically isomorphic class. Interchanging the daughter subtree with the parental subtree and vice versa at an uneven node, say $\psi$, does not affect the product $\kappa(\mathcal{T}_{[\psi,0]}^u)\kappa(\mathcal{T}_{[\psi,1]}^u)$. Thus $\kappa(\mathcal{T}_1^u) = \kappa(\mathcal{T}_2^u)$ and therefore $p(\mathcal{T}_1^u, t) = p(\mathcal{T}_2^u, t)$. So all the trees in a topologically isomorphic class are equiprobable.*

**Corollary 2** *The probability of obtaining a topology from the topologically isomorphic class $\mathbb{F}_{i,s}^u$, conditional on the size of the tree being $s$, at time $t$, is given by*

$$p_s(\mathbb{F}_{i,s}^u, t) = 2^{I\{\mathbb{F}_{j,l}^u \neq \mathbb{F}_{k,s-l}^u\}} \frac{p_l(\mathbb{F}_{j,l}^u, t)p_{s-l}(\mathbb{F}_{k,s-l}^u, t)}{s-1}, \tag{3.9.15}$$

*where $p_r(\cdot, t)$ represents the conditional probability with respect to tree size $r$ and $l$ represents the size of the left-hand subtree.*

**Proof :** Let $\mathcal{T}^u \in \mathbb{F}_{i,s}^u$. The Remark immediately following the proof of Theorem 1 tells us that the probability of each topology in $\mathbb{F}_{i,s}^u$ is uniform, for all $i \in \{1, 2, \ldots, T_s\}$. Therefore, as there are $2^{\epsilon_{\mathcal{T}^u}}$ topologies in $\mathbb{F}_{i,s}^u$, the probability that a random tree is generated with a topology from $\mathbb{F}_{i,s}^u$ at time $t$ is given by,

$$p(\mathbb{F}_{i,s}^u, t) = 2^{\epsilon_{\mathcal{T}^u}}\kappa(\mathcal{T}^u)p(\mathbb{T}_s^u, t). \tag{3.9.16}$$

Now let $\kappa(\mathbb{F}^u_{i,s}) = 2^{\epsilon_{\mathcal{T}^u}}\kappa(\mathcal{T}^u)$, so equation (3.9.16) becomes,

$$p(\mathbb{F}^u_{i,s}, t) = \kappa(\mathbb{F}^u_{i,s})p(\mathbb{T}^u_s, t). \tag{3.9.17}$$

If we condition on trees of size $s$ at time $t$, we obtain,

$$\begin{aligned} p_s(\mathbb{F}^u_{i,s}, t) &= \frac{p(\mathbb{F}^u_{i,s}, t)}{p(\mathbb{T}^u_s, t)} \\ &= \kappa(\mathbb{F}^u_{i,s}). \end{aligned} \tag{3.9.18}$$

Suppose that $\mathcal{T}^u_{[0,0]} \in \mathbb{F}^u_{j,l}$ and $\mathcal{T}^u_{[0,1]} \in \mathbb{F}^u_{k,s-l}$, and that $\mathcal{T}^u_{[0,0]}$, $\mathcal{T}^u_{[0,1]}$ have $\epsilon_0$ and $\epsilon_1$ uneven branch points respectively. Then, $\mathbb{F}^u_{j,l}$ has $2^{\epsilon_0}$ topologies and $\mathbb{F}^u_{k,s-l}$ has $2^{\epsilon_1}$ topologies and so,

$$\kappa(\mathbb{F}^u_{j,l}) = 2^{\epsilon_0}\kappa(\mathcal{T}^u_{[0,0]}), \tag{3.9.19}$$

and

$$\kappa(\mathbb{F}^u_{k,s-l}) = 2^{\epsilon_1}\kappa(\mathcal{T}^u_{[0,1]}). \tag{3.9.20}$$

Substituting the above into equation (3.9.8) gives

$$\frac{\kappa(\mathbb{F}^u_{i,s})}{2^{\epsilon_{\mathcal{T}^u}}} = \frac{1}{2^{\epsilon_0+\epsilon_1}}\frac{\kappa(\mathbb{F}^u_{j,l})\kappa(\mathbb{F}^u_{k,s-l})}{s-1}. \tag{3.9.21}$$

However, recall that,

$$\epsilon_{\mathcal{T}^u} = I\{\mathbb{F}^u_{j,l} \neq \mathbb{F}^u_{k,s-l}\} + \epsilon_0 + \epsilon_1, \tag{3.9.22}$$

and so after some re-arrangement, equation (3.9.21) becomes,

$$\kappa(\mathbb{F}^u_{i,s}) = 2^{I\{\mathbb{F}^u_{j,l} \neq \mathbb{F}^u_{k,s-l}\}}\frac{\kappa(\mathbb{F}^u_{j,l})\kappa(\mathbb{F}^u_{k,s-l})}{s-1}. \tag{3.9.23}$$

Substituting equation (3.9.18) into the above equation we finally obtain

$$p_s(\mathbb{F}^u_{i,s}, t) = 2^{I\{\mathbb{F}^u_{j,l} \neq \mathbb{F}^u_{k,s-l}\}}\frac{p_l(\mathbb{F}^u_{j,l}, t)p_{s-l}(\mathbb{F}^u_{k,s-l}, t)}{s-1}, \tag{3.9.24}$$

and the corollary is proven. ∎

**Remark 2** *Note that equation (3.9.18) tells us that $p_r(\cdot, t)$ is actually independent of $t$.*

**Example 1** *Calculation of probabilities conditioned on trees of size 4.*

To calculate $p_s(\mathbb{F}^u_{i,s})$ we use equation (3.9.23) directly and then use equation (3.9.18) to identify $p_s(\mathbb{F}^u_{i,s})$. Denote the space of class 1 topologies (see Figure 3.7.1) by $\mathbb{F}^u_{1,4}$. The daughter and parental subtrees of this class are identical and we denote them by $\mathbb{F}^u_{1,2}$. Since the two subtrees come from the same topologically isomorphic class, we obtain from equation (3.9.23),

$$\kappa(\mathbb{F}^u_{1,4}) = \frac{\kappa(\mathbb{F}^u_{1,2})^2}{3}.$$

The topologically isomorphic class $\mathbb{F}^u_{1,2}$ has representation,

$$\mathbb{F}^u_{1,2} = \{([\alpha(0)],[0])^{(i)}, \mathbb{F}^u_{1,1}, \mathbb{F}^u_{1,1}\},$$

so,

$$\kappa(\mathbb{F}^u_{1,2}) = \kappa(\mathbb{F}^u_{1,1})^2 = 1.$$

Hence,

$$\kappa(\mathbb{F}^u_{1,4}) = 1/3,$$

and so by equation (3.9.18),

$$p_4(\mathbb{F}^u_{1,4}, t) = \frac{1}{3}.$$

The space of class 2 topologies of size four is denoted by $\mathbb{F}^u_{2,4}$, and has representation,

$$\mathbb{F}^u_{2,4} = \left\{\{([\alpha(0)],[0])^{(i)}, \mathbb{F}^u_{1,3}, \mathbb{F}^u_{1,1}\} \bigcup \{([\alpha(0)],[0])^{(i)}, \mathbb{F}^u_{1,1}, \mathbb{F}^u_{1,3}\}\right\}$$

Since we calculate $\kappa(\mathbb{F}^u_{2,4})$ recursively, we need to calculate $F^u_{1,3}$ first. The space $\mathbb{F}^u_{1,3}$ has representation

$$\mathbb{F}^u_{1,3} = \left\{\{([\alpha(0)],[0])^{(i)}, \mathbb{F}^u_{1,2}, \mathbb{F}^u_{1,1}\} \bigcup \{([\alpha(0)],[0])^{(i)}, \mathbb{F}^u_{1,1}, \mathbb{F}^u_{1,2}\}\right\}$$

so

$$\begin{aligned} \kappa(\mathbb{F}^u_{1,3}) &= 2\frac{\kappa(\mathbb{F}^u_{1,2})\kappa(\mathbb{F}^u_{1,1})}{2} \\ &= 2 \times \frac{1}{2} = 1, \end{aligned}$$

where in the second step we have used $\kappa(\mathbb{F}^u_{1,1}) = \kappa(\mathbb{F}^u_{1,2}) = 1$, as above. Therefore,

$$\kappa(\mathbb{F}^u_{2,4}) = 2\frac{\kappa(\mathbb{F}^u_{1,3})\kappa(\mathbb{F}^u_{1,1})}{3} = 2 \times \frac{1}{3} = \frac{2}{3}.$$

Thus,

$$p_4(\mathbb{F}^u_{2,4}, t) = \frac{2}{3}.$$

Clearly,

$$p_4(\mathbb{F}^u_{1,4}, t) + p_4(\mathbb{F}^u_{2,4}, t) = 1/3 + 2/3 = 1,$$

since there are only two classes of size four.

In Section 3.10 we discuss another simple macroevolutionary model, the proportional-to-distinguishable arrangements (PDA) model, which gives us quite different probability distributions.

## 3.10 Proportional-to-Distinguishable-Arrangements Model

The proportional-to-distinguishable arrangements (PDA) model is defined such that each distinguishable arrangement (DA) of the species of a tree of size $s$ is equally probable. A distinguishable arrangement of $s$ species is an assignment of labels on the $s$ leaf branches that is not equivalent to any other arrangement. Two arrangements are non-distinguishable if by a suitable permutation of the uneven nodes, the labels and topologies can be made identical. Figure 3.10.1 depicts four trees, trees 1, 2 and 4 are DAs, whereas trees 1 and 3 are not DAs. Under the PDA model each of these three distinguishable arrangements are equally likely.

The number of DAs of a given set of $s$ species that generate a phylogenetic tree that belongs to a particular topologically isomorphic class has a direct correspondence to the number of topologies within that class. It can be shown that the number of ways of relabeling a tree of topology $\mathcal{T}$ of size $s$ is given by, [26, Appendix

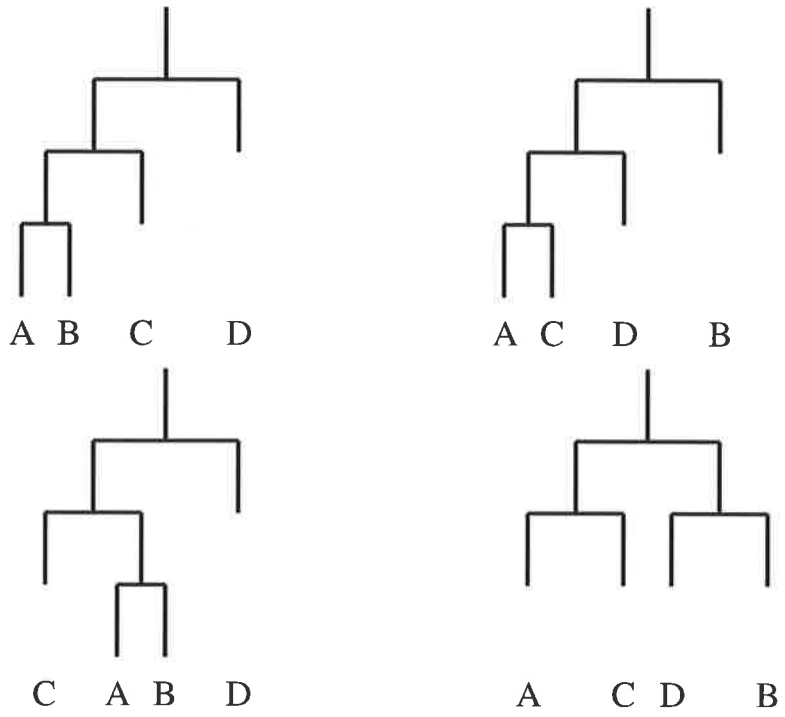


Figure 3.10.1: A distinguishable arrangements example

A],

$$a(\mathcal{T}) = \frac{s!}{2^{s-1}} 2^{\epsilon_{\mathcal{T}}}, \tag{3.10.1}$$

where $\epsilon_{\mathcal{T}}$ is the number of uneven branch points of the topology $\mathcal{T}$. This is just the total number of tree labelings, $s!$, divided by 2 to the power of the number of branch points that are not uneven, $2^{(s-1)-\epsilon_{\mathcal{T}}}$, since swapping labels between subtrees at those branch points do not yield distinct labelings. Interestingly, $N_{\mathbb{F}(\mathcal{T})} = 2^{\epsilon_{\mathcal{T}}}$ and so considering any other topology in $\mathbb{F}(\mathcal{T})$ does not add any more distinguishable arrangements and hence one can write,

$$a(\mathbb{F}(\mathcal{T})) = a(\mathcal{T}). \tag{3.10.2}$$

The above equation demonstrates the relationship between the number of DAs of a particular topology and the number of trees within that topologically isomorphic class.

**Example 2** *The number of DAs for trees of size four*

There are two topologically isomorphic classes of trees of size 4, see Figure 3.7.1. The trees from class 1 have no uneven branch points, thus the number of DAs of this class are given by,

$$a(\mathbb{F}_{1,4}) = \frac{4!}{2^3} = 3.$$

Trees from class 2 have two uneven branch points, hence the number of DAs for class 2 is given by

$$a(\mathbb{F}_{2,4}) = \frac{4!2^2}{2^3} = 12.$$

Thus the total number of DAs for trees of size 4 are 15.

We can use the above example to show how the probability distributions, conditioned on tree size differ between the crBD model and the PDA model. Now, as already stated above, each DA of a given size has an equal probability of occurring, so using the example above, we find that,

$$p_4(\mathbb{F}_{1,4}) = \frac{3}{15} = \frac{1}{5} \tag{3.10.3}$$

$$p_4(\mathbb{F}_{2,4}) = \frac{12}{15} = \frac{4}{5}, \tag{3.10.4}$$

since there are 15 DAs for size four. Comparing these values to those of the crBD model given in Section 3.9, where

$$p_4(\mathbb{F}^u_{1,4}, t) = \frac{1}{3} \tag{3.10.5}$$

$$p_4(\mathbb{F}^u_{2,4}, t) = \frac{2}{3}, \tag{3.10.6}$$

we find that the PDA model tends to allocate higher probabilities to topologically isomorphic classes with higher imbalances, because these classes in general have more uneven branch points and therefore more topologies than the more balanced classes. We see from this example that, under the PDA model, the probability of each topology of size four is the same, namely 1/5.

The number of topologies of size $s$, $N_s$, is given by,

$$N_s = \sum_{i=1}^{s-1} N_i N_{s-i}, \tag{3.10.7}$$

with $N_1 = 1$. It is easy to prove this using induction. The first few terms of the series are,

$$\begin{aligned} N_2 &= N_1 N_1 = 1 \\ N_3 &= N_1 N_2 + N_2 N_1 = 2 \\ N_4 &= N_1 N_3 + N_2 N_2 + N_3 N_1 = 2 + 1 + 2 = 5 \text{ and} \\ N_5 &= N_1 N_4 + N_2 N_3 + N_3 N_2 + N_4 N_1 = 5 + 2 + 2 + 5 = 14. \end{aligned}$$

**Theorem 3**

$$N(x) = \sum_{n=1}^{\infty} N_s x^s = \frac{1}{2}\left(1 - \sqrt{(1-4x)}\right), \tag{3.10.8}$$

*for* $0 \leq x \leq 1/4$.

**Proof :** To prove equation (3.10.8) we first multiply equation (3.10.7) by $x^s$ then sum from 2 to infinity and by noting the boundary condition $N_1 = 1$ to obtain,

$$\begin{aligned} N(x) &= x + \sum_{s=2}^{\infty} \sum_{i=1}^{s-1} N_i N_{s-I} x^s \\ &= x + \sum_{s=2}^{\infty} \sum_{i=1}^{s-1} N_i x^i N_{s-I} x^{s-i}. \end{aligned} \tag{3.10.9}$$

The order of summation in equation (3.10.9) can be swapped to obtain,

$$N(x) = x + \sum_{i=1}^{\infty} N_i x^i \sum_{s=i+1}^{\infty} N_{s-I} x^{s-i}. \tag{3.10.10}$$

One can perform a change of variables in the second summation of equation (3.10.10) with the result that

$$N(x) = x + \sum_{i=1}^{\infty} N_i x^i \sum_{s=1}^{\infty} N_s x^s = N(x) = x + N^2(x). \tag{3.10.11}$$

One can solve equation (3.10.11) using the quadratic formula to obtain,

$$N(x) = \frac{1 \pm \sqrt{(1-4x)}}{2} \tag{3.10.12}$$

Of these two solutions we choose

$$N(x) = \frac{1 - \sqrt{(1-4x)}}{2}, \tag{3.10.13}$$

since we do not want $N(x) > 1$. ■

**Corollary 4**

$$N_s = \frac{(2s-1)!!}{s!}2^{s-1}, s > 1. \tag{3.10.14}$$

**Proof :** Expanding equation (3.10.8) using a Taylor Series about $x = 0$ and applying induction yields equation (3.10.14). ■

More generally then, if we condition on a tree size of $s$, the probability, under the PDA model, of any one particular topology is $1/N_s$, independent of the actual topology. Thus, if there are more topologies in a particular topologically isomorphic class, the higher the probability that a random tree will be generated that has a topology from that class. Let $N_{\mathbb{F}_{i,s}}$ denote the number of topologies in $\mathbb{F}_{i,s}$, then,

$$p_s(\mathbb{F}_{i,s}) = N_{\mathbb{F}_{i,s}}/N_s,$$

and so the more the topologies in $N_{\mathbb{F}_{i,s}}$ the higher $p_s(\mathbb{F}_{i,s})$. Note that in [22] all the trees in $\mathbb{F}_{i,s}$ are considered to be the same topology. Consqeuently, in [22] $p_s(\mathbb{F}_{i,s})$ is the probability that a random tree has the topology represented by the trees in $\mathbb{F}_{i,s}$.

Defined in this manner, the trees in the PDA model do not grow under some stochastic dynamics. Attempts have been made at giving the PDA model an "evolutionary" explanation; evolutionary in the sense of the temporal evolution of a stochastic process [26, 34]. In this section we shall give an alternative model that is simpler than that in [26], with the details given in [26, Appendix A].

To do so, we would like to understand the connection, if any, between the the crBD model and the PDA model. At first glance it appears that they have no relationship to each other, however this is not the case. The transient crBD model of Section 3.9 gave us expressions for the probability that a random tree was generated with a specific topology at time $t$. These topologies consisted entirely of unstable branches; all extinct branches were pruned from the tree. If we consider the asymptotic version, that is, as $t \to \infty$, in the subcritical crBD case trees are generated

with only extinct branches and so from this model we obtain an entirely different set of results for tree shape probabilities. Indeed, the limit of $p_s(\mathcal{T}^e, t)$ in the subcritical crBD model as $t \to \infty$ will be shown to be identical to $p_s(\mathcal{T})$ from the PDA model, for the same underlying topology $\mathcal{T}$.

The probability, $p(|\mathcal{T}^e| = 0)$, that a tree has no extinct branches as $t \to \infty$ is clearly zero, since in the subcritical regime trees with no extinct branches exist only on a set of measure zero. The probability that a tree has topology, $\mathcal{T}^e = ([\alpha(0)], [0])^{(e)}$, is given by,

$$\begin{aligned} p\big(([\alpha(0)], [0])^{(e)}\big) &= \lim_{t\to\infty} \int_0^t \mu \exp\big(-(\lambda+\mu)x\big)dx \\ &= \lim_{t\to\infty} \frac{\mu}{\lambda+\mu}\Big(1 - \exp\big(-(\lambda+\mu)t\big)\Big) \\ &= \frac{\mu}{\lambda+\mu}. \end{aligned} \tag{3.10.15}$$

The probability that at time $t$ a random tree will evolve into a tree with topology $\mathcal{T}^e$ with $|\mathcal{T}^e| \geq 2$ is given by,

$$p(\mathcal{T}^e, t) = \int_0^t \lambda \exp\big(-(\lambda+\mu)x\big) p(\mathcal{T}^e_{[0,0]}, t-x) p(\mathcal{T}^e_{[0,1]}, t-x)dx. \tag{3.10.16}$$

This can be explained as follows. At time 0 the tree begins with a single branch which then undergoes a birth within the interval $(x, x+dx)$ with probability $\lambda \exp\big(-(\lambda+\mu)x\big)dx$. In the interval $(x, t]$ the daughter and parental subtrees at node [0] evolve into trees with topologies $\mathcal{T}^e_{[0,0]}$ and $\mathcal{T}^e_{[0,1]}$ respectively. Since they evolve independently, the probability of this evolution is given by the product of the probabilities of the daughter and the parental subtrees. Finally, because the original branch point can occur anywhere in the interval $(0, t]$ we integrate $x$ over that interval.

**Theorem 5** *For the subcritical crBD model, a random tree eventually evolves to a topology, $\mathcal{T}^e$ of size $s$, with probability,*

$$p(\mathcal{T}^e) = \frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}}, \tag{3.10.17}$$

*in the limit as $t \to \infty$.*

**Proof :** We shall prove equation (3.10.17) using induction. It is clearly true for $s = 1$ because from equation (3.10.15) we have

$$p(\mathcal{T}^e) = \frac{\mu}{\lambda + \mu}. \tag{3.10.18}$$

We shall now define $\bar{p}(\mathcal{T}^e, t)$ for all $t$, such that,

$$\bar{p}(\mathcal{T}^e, t) = \begin{cases} p(\mathcal{T}^e, t) & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases} \tag{3.10.19}$$

Suppose now that equation (3.10.17) is true for all $l \leq s$. Note that for a topology, $\mathcal{T}^e$ of size $s+1$, the daughter subtree $\mathcal{T}^e_{[0,0]}$ is of size $1 \leq l \leq s$ and the parental subtree, $\mathcal{T}^e_{[0,1]}$ is of size $s + 1 - l$. Now taking the limit of the above equation as $t \to \infty$ we have,

$$\bar{p}(\mathcal{T}^e) = \lim_{t\to\infty} \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big)\bar{p}(\mathcal{T}^e_{[0,0]}, t-x)\bar{p}(\mathcal{T}^e_{[0,1]}, t-x)dx. \tag{3.10.20}$$

The functions $\bar{p}(\mathcal{X}^e, t-x)$ are bounded on any compact set, and their limit as $t \to \infty$ exists. In fact,

$$\lim_{t\to\infty} \bar{p}(\mathcal{X}^e, t-x) = \lim_{t\to\infty} p(\mathcal{X}^e, t-x),$$

since $t - x > 0$ and so,

$$\bar{p}(\mathcal{X}^e) = \lim_{t\to\infty} \bar{p}(\mathcal{X}^e, t-x) = \lim_{t\to\infty} p(\mathcal{X}^e, t-x) = \frac{\lambda^{l-1}\mu^l}{(\lambda+\mu)^{2l-1}},$$

for $|\mathcal{X}^e| = l \leq s$, by the induction hypothesis. Therefore the Dominated Convergence Theorem implies that

$$\begin{aligned}
\bar{p}(\mathcal{T}^e) &= \lim_{t\to\infty} \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big)\bar{p}(\mathcal{T}^e_{[0,0]}, t-x)\bar{p}(\mathcal{T}^e_{[0,1]}, t-x)dx \\
&= \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big) \lim_{t\to\infty} \bar{p}(\mathcal{T}^e_{[0,0]}, t-x)\bar{p}(\mathcal{T}^e_{[0,1]}, t-x)dx \\
&= \bar{p}(\mathcal{T}^e_{[0,0]})\bar{p}(\mathcal{T}^e_{[0,1]}) \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big)dx \\
&= \frac{\lambda^{l-1}\mu^l}{(\lambda+\mu)^{2l-1}} \frac{\lambda^{s-l}\mu^{s+1-l}}{(\lambda+\mu)^{2(s+1-l)-1}} \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big)dx \\
&= \frac{\lambda^{s-1}\mu^{s+1}}{(\lambda+\mu)^{2s}} \int_0^\infty \lambda \exp\big(-(\lambda+\mu)x\big)dx.
\end{aligned} \tag{3.10.21}$$

where in the fourth step we used the induction hypothesis as both trees are of size at most $s$. Performing the integration yields

$$p(\mathcal{T}^e_{s+1}) = \frac{\lambda^s \mu^{s+1}}{(\lambda+\mu)^{2(s+1)-1}}, \tag{3.10.22}$$

as required. ■

**Remark 3** *The above theorem shows that the probability of a random tree in the subcritical crBD process, evolving to a topology, $\mathcal{T}^e$, of size $s+1$ is dependent only on its size and not on its topology as $t \to \infty$. This is in contrast to the transient model, where the probability that a random tree evolves to a topology, $\mathcal{T}^u$ by time $t$, depends on $\kappa(\mathcal{T}^u)$.*

**Theorem 6**

$$\sum_{s=1}^{\infty} \frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}} N_s = 1. \tag{3.10.23}$$

**Proof :** Noting that

$$\frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}} = \frac{\lambda+\mu}{\lambda}\left(\frac{\lambda\mu}{(\lambda+\mu)^2}\right)^s,$$

and writing,

$$\begin{aligned}\sum_{s=1}^{\infty} \frac{\lambda+\mu}{\lambda}\left(\frac{\lambda\mu}{(\lambda+\mu)^2}\right)^s N_s &= \frac{\lambda+\mu}{\lambda}\sum_{s=1}^{\infty}\left(\frac{\lambda\mu}{(\lambda+\mu)^2}\right)^s N_s \\ &= \frac{\lambda+\mu}{2\lambda}\left(1-\left(1-4\frac{\lambda\mu}{(\lambda+\mu)^2}\right)^{1/2}\right) \end{aligned} \tag{3.10.24}$$

where equation (3.10.8) was used in the third step. After some simple algebra equation (3.10.24) can be shown to be equal to one thus proving the theorem. ■

**Corollary 7** *The the limit as $t \to \infty$ of the subcritical crBD model conditioned on tree size generates the PDA model as $t \to \infty$.*

**Proof :** The probability, using the crBD model, that a random tree eventually evolves into a topology, $\mathcal{T}^e$ of size $s$ is given by,

$$p(\mathcal{T}^e) = \frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}}.$$

Let the number of topologies of a given size $s$ be $N_s$. Now, since the probability of any topology depends only on size in the subcritical crBD model as $t \to \infty$, we have that

$$p(\mathbb{T}_s^e) = N_s \frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}}.$$

Finally, given that the tree is of size $s$, the probability of obtaining a particular topology, $\mathcal{T}^e$, $p_s(\mathcal{T}^e)$, is

$$\begin{aligned} p_s(\mathcal{T}^e) &= \frac{p(\mathcal{T}^e)}{p(\mathbb{T}_s^e)} \\ &= \frac{\frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}}}{N_s \frac{\lambda^{s-1}\mu^s}{(\lambda+\mu)^{2s-1}}} \\ &= \frac{1}{N_s}. \end{aligned}$$

The above expression is exactly the probability of a random tree having any topology, conditioned on tree size, in the PDA model. ■

The crBD interpretation of the PDA model considers extinct trees; it gives more weighting to less balanced topologies, in comparison to the transient crBD model that considers unstable branches[1]. Topologies that are less balanced belong to topologically isomorphic classes that have a greater number of topologies, since there are a higher number of uneven branch points in these topologies. Therefore, the mean of Colless's imbalance measure is higher in the PDA model than in the transient crBD model. It has been found in many studies (see [22] and references therein) that the actual imbalance of real phylogenetic trees lies somewhere between these two classes of models.

[1]It is for this, and other reasons that we had to define the various mappings, topologies and probability distributions in Section 3.6.

The transient crBD model and the PDA model have been the most extensively studied in the literature. However, the fact that they predict mean imbalances, based on Colless's measure, that are too low and too high with respect to real phylogenetic trees, respectively, [10, 11, 12, 22, 30] suggests that these models are not adequate probability measures for macroevolutionary studies.

## 3.11 Multi-Rate Evolutionary Model

Recall that in [22] it was stated that "most biological taxa have arisen by a branching process of descent with modification", suggesting that a multi-type branching process should be used to generate phylogenetic trees. Aldous [1] also proposed that the multi-type branching process could be used as a model. In fact, since it is generally believed that "*except for mass extinctions and their aftermath, the overall number of species do not tend to increase or decrease exponentially fast*" [1], ct-MMTBPs that are close to criticality might make reasonable models. Pinelis [26] proposed a model called the multi-rate (MR) model and used it to show that the crBD and PDA models, under some fairly stringent conditions, are sub-classes of the MR model. The MR model is a binary-branch point continuous-time Markovian multi-type branching process [2, 9] with some slight modifications that we discuss below.

More formally, Pinelis [26] considers a phase space $\mathcal{S} \subseteq \mathbb{Z}^+$ where each $i \in \mathcal{S}$ is considered to represent a phase or a state that a species can be in. The phases in $\mathcal{S}$ have the capacity to contain any amount of information, for example, size, genotype, geographical location, and behavioural patterns of the species. The tree evolves in the following qualitative manner. In the interval $t$ to $t + dt$ any species in phase $i \in \mathcal{S}$ may,

1. with probability $\mu_{ij}dt$ transform into another state $j$, or

2. with probability $\sigma_{ij}dt$, remain *unchanged* and give rise to one new species in

phase $j$, or finally,

3. with probability $1 - \sum_{j \in \mathcal{S}} (\mu_{ij} + \sigma_{ij})dt$ it does not undergo any change.

The transition rates $\mu_{ij}$ and $\sigma_{ij}$ for all $i, j \in \mathcal{S}$ are non-negative and are referred to as the transformation and speciation rates. The rates are assumed to be constant in time. Any species in pahse $i \in \mathcal{S}$ at time $t$ is evolving independently of all the existing species and of its history.

The flexibility of the multi-rate evolutionary model stems from the flexibility in modelling the state space and the transition rate structure. Pinelis [26] proposes one possible partition of the phase space $\mathcal{S}$ into three subsets, $\mathcal{D}$, $\mathcal{U}$ and $\mathcal{Q}$. Here the set $\mathcal{D}$ can be thought of as consisting of those species that are extinct, in other words for every species with phase $d \in \mathcal{D}$, $\sum_{j \in \mathcal{S}} (\mu_{dj} + \sigma_{dj}) = 0$. In addition the probability that a speciation event from any species in phase $i \in \mathcal{S}$ to a species in phase $d \in \mathcal{D}$, $\sigma_{id}$ is zero. The set $\mathcal{U}$ consists of the species that are classified as unstable, that is they have the capacity to transform and speciate. Thus if $i \in \mathcal{U}$ then $\sigma_{ij} > 0$ for at least one $j \in \mathcal{S}$. Finally, the set $\mathcal{Q}$ is the set of quasi-stable phases, that is, species in one of these phases cannot speciate but are also not extinct. Hence, for $q \in \mathcal{Q}$, $\sigma_{qi} = 0$ for all $i \in \mathcal{S}$, and $\mu_{qi} = 0$ for all $i \in \mathcal{U} \cup \mathcal{D}$. As a result, quasi-stable phases have the capacity to transform only to other quasi-stable phases. The quasi-stable phases are interpreted [26] as representing species that are the most adaptable. These species "wander" [26] the space of quasi-stable phases, changing their attributes in order to suite their current situation. For example, they may be changing their size, feeding patterns, and possibly their genotype to some extent in order to adapt to a world that is changing around them.

At this point a note must be made about the position of the parental and daughter branches at a branch point. Pinelis [26, Appendix A] has chosen the left branch to be the parental branch and the right branch to be the daughter branch. In contrast, we have chosen throughout this entire thesis to represent the daughter branch as the left branch and the parental branch as the right branch. This designation is

entirely arbitrary. However, we chose it ahead of the orientation employed by Pinelis [26] since it is a more natural choice when the Markovian binary tree is introduced and analysed in Chapters 5, 6 and 7.

### 3.11.1 The PDA Model as an MR Model

It is easy to see that the crBD model is a special case of the MR model with two states, the extinct state and the unstable state. What is less clear is whether the PDA model is also a special case of the MR model. Pinelis shows that the PDA model is indeed a special case of the MR model [26, Appendix A] by making some fairly stringent assumptions on his model in order to collapse it to a three phase model. However, the analysis of Pinelis [26, Appendix A] can be performed by assuming from the beginning that it is a three phase process, such that:

- Phase 0 is the phase of all extinct branches,
- Phase 1 is the phase of all quasi-stable branches, and
- Phase 2 is the phase of all unstable branches.

The phases do not serve as markers for a species as in the approach of Pinelis [26, Appendix A], but rather they represent the state that a branch may be in at any particular time.

The analysis below can be found in Pinelis [26, Appendix A]. The transition rates for the process are:

- transformation from phase 2 to phase 0 (represented as $2 \to 0$) which occurs with rate $d$,
- transformation from phase 2 to phase 1 ($2 \to 1$) which occurs with rate $q$,
- a birth from phase 2 with the daughter branch in phase 1 whilst the parental branch remains in phase 2 (represented as $2 \to 1, 2$) which occurs with rate $b\gamma$, and

- a birth from phase 2 generating a daughter branch in phase 2 whilst the parental branch remains in phase 2 ($2 \rightarrow 2, 2$) occurs with rate $b(1-\gamma)$,

where $\gamma \in [0,1]$ and $d+b+q=1$.

To generate an evolutionary model of the PDA assume that the process is sub-critical and $\gamma = 0$. We consider the mapping, $\mathcal{M}^q$, that maps the quasi-stable portion of a tree to some topology, $\mathcal{T}^q$ in the limit as $t \rightarrow \infty$. Thus the trees that we study are those that exist on the space of trees that are finite, almost surely, and consist of only quasi-stable branches.

The probability that a tree will eventually become extinct is given by,

$$p(\emptyset) = d + bp(\emptyset)^2. \tag{3.11.1}$$

Equation (3.11.1) has the obvious interpretation, a tree will become extinct either directly which occurs with probability $d/(d+b+q) = d$ or via a birth followed by the independent eventual extinction of the two subtrees and this occurs with probability $bp(\emptyset)^2$. Solving for $p(\emptyset)$, we obtain

$$p(\emptyset) = \frac{1-\sqrt{1-4bd}}{2b}, \tag{3.11.2}$$

since the other root is clearly greater than one. The equation that characterises the probability of a random tree obtaining a topology $\mathcal{T}^q$ under the MR-PDA model is

$$p(\mathcal{T}^q) = bp(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}) + 2bp(\emptyset)p(\mathcal{T}^q), \tag{3.11.3}$$

where $\mathcal{T}^q_{[0,0]}$ and $\mathcal{T}^q_{[0,1]}$ are the daughter and parental subtrees at node [0]. There are two ways of evolving to a tree with topology $\mathcal{T}^q$. The first term in equation (3.11.3) represents the pathway where there is a birth at node [0] and the daughter subtree evolves to topology $\mathcal{T}^q_{[0,0]}$ and the parental subtree evolves to topology $\mathcal{T}^q_{[0,1]}$ independently. The second term in equation (3.11.3) represents the pathway where a birth occurs at node [0] followed by the eventual extinction of either the parental or daughter subtree whilst the other evolves into a tree that has topology $\mathcal{T}^q$. Re-arranging equation (3.11.3) we obtain,

$$p(\mathcal{T}^q) = \frac{b}{1-2bp(\emptyset)} p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}). \tag{3.11.4}$$

It is worth taking some time to interpret equation (3.11.4) here. Because we have pruned all dead branches/subtrees, a branch point in which either the daughter or the parent subtrees becomes extinct is not treated as an actual branch point. Instead an actual branch point is one in which both the daughter and parental subtrees consist of at least one quasi-stable branch. The factor $1/(1-2bp(\emptyset))$ in equation (3.11.4) reflects this, and is interpreted as giving the expected number of false branch points before the first actual branch point occurs. It can be shown using induction that for a tree of topology $\mathcal{T}^q$ with $\left|\mathcal{T}^q\right| = s$, equation (3.11.4) becomes,

$$p(\mathcal{T}^q) = \left(\frac{b}{1-2bp(\emptyset)}\right)^{s-1} \left(p(1)\right)^s, \tag{3.11.5}$$

where $p(1)$ is the probability that a tree consists of a sole quasi-stable branch. Now $p(1)$ given by,

$$p(1) = q + 2bp(\emptyset)p(1), \tag{3.11.6}$$

since a tree that is of size one occurs if the parent branch directly transforms into a quasi-stable branch, or if after a birth either the daughter or parental subtree eventually becomes extinct and the other subtree is itself of size one. Equation (3.11.6), can be re-arranged to obtain as expected

$$p(1) = \frac{q}{1-2bp(\emptyset)}. \tag{3.11.7}$$

The PDA distribution can then be recovered from equation (3.11.5) by conditioning on the tree size since all topologies of size $s$ are equiprobable. The number of topologies of a given size, $s$, can be obtained using equation (3.10.7).

**Example 3** *Size four trees*

For size four trees there are 5 distinct topologies, four of which belong to topologically isomorphic class $\mathbb{F}^q_{2,4}$ and one of which belongs to class $\mathbb{F}^q_{1,4}$. For a given size,

equation (3.11.5) tells us that all topologies of that size are equally likely. Therefore,

$$\begin{aligned} p_4(\mathbb{F}^q_{2,s}) &= \frac{\sum_{\mathcal{T}^q \in \mathbb{F}^q_{2,4}} p(\mathcal{T}^q)}{\sum_{\mathcal{T}^q, |\mathcal{T}^q|=4} p(\mathcal{T}^q)} \\ &= \frac{4p(\mathcal{T}^q)}{5p(\mathcal{T}^q)} \\ &= \frac{4}{5}. \end{aligned} \tag{3.11.8}$$

A similar calculation shows that, $p(\mathbb{F}^q_{1,4}|s = 4) = 1/5$. More generally then, the probability of obtaining a topology from a particular topologically isomorphic class, $\mathbb{F}^q_{k,s}$, conditional on size $s$ trees is,

$$\begin{aligned} p_s(\mathbb{F}^q_{k,s}) &= \frac{\sum_{\mathcal{T}^q \in \mathbb{F}^q_{k,s}} p(\mathcal{T}^q)}{\sum_{\mathcal{T}^q, |\mathcal{T}^q|=s} p(\mathcal{T}^q)} \\ &= \frac{2^{\epsilon_{\mathcal{T}^q}}}{N_s}. \end{aligned} \tag{3.11.9}$$

In the next section the super-PDA (sPDA) model is discussed. In the sPDA model, random topologies with higher imbalances occur with higher probability.

### 3.11.2 The super-PDA Model

The purpose of this section is to create a model, similar to the PDA, which gives higher mean imbalances than the PDA. The model does this by giving higher weighting to unitary branch points, where a unitary branch point is a branch point in which the daughter subtree is of size 1. In order to define the sPDA model we use the MR model [26, Appendix A]. The MR formulation of the sPDA model, like the MR formulation of the PDA model, is a measure on the asymptotic trees that consist of only quasi-stable branches and the associated mapping $\mathcal{M}^q$. The equation for the probability that a tree eventually has a topology $\mathcal{T}^q$ of size $s$ is given by,

$$p(\mathcal{T}^q) = \left(1 + \frac{\gamma}{(1-\gamma)p(1)}\right)^v \left(\frac{b(1-\gamma)}{1-2b(1-\gamma)p(\emptyset)}\right)^{s-1} (p(1))^s, \tag{3.11.10}$$

where the factor, $\left(1 + \frac{\gamma}{(1-\gamma)p(1)}\right)$, is the weighting given to unitary branch points, the factor $\frac{b(1-\gamma)}{1-2b(1-\gamma)p(\emptyset)}$ is the probability of a branch point, and $p(1)$ is the probability

for a size 1 tree. We shall explain these terms in more detail in what follows. The dependence of the probability on the number of unitary branch points can be thus clearly seen. Thus when conditioning on the size of tree topologies, topologies with a higher number of unitary branch points have a higher likelihood of occurrence, with the completely unbalanced topology having the highest probability. It is for this reason that the mean of Colless's Index of imbalance is greater in the sPDA than in the PDA.

We now give the MR formulation of the sPDA model as given in [26, Appendix A]. In the MR formulation of the PDA model an unstable branch is unable to give birth to a branch that is born quasi-stable; this is reflected by the parameter $\gamma = 0$. However, in the MR formulation of the sPDA model the parameter $\gamma$ is now non-zero and so daughter branches can now be born quasi-stable. In the MR model for the sPDA (MR-sPDA), the probability of obtaining an extinct tree is given by

$$p(\emptyset) = d + b(1-\gamma)p^2(\emptyset), \tag{3.11.11}$$

and has a similar interpretation to equation (3.11.1). The equation for the probability of obtaining a tree of topology $\mathcal{T}^q$ is given by

$$p(\mathcal{T}^q) = b(1-\gamma)p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}) + 2bp(\mathcal{T}^q)p(\emptyset) + b\gamma p(\mathcal{T}^q_{[0,1]})I\{|\mathcal{T}^q_{[0,0]}| = 1\}, \tag{3.11.12}$$

where once again, $\mathcal{T}^q_{[0,0]}$ and $\mathcal{T}^q_{[0,1]}$ are the parent and daughter subtrees at node 0 and $I\{|\mathcal{T}^q_{[0,0]}| = 1\}$ is the indicator function of the event that the daughter topology consists of a single quasi-stable branch. In the MR-sPDA model, a random tree of topology $\mathcal{T}^q$ has three possible routes from which it can evolve. The first is via a non-unitary branch point at [0] such that the daughter and parental subtrees subsequently evolve into topologies $\mathcal{T}^q_{[0,0]}$ and $\mathcal{T}^q_{[0,1]}$ independently; the first term of equation (3.11.12) reflects this route. The second is via a non-unitary branch point at [0] such that either the daughter or the parental subtree eventually becomes extinct, whilst the other evolves into a tree of topology $\mathcal{T}^q$, this is reflected by the second term of equation (3.11.12). The third and final route is via a unitary branch

point at [0] such that the parental subtree then evolves to a tree with topology $\mathcal{T}^q_{[0,1]}$, this is reflected in term three of equation (3.11.12). However, the third term is non-zero only if the daughter subtree at node [0] of the random tree consists of only one quasi-stable branch. Equation (3.11.12) can be re-arranged to obtain,

$$\begin{aligned} p(\mathcal{T}^q) &= \left(\frac{b(1-\gamma)}{1-2b(1-\gamma)p(\emptyset)} + \right. \\ &\quad \left. I\{|\mathcal{T}^q_{[0,0]}| = 1\}\frac{b\gamma}{(1-2b(1-\gamma)p(\emptyset))p(|\mathcal{T}^q_{[0,0]}| = 1)}\right) p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}) \\ &= \left(1 + \frac{\gamma}{(1-\gamma)p(|\mathcal{T}^q_{[0,0]}| = 1)} I\{|\mathcal{T}^q_{[0,0]}| = 1\}\right) \beta p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}), \end{aligned} \tag{3.11.13}$$

where $\beta = \big(b(1-\gamma)\big)/\big(1-2b(1-\gamma)p(\emptyset)\big)$. Equation (3.11.13) can also be written in another way, namely,

$$p(\mathcal{T}^q) = \left(1 + \frac{\gamma}{(1-\gamma)p(1)}\right)^{I\{|\mathcal{T}^q_{[0,0]}|=1\}} \beta p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}), \tag{3.11.14}$$

where $p(1) = p(|\mathcal{T}^q_{[0,0]}| = 1)$ is the probability that a subtree consists of a single quasi-stable branch. The probability of a one branch subtree is

$$p(1) = q + 2b(1-\gamma)p(1)p(\emptyset) + b\gamma p(\emptyset). \tag{3.11.15}$$

This equation can be simply understood by noticing that a branch may become quasi-stable either directly, with probability $q$, or a birth may occur and then either the daughter or parental branch evolves into an extinct subtree whereas the other eventually becomes quasi-stable, and this happens with probability $2b(1-\gamma)p(1)p(\emptyset)$, since either the daughter or parent may become extinct, and finally, a birth may occur such that the daughter is born quasi-stable and the parent subsequently becomes extinct, which occurs with probability, $b\gamma p(\emptyset)$. We can re-arrange equation (3.11.15) to obtain

$$p(1) = \frac{q + b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)}. \tag{3.11.16}$$

This equation can be interpreted in a similar way to equation (3.11.4). Once again the effect of pruning all the extinct branches and hence subtrees at a branch point

means that we do not treat such branch points as actual branch points. In this case, what is occurring is that at each false branch point either the parental branch or the daughter branch is pruned, leaving us with one unstable branch. Since the process is sub-critical the unstable branch must eventually undergo either a direct transition to become quasi-stable or alternatively undergo a false branch point such that the parental branch becomes extinct and the daughter is born quasi-stable. The factor $1/\big(1-2b(1-\gamma)p(\emptyset)\big)$ reflects the fact that this occurs and is interpreted as being the expected number of false branch points before a transition that generates a quasi-stable branch occurs.

As with the PDA model one can use induction to show that equation (3.11.14) becomes

$$p(\mathcal{T}^q) = \left(1+\frac{\gamma}{(1-\gamma)p(1)}\right)^{v} \beta^{s-1}\big(p(1)\big)^s, \tag{3.11.17}$$

if the random tree is of size $s$, and where $v$ denotes the number of unitary splits in the topology $\mathcal{T}^q$. Equation (3.11.17) demonstrates the sPDA property, that is, that topologies with a larger number of unitary splits have a higher probability of occurrence.

The purpose of this chapter has been to describe an alternative state space for branching processes, a state space that lends itself to macroevolutionary modelling. We have also described a number of measures from the space of trees to the space of tree topologies and their associated mappings. We then followed this by detailing one of the most used measures of imbalance in phylogenetics research, Colless's Index. We then began to analyze a number of well known models, the crBD and the PDA model in particular. This chapter was concluded by an analysis of the most sophisticated model to date, the MR model of Pinelis, [26]. This provides the background for beginning our introduction and analysis of the Markovian binary tree, the model that we propose which is equally sophisticated and yet more versatile. To begin to understand the language with which we define the MBT, we need to first discuss some further background material. This is done in Chapter 4 where some important concepts in the area of Matrix-Analytic methods are introduced.

This then allows us in Chapter 5 to define the Markovian binary tree (MBT). In that chapter we consider the modelling flexibility of the MBT. In particular, we show that the PDA, sPDA and the general MR model can be written in terms of the MBT model.

# Chapter 4

# Matrix Analytic Methods: an Introduction

## 4.1 Introduction

The theory of matrix analytic methods forms the foundation of much of the work that is to follow in this thesis. We represent the binary-branch point continuous-time Markovian multi-type branching process (ctMMTBP) as a level-dependent quasi-birth-and-death process (QBD) and call it the Markovian binary tree (MBT). This representation, in contrast to the classical branching process representation, opens the door to the possibility of performing efficient numerical analysis and obtaining some useful measures with which to model phenomena such as macroevolution. In order to motivate the QBD we begin by discussing one of the simplest of all stochastic processes, the Poisson process. The Poisson process, with exponential inter-event times, has enjoyed enormous popularity in applied probability due to its simplicity and wide applicability. However, there are times where such a process is just not flexible enough to provide a useful model. Therefore, the need to develop more complex stochastic models that still retain a certain degree of the mathematical elegance of the exponential distribution, led to the development of

the phase-type distribution (PH). The PH-distribution was utilised to generate the phase-type renewal process from which the more general Markovian arrival process (MAP) was spawned. To obtain the MBT representation of the binary-branch point ctMMTBP, we embed the MAP into the branching process in order to generate a phase process on each living branch of the MBT. This allows us to re-interpret the transition structure of the binary-branch point ctMMTBP, a structure that is devoid of correlations between particle lifetime and the types of offspring that are spawned, to a transition structure that has the flexibility of allowing such correlations. This re-interpretation is based on distinguishing between transitions that are observed and transitions that are hidden.

The Poisson process has played a central role in modelling real world phenomena for many years. For example, it can be used to model radioactive decay, or the arrival of customers to a queue. Let $X(t)$ denote the random variable that counts the number of events that occur in a time interval $(0, t]$, where the arrivals follow a Poisson process. The state space for $X(\cdot)$ is $\{0\} \cup \mathbb{Z}^+$. The Poisson process owes its popularity to the fact that it possesses a number of useful properties:

1. For all $t_0 = 0 < t_1 < t_2 < \ldots < t_n < \ldots$, the random variables $X(t_{n+1}) - X(t_n)$ are independent,

2. the time interval between two successive events is exponentially distributed, and

3. the random variables $X(t+s) - X(s)$ have a Poisson distribution and depend only on the interval length, $t$.

The probability that an event occurs at a time $s \leq t$ is given by,

$$P[s \leq t] = 1 - \exp(-\lambda t), \tag{4.1.1}$$

where $1/\lambda$ is the mean of the process. The exponential distribution has probability density function given by

$$p(t) = \lambda \exp(-\lambda t). \tag{4.1.2}$$

Furthermore, it is easy to show that the exponential distribution is memoryless, that is,

$$P[s \leq t + t_0 | s > t_0] = P[s \leq t]. \tag{4.1.3}$$

The memoryless property lies at the heart of Markov processes and explains their success.

Section 4.2 studies the phase-type renewal process. In Section 4.3 the MAP is analysed and the correlations that may develop within the process are emphasised. Section 4.4 discusses the level-independent QBD and Section 4.5 discusses some algorithms that are of importance: the algorithm of Neuts, algorithm $U$ and the level-independent logarithmic reduction algorithm. The first two of these algorithms will be adapted to the MBT in Chapter 7. Section 4.6 introduces the level-dependent QBD. Finally, Section 4.7 discusses the level-dependent logarithmic reduction algorithm.

## 4.2 Phase-Type Renewal Processes

The ease with which the exponential distribution can be applied prompted a search for a generalisation that still retained many of its useful properties. The phase-type (PH) distribution provides the matrix generalisation of the exponential distribution. A phase type random variable is the time to absorption in state 0 of a Markov process on the phases $\{0, 1, \ldots, n\}$ and a phase type distribution is the distribution of such a random variable. An excellent introduction is given in [18] and [23].

The initial probability vector of the process is given by $(\tau_0, \boldsymbol{\tau})$ with $\tau_0 + \boldsymbol{\tau}\boldsymbol{e} = 1$, where $\boldsymbol{e}$ is a vector of ones of the appropriate dimension and $\boldsymbol{\tau}$ is a $1 \times n$ non-negative vector. The infinitesimal generator is given by

$$Q = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{d} & D_0 \end{bmatrix}, \tag{4.2.1}$$

where $\boldsymbol{d}$ is an $n \times 1$ vector and $D_0$ is an $n \times n$ matrix. Because $Q$ is the infinitesimal generator matrix, we require that the diagonal elements of $D_0$, $(D_0)_{ii}$, be strictly

negative for all $i \in \{1, \ldots, n\}$ and that $d_i \geq 0$ and $(D_0)_{ij} \geq 0$ for $1 \leq i \neq j \leq n$. Finally, we require that the process is conservative, that is,

$$D_0 \boldsymbol{e} + \boldsymbol{d} = \boldsymbol{0}. \tag{4.2.2}$$

As can be seen from the form of $Q$, once the Markov process enters phase 0 it can never exit phase 0.

**Definition 1** *The distribution of time $X(t)$ until the process is absorbed in the state 0 is a phase-type distribution and is denoted by $PH(\boldsymbol{\tau}, D_0)$.*

$PH(\boldsymbol{\tau}, D_0)$ has distribution function,

$$F(x) = 1 - \boldsymbol{\tau} \exp(D_0 x)\boldsymbol{e}. \tag{4.2.3}$$

It is clear then, that equation (4.2.3) provides the generalisation of the exponential distribution as given in equation (4.1.1). It is shown in [18] that absorption into phase 0 occurs from any phase $i \in \{1, \ldots, n\}$ with probability 1 if and only if the matrix $D_0$ is non-singular, that is, if $D_0$ is invertible. Then $(-D_0)^{-1}_{ij}$ is the expected total time spent in phase $j$ during the time until absorption, given an initial phase of $i$.

Consider the Markov process defined above that commences at time $t_0 = 0$ in a phase determined by the distribution $\boldsymbol{\tau}$. Instantaneously upon absorption at $t_1$, restart the Markov process by choosing a new phase from that same distribution, $\boldsymbol{\tau}$. The process then proceeds until absorption, which now occurs at time $t_2$ and from here restart it again according to that same distribution $\boldsymbol{\tau}$. The set $\{0 = t_0 < t_1 < t_2 < \ldots\}$ of reinitialised time points forms a renewal process. The inter-renewal distribution is given by $PH(\tau, D_0)$. We call this process a phase-type renewal process.

Let the above Markov process be denoted by $\{\phi(t) \big| t \geq 0\}$, where $\phi(t) \in \{1, 2, \ldots, n\}$. We call this Markov process the phase process. The infinitesimal generator of the phase process is given by,

$$D = D_0 + \boldsymbol{d} \cdot \boldsymbol{\tau}. \tag{4.2.4}$$

The form of $D$ indicates that there are two ways in which the process can move from phase $i$ to phase $j$, and they are, either

1. directly, $i \rightarrow j$ which occurs with probability $(D_0)_{ij}/(-D_0)_{ii}$, or

2. indirectly, via absorption in 0 from $i$ and then being immediately restarted in phase $j$, which occurs with probability $d_i\tau_j/(-D_0)_{ii}$.

Let $N(t)$ be the random variable that denotes the number of renewals that have occurred up to and including time $t$. The two-dimensional representation of the phase-type renewal process is given by $\{(N(t), \phi(t))|t \geq 0\}$ which is defined on the state space, $\{0\} \bigcup \mathbb{Z}^+ \times \{1, 2, \ldots, n\}$. The $Q$-matrix for the entire process is

$$Q = \begin{bmatrix} D_0 & \boldsymbol{d} \cdot \boldsymbol{\tau} & 0 & \ldots \\ 0 & D_0 & \boldsymbol{d} \cdot \boldsymbol{\tau} & \ldots \\ 0 & 0 & D_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.2.5}$$

The process can be partitioned into levels: if $l$ renewals have occurred then the process is said to be in level $l$, $\mathcal{L}(l)$, and the set of states at each level are

$$\mathcal{L}(l) = \{(l, 1), (l, 2), \ldots, (l, n)\},$$

for all $l \geq 1$.

The process is homogeneous with respect to level, so that the future evolution of the process does not depend on its current level.

The probability generating function of the number of events up to time $t$ in a phase-type renewal process is [18],

$$P(z, t) = \exp\big((D_0 + z\boldsymbol{d} \cdot \boldsymbol{\tau})t\big),$$

and the corresponding probability generating function for the Poisson process is

$$P(z, t) = \exp\big((z - 1)\lambda t\big).$$

As a result the phase type renewal-process can be thought of as providing the matrix generalisation of the Poisson process. The exponential inter-renewal distribution of the Poisson process has been replaced by the phase-type distribution of the phase-type renewal process.

In Section 4.3 we generalise the phase-type renewal process to the Markovian arrival process (MAP). In particular, we shall study the type of MAP called the transient MAP [19]. Transient MAPs, unlike ordinary MAPs, terminate after a finite number of arrivals, almost surely. For the remainder of this chapter and thesis we shall be dealing with transient MAPs and referring to them generically as MAPs.

## 4.3 Markovian Arrival Processes

The core process that governs the growth of a Markovian binary tree (MBT) is the Markovian arrival process (MAP) [20, 25]. The MAP is a continuous time Markov process with two dimensional representation $\{(N(t), \phi(t)) : t \in \mathbb{R}^+\}$, on the state space $\mathbb{Z}^+ \times \{0, 1, \ldots, n\}$, where $n$ is a finite integer. There are two types of transitions: hidden transitions and observable transitions. The random variable $N(t)$ counts the number of observable transitions that have occurred up to time $t$. The random variable $\phi(t)$, which denotes the phase of the process, evolves as a continuous-time Markov chain.

MAPs that have almost-surely infinitely-many points have been studied for a long time and were first introduced by Neuts [25]. A later paper [20] introduced a more economical notation and studied a number of other properties. In [19], Latouche, Remiche and Taylor extended the concept so that a MAP can have almost-surely finitely-many points, known as the transient MAP. These transient MAPs cease to evolve at some catastrophe time $T$. It is the class of transient MAPs that we shall use to generate MBTs.

The transition rate matrix for the MAP is

$$Q = \begin{bmatrix} D_0^* & D_1^* & 0 & 0 & \dots \\ 0 & D_0^* & D_1^* & 0 & \ddots \\ 0 & 0 & D_0^* & D_1^* & \ddots \\ 0 & 0 & 0 & D_0^* & \ddots \\ \dots & \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

where,

$$D_0^* = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{d} & D_0 \end{bmatrix}, \text{ and } D_1^* = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & D_1 \end{bmatrix},$$

and the matrices $D_0$, $D_1$ and $\boldsymbol{d}$ have the properties that, for all $i, j = 1, \dots, n$, $(D_0)_{ij} \geq 0$ if $i \neq j$, $(D_0)_{ii} < 0$, $(D_1)_{ij} \geq 0$ and $d_i \geq 0$ with at least one $k$ for which $d_k > 0$. It is assumed that the matrix $D_0 + D_1$ is irreducible. If this is not the case then there are redundant phases.

Using $\boldsymbol{e}$ to denote a vector of ones of the appropriate dimension, the matrices $D_0^*$ and $D_1^*$ are such that,

$$D_0^* \boldsymbol{e} + D_1^* \boldsymbol{e} = \mathbf{0},$$

which is equivalent to

$$D_0 \boldsymbol{e} + D_1 \boldsymbol{e} + \boldsymbol{d} = \mathbf{0}.$$

The $i$-th entry of the vector $\boldsymbol{d}$ contains the rate at which the process ceases to evolve when the phase process is in phase $i$. In the traditional formulation of the MAP, due to Neuts [25], $\boldsymbol{d} = \mathbf{0}$ and phase 0 is removed. The transient MAPs of Latouche, Remiche and Taylor [19] have $\boldsymbol{d} \geq \mathbf{0}$ componentwise with at least one $k$ for which $d_k > 0$.

Let $(d_0)_i = -(D_0)_{ii}$ for all $i \in \{1, 2, \dots, n\}$. Suppose that the current state of a MAP is $(m, i)$, that is, there have been $m$ observable transitions and the phase process is in phase $i$. The process will remain in this state for an exponentially distributed period of time with mean $1/(d_0)_i$. At the next transition there are three possibilities:

- for $j \neq i$, with probability $(D_0)_{ij}/(d_0)_i$ there will be a hidden transition to phase $j$ and so the new state is $(m, j)$.
- With probability $(D_1)_{ij}/(d_0)_i$ there will be an observable transition to phase $j$, so that the new state is $(m+1, j)$.
- Finally, with probability $d_i/(d_0)_i$ there will be a transition into phase 0. The new state of the process is then $(m, 0)$. We call such an event a catastrophe.

We model the occurrence of a catastrophe by saying that the process enters phase 0. Latouche, Remiche and Taylor [19] envisaged that this transition could be either observable or hidden. However it makes sense in our context for the catastrophic transition to always be hidden.

MAPs have the interesting property that the phase process can generate correlations between the time between two observable events and the phase of the process immediately after the second of those observable events. These correlations arise because the transition rates depend on the underlying phase. What interests us here, is the distribution of the phase immediately after an observable event. As a first step, we determine this distribution and then illustrate the correlations through a concrete example.

The probability $(P_h(t))_{ij}$ that, at time $t$, the phase of the process is $j$ and the intervening phase transitions are all hidden given that the process began in phase $i$ is given by

$$P_h(t) = \exp(D_0 t).$$

The density $(R_o(t))_{ij}$ that the first observable event occurs at time $t$ and changes the phase into phase $j$, given that the process began in phase $i$, is given in matrix form by

$$R_o(t) = \exp(D_0 t) D_1. \tag{4.3.1}$$

The probability, $(P_{obs}(t))_{ij}$, that the phase of the process is $j$ immediately after the first observable event, given that the phase process began in phase $i$ and the first

observable event occurs at time $t$, is

$$(P_{obs}(t))_{ij} = \frac{(R_o(t))_{ij}}{(R_o(t)\boldsymbol{e})_i}. \tag{4.3.2}$$

Equation (4.3.2) illustrates the fact that the distribution of phase immediately after the first observable event is dependent on the time $t$ of that first observable event.

**Example 4** *Four phase MAP*

Consider the MAP on phases $\{1, 2, 3, 4\}$ defined by

$$D_0 = \begin{bmatrix} -20.0000 & 0.0000 & 0.0000 & 0.1000 \\ 0.1000 & -20.0000 & 19.4000 & 0.1000 \\ 0.1000 & 19.0000 & -20.0000 & 0.1000 \\ 0.1000 & 0.1000 & 0.1000 & -20.0000 \end{bmatrix},$$

and,

$$D_1 = \begin{bmatrix} 0.0000 & 0.1000 & 0.0100 & 19.3000 \\ 0.0000 & 0.0100 & 0.1000 & 0.0000 \\ 0.1000 & 0.2000 & 0.1000 & 0.2000 \\ 15.0000 & 0.1000 & 0.1000 & 0.1000 \end{bmatrix}.$$

An easy calculation shows us that

$$P_{obs}(0.01) = \begin{bmatrix} 0.0008 & 0.0052 & 0.0045 & 0.9936 \\ 0.1249 & 0.1849 & 0.4542 & 0.2260 \\ 0.1766 & 0.3070 & 0.1806 & 0.3357 \\ 0.9791 & 0.0065 & 0.0065 & 0.0078 \end{bmatrix} \text{ and}$$

$$P_{obs}(0.1) = \begin{bmatrix} 0.0077 & 0.0052 & 0.0006 & 0.9866 \\ 0.2295 & 0.2063 & 0.1987 & 0.3655 \\ 0.2290 & 0.2099 & 0.1946 & 0.3665 \\ 0.9667 & 0.0069 & 0.0068 & 0.0195 \end{bmatrix}.$$

The most striking indication of the dependence of the phase distribution immediately after the observable event on the time of the event comes when one looks

at $(P_{obs}(0.01))_{23} = 0.4542$, $(P_{obs}(0.01))_{24} = 0.2260$ and $(P_{obs}(0.1))_{23} = 0.1987$, $(P_{obs}(0.1))_{24} = 0.3655$. Thus if the process begins in phase 2, the earlier the first observable event the more likely that the process will enter phase 3. In contrast, the longer the inter-observable event lifetime the more likely that the process will enter phase 4 at the first observable event. Consider the observable transition from phase 1 to phase 1 at $t = 0.01$ we have $P_{obs}(0.01)_{11} = 0.0008$ and at $t = 0.1$ we have $P_{obs}(0.1)_{11} = 0.0077$, the probability of such an event increases almost 10-fold in that space of time. The reason for this is that from phase 1 there cannot be a direct observable transition back into phase 1, so as time increases the probability that a hidden transition say to phase 4 can occur followed by an observable transition from phase 4 to phase 1 increases in likelihood.

Example 4 illustrates the important property that the lifetime distribution and the distribution of phase after an observable event are correlated. The level of correlation depends on the nature of the matrices $D_0$ and $D_1$. As will be seen in Chapter 5, by constructing the MBT using a MAP process on each branch, we can model correlations between the initial phase, branch lifetime and the initial phases of the offspring.

Let $P_{ij}(t)$ be the probability that there will be an observable transition to phase $j$ within the interval $[0, t]$ given that the process began in phase $i$ and $P(t) = [P_{ij}(t)]$. Then,

$$\begin{aligned} P(t) &= \int_0^t \exp(D_0 x) D_1 dx \\ &= (-D_0)^{-1}(I - \exp(D_0 t))D_1. \end{aligned} \tag{4.3.3}$$

Figure 4.3.1 graphs the time-dependent changes to the probability for an observable transition into the four phases given that the process began in phase 2 for the MAP from Example 4. This figure demonstrates quite clearly how the probability of an observable event occurring depends on time. For example, an observable transition from phase 2 to phase 3 is the most probable up until approximately time $t = 0.06$ at which point phase 4 becomes the most probable. Furthermore, up until

approximately $t = 0.11$ an observable transition into phase 1 is the least probable transition, because the transition $2 \rightarrow 1$ cannot occur directly but requires a number of hidden transitions, say a $2 \rightarrow 4$ transition followed by a $4 \rightarrow 1$ transition. The probability of these transitions increases over time thus increasing the probability of a $2 \rightarrow 1$ observable transition; the observable $2 \rightarrow 1$ transition becomes the second most likely observable transition as time increases. These observed behaviours are due to the complex internal hidden transitions that are generated by the $D_0$ matrix.

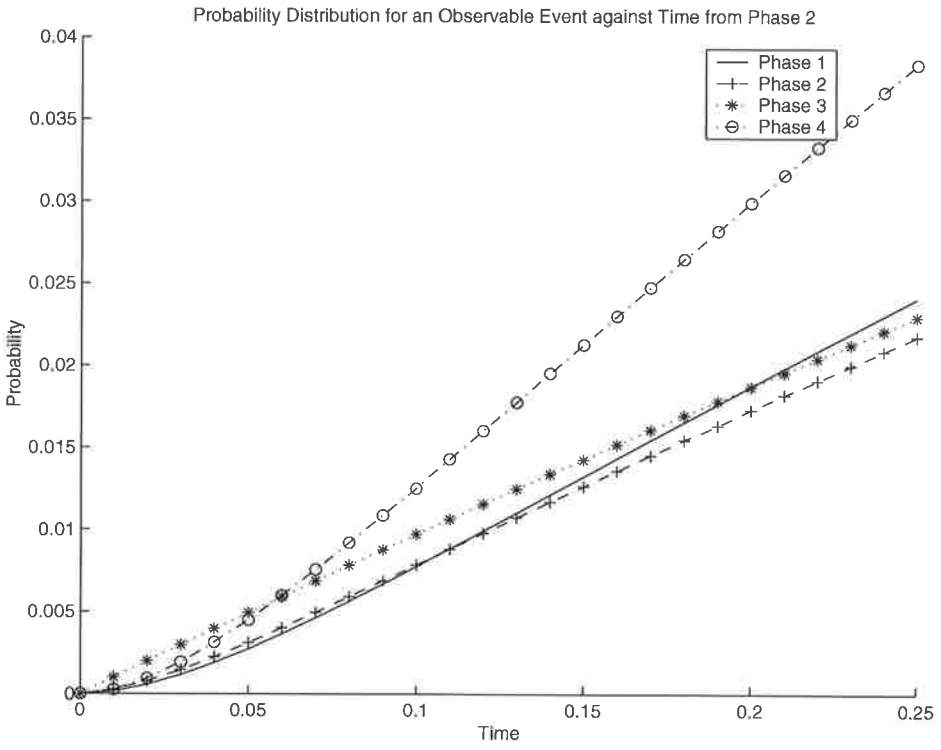


Figure 4.3.1: Probability distribution for an observable event against time given that the process began in phase 2

It is clear from equation (4.3.3) that the time until the next observable transition is not exponentially distributed. Let $T$ be the mean time until an observable event

occurs given that the process began in phase $i$, then

$$T = \int_0^\infty t \exp(D_0 t) D_1 \mathbf{e} dt. \tag{4.3.4}$$

After integration by parts this can be shown to be

$$T = (-D_0)^{-2} D_1 \mathbf{e}. \tag{4.3.5}$$

Once again referring to Example 4, the mean time between observable transitions conditional on the initial phase is,

$$T = \begin{bmatrix} 0.0492 \\ 0.8451 \\ 0.8367 \\ 0.0475 \end{bmatrix}. \tag{4.3.6}$$

This is what we would have expected since phases 1 and 4 have a high probability of triggering observable transitions at earlier times, whereas phases 2 and 3 are more likely to undergo more hidden transitions before an observable one.

Having introduced the MAP we now are ready to begin a brief exposition on quasi-birth-and-death processes (QBDs). The importance of the QBD in this work is that its provides an alternative framework to that of the branching process framework within which to define the Markovian binary tree. In particular, it will be seen that despite the fact that the QBD and the traditional branching process representations of the MBT refer to the same underlying process, the QBD interpretation confers to a process otherwise devoid of interesting lifetime-offspring correlations a much richer structure.

## 4.4 Level Independent Quasi-Birth-and-Death Processes

The level-independent QBD process is a Markov process, $\{X(t) \big| t \in \{0\} \cup \mathbb{R}^+\}$ on the two dimensional state space $\{(m,i) \big| m \in \{0\} \cup \mathbb{Z}^+,\ 1 \leq i \leq n\}$. If the process

is in state $(m, i)$ at some time $t$, we say that the process is currently in level $\mathcal{L}(m)$ and in phase $i$. The state space is therefore partitioned into levels, $\mathcal{L}(m)$, where, $\mathcal{L}(m) = \{(m,1),(m,2),\ldots,(m,n)\}$ for all $m \geq 0$. Consequently, for all $m$ there are exactly $n$ states in $\mathcal{L}(m)$. The infinitesimal generator of the process is given by

$$Q = \begin{bmatrix} B & A_0 & 0 & 0 & \ldots \\ A_2 & A_1 & A_0 & 0 & \ldots \\ 0 & A_2 & A_1 & A_0 & \ldots \\ 0 & 0 & A_2 & A_1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{4.4.1}$$

where $B$, $A_0$, $A_1$ and $A_2$ are all $n \times n$ matrices. All the elements of the matrices are non-negative except for $(A_1)_{ii}$ and $B_{ii}$ which are strictly negative for all $i \in \{1,2,\ldots,n\}$. In addition, the matrices $A_0$, $A_1$ and $A_2$ obey $(A_2 + A_1 + A_0)\boldsymbol{e} = \mathbf{0}$ where $\boldsymbol{e}$ is a vector of ones of the appropriate dimension. It is assumed that $A_2 + A_1 + A_0$ is irreducible.

Suppose that the process is currently in state $(m, i)$. At the next transition the process may either

- move down to $\mathcal{L}(m-1)$ by entering state $(m-1, j)$ with rate $(A_2)_{ij}$, or
- remain in the same level by entering state $(m, j)$, $j \neq i$, with rate $(A_1)_{ij}$, or finally,
- move up to $\mathcal{L}(m+1)$ by entering state $(m+1, j)$ with rate $(A_0)_{ij}$.

In keeping with the traditional literature we call transitions that move down a level, left transitions and transitions that move up a level, right transitions.

It is sometimes important to determine the probability of eventually moving from $\mathcal{L}(m)$ down to $\mathcal{L}(m-1)$. This probability is given by the matrix $G$. Let $\gamma(\ell) = \inf\{t > 0 \big| X(t) \in \mathcal{L}(\ell)\}$ be the first passage time into $\mathcal{L}(m)$. The probability that the process eventually reaches $\mathcal{L}(m-1)$ for the first time and does so in phase

$k$, given that it commenced in state $(m, i)$ is

$$G_{ik} = P[\gamma(m-1) < \infty \ \& \ X(\gamma(m-1)) = (m-1, k) \big| X(0) = (m, i)]. \tag{4.4.2}$$

Notice that since the process is independent of level, $G_{ik}$ does not depend on $m$. In a similar vein to [18] we shall adopt a slight abuse of notation and write,

$$G = P[\gamma(m-1) < \infty \ \& \ X(\gamma(m-1)) \big| X(0) \in \mathcal{L}(m)], \tag{4.4.3}$$

for equation (4.4.2) in order to "avoid storms of subscripts" [18], thereby making the equations easier to read. Since this process is homogeneous with respect to level, $G$ is independent of the level, $\mathcal{L}(m)$, from which the process commences, provided that $m > 0$.

It can be shown that $G$ is the minimal non-negative solution to the matrix quadratic equation [24]

$$F = (-A_1)^{-1}A_2 + (-A_1)^{-1}A_0F^2. \tag{4.4.4}$$

Since $G$ is the minimal non-negative solution to equation (4.4.4) we often write (4.4.4) as

$$G = (-A_1)^{-1}A_2 + (-A_1)^{-1}A_0G^2. \tag{4.4.5}$$

Equation (4.4.5) has a very neat physical interpretation. If the process begins in $\mathcal{L}(m)$ then there are two ways of moving down to $\mathcal{L}(m-1)$. The first way is a direct transition to $\mathcal{L}(m-1)$ which occurs with probability $(-A_1)^{-1}A_2$. In the second way, the process undergoes a right transition and moves up to $\mathcal{L}(m+1)$ with probability $(-A_1)^{-1}A_0$. From here the process eventually moves down to $\mathcal{L}(m-1)$ from $\mathcal{L}(m+1)$ by first moving to $\mathcal{L}(m)$ and then moving to $\mathcal{L}(m-1)$; each of these transitions has probability, $G$. Due to the independence of each of these three events, the probability of this second way is $(-A_1)^{-1}A_0G^2$.

By repeatedly substituting the left hand side into the right hand side, equation (4.4.5) can also be expressed as

$$G = \sum_{l=0}^{\infty} U^l(-A_1)^{-1}A_2 = (I-U)^{-1}(-A_1)^{-1}A_2, \tag{4.4.6}$$

where $U = (-A_1)^{-1}A_0G$. Equation (4.4.6) also has a neat probabilistic interpretation. The matrix $U$ is the probability of first return to $\mathcal{L}(m)$ under the taboo that the process does not go below $\mathcal{L}(m)$ given that it began in $\mathcal{L}(m)$. Mathematically,

$$U = P[\gamma(m) < \gamma(m-1) \ \& \ \gamma(m) < \infty \ \& \ X(\gamma(m)) \big| X(0) \in \mathcal{L}(m)]. \tag{4.4.7}$$

Similarly to $G$, the matrix $U$ does not depend on the initial level.

Each individual term, $U^l(-A_1)^{-1}A_2$, of equation (4.4.6) gives the probability that the process will return to $\mathcal{L}(m)$ $l$ times, before it eventually undergoes a left transition and enters $\mathcal{L}(m-1)$, given that it began in $\mathcal{L}(m)$. The total probability of eventually entering $\mathcal{L}(m-1)$ is given by the sum of these terms for all $l$.

In addition to $G$ being the minimal non-negative solution to equation (4.4.4) it can also be shown [18] that $G$ is the minimal non-negative solution to

$$F = \sum_{k=0}^{\infty} \left((-A_1)^{-1}A_0F\right)^k (-A_1)^{-1}A_2. \tag{4.4.8}$$

A number of important numerical schemes have been developed to solve for $G$, (see [7, 17, 16, 18, 24, 23, 27]). The two of most interest in this work are the Neuts algorithm which is also called the method of modified substitutions, [24, 23] and algorithm $U$ [18]. The Neuts algorithm, which is discussed in Section 4.5.1, is an iterative scheme based on equation (4.4.4) and algorithm $U$, which is discussed in Section 4.5.2, is an iterative scheme based on equation (4.4.8). These algorithms will form the basis of algorithms we develop for the MBT and then the more general Markovian tree in Chapters 7 and 8, respectively.

## 4.5 Level Independent Algorithms

### 4.5.1 The Algorithm of Neuts

The first algorithm that we discuss here is called the algorithm of Neuts, or the method of modified substitutions [24, 23]. The algorithm is developed by considering

equation (4.4.4),

$$F = (-A_1)^{-1}A_2 + (-A_1)^{-1}A_0F^2. \tag{4.5.1}$$

Neuts [24, 23] showed that the sequence of matrices defined by,

$$G(0) = (-A_1)^{-1}A_2 \tag{4.5.2}$$

$$G(l) = (-A_1)^{-1}A_2 + (-A_1)^{-1}A_0G^2(l-1), \tag{4.5.3}$$

for $l \geq 1$, are non-decreasing and converge to the minimal non-negative solution, $G$, of equation (4.5.1).

Consider the set $\mathbb{S}^l$, which contains all sample paths of the process that begin in $\mathcal{L}(m)$ and which eventually visit, $\mathcal{L}(m-1)$, such that the maximum level reached is $\mathcal{L}(m+l)$ with the added restriction that each of the sample paths has at most $2^l$ left transitions.

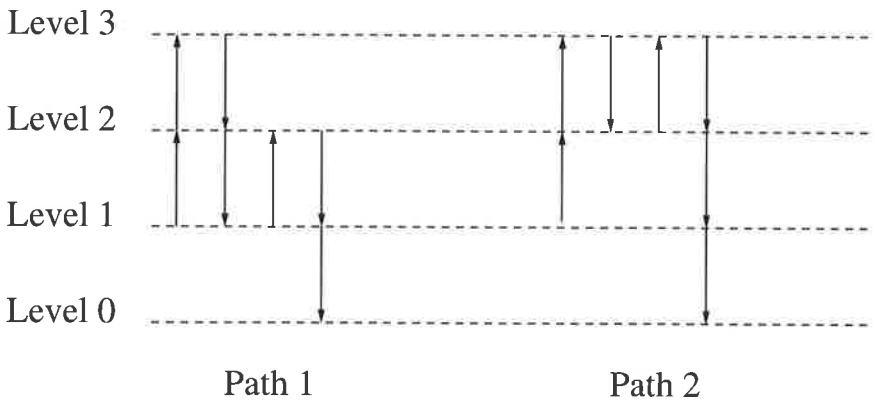


Figure 4.5.1: Two sample paths in $\mathbb{S}^2$

The Neuts algorithm at iteration $l$ only considers sample paths from the set $\mathbb{S}^l$. However, the space of sample paths included at the $l$-th iteration is a strict subset of this set. To illustrate this, Figure 4.5.1 depicts two sample paths from $\mathbb{S}^2$. Sample path 1 is included at the step $l = 2$ of the Neuts algorithm but sample path 2 is not, sample path 2 is considered at the next step of the algorithm. The space of sample paths included at each step of the algorithm is not easily described. We will show

in Chapter 7 a description is possible by a suitable transformation to the space of binary trees.

### 4.5.2 Algorithm $U$

Recall that $U$ gives the probability that the process will eventually return to $\mathcal{L}(m)$, given that it began in $\mathcal{L}(m)$, under the taboo that it does not move down to $\mathcal{L}(m-1)$, as expressed in equation (4.4.7), and $G$ is the probability that the process will eventually enter $\mathcal{L}(m-1)$ given that it began in $\mathcal{L}(m)$. Algorithm $U$ is defined as,

$$F(0) = (-A_1)^{-1}A_2 \tag{4.5.4}$$

$$M(l) = (-A_1)^{-1}A_0F(l-1) \tag{4.5.5}$$

$$F(l) = \big(I - M(l)\big)^{-1}(-A_1)^{-1}A_2, \tag{4.5.6}$$

for $l \geq 1$.

Now consider the sequences $\{U(l)\}$ and $\{G(l)\}$ for $l \geq 1$, defined by

$$U(l) = P[\gamma(m) < \gamma(m-1) \ \& \ \gamma(m) < \gamma(l+m+1) \ \& \ X(\gamma(m))|X(0) \in \mathcal{L}(m)],$$

and

$$G(l) = P[\gamma(m-1) < \gamma(m+l+1) \ \& \ X(\gamma(m-1))|X(0) \in \mathcal{L}(m)]. \tag{4.5.7}$$

The matrix $U(l)$ is the probability that the process will return to $\mathcal{L}(m)$ under the taboo that it doesn't visit $\mathcal{L}(m-1)$ and it can reach at most $\mathcal{L}(m+l)$ given that it began in $\mathcal{L}(m)$, and $G(l)$ is the probability that the process eventually enters $\mathcal{L}(m-1)$ and it can reach at most $\mathcal{L}(m+l)$ given that it began in $\mathcal{L}(m)$.

Latouche and Ramaswami [18] give a simple, elegant physically-motivated proof that shows that the sequences $U(l)$ and $G(l)$ are monotonically increasing and converge to the matrices $U$ and $G$ respectively and that $G(l)$ and $U(l)$ are identical to $F(l)$ and $M(l)$ for all $l \geq 0$ respectively. They also show that $G$ is the minimal non-negative solution to equation (4.4.8).

Algorithm $U$ converges linearly with respect to level. It converges at a faster rate than the Neuts algorithm, because at the $l$-th iteration of algorithm $U$ all the sample paths from the $l$-th iteration of the Neuts algorithm are included, in addition to many more, because algorithm $U$ places no restriction on the number or pattern of left transitions.

### 4.5.3 The Level-Independent Logarithmic Reduction Algorithm

The space of sample paths that are measured at each step of algorithm $U$ increase linearly with respect to the taboo level. Thus at the $l$-th step the sample paths consist of those paths that commence at $\mathcal{L}(m)$ and end in $\mathcal{L}(m-1)$ under the taboo that the maximum level reached is $\mathcal{L}(m+l)$ without placing any other restrictions on the number or positions of the left or right transitions. An algorithm that converges quadratically with respect to level was developed by Latouche and Ramaswami [17]. This algorithm is called the level-independent logarithmic reduction algorithm (LILRA). It converges quadratically because the maximum level that the sample paths can reach does not increase linearly per step but geometrically.

Consider the matrix, $H^{[l]}$, whose definition is,

$$H^{[l]} = P[\gamma(m+2^l) < \gamma(m-2^l) \ \& \ X\big(\gamma(m+2^l)\big) \big| X(0) \in \mathcal{L}(m)], \tag{4.5.8}$$

for $m \geq 2^l$. The matrix $H^{[l]}$ has a simple physical interpretation: it is the probability that the process will enter level $\mathcal{L}(m+2^l)$ before $\mathcal{L}(m-2^l)$ given that it commenced in $\mathcal{L}(m)$. The matrix, $L^{[l]}$, is defined as follows,

$$L^{[l]} = P[\gamma(m-2^l) < \gamma(m+2^l) \ \& \ X\big(\gamma(m-2^l)\big) \big| X(0) \in \mathcal{L}(m)]. \tag{4.5.9}$$

The physical interpretation of $L^{[l]}$ is very similar to $H^{[l]}$: it is the probability that the process reaches $\mathcal{L}(m-2^l)$ before ever entering $\mathcal{L}(m+2^l)$ given that the process began in $\mathcal{L}(m)$. The process is level-independent so it does not matter from which

level, $\mathcal{L}(m)$ we commence provided that $m \geq 2^l$. We can therefore write equations (4.5.8) and (4.5.9) as,

$$H^{[l]} = P[\gamma(2^{l+1}) < \gamma(0) \ \& \ X\big(\gamma(2^{l+1})\big)\big|X(0) \in \mathcal{L}(2^l)], \tag{4.5.10}$$

and,

$$L^{[l]} = P[\gamma(0) < \gamma(2^{l+1}) \ \& \ X\big(\gamma(0)\big)\big|X(0) \in \mathcal{L}(2^l)]. \tag{4.5.11}$$

Define the matrix $U^{[l]}$ to be the probability that commencing from $\mathcal{L}(2^{l+1})$ the process returns to that level after visiting $\mathcal{L}(2^{l+1}+2^l)$ or $\mathcal{L}(2^l)$ but before visiting $\mathcal{L}(2^{l+2})$ or $\mathcal{L}(0)$ and is given by,

$$U^{[l]} = H^{[l]}L^{[l]} + L^{[l]}H^{[l]}, \tag{4.5.12}$$

for $l \geq 0$. Let us commence from $\mathcal{L}(1)$ so $G$ is,

$$G = P[\gamma(0) < \infty \ \& \ X(\gamma(0))\big|X(0) \in \mathcal{L}(1)],$$

and can be determined using the level-independent logarithmic reduction algorithm [17] that is stated below.

**Theorem 8** *The matrix $G$ is given by,*

$$G = \sum_{l\geq 0}\left(\prod_{0\leq i\leq l-1} H^{[i]}\right)L^{[l]}, \tag{4.5.13}$$

*where,*

$$H^{[0]} = (-A_1)^{-1}A_0, \tag{4.5.14}$$

$$L^{[0]} = (-A_1)^{-1}A_2, \tag{4.5.15}$$

$$H^{[l+1]} = (I - U^{[l]})^{-1}\big(H^{[l]}\big)^2, \tag{4.5.16}$$

$$L^{[l+1]} = (I - U^{[l]})^{-1}\big(L^{[l]}\big)^2, \tag{4.5.17}$$

*for $l \geq 0$.*

The proof can be found in [17] and [18]. The $l$-th term in (4.5.13) contains all those sample paths that commence in $\mathcal{L}(1)$ under the taboo that they never visit $\mathcal{L}(2^{l+1})$ before entering $\mathcal{L}(0)$. This is a powerful approach when one considers that at the $l$-th step of algorithm $U$ only those sample paths that reach at most $\mathcal{L}(l+1)$ are included.

The remainder of this chapter is dedicated to the level-dependent QBD (LDQBD).

## 4.6 Level-Dependent Quasi-Birth-and-Death Processes

A level-dependent QBD (LDQBD) is defined in the following manner [3]: let $X(t)$ be a two-dimensional Markov process on the state space $\left\{(m,\Phi) : m \geq 0, 1 \leq \Phi \leq M_m\right\}$. If $X(t) = (m,\Phi)$ we say that the QBD is in $\mathcal{L}(m)$ and in phase $\Phi \in \{1,\ldots,M_m\}$. The number of states of the process in each level is given by the number of phases in each level, so if there are $M_m$ phases at level $\mathcal{L}(m)$ then there are $M_m$ distinct states.

The infinitesimal generator of the LDQBD is,

$$Q = \begin{bmatrix} Q_1^{(0)} & Q_0^{(0)} & 0 & 0\ldots & \\ Q_2^{(1)} & Q_1^{(1)} & Q_0^{(1)} & 0\ldots & \\ 0 & Q_2^{(2)} & Q_1^{(2)} & Q_0^{(2)} & \ldots \\ 0 & 0 & Q_2^{(3)} & Q_1^{(3)} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{4.6.1}$$

where the matrix $Q_2^{(m)}$ is of dimension $M_m \times M_{m-1}$, $Q_1^{(m)}$ is $M_m \times M_m$ and $Q_0^{(m)}$ is $M_m \times M_{m+1}$. The entries in each matrix for $m \geq 0$ are strictly non-negative, except the diagonal elements of $Q_1^{(m)}$ which are strictly less than zero for all $m \geq 0$. In addition, we assume that the process is irreducible and that $Q_2^{(m)}\boldsymbol{e}+Q_1^{(m)}\boldsymbol{e}+Q_0^{(m)}\boldsymbol{e} = \mathbf{0}$ where each $\boldsymbol{e}$ is of the appropriate dimension.

The equation for the family of matrices, $\{G_m, m \geq 1\}$ is slightly more complicated in the level dependent domain. The probability of eventually entering state $(m-1, j)$ starting from $(m, i)$ is $(G_m)_{ij}$, where

$$(G_m)_{ij} = P[\gamma(m-1) < \infty \ \& \ X\big(\gamma(m-1)\big) = (m-1, j) \big| X(0) = (m, i)]. \quad (4.6.2)$$

The family of matrices $\{G_m, m \geq 1\}$ is the minimal non-negative solution to the family of equations,

$$G_m = (-Q_1^{(m)})^{-1} Q_2^{(m)} + (-Q_1^{(m)})^{-1} Q_0^{(m)} G_{m+1} G_m, \quad (4.6.3)$$

for $m \geq 1$. Physically then, the process may move down to $\mathcal{L}(m-1)$ from $\mathcal{L}(m)$ directly with probability $(-Q_1^{(m)})^{-1} Q_2^{(m)}$ or by first moving to $\mathcal{L}(m+1)$ with probability $(-Q_1^{(m)})^{-1} Q_0^{(m)}$ and then eventually moving down to $\mathcal{L}(m)$ with probability $G_{m+1}$, and then eventually moving down to $\mathcal{L}(m-1)$ with probability $G_m$. Since each of these events is independent we have that the probability of the indirect approach is $(-Q_1^{(m)})^{-1} Q_0^{(m)} G_{m+1} G_m$.

## 4.7 Level-Dependent Algorithms

### 4.7.1 The Level-Dependent Logarithmic Reduction Algorithm

The level-dependent logarithmic reduction algorithm (LDLRA) was developed by Bright and Taylor [3] and Ramaswami and Taylor [28]. For $m \geq 2^l$ define the matrix $U_m^l$ to be

$$H_m^{[l]} = P[\gamma(m+2^l) < \gamma(m-2^l) \ \& \ X\big(\gamma(m+2^l)\big) \big| X(0) \in \mathcal{L}(m)]. \quad (4.7.1)$$

Physically, $H_m^{[l]}$, gives the probability that beginning in $\mathcal{L}(m)$ the process will eventually reach $\mathcal{L}(m+2^l)$ under the taboo that it does not ever visit $\mathcal{L}(m-2^l)$. The matrix $L_m^{[l]}$, for $m \geq 2^l$, is defined to be,

$$L_m^{[l]} = P[\gamma(m-2^l) < \gamma(m+2^l) \ \& \ X\big(\gamma(m-2^l)\big) \big| X(0) \in \mathcal{L}(m)]. \quad (4.7.2)$$

$L_m^{[l]}$ can be interpreted as being the probability that beginning in $\mathcal{L}(m)$ the process eventually visits $\mathcal{L}(m-2^l)$ before it ever visits $\mathcal{L}(m+2^l)$. The matrices, $H_m^{[l]}$ and $L_m^{[l]}$ form the basis for the LDLRA.

The factor,

$$H_m^{[l]} L_{m+2^l}^{[l]} + L_m^{[l]} H_{m-2^l}^{[l]}, \tag{4.7.3}$$

is interpreted as being the probability that the process will return to $\mathcal{L}(m)$ after visiting either $\mathcal{L}(m+2^l)$ or $\mathcal{L}(m-2^l)$ under the taboo that it does not visit $\mathcal{L}(m+2^{l+1})$ or $\mathcal{L}(m-2^{l+1})$. The algorithm to determine $G_m$ for $m \geq 1$ is as follows [28],

**Theorem 9** *The family of matrices defined by equation (4.6.2) is given, for* $m \geq 1$ *by,*

$$G_m = \sum_{l=0}^{\infty} \left( \prod_{i=0}^{l-1} H_{m-1+2^i}^{[i]} \right) L_{m-1+2^l}^{[l]}, \tag{4.7.4}$$

*where, for* $\ell \geq 0$

$$H_\ell^{[0]} = (-Q_1^{(\ell)})^{-1} Q_0^{(\ell)}, \tag{4.7.5}$$

$$L_\ell^{[0]} = (-Q_1^{(\ell)})^{-1} Q_2^{(\ell)}, \tag{4.7.6}$$

$$H_\ell^{[l+1]} = \sum_{k=0}^{\infty} \left( H_\ell^{[l]} L_{\ell+2^l}^{[l]} + L_\ell^{[l]} H_{\ell-2^l}^{[l]} \right)^k H_\ell^{[l]} H_{\ell+2^l}^{[l]}, \tag{4.7.7}$$

$$L_\ell^{[l+1]} = \sum_{k=0}^{\infty} \left( H_\ell^{[l]} L_{m\ell+2^l}^{[l]} + L_\ell^{[l]} H_{\ell-2^l}^{[l]} \right)^k L_\ell^{[l]} L_{\ell-2^l}^{[l]}. \tag{4.7.8}$$

An elegant physically motivated proof can be found in [28].

This chapter has been primarily concerned with many of the fundamental aspects of matrix analytic methods. The MAP and the QBD play a prominent role in what is to follow; we use both to define the MBT. The MBT is an alternative representation of the binary-branch point ctMMTBP. In particular, the MBT representation of the binary-branch point ctMMTBP gives to the process a structure where correlations between the branch lifetimes and branch offspring distributions arise naturally. Furthermore, this representation allows us to exploit a much richer algorithmic basis from which to obtain some interesting measures that are of use in

biology; an algorithmic basis that is quite under-developed in the branching process literature.

# Chapter 5

# Markovian Binary Trees

## 5.1 Markovian Binary Tree: Definition

In this section we define and construct the Markovian binary tree as a particular example of a level-dependent QBD, which was discussed in Chapter 4. A Markovian binary tree (MBT) is a level-dependent QBD process with states $X(t) = (N(t), \phi_1(t), \ldots, \phi_{N(t)}(t))$ defined on $\bigcup_{k=0}^{\infty} \{\{k\} \times \{1, \ldots, n\}^k\}$. The random variable $N(t)$ denotes the number of branches alive at time $t$. For $k = 1, \ldots, N(t)$, $\phi_k(t)$ gives the phase of the $k$-th branch at time $t$. The phase process in each of the branches evolves as the phase process of an $n$-phase MAP, as described in Section 4.3. If there are $k$ branches alive then there are $n^k$ possible states of the phase process for the entire MBT.

We define the Kronecker product and sum of two matrices. If a matrix $Z$ is of dimension $a \times b$ and a matrix $Y$ is of dimension $c \times e$, then the Kronecker product of the matrices $Z \otimes Y$, is given by the $ac \times be$ matrix,

$$Z \otimes Y = \begin{bmatrix} z_{11}Y & z_{12}Y & \ldots & z_{1b}Y \\ z_{21}Y & z_{22}Y & \ldots & z_{2b}Y \\ \vdots & \vdots & \vdots & \vdots \\ z_{a1}Y & z_{a2}Y & \ldots & z_{ab}Y \end{bmatrix}.$$

The Kronecker sum of two matrices $Z$ and $Y$ which are both of dimension $a \times a$ is given by

$$Z \oplus Y = Z \otimes I + I \otimes Y,$$

where $I$ is the $a \times a$ identity matrix.

We shall state and explain the infinitesimal generator matrix for the process and then discuss its qualitative behaviour. The transition rate matrix for the MBT is

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \mathbf{A}_2^{(1)} & A_1^{(1)} & A_0^{(1)} & 0 & 0 & 0 & \ddots \\ \mathbf{0} & A_2^{(2)} & A_1^{(2)} & A_0^{(2)} & 0 & 0 & \ddots \\ \mathbf{0} & 0 & A_2^{(3)} & A_1^{(3)} & A_0^{(3)} & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Let $I^{(k)}$ by the $n^k \times n^k$ identity matrix, with $I^{(0)} = 1$. The $n^k \times n^{k-1}$ matrix $A_2^{(k)}$ for $k \geq 1$ is given by

$$A_2^{(k)} = \sum_{j=0}^{k-1} I^{(j)} \otimes \boldsymbol{d} \otimes I^{(k-1-j)}. \tag{5.1.1}$$

Equation (5.1.1) embodies the fact that only one branch can become extinct at any moment of time. The $n^k \times n^k$ matrix $A_1^{(k)}$ for $k \geq 1$ satisfies the recursion

$$A_1^{(k)} = A_1^{(k-1)} \oplus D_0, \tag{5.1.2}$$

with $A_1^{(0)} = 0$. Once again the nature of equation (5.1.2) indicates that there is no interaction between any of the $k$ phase processes that are currently evolving.

Finally the $n^k \times n^{k+1}$ matrices $A_0^{(k)}$ for all $k \geq 1$ are given by

$$A_0^{(k)} = \sum_{j=0}^{k-1} I^{(j)} \otimes B \otimes I^{(k-1-j)}, \ k \geq 1 \tag{5.1.3}$$

where the $n \times n^2$ matrix $B$ governs the observable transitions of the process. The expression in equation (5.1.3) has $k$ terms reflecting the fact that there are $k$ actively evolving branches and that any one of these branches can give rise to a new daughter branch via the action of the matrix $B$. The independence of the $k$ evolving branches

is clearly seen by the fact that there is no interaction between the $k$ copies of the $B$ matrix.

The state space of the process can be partitioned into levels, just as in the QBDs of Chapter 4. The space of states in level $k \in \{0\} \cup \mathbb{Z}+$, denoted by $\mathcal{L}(k)$, are those states that consist of $k$ actively evolving branches.

**Example 5** *A model with three non-absorbing phases,* $\{1, 2, 3\}$.

The natural extension of the MAP is to force both branches at a branch point to be in the same phase immediately after the branch point. The matrix $B$ for such a model has the form,

| | $(1,1)$ | $(1,2)$ | $(1,3)$ | $(2,1)$ | $(2,2)$ | $(2,3)$ | $(3,1)$ | $(3,2)$ | $(3,3)$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $(D_1)_{11}$ | 0 | 0 | 0 | $(D_1)_{12}$ | 0 | 0 | 0 | $(D_1)_{13}$ |
| 2 | $(D_1)_{21}$ | 0 | 0 | 0 | $(D_1)_{22}$ | 0 | 0 | 0 | $(D_1)_{23}$ |
| 3 | $(D_1)_{31}$ | 0 | 0 | 0 | $(D_1)_{32}$ | 0 | 0 | 0 | $(D_1)_{33}$ |

. (5.1.4)

The rows indicate the phase that the branch was in immediately before it underwent an observable transition and the columns give the birth phase of the daughter branch (the left digit) and the phase that the parental branch (right digit) is in immediately after the observable transition. Thus a branch in phase $i$ immediately before the branch point will generate a daughter and parent branch that are in an identical phase immediately following the branch point. Hence only $B_{i,jj}$ for $j = 1, 2, 3$ are non-zero and correspond to the elements $(D_1)_{ij}$. The general case where the parent and daughter branches can be in any phase immediately after the branch point has the transition structure given by

$$B = \begin{bmatrix} B_{1,11} & B_{1,12} & B_{1,13} & B_{1,21} & B_{1,22} & B_{1,23} & B_{1,31} & B_{1,32} & B_{1,33} \\ B_{2,11} & B_{2,12} & B_{2,13} & B_{2,21} & B_{2,22} & B_{2,23} & B_{2,31} & B_{2,32} & B_{2,33} \\ B_{3,11} & B_{3,12} & B_{3,13} & B_{3,21} & B_{3,22} & B_{3,23} & B_{3,31} & B_{3,32} & B_{3,33} \end{bmatrix}. \quad (5.1.5)$$

The element $B_{i,jk}$ of the matrix $B$ gives the rate at which the parent branch in phase $i$ spawns a branch point (observable transition) such that the daughter branch is in

phase $j$ whilst the parental branch is in phase $k$ immediately after the branch point. When pictorially representing a branch point, the parental branch is drawn as the right branch and the daughter branch is drawn as the left branch.

The MBT is a special case of the continuous-time Markovian multi-type branching process as we shall show in Section 5.2. The point of departure from the ctMMTBP is an issue of interpretation. Because the branches of an MBT are governed by MAPs, hidden transitions do not correspond to tree nodes. Observable transitions that spawn daughter branches, on the other hand, do correspond to nodes. In the ctMMTBP, the hidden transitions of the MBT are called singular transitions, and result in the transformation of one particle type into a distinct particle type. In the ctMMTBP such singular transitions do correspond to nodes. By interpreting the dynamics of a branch using a MAP, the time intervals between branch points (or nodes) need not be exponentially distributed. Furthermore, the interval between branch points can influence the offspring distribution, in particular, see equations (4.3.2), (4.3.3) and (4.3.4) in Chapter 4,.

Suppose that at time $t$ the process is in a state with $m$ branches and let branch $k \leq m$ be in phase $r$. Suppose the current state of the process is,

$$\begin{array}{cccccccccc} (m, & a, & \ldots, & b, & r, & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & k+1 & \ldots & m \end{array}$$

where the number beneath each branch denotes the label of that branch. The following transitions are then possible:

- A hidden transition to phase $j \neq r$, occurs with rate $(D_0)_{rj}$. This transition causes the state of the MBT to become

$$\begin{array}{cccccccccc} (m, & a, & \ldots, & b, & j, & c, & \ldots, & d), \\ & 1 & \ldots & k-1 & k & k+1 & \ldots & m \end{array}$$

- An observable transition that spawns a daughter branch in phase $i$ whilst the parental branch is in phase $j$ immediately following the transition, occurs with

rate $B_{r,ij}$. The new state of the MBT is

$$\begin{array}{cccccccccc} (m+1, & a, & \ldots, & b, & i, & j & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & k+1 & k+2 & \ldots & m+1 \end{array}$$

The tree is oriented such that the parental branch is the right branch and is associated with the $k+1$-th label and the daughter branch is the left branch and is associated with the $k$-th label. The branches that were previously labelled $k+1, \ldots, m$ have been re-labelled to $k+2, \ldots, m+1$.

- Finally a catastrophe occurs on branch $k$ at a rate $d_r$. This causes branch $k$ to cease to exist and the new state is

$$\begin{array}{cccccccc} (m-1, & a, & \ldots, & b, & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & \ldots & m-1 \end{array}$$

where the branches that were previously labelled $k+1, \ldots, m$ have been relabelled to $k-1, \ldots, m-1$.

### 5.1.1 An Alternative Representation of the States of the Process

Using the above representation for the states of the process we find that $\mathcal{L}(m)$ is populated by a total of $n^m$ states, since there are $m$ branches and each branch can be in any one of $n$ possible phases. However, states of the process can be represented in a more compact form, whereby at any time $t$, the state of the process is given by the number of branches that are in each phase. For example, if we have $m$ living branches, then the states of the process are given by,

$$(m_1, m_2, \ldots, m_n),$$

where $m_i$ is the number of branches in phase $i$, and such that,

$$\sum_{i=1}^{n} m_i = m.$$

Using this representation then, the number of states with $m$ living branches is

$$\binom{m+n-1}{m}.$$

This is just the number of ways of placing $m$ like objects into $n$ different cells, as can be found in any textbook on combinatorics, for example see [29].

The number of states in this representation is clearly less than $n^m$, however under this representation we lose the ability to distinguish between parent and daughter branches; all branches are essentially treated alike, modulo their current phase and thus we lose any concept of ancestry. The major portion of this thesis is concerned with analysing tree topologies and this state space representation will not allow us to identify topologies as we have lost all relationships between branches and merely know the number of branches in each phase. As a result, in the remainder of this thesis we use the original representation since this representation does allow us to keep track of the history of each branch and hence of the topology of the tree.

## 5.2 An MBT is a special case of a ctMMTBP

### 5.2.1 Definition

The MBT is a special case of the continuous-time Markovian multi-type branching process where each and every branch point may have 0, 1, or 2 offspring. We now write the MBT as a branching process. Recall from Section 2.4 that the probability generating functions of the offspring distribution for each particle $i$ are given by,

$$f^{(i)}(\boldsymbol{s}) = \sum_{j_1,\ldots,j_n \in \mathbb{Z}^+} p^{(i)}(\boldsymbol{j}) s_1^{j_1} \ldots s_n^{j_n}. \tag{5.2.1}$$

The MBT is a special case of the ctMMTBP with offspring probability generating functions that are quadratic. The generating functions are therefore,

$$f^{(i)}(\boldsymbol{s}) = \frac{d_i}{(d_0)_i} + \sum_{k=1,k\neq i}^{n} \frac{(D_0)_{ik}}{(d_0)_i} s_k + \sum_{k,j=1}^{n} \frac{B_{i,jk}}{(d_0)_i} s_j s_k, \tag{5.2.2}$$

for all $i \in \{1, 2, \ldots, n\}$. Suppose that the process consists of one particle of type $i$, we can describe the qualitative behaviour of the process quite simply. The particle of type $i$ will live for an exponentially distributed amount of time with mean $1/(d_0)_{ii}$ at which point the particle will either,

1. die without giving birth to any new particle and this occurs with probability $d_i/(d_0)_i$, or
2. die and transform into a single particle of type $j \neq i$ and this occurs with probability, $(D_0)_{ij}/(d_0)_i$, or finally,
3. die and give rise to two new particles that have types $j$ and $k$ and this occurs with probability, $B_{i,jk}/(d_0)_{id}$.

### 5.2.2 Regularity and the Mean Number of Branches

Since all the derivatives of $f^{(i)}(\boldsymbol{s})$ are clearly finite, the process does not explode [2]. The matrix of the expected number of branches in the MBT case is, by Section 2.4.2

$$M(t) = \exp(At), \tag{5.2.3}$$

where

$$A_{ij} = (d_0)_i k_{ij}, \tag{5.2.4}$$

and

$$k_{ij} = \frac{1}{(d_0)_i}\left((1-\delta_{ij})(D_0)_{ij} + \sum_{k=1}^{n}(B_{i,jk} + B_{i,kj})\right) - \delta_{ij}, \tag{5.2.5}$$

where $\delta_{ij}$ is the Kronecker delta , equal to one if $i = j$ and zero otherwise. From equations (5.2.4) and (5.2.5) it can be deduced that

$$A = D_0 + BC, \tag{5.2.6}$$

where $C$ is an $n^2 \times n$ counting matrix, with

$$C_{ij,k} = I\{i = k\} + I\{j = k\}. \tag{5.2.7}$$

For a process with three non-absorbing phases, $\{1, 2, 3\}$, the $C$ matrix has the form,

| | 1 | 2 | 3 |
|---|---|---|---|
| 11 | 2 | 0 | 0 |
| 12 | 1 | 1 | 0 |
| 13 | 1 | 0 | 1 |
| 21 | 1 | 1 | 0 |
| 22 | 0 | 2 | 0 |
| 23 | 0 | 1 | 1 |
| 31 | 1 | 0 | 1 |
| 32 | 0 | 1 | 1 |
| 33 | 0 | 0 | 2 |

(5.2.8)

Therefore the expected number of branches at time $t$ is given by

$$M(t) = \exp[(D_0 + BC)t]. \tag{5.2.9}$$

From Chapter 2.4 we saw that the process was sub-critical, critical or super-critical depending on the dominant eigenvalue of the matrix $A$. Thus

- if $\lambda_A < 0$ the process is sub-critical,
- if $\lambda_A = 0$ the process is critical, and finally
- if $\lambda_A > 0$ the process is super-critical.

### 5.2.3 Probability of Eventual Extinction

Recall that the minimal non-negative solution of

$$\boldsymbol{u}(\boldsymbol{s}) = \mathbf{0}, \tag{5.2.10}$$

is the probability of ultimate extinction of the process, see equation (2.4.22). Now if $\lambda_A \leq 0$ the process will become extinct almost surely, and if $\lambda_A > 0$ then $\boldsymbol{q} < \boldsymbol{e}$

component-wise. For the MBT we have,

$$\begin{aligned} 0 &= u^{(i)}(\boldsymbol{s}) \\ &= (d_0)_i\big(f^{(i)}(\boldsymbol{s}) - s_i\big) \qquad (5.2.11) \\ &= (d_0)_i\left(\frac{d_i}{(d_0)_i} + \sum_{k=1,k\neq i}^{n}\frac{(D_0)_{ik}}{(d_0)_i}s_k + \sum_{k,j=1}^{n}\frac{B_{i,jk}}{(d_0)_i}s_j s_k - s_i\right) \\ &= d_i + \sum_{k=1}^{n}(D_0)_{ik}s_k + \sum_{k=1}^{n}\sum_{j=1}^{m}B_{i,kj}s_k s_j \qquad (5.2.12) \end{aligned}$$

This can be re-written in matrix form as,

$$\mathbf{0} = \boldsymbol{d} + D_0\boldsymbol{s} + B(\boldsymbol{s}\otimes\boldsymbol{s}). \qquad (5.2.13)$$

The probability of eventual extinction is the minimal non-negative solution to equation (5.2.13), [2]. We shall return to equation (5.2.13) in Chapter 7. We next show that the macroevolutionary models that were discussed in Chapter 3 can all be subsumed by the MBT model.

## 5.3 MBTs and Simple Macroevolutionary Models

We begin by discussing the simplest of the models, the constant rates birth-and-death model (crBD).

### 5.3.1 Constant Rates Birth-and-Death Model

This model is characterised by the fact that there is only one particle type. A particle will speciate with probability $\lambda/(\lambda+\mu)$ or will become extinct with probability $\mu/(\lambda+\mu)$. The probability generating function for the process is given by,

$$p(s) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}s^2. \qquad (5.3.1)$$

To develop the MBT version of the crBD we associate $\lambda/(\lambda+\mu)$ with $B_{1,11}/(d_0)_1$ and $\mu/(\lambda+\mu)$ with $d_1/(d_0)_1$ and apply the $\mathcal{M}^u$ mapping, since as before we only

consider unstable branches in line with the analysis of [26, Appendix A]. Finally, we write the MBT transition rates

$$(D_0)_{11} = -(\lambda + \mu) \tag{5.3.2}$$

$$B_{1,11} = \lambda \tag{5.3.3}$$

$$d_1 = \mu. \tag{5.3.4}$$

### 5.3.2 Proportional-to-Distinguishable Arrangements Model

The PDA model in its original incarnation stated that each distinguishable arrangement of the labels on leaf branches of a given size are equally likely. The PDA model was shown in Chapter 3, Section 3.10 to correspond to the asymptotic subcritical constant rates birth-and-death model. If in the MBT model the rates are given by equations (5.3.2)-(5.3.4) and if $\mu \geq \lambda$, then as $t \to \infty$ the distribution of probabilities over the extinct tree topologies is exactly given by the PDA model. Here we mention a subtle point, in a subcritical branching process as $t \to \infty$, the only tree topologies that have a non-zero probability are those that have only extinct branches, application of the mapping $\mathcal{M}^e$ does not remove any non-extinct branches as there are none. Pinelis [26] provided an alternative definition of the PDA, called the multi-rate-PDA (MR-PDA). In the subcritical domain of this model, the mapping $\mathcal{M}^q$ needs to be applied, and after all extinct branches are pruned, all finite quasi-stable topologies of a given size are equally likely. To demonstrate the versatility of the MBT we shall derive an MBT with an identical distribution to that of the MR-PDA model as $t \to \infty$.

Recall that in the MR-PDA setting, the probability of a particular random tree of topology $\mathcal{T}^q$ is given by,

$$p(\mathcal{T}^q) = \left(\frac{b}{1 - 2bp(\emptyset)}\right)^{s-1} \left(p(1)\right)^s, \tag{5.3.5}$$

if $|\mathcal{T}^q| = s$. The MBT does not contain quasi-stable phases, instead it has two types of phase:

- phase 0 which represents extinct species, and
- phases 1 to $n$ which represent live species.

In order to transform the MR-PDA into the MBT domain we therefore

- map the quasi-stable phase of the MR to the absorbing phase of the MBT, phase 0, and
- exclude the extinct state of the MR-PDA.

What the second condition implies is that we map the space of finite non-extinct quasi-stable trees to the space of extinct MBT trees and discard all the MR extinct trees.

In the MBT domain the rate at which a branch becomes absorbed is given by,

$$d_1 = \frac{q}{1 - p(\emptyset)}, \tag{5.3.6}$$

and the rate at which a branch gives birth to an identical daughter branch is given by,

$$B_{1,11} = b\big(1 - p(\emptyset)\big). \tag{5.3.7}$$

It can be shown that

$$(D_0)_{11} = -d_1 - B_{1,11} = -\big(1 - 2bp(\emptyset)\big), \tag{5.3.8}$$

by using the fact that $p^2(\emptyset) + 2p(\emptyset)\big(1 - p(\emptyset)\big) + \big(1 - p(\emptyset)\big)^2 = 1$ and

$$p(\emptyset) = d + bp^2(\emptyset),$$

which is equation (3.11.1) from Chapter 3. What is interesting about equation (5.3.8) is the fact that $(d_0)_1 = -(D_0)_{11} < 1$ . This tells us that by discarding extinct trees in the MR sense we are essentially slowing down the clock of the process, since these false events never occur in the MBT.

By analyzing the process as $t \to \infty$ we can write the probability that a random tree has topology $\mathcal{T}^e$ as

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1}B\big(\boldsymbol{p}(\mathcal{T}^e_{[0,0]}) \otimes \boldsymbol{p}(\mathcal{T}^e_{[0,1]})\big), \tag{5.3.9}$$

since $(-D_0)^{-1}B$ is the probability that the node [0] will eventually become an internal node, and $\boldsymbol{p}(\mathcal{T}^e_{[0,0]})$ and $\boldsymbol{p}(\mathcal{T}^e_{[0,1]})$ are the probabilities that the daughter branch and the parent branch eventually evolve into trees of topology $\mathcal{T}^e_{[0,0]}$ and $\mathcal{T}^e_{[0,1]}$ respectively. In this case the process has only two phases, an extinct phase, corresponding to quasi-stability from the MR-PDA model, and one unstable phase, so that equation (5.3.9) is a scalar, that can easily be shown to be equal to

$$\begin{aligned} p(\mathcal{T}^e) &= \frac{1}{(d_0)_1}B_{1,11}p(\mathcal{T}^e_{[0,0]})p(\mathcal{T}^e_{[0,1]}) \\ &= \frac{1}{1-2bp(\emptyset)}b\big(1-p(\emptyset)\big)p(\mathcal{T}^e_{[0,0]})p(\mathcal{T}^e_{[0,1]}). \end{aligned} \tag{5.3.10}$$

Now, prior to any true binary branch point, that is, a branch point that generates two branches that do not eventually become extinct, there may be any number of branch points in which one of the two branches eventually generates an extinct subtree. This possibility is reflected by the fact that for each true branch point of the topology we must pre-multiply by $1/(1-2bp(\emptyset))$, which is interpreted as giving the mean number of false branch points that occur before the occurence of the true branch point. Now, in the MBT-PDA interpretation we do not allow for extinct trees in the sense of the MR-PDA model. In the MBT-PDA model, extinct trees are precisely those trees that correspond to the quasi-stable portion of MR-PDA trees.

Using induction on equation (5.3.10) it can be shown that the above equation is equal to

$$p(\mathcal{T}^e) = \left(\frac{1}{1-2bp(\emptyset)}\right)^{s-1}\Big(b\big(1-p(\emptyset)\big)\Big)^{s-1}\big(p(1)\big)^s, \tag{5.3.11}$$

where $p(1)$ is the probability that an individual branch eventually becomes absorbed into phase 0 before a birth, and is given by

$$p(1) = \frac{1}{(d_0)_1}d_1 = \frac{q}{\big(1-2bp(\emptyset)\big)\big(1-p(\emptyset)\big)}. \tag{5.3.12}$$

Combining equations (5.3.11) and (5.3.12) we obtain,

$$p(\mathcal{T}^e) = \frac{1}{1-p(\emptyset)} \left( \frac{b}{1-2bp(\emptyset)} \right)^{s-1} \left( \frac{q}{(1-2bp(\emptyset))} \right)^{s}. \tag{5.3.13}$$

This equation is identical to equation (3.11.5) from Chapter 3 except for the factor $1/(1-p(\emptyset))$. Recall that in the MR-PDA model, any branches that eventually became extinct are pruned from the evolving tree in order to generate the correct topology. As an alternative way of thinking about this, we apply the $\mathcal{M}^q$ mapping from the space of realizations to the space of topologies. On the other hand, to correctly obtain the measure for each topology, we cannot disregard the existence of extinct subtrees; they must be taken into account. For the MR-PDA model, the set of trees as $t \to \infty$ is populated by two subsets of non-zero measure, the space of topologies representing extinct trees and the space of topologies representing quasi-stable trees. The space of extinct trees, in the MR sense, are absent in the MBT, and so to obtain the correct distribution we must divide the MR-PDA probabilities by $1-p(\emptyset)$ to obtain equation (5.3.13).

### 5.3.3 The super-PDA model

The multi-rate interpretation of the sPDA (MR-sPDA) model was discussed in Section 3.11.2. The probability that a random tree is generated with topology $\mathcal{T}^q$ is dependent on the number of unitary splits of that topology. The MR-sPDA allows daughter branches to be born in the quasi-stable state, whereas the MR-PDA does not. The MR-sPDA model can also be re-written in terms of an MBT. The space of topologies that we are concerned with in the MBT version of the sPDA model is precisely that space which consists of only the quasi-stable topologies of the MR-sPDA model. This is just the generalization of our analysis for the MBT version of the MR-PDA in the previous section.

To transform the MR-sPDA to the MBT domain,

1. we wish to map the quasi-stable phase to the absorbing phase of the MBT.

The problem with doing this directly is that in the MR-sPDA model, a branch can give birth to a daughter in the quasi-stable phase, whereas, in the MBT we do not allow births directly into the absorbing phase. Consequently, we map the quasi-stable phase to a holding phase. Once a branch enters the holding phase it will then be absorbed with probability one.

2. we exclude the extinct MR-sPDA state.

The second point implies that at each transition we exclude the possibility that a branch and hence subtree will become extinct (in the multi-rate sense). Thus we condition on the space of trees where extinction does not occur, that is, the topologies with a positive number of quasi-stable branches as $t \to \infty$; see the discussion in Section 5.3.2. We map the space of these trees to the space of sub-critical MBT trees whose branches have all been absorbed in phase 0.

| MR-sPDA | Intermediate | MBT-sPDA |
|---|---|---|
| $q$ | $\frac{q}{1-p(\emptyset)}$ | $\frac{q}{(1-2b(1-\gamma)p(\emptyset))(1-p(\emptyset))}$ |
| $b\gamma p(\emptyset)$ | $\frac{b\gamma p(\emptyset)}{1-p(\emptyset)}$ | $\frac{b\gamma p(\emptyset)}{(1-2b(1-\gamma)p(\emptyset))(1-p(\emptyset))}$ |
| $b\gamma(1-p(\emptyset))$ | $b\gamma$ | $\frac{b\gamma}{1-2b(1-\gamma)p(\emptyset)}$ |
| $2b(1-\gamma)p(\emptyset)(1-p(\emptyset))$ | $2b(1-\gamma)p(\emptyset)$ | — |
| $b(1-\gamma)(1-p(\emptyset))^2$ | $b(1-\gamma)(1-p(\emptyset))$ | $\frac{b(1-\gamma)(1-p(\emptyset))}{1-2b(1-\gamma)p(\emptyset)}$ |

Table 5.3.1: Branch point transitions from MR-sPDA to MBT-sPDA

Table 5.3.1 has three columns. The first column tabulates the most important branch point transition probabilities of the MR-sPDA model:

1. the rate at which a branch transitions into the quasi-stable state,

2. the rate at which a branch point occurs such that the daughter branch is born quasi-stable and the parent branch eventually becomes extinct is $b\gamma p(\emptyset)$,

3. the rate at which a branch point occurs such that the daughter is born quasi-stable and the parent branch does not eventually become extinct is $b\gamma(1-p(\emptyset))$,

4. the rate at which a branch point occurs such that either the daughter or the parent eventually becomes extinct whilst the other does not is $2b(1-\gamma)p(\emptyset)(1-p(\emptyset))$, and finally,

5. the rate at which a branch point occurs such that neither the daughter nor the parent eventually become extinct is $b(1-\gamma)(1-p(\emptyset))^2$

As we have stated already, in the MBT environment we have elected to exclude any transitions that lead to extinction in the multi-rate sense. Extinction in the MBT environment coincides with quasi-stability in the MR environment. Consequently, we cannot allow any transitions in the MBT environment that could potentially lead to tree extinction in the MR sense. Thus, following the explanation given in Section 5.3.2, we divide all these rates by $1-p(\emptyset)$ which then gives us the second column in the table.

Consider now the transition $2b(1-\gamma)p(\emptyset)$ from the second column. In the MR world, this leads to a branch point such that one of the branches generates a subtree that eventually becomes extinct whereas the second one does not. In the MBT world such a branch point, or such a transition does not exist. Why? Because we have chosen to not allow extinction in the MR sense. We divide all the rates in the second column by $1-2b(1-\gamma)p(\emptyset)$ because as we shall see this plays the role of one of the diagonal elements of $D_0$. This factor is less than one and similarly to the PDA example of the previous section it acts to set the "clock" of the process.

Because we are dividing the rates of the second column by what amounts to one of the diagonal elements of $D_0$, the third column actually gives us the transition probabilities for the MBT model of the process. We explain each transition in turn.

But before we do, note that, the absorbing phase for the MBT is 0, phase 1 is the holding phase and phase 2 is the unstable phase. Now, there are three ways in which a branch can become extinct in the MBT sense, that is, enter phase 0. The first is via a direct transition from phase 2 with probability $\frac{q}{(1-2b(1-\gamma)p(\emptyset))(1-p(\emptyset))}$. The second is given by $\frac{b\gamma p(\emptyset)}{(1-2b(1-\gamma)p(\emptyset))(1-p(\emptyset))}$, whose interpretation is quite simple, it is the probability that a branch in phase 2 will eventually undergo a transition to phase 0, without any associated branch point. Of course, in the MR-sPDA environment this was associated with an internal branch point, such that the parental branch became extinct and the daughter branch directly became quasi-stable. This internal branch point does not exist in the MBT, it is a leaf node instead. The third way is given by $\frac{1}{1-2b(1-\gamma)p(\emptyset)}b\gamma$, which gives the probability that a branch, which was in phase 2 immediately before the branch point, will undergo a branch point such that the daughter will be born in the holding phase, phase 1. This daughter branch will then become extinct with probability 1. The parental branch remains in phase 2. The final entry in the third column, $\frac{b(1-\gamma)(1-p(\emptyset))}{1-2b(1-\gamma)p(\emptyset)}$ gives us the probability that a branch point occurs when the parent branch is in phase 2 such that the parental branch remains in phase 2 immediately after the branch point and the daughter branch is also spawned into phase 2. Given the above transition probabilities we can easily write the rate matrices for the process:

$$D_0 = \begin{bmatrix} -1 & 0 \\ 0 & -\big(1-2b(1-\gamma)p(\emptyset)\big) \end{bmatrix}, \tag{5.3.14}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & b\gamma & 0 & b(1-\gamma)\big(1-p(\emptyset)\big) \end{bmatrix}, \tag{5.3.15}$$

and finally,

$$\boldsymbol{d} = \begin{bmatrix} 1 \\ \frac{q+b\gamma p(\emptyset)}{1-p(\emptyset)} \end{bmatrix}. \tag{5.3.16}$$

We elect to study the sub-critical process as $t \to \infty$, so a random tree will eventually become extinct with probability one. Recall from Chapter 3 that any

topology, and in particular, any extinct topology can be written as,

$$\mathcal{T}^e = \{([\alpha(0)],[0])^i, \mathcal{T}^e_{[0,0]}, \mathcal{T}^e_{[0,1]}\},$$

so the probability that a random tree will be mapped to a tree of topology $\mathcal{T}^e$ is just given by,

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1}B\big(\boldsymbol{p}(\mathcal{T}^e_{[0,0]}) \otimes \boldsymbol{p}(\mathcal{T}^e_{[0,1]})\big), \tag{5.3.17}$$

where $(-D_0)^{-1}B$ is the probability that the root branch will eventually undergo a branch point, and $\boldsymbol{p}(\mathcal{T}^e_{[0,0]})$ and $\boldsymbol{p}(\mathcal{T}^e_{[0,1]})$ are the probabilities that the daughter and parental subtrees will eventually have topologies $\mathcal{T}^e_{[0,0]}$ and $\mathcal{T}^e_{[0,1]}$. Since each event is independent we multiply them together, and the Kronecker product reflects the fact that we need to keep track of the independent evolution of the two branches emanating from the first branch point.

More explicitly then,

$$\begin{bmatrix} p_1(\mathcal{T}^e) \\ p_2(\mathcal{T}^e) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{1-2b(1-\gamma)p(\emptyset)} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & b\gamma & 0 & b(1-\gamma)\big(1-p(\emptyset)\big) \end{bmatrix} \begin{bmatrix} p_1(\mathcal{T}^e_{[0,0]})p_1(\mathcal{T}^e_{[0,1]}) \\ p_1(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}) \\ p_2(\mathcal{T}^e_{[0,0]})p_1(\mathcal{T}^e_{[0,1]}) \\ p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}) \end{bmatrix}. \tag{5.3.18}$$

Any tree that commences in phase 1 can only ever reach a size of one, thus $p_1(\mathcal{T}^e) = 0$ if $\big|\mathcal{T}^e\big| \geq 2$. As a result only $p_2(\mathcal{T}^e)$ is non-zero for $\big|\mathcal{T}^e\big| \geq 2$, thus

$$p_2(\mathcal{T}^e) = \frac{1}{1-2b(1-\gamma)p(\emptyset)}\left(b\gamma p_1(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}) + b(1-\gamma)\big(1-p(\emptyset)\big)p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]})\right). \tag{5.3.19}$$

The first term of equation (5.3.19) is non-zero only if $\mathcal{T}^e_{[0,0]}$ is a single branch topology since it commences in phase 1, and in this case $p_1(\mathcal{T}^e_{[0,0]}) = p_1(\big|\mathcal{T}^e_{[0,0]}\big| = 1) = 1$. We therefore re-write the above equation to reflect this,

$$\begin{aligned} p_2(\mathcal{T}^e) \;=\; & \frac{1}{1-2b(1-\gamma)p(\emptyset)}\left(b\gamma I\{\big|\mathcal{T}^e_{[0,0]}\big| = 1\}p_2(\mathcal{T}^e_{[0,1]}) + \right. \\ & \left. b(1-\gamma)\big(1-p(\emptyset)\big)p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]})\right). \end{aligned} \tag{5.3.20}$$

Re-arranging equation (5.3.20) leaves

$$p_2(\mathcal{T}^e) = \frac{b(1-\gamma)\big(1-p(\emptyset)\big)}{1-2b(1-\gamma)p(\emptyset)}\left(\frac{\gamma}{(1-\gamma)\big(1-p(\emptyset)\big)}I\{\big|\mathcal{T}^e_{[0,0]}\big|=1\}p_2(\mathcal{T}^e_{[0,1]})\right.$$
$$\left.+p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]})\right). \tag{5.3.21}$$

The first term in the brackets of equation (5.3.21) can be re-written as

$$\frac{\gamma}{(1-\gamma)\big(1-p(\emptyset)\big)p_2(\big|\mathcal{T}^e_{[0,0]}\big|=1)}I\{\big|\mathcal{T}^e_{[0,0]}\big|=1\}p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}), \tag{5.3.22}$$

because of the indicator function $I\{\big|\mathcal{T}^e_{[0,0]}\big|=1\}$. Combining equations (5.3.21) and (5.3.22) we obtain,

$$p_2(\mathcal{T}^e) = \frac{b(1-\gamma)\big(1-p(\emptyset)\big)}{1-2b(1-\gamma)p(\emptyset)}\left(\frac{\gamma}{(1-\gamma)\big(1-p(\emptyset)\big)p_2(\big|\mathcal{T}^e_{[0,0]}\big|=1)}I\{\big|\mathcal{T}^e_{[0,0]}\big|=1\}+1\right)$$
$$\times p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}). \tag{5.3.23}$$

We can now extract the indicator function in equation (5.3.23) and place it as the exponent of the expression in brackets,

$$p_2(\mathcal{T}^e) = \frac{b(1-\gamma)\big(1-p(\emptyset)\big)}{1-2b(1-\gamma)p(\emptyset)}\left(\frac{\gamma}{(1-\gamma)\big(1-p(\emptyset)\big)p_2(\big|\mathcal{T}^e_{[0,0]}\big|=1)}+1\right)^{I\{|\mathcal{T}^e_{[0,0]}|=1\}}$$
$$\times p_2(\mathcal{T}^e_{[0,0]})p_2(\mathcal{T}^e_{[0,1]}). \tag{5.3.24}$$

Suppose the tree of topology $\mathcal{T}^e$ has size $\big|\mathcal{T}^e\big| = s$ with $v$ unitary branch points, then using an inductive argument, equation (5.3.24) can be shown to be equivalent to,

$$p_2(\mathcal{T}^e) = \left(\frac{b(1-\gamma)\big(1-p(\emptyset)\big)}{1-2b(1-\gamma)p(\emptyset)}\right)^{s-1}\left(\frac{\gamma}{(1-\gamma)\big(1-p(\emptyset)\big)p_2(\big|\mathcal{T}^e_{[0,0]}\big|=1)}+1\right)^{v}\big(p_2(1)\big)^s. \tag{5.3.25}$$

Clearly,

$$p_2(\big|\mathcal{T}^e_{[0,0]}\big|=1) = p_2(1) = \frac{q+b\gamma p(\emptyset)}{\big(1-p(\emptyset)\big)\big(1-2b(1-\gamma)p(\emptyset)\big)},$$

and incorporating this into equation (5.3.25) the final result is,

$$p_2(\mathcal{T}^e) = \frac{1}{1-p(\emptyset)}\beta^{s-1}\left(\frac{\gamma}{(1-\gamma)}\frac{1}{\frac{q+b\gamma p(\emptyset}{1-2b(1-\gamma)p(\emptyset)}}+1\right)^{v}\left(\frac{q+b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)}\right)^{s}, \tag{5.3.26}$$

where $\beta = \frac{b(1-\gamma)}{1-2b(1-\gamma)p(\emptyset)}$. This is precisely equation (3.11.17), except for the factor of $1/(1-p(\emptyset))$. Again, for an explanation, see the text immediately following equation (5.3.13) in Section 5.3.2.

## 5.4 The MBT and the Multi-Rate Model

Simple evolutionary models such as the constant-rates BD, the PDA and the sPDA can be shown to be special cases of the MR model. All three of these models are characterised, in the MR context, by the fact that they have only one unstable phase, one quasi-stable phase and one extinct phase. These models are scalar (as there is only one phase per class) and can be analysed. However, in order to provide for better macroevolutionary models clearly more complex MRs are required, since these simple models do not yield imbalances that are consistent with observation. Non-scalar MR models have the major drawback that they cannot be conveniently analysed because they do not have a representation that allows such an analysis to be performed, see Chapter 3.

Even if such a representation existed, the MR model has one other complication, the distinction between quasi-stability and extinction makes any transient analysis of the process very difficult to perform. Branches that become extinct must be pruned from the tree. The consequences of pruning are such that the death of only a few branches may have profound effects on the entire topology of the tree. Figure 5.4.1 depicts a tree that evolves under a MR model and at two different times we perform the necessary pruning of the extinct branches to ascertain the tree topology. Tree, $\mathcal{T}^a$, is the complete tree with unstable, extinct and quasi-stable branches, tree $\mathcal{T}_1^u$ is the unstable topology of the tree after a pruning is performed at time $t_1$ and tree $\mathcal{T}_2^u$ is the unstable topology of the tree after pruning is performed at $t_2$. Topologies $\mathcal{T}^a$ and $\mathcal{T}_1^u$ are very different; pruning the extinct branches has the effect of producing longer branches, since any transition that generates an extinct branch is not treated as a branch point after pruning. Thus the rate of undergoing a true

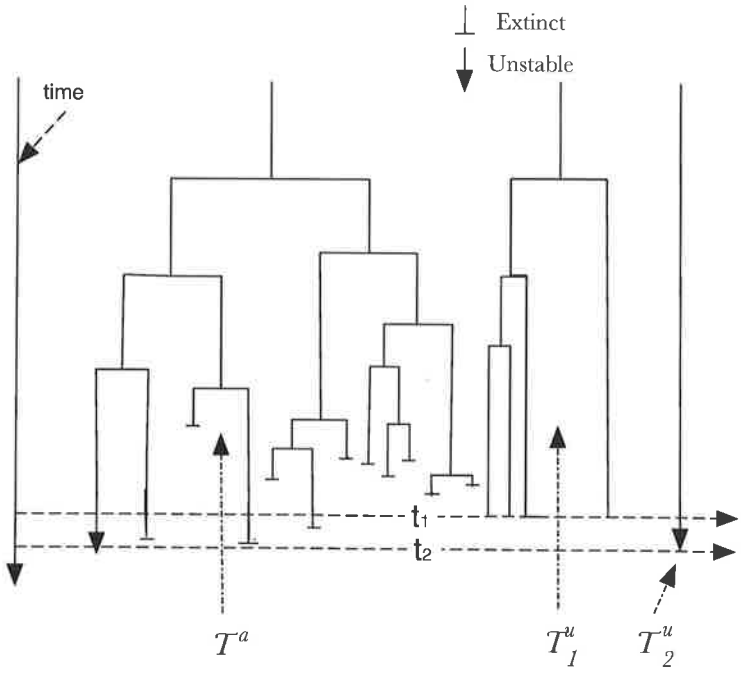


Figure 5.4.1: An example of the pruning required for MR trees.

branch point, that is, a branch point that yields two non-extinct branches, is clearly less than the rate of undergoing just a branching event. Topologies $\mathcal{T}_1^u$ and $\mathcal{T}_2^u$ are also very noticeably different; the death of three branches between $t_1$ and $t_2$, has changed a four branch topology from the second topologically isomorphic class of size four trees to a single branch topology. In other words, after pruning all extinct branches at $t_2$ the tree has not undergone even one true branch point!

This highlights the critical problem that the MR faces by pruning extinct branches: in order to prune correctly at any time, a complete historical knowledge of the tree is required because of the profound effects of even a small number of extinctions. From a mathematical perspective then, to avoid these issues one studies the process in the limit as $t \to \infty$ and then prunes the extinct branches.

Now, because the MR lacks a representation from which a useful analysis can be performed in the limit as $t \to \infty$, we shall express the model in two different ways that render it more transparent to analysis. The MR model can be transformed

into an MBT-like representation or directly into a MBT representation. We shall discuss both methods and then show that the MBT representation of the MR is the preferred option.

### 5.4.1 The MBT Representation of the MR Model

As a first step we shall directly transform the MR model into a standard MBT representation, called the MBT-MR model. The impetus behind such an approach stems from the success with which the MR-PDA and the MR-sPDA models were transformed to MBT processes. Because the procedure we use in this section is a generalisation of the methods employed in Sections 5.3.2 and 5.3.3, the MBT-MR model reduces to those models when the phase space consists of one quasi-stable and one unstable phase.

The model is correctly formulated by

- discarding all the extinction phases of the MR model as they are not counted in the topologies, and
- mapping all the quasi-stable phases to a holding phase, say 1. From this phase absorption subsequently ensues with probability one.

As a result, the space we are concerned with here is the space of MR trees that are finite with all branches being quasi-stable. Once again, since extinction in the MR sense is not allowed in the MBT model, to transform the rates from the MR model to the MBT domain we must therefore divide each rate by the appropriate $1-p_i(\emptyset)$, see Sections 5.3.2 and 5.3.3. This implies that we are interested in the distribution of the space of random trees from the MR domain whose quasi-stable portions are mapped to $\mathcal{T}^q$, and not the measure of the subset of trees that correspond to extinct trees in the MR sense.

As before, let $\mathcal{U}$ be the space of unstable phases, and $\mathcal{Q}$ be the space of quasi-

stable phases. The absorption rate from any phase $i \in \mathcal{U}$ is given by,

$$d_i = \frac{\sum_{q\in\mathcal{Q}} \mu_{iq} + \sum_{q\in\mathcal{Q}} \sigma_{iq} p_i(\emptyset)}{1 - p_i(\emptyset)}. \tag{5.4.1}$$

This equation tells us that if a branch is in phase $i \in \mathcal{U}$ then it will become absorbed if it undergoes

1. a direct transformation from $i$ to any $q \in \mathcal{Q}$, or,

2. a branch point occurs such that the daughter is spawned in some phase, $q \in \mathcal{Q}$, whilst the parent branch, still in phase $i$ immediately after the branch point, subsequently generates a subtree that becomes extinct in the MR sense.

We sum over all $\mathcal{Q}$ since we are not interested in which quasi-stable phase it enters but merely that it enters one. Thus the absorption vector $\boldsymbol{d}$ is given by,

$$\boldsymbol{d} = \begin{bmatrix} 1 \\ \boldsymbol{d}_{\mathcal{U}} \end{bmatrix}, \tag{5.4.2}$$

where $\boldsymbol{d}_{\mathcal{U}}$ is the absorption vector for the unstable phases and is of dimension $N(\mathcal{U}) \times 1$, and where the first element represents the fact that a branch in phase 1 will be absorbed with probability one.

The rate at which branch points occur from $i \in \mathcal{U}$ to a daughter branch in some phase $u \in \mathcal{U}$ is given by,

$$B_{i,ui} = \frac{\sigma_{iu}\big(1 - p_u(\emptyset)\big)\big(1 - p_i(\emptyset)\big)}{1 - p_i(\emptyset)}. \tag{5.4.3}$$

A parent branch in phase $i$ gives birth to a daughter branch in phase $u$ with rate $\sigma_{iu}$ and since we wish that both branches independently never become extinct, $\sigma_{iu}$ is multiplied by the probability of non-extinction which is $\big(1 - p_u(\emptyset)\big)\big(1 - p_i(\emptyset)\big)$. The rate at which a parent branch in phase $i \in \mathcal{U}$ will give birth to a daughter branch in phase 1, is given by

$$B_{i,1i} = \frac{\sum_{q\in\mathcal{Q}} \sigma_{iq}\big(1 - p_i(\emptyset)\big)}{1 - p_i(\emptyset)}. \tag{5.4.4}$$

The total tate at which a parent branch gives birth to a quasi-stable daughter branch occurs with rate $\sum_{q\in\mathcal{Q}}\sigma_{iq}$. We require that the parent never becomes extinct (in the MR sense) and to ensure this happens we multiply by $\big(1-p_i(\emptyset)\big)$.

To illustrate the transition structure, consider an example where there are two unstable states denoted by $\mathcal{U}=\{2,3\}$ in the MR, which are mapped to phases 2 and 3 in the MBT, and two MR quasi-stable states $\mathcal{Q}=\{4,5\}$ mapped to the phase 1 in the MBT. In this instance the matrix $B$ is of the form,

| | 12 | 13 | 21 | 22 | 23 | 31 | 32 | 33 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | $\sigma_{24}+\sigma_{25}$ | 0 | 0 | $\sigma_{22}(1-p_2(\emptyset))$ | 0 | 0 | $\sigma_{23}(1-p_3(\emptyset))$ | 0 |
| 3 | 0 | $\sigma_{34}+\sigma_{35}$ | 0 | 0 | $\sigma_{32}(1-p_2(\emptyset)$ | 0 | 0 | $\sigma_{33}(1-p_3(\emptyset)$ |

where for space reasons we have excluded the 11 column, which has only zero entries. Let $(D_0)_{iu}$, for $i,u\in\mathcal{U}$, $i\neq u$, be the rate at which a hidden transition occurs from phase $i$ to phase $u$. This is given by,

$$(D_0)_{iu}=\frac{\mu_{iu}}{1-p_i(\emptyset)}+\frac{\sigma_{iu}\big(1-p_u(\emptyset)\big)p_i(\emptyset)}{1-p_i(\emptyset)}. \tag{5.4.5}$$

This equation also has a similar interpretation: a branch in phase $i$ can undergo a hidden transition into phase $u$ either directly via $\mu_{iu}$ or indirectly via giving birth to a daughter branch in phase $u$, whilst the parent branch eventually becomes extinct (in the MR sense); once the parent is pruned all that remains is a single branch in phase $u$. The diagonal elements of the matrix $D_0$ are given by,

$$(D_0)_{ii}=-\left(d_i+\sum_{u\in\mathcal{U}}(D_0)_{iu}+\sum_{u\in\{1,\,\mathcal{U}\}}B_{i,ui}\right), \tag{5.4.6}$$

for $i\in\mathcal{U}$ and $(D_0)_{11}=-1$. The total rate at which a branch may undergo a birth is given by the sum of all the possible eventual outcomes of a birth. These outcomes are,

- both subtrees become extinct, or

- the daughter branch is quasi-stable and the parental subtree does not become extinct, or
- the daughter branch is quasi-stable and the parental subtree does become extinct, or
- the daughter branch eventually becomes extinct and the parental subtree does not,
- the daughter branch does not become extinct while the parental subtree does, or finally,
- neither subtree becomes extinct.

Thus, it is relatively straight-forward to show that the rate of giving birth, is given by,

$$\begin{aligned}\sum_{q\in\mathcal{Q}}\sigma_{iq}+\sum_{u\in\mathcal{U}}\sigma_{iu} &= \sum_{u\in\mathcal{U}}\sigma_{iu}p_u(\emptyset)p_i(\emptyset)+\sum_{q\in\mathcal{Q}}\sigma_{iq}\big(1-p_i(\emptyset)\big)+\sum_{q\in\mathcal{Q}}\sigma_{iq}p_i(\emptyset)\\ &\quad+\sum_{u\in\mathcal{U}}\big(\sigma_{iu}p_u(\emptyset)\big(1-p_i(\emptyset)\big)+\sigma_{iu}\big(1-p_u(\emptyset)\big)p_i(\emptyset)\big)\\ &\quad+\sum_{u\in\mathcal{U}}\sigma_{iu}\big(1-p_u(\emptyset)\big)\big(1-p_i(\emptyset)\big). \qquad (5.4.7)\end{aligned}$$

Now using equation (5.4.7) it can be shown that (5.4.6) is equal to,

$$(D_0)_{ii}=-\left(1-\left(\sum_{u\in\mathcal{U},\ u\neq i}\sigma_{iu}p_u(\emptyset)+2\sigma_{ii}p_i(\emptyset)\right)\right). \qquad (5.4.8)$$

It is also easy to see that if there is only one unstable phase, then equation (5.4.8) reduces to equation (5.3.14).

The probability that a random tree eventually has a topology of $\mathcal{T}^e$ is given by

$$\boldsymbol{p}(\mathcal{T}^e)=(-D_0)^{-1}B\left(\boldsymbol{p}(\mathcal{T}^e_{[0,0]})\otimes\boldsymbol{p}(\mathcal{T}^e_{[0,1]})\right). \qquad (5.4.9)$$

Equation (5.4.9) can be solved analytically using a simple recursion on the subtree topologies. Let $\mathcal{T}^e$ be a topology of size $s$, then the recursion to solve for $\mathcal{T}^e$ is

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1}\boldsymbol{d} \text{ if } \left|\mathcal{T}^e\right|=1 \qquad (5.4.10)$$

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1}B\left(\boldsymbol{p}(\mathcal{T}^e_{[0,0]})\otimes\boldsymbol{p}(\mathcal{T}^e_{[0,1]})\right) \text{ otherwise,} \qquad (5.4.11)$$

where $\left|\mathcal{T}^e_{[0,0]}\right| = j$ and $\left|\mathcal{T}^e_{[0,1]}\right| = s - j$.

It can be seen that the direct transformation from the MR to the standard MBT model generates rather complicated transition rates between phases. The alternative is to transform the MR into an MBT-like model. Such a transformation is performed in the next section where it will be seen that complicated transition rates are small price to pay compared to the difficulty in solving the equations that give the probability of any topology, even for the one branch topology!

### 5.4.2 The MBT-like Representation of the MR model

In the MBT-like representation we

1. group all the quasi-stable states into phase $-1$. This phase is effectively a second absorbing phase, the rate of absorption into this phase from any unstable phase $i \in \mathcal{U}$ is $\sum_{q\in\mathcal{Q}} \mu_{iq}$, and

2. group all the extinct states into the second absorbing phase, 0. The rate of absorption into this phase from $i \in \mathcal{U}$ is given by $\sum_{d\in\mathcal{D}} \mu_{id}$. Any branch that eventually enters phase 0 is pruned from the tree.

Thus the reason why we call this an MBT-like structure is due to the presence of two absorbing phases whereas an MBT is defined with only one absorbing phase. We shall call this model the MBT-like-MR (MBTl-MR) model. Furthermore, the $-1$ phase is not treated as a traditional absorbing phase, since a daughter may be spawned in phase $-1$. Let $N(\mathcal{U})$ be the number of states in $\mathcal{U}$. The transformed model has the following hidden transition structure,

$$\begin{array}{c|ccc} & -1 & 0 & \mathcal{U} \\ \hline -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathcal{U} & \boldsymbol{a} & \boldsymbol{d} & D_0 \end{array}, \tag{5.4.12}$$

where $a_i = \sum_{q\in\mathcal{Q}} \mu_{iq}$, $d_i = \sum_{d\in\mathcal{D}} \mu_{id}$ and where $D_0$ is the matrix of internal (hidden) transitions between the states that are in $\mathcal{U}$. The dimension of $D_0$ is $N(\mathcal{U}) \times N(\mathcal{U})$. The non-diagonal elements of $D_0$ are, $(D_0)_{ij} = \mu_{ij}$ if $i \neq j$ and the diagonal elements are given by,

$$(D_0)_{ii} = -\left(\sum_{q\in\mathcal{Q}} \mu_{iq} + \sum_{d\in\mathcal{D}} \mu_{id} + \sum_{u\in U} \mu_{iu} + \sum_{u\in\{-1,\mathcal{U}\}} \sigma_{ui}\right). \tag{5.4.13}$$

At this point we would like to point out that from an analytical perspective, it would be much easier to just use the $-1$ phase like a holding phase, as we did in the MBT version of the MR-sPDA model. However, because we wish to keep a distinction between multi-rate quasi-stability and extinction we cannot do this. We shall get around this by introducing the $(N(\mathcal{U})+1) \times (N(\mathcal{U})+1)$ matrix, $D$, such that,

$$\bar{D} = \begin{bmatrix} 0 & \mathbf{0}^t \\ \mathbf{0} & D_0 \end{bmatrix}, \tag{5.4.14}$$

where $\mathbf{0}^t$ is the transpose of the zero vector $\mathbf{0}$. The first row now corresponds to the phase, $-1$. Now any branch that enters phase $-1$ will never exit it. Let,

$$\hat{\boldsymbol{d}} = \begin{bmatrix} 0 \\ \boldsymbol{d} \end{bmatrix}, \tag{5.4.15}$$

be the vector that gives the rates of extinction from each phase. Notice how $d_{-1}$ is set to zero which distinguishes $-1$ from phase 0, because a branch in $-1$ remains in this phase and it cannot become extinct, that is, enter phase 0. We shall also see that there is no communication between $-1$ and 0 via the matrix $B$, the matrix governing observable transitions. Consider, now the observable transitions. The observable transitions are those transitions that generate a distinct daughter branch. The transitions where a parent branch, in phase $i$, spawns a daughter branch in phase $j \in \{-1, \mathcal{U}\}$ whilst the parental branch remains in phase $i$, occur at rate $\sigma_{ij}$ if $j \in \mathcal{U}$ and $\sigma_{ij} \equiv \sum_{q\in\mathcal{Q}} \sigma_{iq}$ if $j = -1$. Therefore,

$$B_{i,ji} = \sigma_{ij},\ j \in \{-1, \mathcal{U}\}, \tag{5.4.16}$$

and $B_{i,jk} = 0$ for $k \neq i$. Furthermore, $B_{-1,ij} = 0$ for all $i, j \in \{-1, \mathcal{U}\}$. So, the matrix $B$ is of dimension $(N(\mathcal{U}) + 1) \times (N(\mathcal{U}) + 1)^2$. This process is conservative,

$$D\boldsymbol{e} + B\boldsymbol{e} + \hat{\boldsymbol{d}} = \mathbf{0}. \tag{5.4.17}$$

The internal dynamics of each branch is governed by the matrix $D_0$ just as in the ordinary MBT. The matrix $D_0$ is invertible so an analysis as $t \to \infty$ is possible, however the matrix $\bar{D}$ is not invertible. We introduce the matrix $D$, which is defined by,

$$D = \begin{bmatrix} 0 & \mathbf{0}^t \\ \mathbf{0} & (D_0)^{-1} \end{bmatrix}. \tag{5.4.18}$$

The purpose of $D$ is to keep the dimensions of the matrices in the expressions to follow from becoming too skewed.

The space of extinct trees consists of those trees whose branches have all been absorbed in phase 0. Let $p_i(\emptyset)$, for all $i \in \{-1, \mathcal{U}\}$, be the probability that a tree commencing in phase $i$ eventually becomes extinct. The equation for $\boldsymbol{p}(\emptyset)$ is,

$$\boldsymbol{p}(\emptyset) = (-D)^{-1}\hat{\boldsymbol{d}} + (-D)^{-1}B\boldsymbol{p}(\emptyset) \otimes \boldsymbol{p}(\emptyset). \tag{5.4.19}$$

A tree may become extinct by direct absorption into 0 from the root branch, with probability given by the first term of equation (5.4.19) or by first undergoing a birth and then having both subtrees subsequently becoming extinct independently and this is given by the second term in equation (5.4.19). Note that the probability that a tree becomes extinct given that its root branch began in phase $-1$, $p_{-1}(\emptyset)$, is clearly equal to 0, since the phases $-1$ and 0 never communicate. It will be seen in Chapter 7 that $\boldsymbol{p}(\emptyset)$ can be found algorithmically.

Recall, from Chapter 3 that the topology of the quasi-stable portion of a tree was denoted by $\mathcal{T}^q$, in this chapter, we continue with nomenclature here. Let $p_i(\mathcal{T}^q)$, $i \in \{-1, U\}$, be the probability that a random tree commencing life in phase $i$ will eventually attain a topology of $\mathcal{T}^q$. Note that since we are studying the distribution as $t \to \infty$ and we are assuming that the process is subcritical, the only trees that

have a non-zero probability of occurring are those that have a finite number of extinct and quasi-stable branches and zero unstable branches. Now, there exists only one topology with $\left|\mathcal{T}^q\right| = 1$, and so we often write, $\boldsymbol{p}(1)$, for the probability that a random tree evolves to eventually attain a one branch (quasi-stable) topology. We can write the expression for this probability, and it is given by,

$$\boldsymbol{p}(1) = (-D)^{-1}\hat{\boldsymbol{a}} + (-D)^{-1}B(\boldsymbol{p}(\emptyset)\otimes\boldsymbol{p}(1)) + (-D)^{-1}B(\boldsymbol{p}(1)\otimes\boldsymbol{p}(\emptyset)) + \boldsymbol{e}_{-1}, \quad (5.4.20)$$

where

$$\hat{\boldsymbol{a}} = \begin{bmatrix} 0 \\ \boldsymbol{a} \end{bmatrix},$$

and $\boldsymbol{e}_{-1}$ is a vector of zeros, except for the $-1$ position which is one. The interpretation for equation (5.4.20) is straightforward. A tree can attain a one branch quasi-stable topology in two ways, the first is directly via a phase transition to $-1$, and this has probability $(-D)^{-1}\hat{\boldsymbol{a}}$, and the second is via a birth such that either the daughter or the parent subtrees eventually become extinct and the other eventually attains a one branch topology, and this has probability $(-D)^{-1}B\left(\boldsymbol{p}(\emptyset)\otimes\boldsymbol{p}(1)\right) + (-D)^{-1}B\left(\boldsymbol{p}(1)\otimes\boldsymbol{p}(\emptyset)\right)$. Now if the tree commences in phase $-1$, it is with probability 1 that the tree will be of size 1, hence the presence of $\boldsymbol{e}_{-1}$. More generally, $p_{-1}(\mathcal{T}^q) = I\{\left|\mathcal{T}^q\right| = 1\}$ since any branch commencing in phase $-1$ cannot give birth.

The equation for the probability that a random tree will attain a topology of $\mathcal{T}^q$ for $\left|\mathcal{T}^q\right| \geq 2$ is given by,

$$\begin{aligned}\boldsymbol{p}(\mathcal{T}^q) &= (-D)^{-1}B\left(\boldsymbol{p}(\mathcal{T}^q_{[0,0]})\otimes\boldsymbol{p}(\mathcal{T}^q_{[0,1]})\right) + \\ &\quad +(-D)^{-1}B\left(\boldsymbol{p}(\mathcal{T}^q)\otimes\boldsymbol{p}(\emptyset)\right) + (-D)^{-1}B\left(\boldsymbol{p}(\emptyset)\otimes\boldsymbol{p}(\mathcal{T}^q)\right). \quad (5.4.21)\end{aligned}$$

This equation illustrates the fact that a tree with eventual topology $\mathcal{T}^q$ achieves this by giving birth

1. to two non-extinct subtrees of topologies $\mathcal{T}^q_{[0,0]}$ and $\mathcal{T}^q_{[0,1]}$ such that $\mathcal{T}^q = \{([\alpha(0)],[0])^{(i)}, \mathcal{T}^q_{[0,0]}, \mathcal{T}^q_{[0,1]}\}$, or,

2. such that the parent or daughter subtree becomes extinct, whilst the other evolves into a tree of topology $\mathcal{T}^q$.

One must be very careful in expanding equation (5.4.21). Recall in Chapter 3, the equation for the probability that a random tree eventually has topology $\mathcal{T}^q$ in the MR-sPDA model is given by,

$$p(\mathcal{T}^q) = b(1-\gamma)p(\mathcal{T}^q_{[0,0]})p(\mathcal{T}^q_{[0,1]}) + 2bp(\mathcal{T}^q)p(\emptyset) + b\gamma p(\mathcal{T}^q_{[0,1]})I\{\left|\mathcal{T}^q_{[0,0]}\right| = 1\}. \quad (5.4.22)$$

Notice that the last term in this equation gives the probability that the daughter branch at the first branch point is born quasi-stable and hence can only have a one branch topology, so if $\mathcal{T}^q$ has a daughter subtree that has a size greater than 2, this term is zero. Now, since equation (5.4.21) is the matrix generalization of the above equation, when we expand it there will be a term from $(-D)^{-1}B\hat{\boldsymbol{p}}(\mathcal{T}^q_{[0,0]}) \otimes \hat{\boldsymbol{p}}(\mathcal{T}^q_{[0,1]})$ which will be non-zero only if the daughter subtree is of size 1.

The effects of disregarding extinct branches when mapping to the topological domain, can be seen from equations (5.4.20) and (5.4.21). The last two terms from both equations represent the scenario that, following a birth, either the daughter or the parental subtree becomes extinct, and is subsequently pruned, whilst the other subtree evolves to a topology of $\mathcal{T}^q$.

The last two terms in equation (5.4.20) cannot be gathered because the Kronecker product is non-commutative and so an analysis such as that performed for the simple sPDA model cannot be applied, see [26] and Chapter 3. The equations for the dynamics of this process are difficult to solve analytically even in the single branch topology case because of the complicating effects of pruning extinct branches. Algorithms can be developed to solve equations (5.4.20) and (5.4.21)

The problems associated with pruning were not an issue in the standard MBT representation for the MR model since all the extinct states from the MR model were discarded and quasi-stability was treated as extinction. The distribution of the standard MBT model is identical to the distribution, $p(\mathcal{T}^q)$ of the MR model

restricted to the space of random trees whose quasi-stable portions are mapped to $\mathbb{T}$.

To illustrate the simplicity in using the MBT-MT model, consider the three branch topology depicted in Figure 5.4.2. The probability of attaining this topology

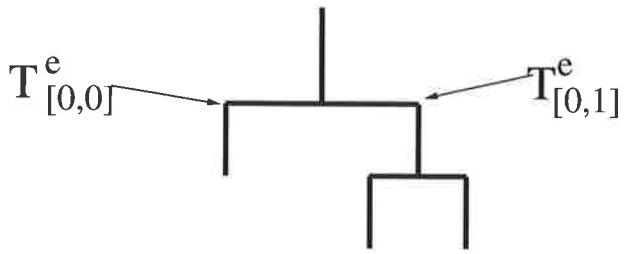


Figure 5.4.2: A three branch topology

is given by,

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1} B \left( \boldsymbol{p}(\mathcal{T}^e_{[0,0]}) \otimes \boldsymbol{p}(\mathcal{T}^e_{[0,1]}) \right), \tag{5.4.23}$$

where $\left|\mathcal{T}^e_{[0,0]}\right| = 1$ and $\left|\mathcal{T}^e_{[0,1]}\right| = 2$. Since the daughter topology consists of a single branch we have that

$$\boldsymbol{p}(\mathcal{T}^e_{[0,0]}) = (-D_0)^{-1}\boldsymbol{d}, \tag{5.4.24}$$

and since the topology of the parental subtree is a two branch topology, using equations (5.4.11) and (5.4.10) we obtain,

$$\boldsymbol{p}(\mathcal{T}^e_{[0,1]}) = (-D_0)^{-1} B \Big[ (-D_0)^{-1}\boldsymbol{d} \otimes (-D_0)^{-1}\boldsymbol{d} \Big]. \tag{5.4.25}$$

Substituting the expressions for $\boldsymbol{p}(\mathcal{T}^e_{[0,0]})$ and $\boldsymbol{p}(\mathcal{T}^e_{[0,1]})$, into equation (5.4.23) we finally obtain,

$$\boldsymbol{p}(\mathcal{T}^e) = (-D_0)^{-1} B \left( (-D_0)^{-1}\boldsymbol{d} \otimes (-D_0)^{-1} B \Big[ (-D_0)^{-1}\boldsymbol{d} \otimes (-D_0)^{-1}\boldsymbol{d} \Big] \right) \tag{5.4.26}$$

This example demonstrates the power of transforming the MR into the MBT-MR model. The three branch topology has an analytical solution whereas the single branch topology in the MBTL-MR model requires algorithmic methods to solve!

The implicit nature of equations (5.4.20) and (5.4.21) represents the interaction between the full topology and extinct subtrees are absent in the MBT-MR equations; this is the result of neglecting the extinct states in the MBT-MR model. This difference is most striking because equation (5.4.11) can be solved using recursion and matrix multiplication for all $s \geq 2$, whereas the most simple equation in the MBTL-MR model, equation (5.4.20), is implicit and requires an algorithmic scheme to solve! Finally, the asymptotic distribution of the much more simple MBT-MR model is still identical to that of the MR model.

The MR-PDA, and MR-sPDA models consider only topologies that consist entirely of a finite number of quasi-stable branches, almost surely. The crBD model considers topologies that consist entirely of unstable branches, and the crBD interpretation of the PDA model (crBD-PDA) considers topologies that consist of a finite number of extinct branches, almost surely. The MBT representation of all the above models considers topologies that consist of a finite number of extinct branches, almost surely, except for the crBD model which considers unstable topologies. In order to be completely unambiguous about the types of topologies we are considering when having discussed each model, we have been meticulous in denoting the type of the topology of a tree as $\mathcal{T}^z$ and the type of a topologically isomorphic class of size $s$ as $\mathbb{F}_{i,s}^z$, where $z = a, u, q, e$. From this point and throughout the remainder of the thesis we shall no longer make such a designation as all models will be discussed in terms of the MBT in which case the types of the topologies that are being considered are unambiguous.

The first five chapters have seen us discuss the need to provide useful models of the macroevolutionary process. The most simple models have been shown to be deficient in a number of areas, particularly when one factors in the need to generate topologies with imbalances that mimic those of phylogenetic trees. The most complex model to date, has been the MR model of Pinelis [26]. In theory it provides a significant amount of flexibility that can account for the variable imbalances found in nature. However, this model suffers from some quite fundamental problems. The

MR model of Pinelis, which is a special case of a ctMMTBP, cannot be analysed in a practical fashion. There is a well developed theoretical approach to the ctMMTBP, however very little of this theory provides any practical use in a modelling context. Furthermore, the MR model requires one to prune extinct branches from the topology, an extremely difficult task to achieve in any transient analysis, and still of considerable difficulty in the limit as $t \to \infty$. Thus, in order,

1. to provide a physically reasonable, flexible model of macroevolution, and

2. to provide a model that is amenable to practical algorithmic analysis,

we have given an alternative representation of the binary-branch point ctMMTBP in the language of QBDs, known as the MBT. In doing this, we have given to the binary-branch point ctMMTBP, which is a structure that has very little correlation between particle lifetime and offspring distribution, a very rich correlation structure. The correlations are caused by considering the evolution of each branch in terms of a MAP.

We have thus far shown that the MBT model subsumes all the simple macroevolutionary models. Furthermore, the flexibility of the MBT is such that we have provided two alternative representations for the MR model. The first representation is in terms of an MBT-like model and the second representation is in terms of the standard MBT model.

Our next step is to begin demonstrating the power of the algorithmic approach and the flexibility of the MBT as a model. In Chapter 6 we provide an algorithm that calculates the distribution of imbalance conditioned on tree size. Using this distribution we than calculate the mean of Colless' index of imbalance. We show that there exists a one parameter family of MBT models that can generate mean imbalances that have a range spanning from a lower mean imbalance than the crBD model all the way to the theoretical maximal mean imbalance. The correlations that abound in the MBT domain, due to the MAP, produce interesting effects that can be exploited in developing a suitable macroevolutionary model.

# Chapter 6

# Probability Distribution of Imbalance

## 6.1 Introduction

Chapter 3 provided an introduction to the types of models that have been utilised in current phylogenetic research. As stated in that chapter, the imbalance of phylogenetic trees is playing an increasingly important role. The reason is that the imbalance can be used to infer information on how the rates of speciation or extinction have changed over time, because higher or lesser imbalances are likely to have been generated by variations in these rates. The simplest macroevolutionary models predict imbalances that are either too low, for example the constant rates BD (crBD) model [10, 22, 30], or too high, for example the PDA model [22, 26, 30]. For a graphical representation of this point for varying tree size, see Figure 5 from Rogers [30]. The reason why neither of these simple models can predict the correct imbalances is that they do not have the capacity to allow the rates to vary across a tree; the rates of speciation and extinction are fixed. As a result, the ability to vary rates over time and throughout a topology gives one the propensity to generate topologies with the imbalances that are more closely aligned with those found in

nature. More recently, there has been a greater emphasis on developing models that allow for rate variation. The most sophisticated model to date, that proposed by Pinelis [26], is a special case of the ctMMTBP.

The MBT model, as developed in Chapter 5, is also based on the ctMMTBP. However, a novel interpretation and representation allows us to develop complex interactions between the phases of each branch, and as a result, allows for significant rate variation in different parts of the tree and at different times. Furthermore, the MBT model subsumes all of the other important macroevolutionary models that have been discussed here, including the multi-rate model of Pinelis.

The purpose of this chapter is to demonstrate further this flexibility by developing an algorithm that can calculate the mean imbalance for the MBT when we condition on the size of the trees we are interested in. This follows the work of Rogers [30] who developed algorithms to determine the moments of the imbalance for the constant rates BD and PDA models, when conditioned on tree size. We will further show that we can find a very simple one-parameter family of MBTs that has the flexibility to produce the entire range of theoretically possible mean imbalances.

In Section 6.2 we discuss the algorithm for determining the mean imbalance of an MBT model given its size. In Section 6.3 we discuss the mean imbalance for the simple models, crBD, PDA, s-PDA and Completely Unbalanced (a special case of the sPDA model), and then using the algorithm from Section 6.2 we show that there exists a one-parameter family of MBTs that has the flexibility to generate any mean imbalance. Finally, in Section 6.4 we discuss the computational complexity of the algorithm.

## 6.2 The Imbalance Algorithm

In this section, we wish to develop an algorithm that can calculate the mean imbalance, using Colless' index of imbalance, for the MBT model. To do this, we must first calculate the distribution of imbalance, conditioned on tree size.

We know that for any given size $s$, Colless' index of imbalance must lie in the set, $\{0, 1, \ldots, (1/2)(s-1)(s-2)\}$. Let $\mathbb{C}_{i,s}$ denote the set of trees of size $s$ that have imbalance $i$ for all $i \in \{0, 1, \ldots, (1/2)(s-1)(s-2)\}$. Using the sets $\mathbb{C}_{i,s}$ we partition the space of trees of size $s$, $\mathbb{T}_s$ into

$$\mathbb{T}_s = \bigcup_{i=0}^{\frac{1}{2}(s-1)(s-2)} \mathbb{C}_{i,s}.$$

If $\boldsymbol{p}[\mathbb{C}_{i,s}]$ represents the probability of the space $\mathbb{C}_{i,s}$, in other words the probability that a random tree has a topology that is in $\mathbb{C}_{i,s}$, then the mean of Colless' index of imbalance is given by,

$$\mathbb{E}[I_c|s] = \sum_{i=0}^{1/2(s-1)(s-2)} \frac{i\boldsymbol{p}[\mathbb{C}_{i,s}]}{\sum_{j=0}^{\frac{1}{2}(s-1)(s-2)} \boldsymbol{p}[\mathbb{C}_{j,s}]}. \tag{6.2.1}$$

since each $\mathbb{C}_{i,s}$ is disjoint. Consequently, to determine the mean of Colless' index of imbalance we need to first determine the probabilities, $\boldsymbol{p}[\mathbb{C}_{i,s}]$.

Recall, that we can represent a tree of topology $\mathcal{T}$ as,

$$\mathcal{T} = \{([\alpha(0)], [0])^{(i)}, \mathcal{T}_{[0,0]}, \mathcal{T}_{[0,1]}\},$$

where $\mathcal{T}_{[0,0]}$ and $\mathcal{T}_{[0,1]}$ are the topologies of the daughter and parental subtrees, respectively. The imbalance of a tree of shape $\mathcal{T}$ can be determined from the shapes of its constituent daughter and parental subtrees, $\mathcal{T}_{[0,0]}$ and $\mathcal{T}_{[0,1]}$, in other words,

$$I_c(\mathcal{T}) = I_c(\mathcal{T}_{[0,0]}) + I_c(\mathcal{T}_{[0,1]}) + \big||\mathcal{T}| - 2|\mathcal{T}_{[0,0]}|\big|. \tag{6.2.2}$$

Let $\mathbb{F}_{t,s}$ be the $t$-th topologically isomorphic class of size $s$, where $t = 1, 2, \ldots, T_s$. It is clear, then, that any topology from this class has the same imbalance, because rotating the subtrees at each internal node does not change equation (6.2.2) since $\big||\mathcal{T}| - 2|\mathcal{T}_{[0,0]}|\big| = \big||\mathcal{T}| - 2|\mathcal{T}_{[0,1]}|\big|$ since $|\mathcal{T}_{[0,0]}| = |\mathcal{T}| - |\mathcal{T}_{[0,1]}|$. Furthermore, it is also easy to see that each imbalance class, $\mathbb{C}_{i,s}$ is the union of the topologically isomorphic classes of size $s$ that have an imbalance of $i$. Let $\mathcal{C}_{i,s}$ be the set of indices $t$ that have the property,

$$\mathcal{C}_{i,s} = \{t|I_c[\mathbb{F}_{t,s}] = i\},$$

where $I_c[\mathbb{F}_{t,s}]$ is the imbalance of one member, and hence all the members, of the topologically isomorphic class $\mathbb{F}_{t,s}$. We then have that,

$$\mathbb{C}_{i,s} = \bigcup_{t \in \mathcal{C}_{i,s}} \mathbb{F}_{t,s}.$$

We can generate the set of trees in a non-empty class, $\mathbb{C}_{i,s}$, recursively by combining trees from the non-empty classes $\mathbb{C}_{l,j}$ and $\mathbb{C}_{k,s-j}$ at node [0] provided that, $i = l + k + |s - 2j|$ with $\mathbb{C}_{0,1}$ consisting of only the single branch topology. As we shall see the algorithm to determine the mean imbalance given tree size is based on this recursion. The set $\mathbb{C}_{i,s}$ can be written as,

$$\begin{aligned} \mathbb{C}_{i,s} &= \left\{\{([\alpha(0)],[0])^{(i)}, \mathbb{C}_{l,j}, \mathbb{C}_{k,s-j}\} \middle| i = l + k + |s - 2j|\right\} \\ &\quad \bigcup \left\{\{([\alpha(0)],0)^{(i)}, \mathbb{C}_{k,s-j}, \mathbb{C}_{l,j}\} \middle| i = l + k + |s - 2j|\right\}, \end{aligned} \tag{6.2.3}$$

if $\mathbb{C}_{l,j} \neq \mathbb{C}_{k,s-j}$ and

$$\mathbb{C}_{i,s} = \left\{\{([\alpha(0)],[0])^{(i)}, \mathbb{C}_{l,j}, \mathbb{C}_{k,s-j}\} \middle| i = l + k + |s - 2j|\right\} \tag{6.2.4}$$

otherwise. Figure 6.2.1 depicts all the non-empty imbalance classes for trees of size 1 through 5. Commencing from $\mathcal{C}_{0,1}$, we see that $\mathcal{C}_{0,2}$ is constructed from two $\mathcal{C}_{0,1}$. Proceeding in this manner one finds that $\mathcal{C}_{0,3}$ is empty and $\mathcal{C}_{1,3}$ is non-empty since the combination $\mathcal{C}_{0,1}$ with $\mathcal{C}_{0,2}$ gives us an imbalance of $I_c = 0 + 0 + |3 - 2| = 1$. Similarly, $\mathbb{C}_{3,4}$ is constructed by combining $\mathbb{C}_{0,1}$ and $\mathbb{C}_{1,3}$ and so on.

Using equations (6.2.3) and (6.2.4) the probability that a random tree has a topology from $\mathbb{C}_{i,s}$, $\boldsymbol{p}[\mathbb{C}_{i,s}]$, is given by

$$\boldsymbol{p}[\mathbb{C}_{i,s}] = (-D_0)^{-1} B \sum_{\{l,k,j\} \in \mathcal{S}_i^s} \left(\boldsymbol{p}[\mathbb{C}_{l,j}] \otimes \boldsymbol{p}[\mathbb{C}_{k,s-j}] + I\{\mathbb{C}_{l,j} \neq \mathbb{C}_{k,s-j}\} \boldsymbol{p}[\mathbb{C}_{k,s-j}] \otimes \boldsymbol{p}[\mathbb{C}_{l,j}]\right). \tag{6.2.5}$$

Suppose we wish to determine the probability distribution of imbalance classes for trees of size $s$. We use the following recursive procedure:

Set

$$\boldsymbol{p}[\mathbb{C}_{0,1}] = (-D_0)^{-1} \boldsymbol{d},$$

since it consists of only one branch. Then, loop through all tree sizes, $2 \leq t \leq s$.

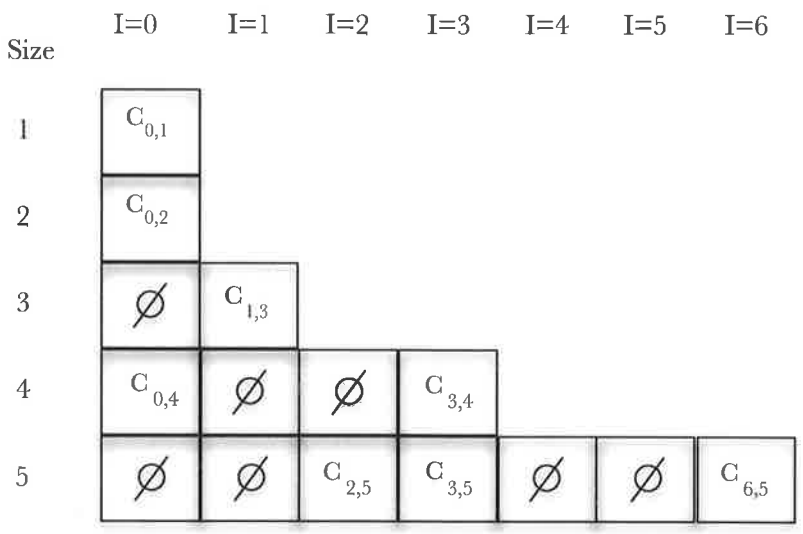


Figure 6.2.1: An illustration of the imbalance algorithm

- For each tree size, $t$, loop through all the imbalances, $0 \leq i \leq 1/2(t-1)(t-2)$.
  - Set $\boldsymbol{p}[\mathbb{C}_{i,t}] = 0$.
  - Loop through all possible daughter subtree sizes, $1 \leq l \leq t-1$.
    * Loop through all possible daughter subtree imbalances, $0 \leq l_i \leq 1/2(l-1)(l-2)$.
      · Test to see whether there exists a non-empty imbalance class $\mathcal{C}_{r_i,t-l}$ such that,

$$r_i = i - l_i - \left|t - 2l\right|,$$

with $0 \leq r_i \leq 1/2(t-l-1)(t-l-2)$ and if $\mathcal{C}_{l_i,l}$ is also non-empty then set

$$\begin{aligned}\boldsymbol{p}[\mathbb{C}_{i,t}] \;=\; & \boldsymbol{p}[\mathbb{C}_{i,t}] + (-D_0)^{-1}B\big(\boldsymbol{p}[\mathbb{C}_{l_i,l}] \otimes \boldsymbol{p}[\mathbb{C}_{r_i,t-l}] + \\ & I\{\mathbb{C}_{l_i,l} \neq \mathbb{C}_{r_i,t-l}\}\boldsymbol{p}[\mathbb{C}_{r_i,t-l}] \otimes \boldsymbol{p}[\mathbb{C}_{l_i,l}]\big).\end{aligned}$$

In Section 6.3 we begin by computing the mean imbalance for the simple macroevolutionary models we discussed in Chapter 3, for size 5 trees. We conclude that section

be demonstrating the flexibility of the MBT by developing a one parameter family of MBTs that can generate any desired mean imbalance, conditional on size five trees. This model also generates some interesting and surprising features for higher size trees.

## 6.3 Some Results for Simple Models

In Section 6.2 the mean of Colless' index of imbalance, equation (6.2.1), was expressed in terms of the imbalance classes. In most of the previous work on determining the mean imbalance, [30], a different probability measure was used. Rogers [30] utilised a measure based on partitioning trees into their topologically isomorphic classes and not in terms of the imbalance classes. However, since the imbalance classes consist of the union of topologically isomorphic classes there are fewer imbalance classes, and therefore our imbalance algorithm has to calculate fewer probabilities. Nevertheless, we shall determine the mean imbalance for some simple models using the topologically isomorphic class approach which for these simple models is straightforward, particularly for small trees.

As before let $\mathbb{F}_{t,s}$ be the $t$-th topologically isomorphic class of size $s$, where $t \in \{1, \ldots, T_s\}$, and $T_s$ is the number of possible topologically isomorphic classes for trees of size $s$. Let $I_c(\mathbb{F}_{t,s})$ be the imbalance of this class. Recall that,

$$\mathbb{E}_s[I_c] = \mathbb{E}[I_c|s],$$

the expected value of Colless' index of imbalance conditional on size $s$ trees. This can also be written as

$$\mathbb{E}_s[I_c] = \sum_{t=1}^{T_s} I_c(\mathbb{F}_{t,s}) p_s[\mathbb{F}_{t,s}]. \tag{6.3.1}$$

In Sections 6.3.1-6.3.4 we show how to find the mean of Colless' index of imbalance for some of the simplest macroevolutionary models. We then show in Section 6.3.5, using the imbalance algorithm of Section 6.2, that the MBT has sufficient flexibility

to cover all possible mean imbalances for a given tree size. To do this we generate a simple four phase one-parameter family of MBTs.

### 6.3.1 The Constant Rates BD Model

The space of trees, $\mathbb{F}_{t,s}$, can be generated by joining, $\mathbb{F}_{l,j}$ and $\mathbb{F}_{k,s-j}$, where $l \in \{1,\ldots,T_j\}$ and $k \in \{1,\ldots,T_{s-j}\}$. Note, for each $\mathbb{F}_{t,s}$ there only exists one $l$, $k$ and $j$ such that,

$$\mathbb{F}_{t,s} = \begin{cases} \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{l,j}, \mathbb{F}_{k,s-j}\} \bigcup \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{k,s-j}, \mathbb{F}_{l,j}\} & \text{if } \mathbb{F}_{l,j} \neq \mathbb{F}_{k,s-j} \\ \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{l,j}, \mathbb{F}_{k,s-j}\} & \text{if } \mathbb{F}_{l,j} = \mathbb{F}_{k,s-j} \end{cases}$$

Harding [8] first derived the probability distribution of topologically isomorphic classes for the constant rates birth model. The probability of a random tree of size $s$ attaining a topology from $\mathbb{F}_{t,s}$, is given by

$$p_s[\mathbb{F}_{t,s}] = \begin{cases} \frac{1}{s-1} p_j[\mathbb{F}_{k,j}]^2 & \text{if } j = s/2 \ \& \ k = l \\ \frac{2}{s-1} p_j[\mathbb{F}_{k,j}] p_{s-j}[\mathbb{F}_{l,s-j}] & \text{otherwise} \end{cases} \tag{6.3.2}$$

where $\mathbb{F}_{k,j}$ and $\mathbb{F}_{l,s-j}$ are the topologically isomorphic classes of the two subtrees at node [0]. We also showed these equations to be valid for the constant-rates birth-and-death model in Chapter 3.

**Example 6** *Mean imbalance for size 5 trees*

It is easy to deduce that there are three topologically isomorphic classes for size five trees, since size five trees are constructed from,

- size 1 and size 4 trees, or
- size 2 and size 3 trees.

There is only one class for each of the size two and size three trees, and there are two classes of size four; one consists of joining two size two trees and the other one consists of joining a size three tree with a size 1 tree. As a result, this implies that

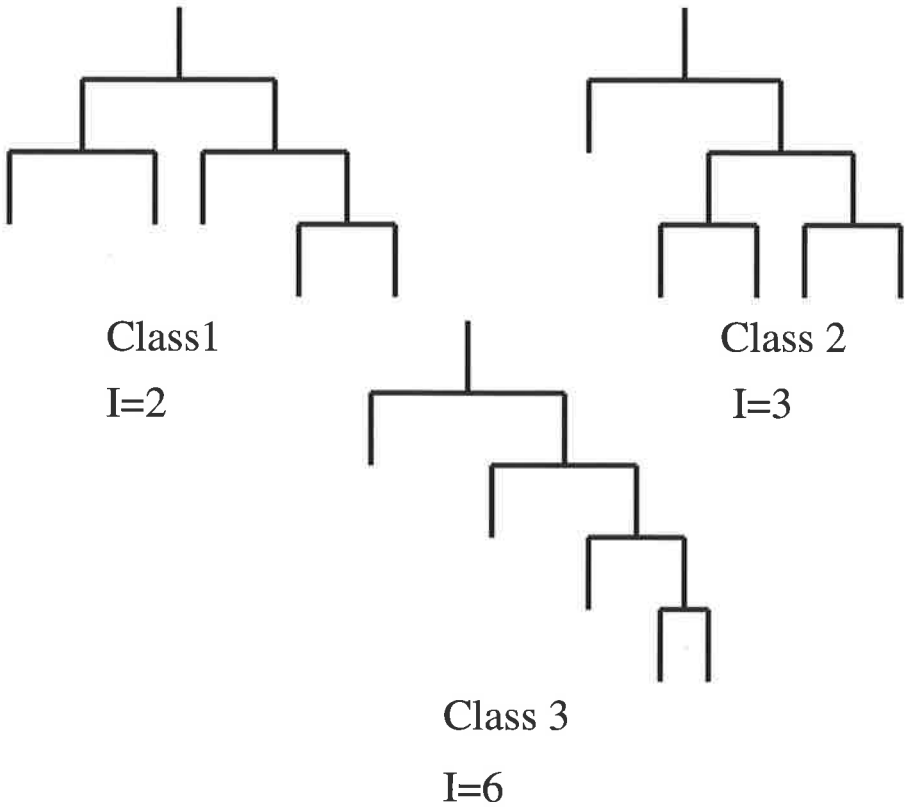


Figure 6.3.1: The three topologically isomorphic classes of size 5

there are three classes for size 5 trees. A member from each class is depicted in Figure 6.3.1.

It is straightforward using equations (6.3.2) to show that,

$$\begin{aligned} p_5[\mathbb{F}_{1,5}] &= \frac{2}{6},\ I_c(\mathbb{F}_{1,5}) = 2, \\ p_5[\mathbb{F}_{2,5}] &= \frac{1}{6},\ I_c(\mathbb{F}_{2,5}) = 3, \\ p_5[\mathbb{F}_{3,5}] &= \frac{3}{6},\ I_c(\mathbb{F}_{3,5}) = 6. \end{aligned}$$

The mean imbalance, equation (6.3.1), is then,

$$\mathbb{E}_5[I_c] = 2 \times \frac{2}{6} + 3 \times \frac{1}{6} + 6 \times \frac{3}{6} = \frac{25}{6} = 4.16667.$$

### 6.3.2 The PDA Model

The PDA model states that each topology of a given size is equally likely. The PDA can be represented as a constant rates birth and death model in the limit as $t \to \infty$, or as an MR model [26], or as an MBT, see Chapter 5. Slowinski [33] first derived the equations for the imbalance distribution for the PDA model. In this section, we derive a much simpler set of equations that generate the distribution. Pinelis [26] showed that there is a correspondence between the number of distinguishable arrangements and the number of topologies given the tree size. Due to this correspondence we derive a set of equations that give the PDA distribution on the topologically isomorphic classes. Recall from Chapter 3 that the number of distinct topologies of a given size can be found using the following recursion equation,

$$N_s = \sum_{i=1}^{s-1} N_i N_{s-i}, \tag{6.3.3}$$

with boundary condition, $N_1 = 1$. Therefore, we can see, for example, that there is 1 topology of size 2, 2 topologies of size 3, 5 topologies of size 4 and 14 topologies of size 5.

Recall from Chapter 3, that the number of topologies in a topologically isomorphic class is given by the number of uneven branch points. Let the number of uneven branch points of the topologies in $\mathbb{F}_{t,s}$ be denoted by $\epsilon_{t,s}$, then the number of topologies in $\mathbb{F}_{t,s}$, $N(\mathbb{F}_{t,s})$ is

$$N(\mathbb{F}_{t,s}) = 2^{\epsilon_{t,s}}. \tag{6.3.4}$$

All the topologies of a given size are equally likely, so the probability that a tree of size $s$ comes from $\mathbb{F}_{t,s}$, is given by,

$$p_s[\mathbb{F}_{t,s}] = \frac{N(\mathbb{F}_{t,s})}{N_s}. \tag{6.3.5}$$

**Example 7** *Mean imbalance for size 5 trees.*

Figure 6.3.1 indicates that, $N(\mathbb{F}_{1,s}) = 2^2 = 4$, $N(\mathbb{F}_{2,s}) = 2^1 = 2$ and $N(\mathbb{F}_{3,s}) = 2^3 = 8$. As a result we have,

$$p_5[\mathbb{F}_{1,5}] = \frac{4}{14} = \frac{2}{7}, \tag{6.3.6}$$

$$p_5[\mathbb{F}_{2,5}] = \frac{2}{14} = \frac{1}{7}, \tag{6.3.7}$$

$$p_5[\mathbb{F}_{3,5}] = \frac{8}{14} = \frac{4}{7}. \tag{6.3.8}$$

So the mean of Colless' index of imbalance is therefore given by,

$$\mathbb{E}_5[I_c|s=5] = 2 \times \frac{2}{7} + 3 \times \frac{1}{7} + 6 \times \frac{4}{7} = \frac{31}{7} = 4.42857. \tag{6.3.9}$$

### 6.3.3 The sPDA Model

The PDA model discussed in Section 6.3.2, states that all topologies of a given size are equally likely. The sPDA model on the other hand, states that the larger the number of unitary branch points a topology has, the more likely it is to occur. A unitary branch point is a branch point where the *daughter* subtree consists of only a single branch. Recall from Chapter 5 that for the MBT version of the sPDA the probability of obtaining a topology of size $s$ with $v$ unitary splits given that the process began in phase 2, was

$$p_2(\mathcal{T}) = \frac{1}{1-p(\emptyset)}\beta^{s-1}\left(\frac{\gamma}{(1-\gamma)}\frac{1}{\frac{q+b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)}}+1\right)^v\left(\frac{q+b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)}\right)^s, \tag{6.3.10}$$

where,

$$\beta = \frac{b(1-\gamma)}{1-2b(1-\gamma)p(\emptyset)}. \tag{6.3.11}$$

In this section we are interested in understanding the distribution given the size of a tree and as a result, for trees of size $s$, the factors, $\beta^{s-1}$,

$$\left(\frac{q+b\gamma p(\emptyset)}{1-2b(1-\gamma)p(\emptyset)}\right)^s,$$

and $1-p(\emptyset)$ are not important because they are identical for all those topologies of size $s$.

Trees in a topologically isomorphic class do not all have the same number of unitary branch points, because rotating the internal nodes of a topology has the effect of swapping parent and daughter branches, and thus changing the number of unitary branch points. As stated previously, the topologically isomorphic class, $\mathbb{F}_{t,s}$, is represented as

$$\mathbb{F}_{t,s} = \begin{cases} \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{k,j}, \mathbb{F}_{l,s-j}\} \bigcup \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{l,s-j}, \mathbb{F}_{k,j}\} & \text{if } \mathbb{F}_{k,j} \neq \mathbb{F}_{l,s-j}, \\ \{([\alpha(0)],[0])^{(i)}, \mathbb{F}_{k,j}, \mathbb{F}_{l,s-j}\} & \text{if } \mathbb{F}_{k,j} = \mathbb{F}_{l,s-j}. \end{cases} \tag{6.3.12}$$

Now since the position of a single branch subtree determines whether the branch point to which it is attached is a unitary branch point or not, we introduce two measures, $N_p(\mathbb{F}_{t,s}, \upsilon)$ which gives the number of topologies in $\mathbb{F}_{t,s}$ that have $\upsilon$ unitary branch points, if the trees in $\mathbb{F}_{t,s}$ are joined as the parental subtree at some branch point, and $N_d(\mathbb{F}_{t,s}, \upsilon)$ which gives the number of topologies in $\mathbb{F}_{t,s}$ that have $\upsilon$ unitary branch points, if the trees in $\mathbb{F}_{t,s}$ are joined as the daughter subtree at some branch point. The number of topologies in $\mathbb{F}_{t,s}$ that have $\upsilon$ unitary branch points when considered as a parental subtree is given by

$$\begin{aligned} N_p(\mathbb{F}_{t,s}, \upsilon) &= \sum_{\omega=1}^{\upsilon} \big( N_d(\mathbb{F}_{k,j}, \omega) N_p(\mathbb{F}_{l,s-j}, \upsilon-\omega) + \\ &\quad I\{\mathbb{F}_{l,s-j} \neq \mathbb{F}_{k,j}\} N_d(\mathbb{F}_{l,s-j}, \omega) N_p(\mathbb{F}_{k,j}, \upsilon-\omega) \big), \end{aligned} \tag{6.3.13}$$

where, $N_p(\mathbb{F}_{1,1}, 0) = 1$ and $N_p(\mathbb{F}_{1,1}, \upsilon) = 0$, for $\upsilon \geq 1$, and where $k \in \{1, 2, \ldots, T_j\}$ and $l \in \{1, 2, \ldots, T_{s-j}\}$. When considered as a daughter subtree, $N_d(\mathbb{F}_{t,s}, \upsilon)$ obeys the same recursion, but has a different set of boundary conditions, given by $N_d(\mathbb{F}_{1,1}, 1) = 1$, and $N_d(\mathbb{F}_{1,1}, \upsilon) = 0$ for $\upsilon \neq 1$. We need the different boundary conditions because single branch daughter subtrees give us one unitary branch point, whereas a single branch parental subtree does not contribute any unitary branch points. The summation in equation (6.3.13) commences from $\omega = 1$ because all daughter subtrees have at least one unitary branch point, and finishes at $\omega = \upsilon$ since considered as a

parental subtree a single branch has no unitary branch points.

We now give an example of calculating $N_d(\mathbb{F}_{t,s}, \upsilon)$ and $N_p(\mathbb{F}_{t,s}, \upsilon)$. There are two topologically isomorphic classes of size four, $\mathbb{F}_{1,4}$ and $\mathbb{F}_{2,4}$. We shall apply equation (6.3.13) to $\mathbb{F}_{2,4}$. We first note that, this class is constructed from size one trees and size three trees. Now, there is only one topologically isomorphic class of size three, $\mathbb{F}_{1,3}$ and it is obvious that there are only two trees in $\mathbb{F}_{1,3}$. One of the trees in $\mathbb{F}_{1,3}$ has only one unitary branch point and the other has two, thus

$$N_p(\mathbb{F}_{1,3}, 1) = N_d(\mathbb{F}_{1,3}, 1) = 1, \text{ and} \tag{6.3.14}$$

$$N_p(\mathbb{F}_{1,3}, 2) = N_d(\mathbb{F}_{1,3}, 2) = 1. \tag{6.3.15}$$

Using the above we can calculate, for example, $N_p(\mathbb{F}_{2,4}, 2)$,

$$\begin{aligned} N_p(\mathbb{F}_{2,4}, 2) &= \sum_{j=1}^{2} \left( N_d(\mathbb{F}_{1,1}, j) N_p(\mathbb{F}_{1,3}, 2-j) + N_d(\mathbb{F}_{1,3}, j) N_p(\mathbb{F}_{1,1}, 2-j) \right) \\ &= N_d(\mathbb{F}_{1,1}, 1) N_p(\mathbb{F}_{1,3}, 1) + N_d(\mathbb{F}_{1,1}, 2) N_p(\mathbb{F}_{1,3}, 0) + \\ &\quad N_d(\mathbb{F}_{1,3}, 1) N_p(\mathbb{F}_{1,1}, 1) + N_d(\mathbb{F}_{1,3}, 2) N_p(\mathbb{F}_{1,1}, 0). \end{aligned}$$

Consider the first term, $N_d(\mathbb{F}_{1,1}, 1) N_p(\mathbb{F}_{1,3}, 1)$. Now $N_d(\mathbb{F}_{1,1}, 1) = 1$ because the single branch forms the daughter subtree and $N_p(\mathbb{F}_{1,3}, 1) = 1$, so,

$$N_d(\mathbb{F}_{1,1}, 1) N_p(\mathbb{F}_{1,3}, 1) = 1 \times 1 = 1.$$

The second term, $N_d(\mathbb{F}_{1,1}, 2) N_p(\mathbb{F}_{1,3}, 0)$, is zero because, $N_d(\mathbb{F}_{1,1}, \upsilon) = 0$ if $\upsilon \neq 1$. The third term, $N_d(\mathbb{F}_{1,3}, 1) N_p(\mathbb{F}_{1,1}, 1)$, is also zero because $N_p(\mathbb{F}_{1,1}, 1) = 0$, since a single branch parent subtree does not generate a unitary branch point. The final term, $N_d(\mathbb{F}_{1,3}, 2) N_p(\mathbb{F}_{1,1}, 0) = 1$ because $N_d(\mathbb{F}_{1,3}, 2) = 1$ from equation (6.3.15) and $N_p(\mathbb{F}_{1,1}, 0) = 1$. As a result, $N(\mathbb{F}_{2,4}, 2) = 2$.

Thus equation (6.3.10) tells us that the probability of a random tree of size $s$ having a topology that is an element of $\mathbb{F}_{t,s}$, in the sPDA model is

$$p_s[\mathbb{F}_{t,s}] = \frac{\sum_{\upsilon=1}^{s-1} N_p(\mathbb{F}_{t,s}, \upsilon) \left(1 + \frac{\gamma}{(1-\gamma)p(1)}\right)^{\upsilon}}{\sum_{l=1}^{T_s} \sum_{\upsilon=1}^{s-1} N_p(\mathbb{F}_{l,s}, \upsilon) \left(1 + \frac{\gamma}{(1-\gamma)p(1)}\right)^{\upsilon}}, \tag{6.3.16}$$

where $p(1)$ is the probability of a size one tree and is given by equation (3.11.16).

**Example 8** *Mean imbalance for size 5 trees.*

Our first step is to calculate the numbers of trees with $v$ unitary branch points for each of the topologically isomorphic classes. In order to avoid repititious calculations that require equation (6.3.13) we shall give the results for the non-zero values:

$$N_p(\mathbb{F}_{1,5}, 2) = 2, \tag{6.3.17}$$

$$N_p(\mathbb{F}_{1,5}, 3) = 2, \tag{6.3.18}$$

$$N_p(\mathbb{F}_{2,5}, 2) = 1, \tag{6.3.19}$$

$$N_p(\mathbb{F}_{2,5}, 3) = 1, \tag{6.3.20}$$

$$N_p(\mathbb{F}_{3,5}, 1) = 1, \tag{6.3.21}$$

$$N_p(\mathbb{F}_{3,5}, 2) = 3, \tag{6.3.22}$$

$$N_p(\mathbb{F}_{3,5}, 3) = 3, \tag{6.3.23}$$

$$N_p(\mathbb{F}_{3,5}, 4) = 1. \tag{6.3.24}$$

We choose, for example, the values, $b = 0.3$, $q = 0.3$, $d = 0.4$ and $\gamma = 0.5$ for the sPDA model. For these values of the parameters, the only solution to equation (3.11.11) for $p(\emptyset)$ from Chapter 3 is given by,

$$p(\emptyset) = \frac{1 - \left(1 - 4b(1-\gamma)d\right)^{1/2}}{2b(1-\gamma)} = 0.4247, \tag{6.3.25}$$

and thus,

$$p(1) = \frac{q + b\gamma p(\emptyset)}{1 - 2b(1-\gamma)p(\emptyset)} = 0.4177 \tag{6.3.26}$$

and,

$$1 + \frac{\gamma}{(1-\gamma)p(1)} = 3.3941. \tag{6.3.27}$$

Using equations (6.3.17)-(6.3.27), the denominator of equation (6.3.16) can be shown to be equal to, 439.8054. As a result, we have that,

$$p_5[\mathbb{F}_{1,5}] = \frac{N_p(\mathbb{F}_{1,5}, 2)(3.3941)^2 + N_p(\mathbb{F}_{1,5}, 3)(3.3941)^3}{439.8054} = 0.2302$$

$$p_5[\mathbb{F}_{2,5}] = 0.1151$$

$$p_5[\mathbb{F}_{3,5}] = 0.6547.$$

Thus, finally, the mean of Colless' index of imbalance for the sPDA model is given by,

$$\mathbb{E}_5[I_c] = 2 \times 0.2302 + 3 \times 0.1151 + 6 \times 0.6547 = 4.7339. \tag{6.3.28}$$

### 6.3.4 The Completely Unbalanced Model

The completely unbalanced (CU) model is just the MR-sPDA or the MBT-sPDA model with $\gamma = 1$. The $B$ matrix for the MBT-sPDA, reproduced here for convenience, is

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & b\gamma & 0 & b(1-\gamma)\big(1-p(\emptyset)\big) \end{bmatrix}, \tag{6.3.29}$$

which becomes, in the case of the CU model,

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & b & 0 & 0 \end{bmatrix}, \tag{6.3.30}$$

when $\gamma = 1$. Consequently, all branches that are in phase 2, can only give birth to daughters in phase 1, while the parents remain in phase 2 immediately after the observable event. Since all branches in phase 1 eventually become extinct almost surely, all the daughter subtrees are single branches. Consequently, only topologies with the maximum allowed imbalances are possible, that is, the completely unbalanced topologies. Figure 6.3.2 depicts such a tree for size 5.

**Example 9** *Mean imbalance for size 5 trees*

Since the only possible topology in this case is the maximally unbalanced topology, this implies that,

$$\mathbb{E}_5[I_c] = 6. \tag{6.3.31}$$

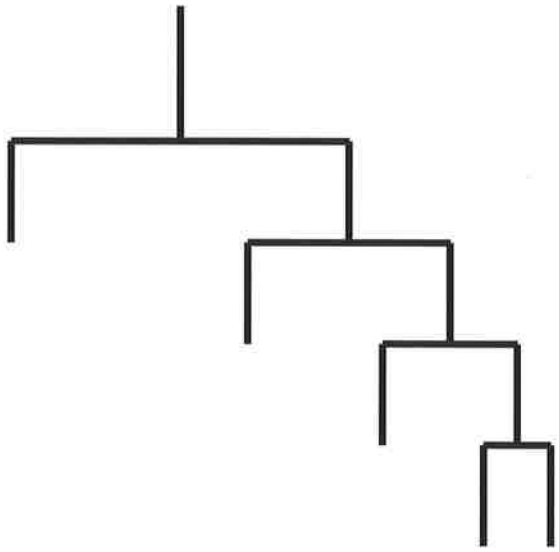

Figure 6.3.2: The completely unbalanced tree of size 5

### 6.3.5 A One Parameter Family of MBTs

In the preceeding sections we obtained the mean of Colless' index of imbalance for the crBD, the PDA, the sPDA and the CU models for trees of size 5. We have shown theoretically, in Chapter 5, that the MBT subsumes these models. This is good, however is this good enough? We need a model that can not only account for these simple models but has the ability to show more complex behaviour and produce mean imbalances that mimic those found in phylogenetic trees. Indeed, it is the purpose of this section to show that a one-parameter family of four-phase MBTs has sufficient flexibility to give us any mean from the entire range of possible mean imbalances, as that parameter varies from 0 to 1. This demonstrates that the MBT does have this required flexibility.

The more complex behaviour of the rates of speciation and extinction in the MBT arise due to the interactions that can be introduced between the underlying phases. We are therefore free to choose transition rate matrices for the process that show very high correlations between the phases. Consequently then, the initial phase of an MBT can have profound effects on its subsequent evolution. To demonstrate how

profound these effects can be, we study size 5 trees; at size 5, interesting behaviour begins to emerge.

We commence by defining some terminology. A hidden transition from phase $a$ to phase $b$ is denoted by, $a \rightarrow b$, and an observable transition in which a branch is in phase $a$ immediately before the branch point, produces a daughter branch in phase $i$ immediately after the branch point, and a parental branch that is in phase $j$ immediately after the branch point is denoted by $a \rightarrow i, j$.

### Low Imbalance Model

Consider a simple four-phase model that consists of phase 0, the absorbing phase, and phases $1, 2, 3$, and which has transition rate matrices given by,

$$D_0 = \begin{bmatrix} -10 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0.000001 & -10 \end{bmatrix}, \tag{6.3.32}$$

and

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 9.99999 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9.99999 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{6.3.33}$$

The above model has the following qualitative behaviour:

- a $1 \rightarrow 2, 2$ observable transition occurs with probability 0.999999,
- a $1 \rightarrow 0$ hidden transition occurs with probability 0.000001,
- a $2 \rightarrow 3, 3$ observable transition occurs with probability 0.999999,
- a $2 \rightarrow 0$ hidden transition occurs with probability 0.000001,
- a $3 \rightarrow 2$ hidden transition occurs with probability 0.000001,
- a $3 \rightarrow 0$ hidden transition occurs with probability 0.999999.

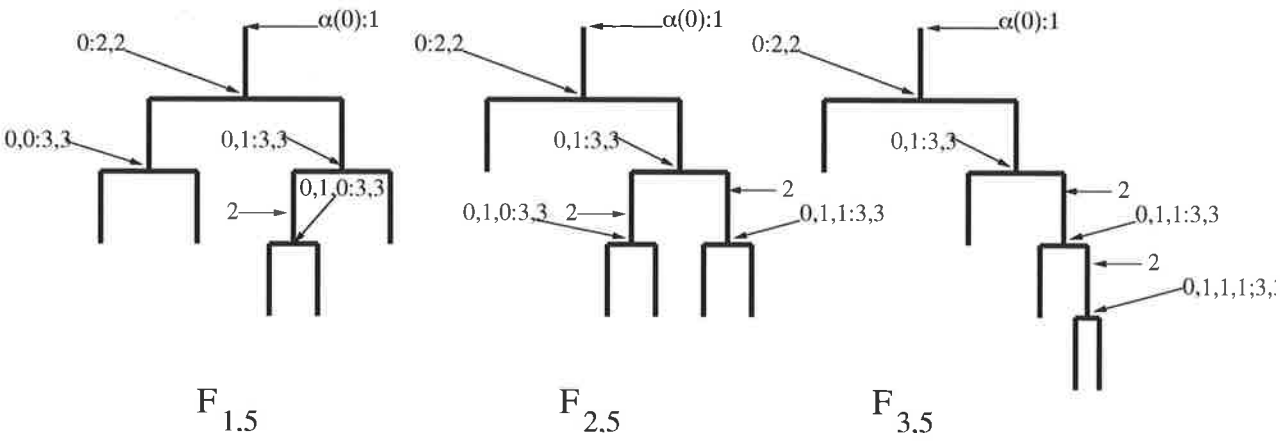


Figure 6.3.3: Low imbalance MBT model

These transitions have been chosen specifically for the interesting tree shape dynamics for trees of size 5 that ensue.

Figure 6.3.3 depicts a representative from each of the three topologically isomorphic classes of size five. The labels $\psi : j, k$ mean that immediately after node $[\psi]$ the daughter branch is in phase $j$ and the parental branch is in phase $k$. The most probable transitions that must occur in order to generate these topologies have also been depicted. Note, that in this case, each and every topology within a topologically isomorphic class is generated by the exact same transitions as depicted in the figure, just in a different order. The representative topology from $\mathbb{F}_{1,5}$ has only one transition that is low probability,

- a hidden transition from phase 3 to phase 2 along branch $([0,1],[0,1,0])^{(i)}$.

The representative topology from $\mathbb{F}_{2,5}$ has three low probability transitions and these are,

- branch $([0],[0,0])^{(e)}$ becomes extinct from phase 2,
- a hidden transition from phase 3 to phase 2 along branch $([0,1],[0,1,0])^{(i)}$, and along branch $([0,1],[0,1,1])^{(i)}$.

The representative topology from $\mathbb{F}_{3,5}$ also has three low probability transitions,

- branch $([0],[0,0])^{(e)}$ becomes extinct from phase 2,
- a hidden transition from phase 3 to phase 2 along branch $([0,1],[0,1,1])^{(i)}$, and along branch $([0,1,1],[0,1,1,1])^{(i)}$.

Consequently, when considering size five trees, one would expect that $p_1[\mathbb{F}_{1,5}]$ should be much greater than either $p_1[\mathbb{F}_{2,5}]$ or $p_1[\mathbb{F}_{3,5}]$, since there is only one low probability transition in each of the topologies in $\mathbb{F}_{1,5}$ and three low probability transitions in $\mathbb{F}_{2,5}$ and $\mathbb{F}_{3,5}$. When we apply our imbalance algorithm to this model we find that the mean of Colless' index of imbalance for size five trees when starting in phase 1 is very near to $I_c[\mathbb{F}_{1,5}] = 2$ as we would expect. What is interesting about this particular value is that it is considerably lower than that of the crBD model as calculated in Section 6.3.1. That this is achievable, is testimony to the dependency that can be generated amongst phases in an MBT. Thus, phase 1 is almost forced to generate two phase 2 branches, phase 2 is almost forced to generate phase 3 branches and phase 3 branches are essentially forced to be absorbed. Such dependency is absent in the crBD, PDA and sPDA models.

### High Imbalance Model

We now show that we can find an MBT that generates maximally imbalanced size 5 topologies. Consider the MBT with

$$D_0 = \begin{bmatrix} -10 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0.000001 & -10 \end{bmatrix}, \tag{6.3.34}$$

and

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 4.999995 & 0 & 4.999995 \\ 9.99999 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{6.3.35}$$

Its qualitative behaviour is such that,

- a $1 \rightarrow 3,1$ observable transition occurs with probability .4999995,

- a $1 \rightarrow 3,3$ observable transition occurs with probability 0.4999995,
- a $1 \rightarrow 0$ hidden transition occurs with probability 0.000001,
- a $2 \rightarrow 1,1$ observable transition occurs with probability 0.999999,
- a $2 \rightarrow 0$ hidden transition occurs with probability 0.000001,
- a $3 \rightarrow 2$ hidden transition occurs with probability 0.000001,
- a $3 \rightarrow 0$ hidden transition occurs with probability 0.999999.

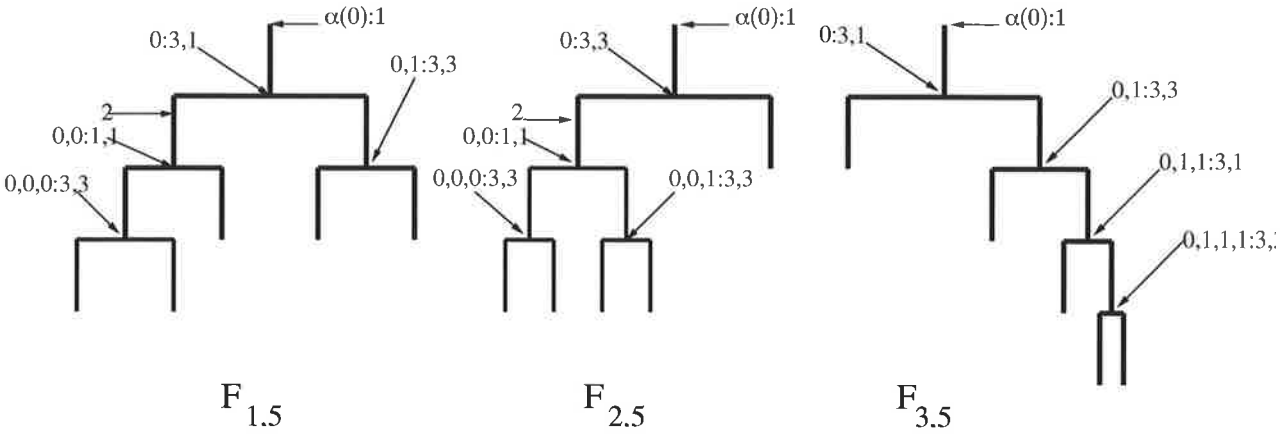


Figure 6.3.4: Maximally imbalanced MBT model

Figure 6.3.4 depicts the most probable representative topology from each of the three size five topologically isomorphic classes of size 5. The transitions along each branch have been included in the figure, and as before, if $[\psi]$ is some node, then $\psi : k, l$ tells us that immediately following $[\psi]$, the daughter branch is in phase $k$ and the parental branch is in phase $l$. The topology from $\mathbb{F}_{1,5}$ has two low probability transitions:

- there is a hidden transition from $3 \rightarrow 2$ along branch $([0],[0,0])^{(i)}$ and
- branch $([0,0],[0,0,1])$ becomes extinct from phase 1, that is, a $1 \rightarrow 0$ hidden transition.

The representative topology from $\mathbb{F}_{2,5}$ has one low probability transition, the hidden transition from $3 \rightarrow 2$ along branch $([0],[0,0])^{(i)}$.

The only topology from all the size five topologies that does not have any low probability transitions is the completely unbalanced topology; the third in the figure. This particular topology is by far the most likely topology because all the other size 5 topologies require at least one low probability transition. This includes all topologies in $\mathbb{F}_{3,5}$ except for the one depicted in the figure. When considering trees of size 5, we would expect that $p_1[\mathbb{F}_{3,5}]$ dominates because of the completely unbalanced topology depicted in the figure. So the mean imbalance should be very close to $I_c[\mathbb{F}_{3,5}] = 6$, when starting in phase 1. In fact, when one applies the imbalance algorithm from Section 6.2 this is indeed the case, $\mathbb{E}[I_c|5] = 6$.

### A One-Parameter Family of MBTs

We are now in a position to discuss the one-parameter family of MBT models that transform from our low imbalance model to our high imbalance model. The infinitesimal rate matrices for this model are,

$$D_0 = \begin{bmatrix} -10 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0.000001 & -10 \end{bmatrix}, \tag{6.3.36}$$

and

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 9.99999(1-\zeta) & 0 & 4.999995\zeta & 0 & 4.999995\zeta \\ 9.99999\zeta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9.99999(1-\zeta) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.3.37}$$

where $\zeta \in [0,1]$.

The graph in Figure 6.3.5 plots the relationship between the mean of Colless' index of imbalance, conditioned on size 5 trees, and the value of $\zeta$ for trees commencing in phases 1 and phases 2. Although, we do not analyse the transitions

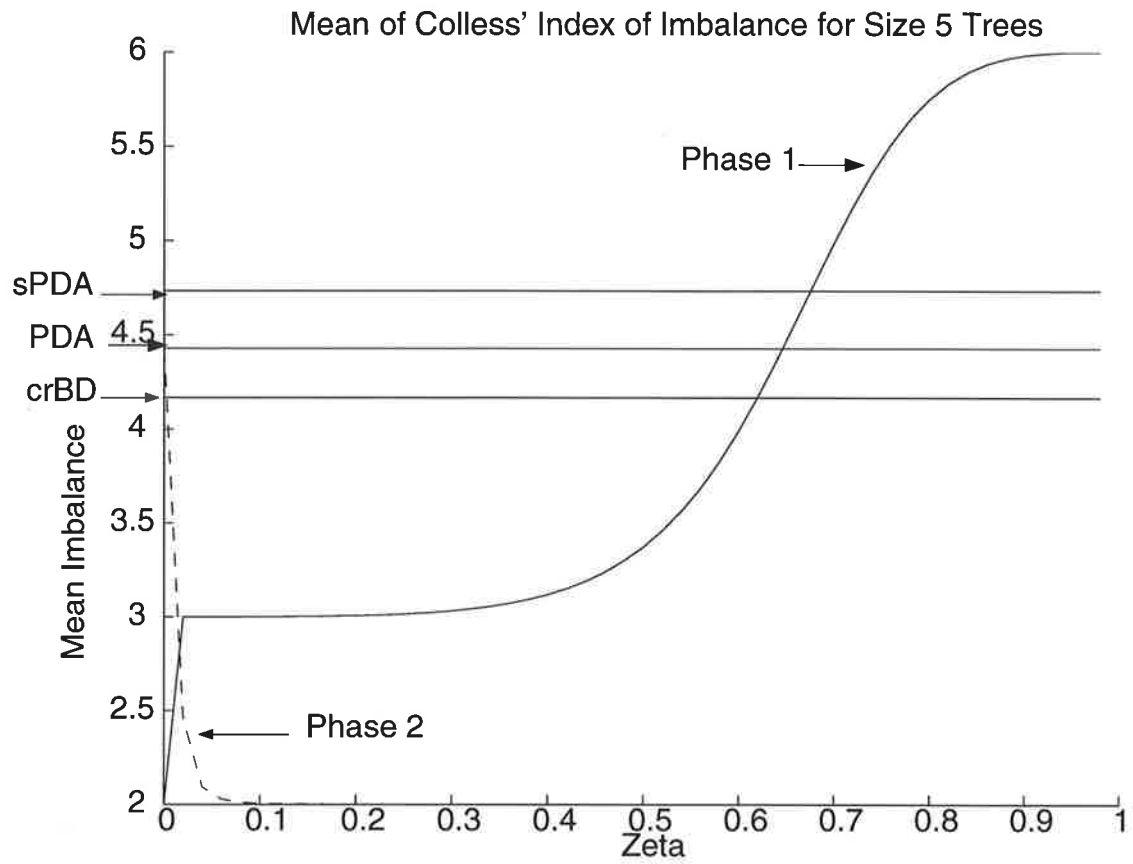


Figure 6.3.5: Mean of Colless' index of imbalance for size 5 trees

that generate size 5 trees commencing from phase 2 we plot the mean imbalance for trees commencing from this phase because it demonstrates just how crucial the initial phase is. Also plotted are the mean imbalances for the crBD, PDA, and sPDA models.

When $\zeta = 0$, the MBT reduces to the low imbalance MBT we discussed at the beginning of this section. So, considering size 5 trees commencing from phase 1 we find that the mean imbalance is very close to 2 as expected. At the other extreme is a model very similar to the high imbalance MBT we discussed earlier. In this case, the mean imbalance is very close to 6. When $\zeta$ is small but not zero we see that there is a spike in the mean imbalance to three. Figure 6.3.6 magnifies this region and depicts what happens as $\zeta$ becomes non-zero: as $\zeta \to 1 \times 10^{-5}$ the mean

imbalance approaches 3. In other words, this means that the probability of obtaining topologies from $C_{3,5} = \mathbb{F}_{2,5}$ is approaching one. To explain why this occurs, we study

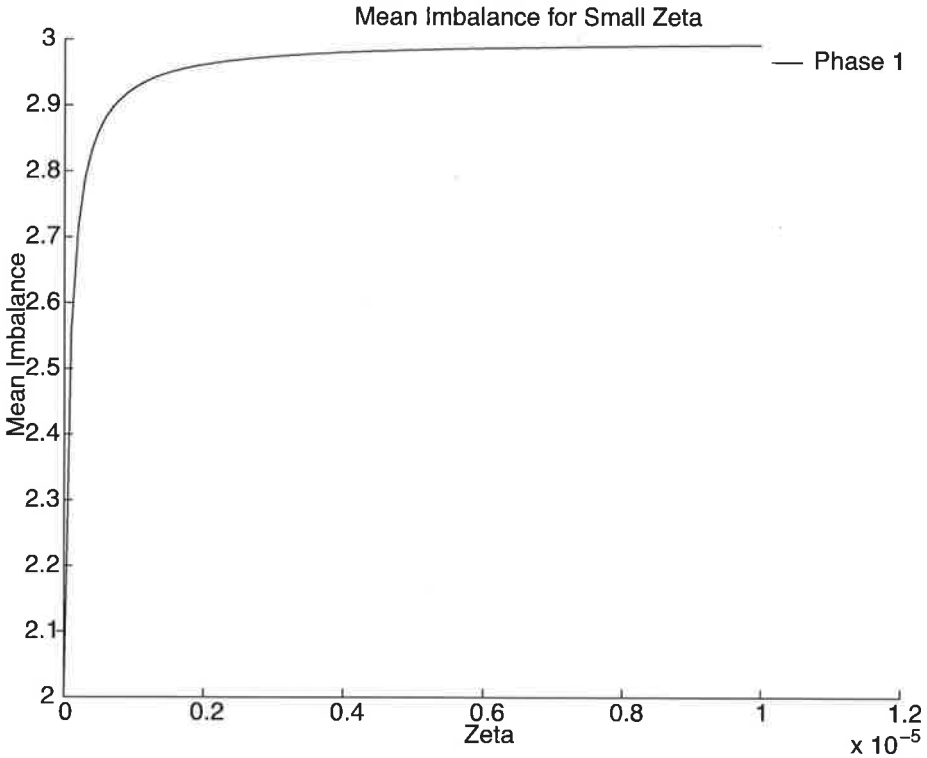


Figure 6.3.6: Magnification of the mean imbalance for small $\zeta$.

Figure 6.3.7, which once again depicts one tree from each of the three topologically isomorphic classes of size 5 and the transitions that can generate these trees have also been depicted. When $\zeta$ is small there are four low probability transitions,

- the hidden transition, $3 \rightarrow 2$,
- the observable transition, $1 \rightarrow 3, 1$,
- the observable transition, $1 \rightarrow 3, 3$, and
- the observable transition, $2 \rightarrow 1, 1$.

All topologies from $\mathbb{F}_{3,5}$ have 4 low probability transitions, and these are,

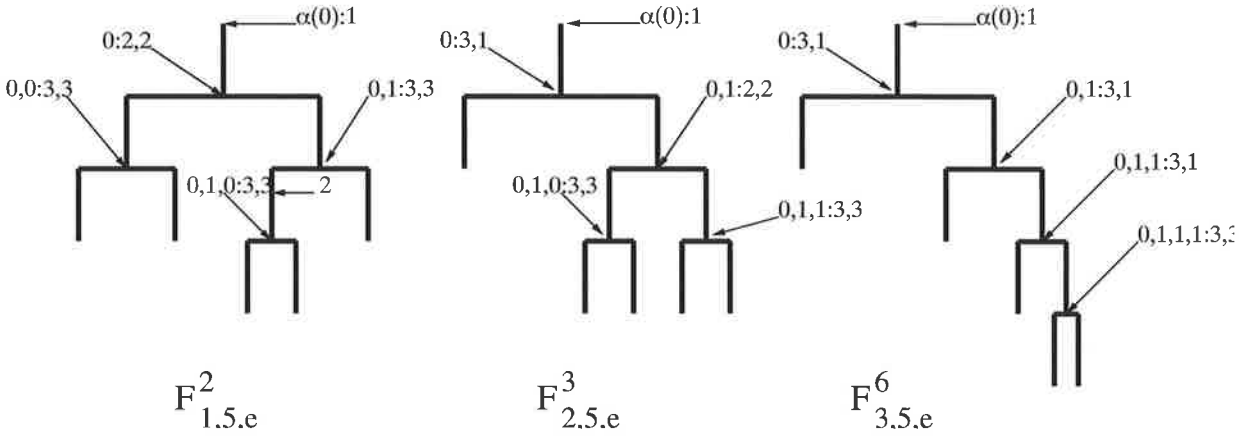


Figure 6.3.7: One topology from each topologically isomorphic class of size 5.

- the observable transitions at $[0]$, $[0,1]$, $[0,1,1]$ from $1 \rightarrow 3,1$, and
- the observable transition at $[0,1,1,1]$ from $1 \rightarrow 3,3$.

All the topologies from $\mathbb{F}_{1,5}$ only have one low probability transition, a hidden transition from $3 \rightarrow 2$. In the tree depicted in Figure 6.3.7 this occurs on branch $([0,1],[0,1,0])^{(i)}$, but it could occur on any of the other three leaf branches. The topology from $\mathbb{F}_{2,5}$ that is depicted also only has one low probability transition; the observable transition at $[0]$ from $1 \rightarrow 3,1$. The other topology from $\mathbb{F}_{2,5}$ has two more low probability transitions. We can therefore discard the topologies from $\mathbb{F}_{3,5}$ as occurring with extremely low probability and concentrate only on those from $\mathbb{F}_{1,5}$ and from $\mathbb{F}_{2,5}$; in $\mathbb{F}_{2,5}$ the topology that we have depicted dominates. So, for low $\zeta > 0$, why does the topology from $\mathbb{F}_{2,5}$ and hence $\mathbb{F}_{2,5}$ predominate over $\mathbb{F}_{1,5}$?

The answer to this question resides in the relative sizes of the probabilities of the $3 \rightarrow 2$ followed by a $2 \rightarrow 3,3$ transition and the $1 \rightarrow 3,1$ transition. The probabilities of each of these transitions are given by their respective elements from the matrix product $(-D_0)^{-1}B$. For $3 \rightarrow 2$, $2 \rightarrow 3,3$, we have, $(-D_0)^{-1}_{32}B_{2,33} = 1 \times 10^{-9}9.99999(1-\zeta) = 9.99999 \times 10^{-9}(1-\zeta)$ and for $1 \rightarrow 3,1$ we have $(-D_0)^{-1}_{11}B_{1,31} =$

$(0.1)(4.999995\zeta) = 0.4999995\zeta$. Now the ratio of these two transitions is

$$\frac{(-D_0)^{-1}_{11}B_{1,31}}{(-D_0)^{-1}_{32}B_{2,33}} = 50000000\frac{\zeta}{1-\zeta}. \tag{6.3.38}$$

We can see from the above equation that when $\zeta = 3.3 \times 10^{-8}$ the orders of the two probabilities are about the same, however, if $\zeta = 1 \times 10^{-5}$ we have that,

$$\frac{(-D_0)^{-1}_{11}B_{1,31}}{(-D_0)^{-1}_{32}B_{2,33}} \approx 500. \tag{6.3.39}$$

Hence, for $\zeta = 1 \times 10^{-5}$, the $1 \rightarrow 3,1$ transition is roughly five hundred times more likely than the $3 \rightarrow 2 \rightarrow 3,3$ transition and this is reflected in the fact that the probability of the $\mathbb{F}_{2,5}$ class is 0.99206 as opposed to 0.007936 for the $\mathbb{F}_{1,5}$ class; about 126 times more likely.

As $\zeta \rightarrow 1$,

$$B \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 4.999995 & 0 & 4.999995 \\ 9.99999 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

which is identical to equation (6.3.35) and as a result, $\mathbb{E}[I_c|5] \rightarrow 6$; as shown in the graph of Figure 6.3.5. There are essentially three competing transitions from phase 1 as $\zeta$ gets larger and they are,

- the $1 \rightarrow 3,1$ which occurs with rate $4.999995\zeta$,
- the $1 \rightarrow 3,3$ which occurs with rate $4.999995\zeta$, and
- the $1 \rightarrow 2,2$ which occurs with rate $9.99999(1-\zeta)$.

The first transition tends to favour less balanced topologies, the second transition favours small size topologies and the last transition tends to favour more balanced topologies. For small $\zeta$ the third transition dominates and so more balanced topologies result, however as $\zeta$ approaches one the first two transitions dominate, thus generating small highly imbalanced topologies, until $\zeta = 1$ when a mean imbalance of 6 is achieved.

The above one parameter family of MBTs was constructed in order to demonstrate the flexibility of the MBT. Specifically, for size 5 trees we constructed matrices that generated the desired correlations amoungst the phases in order to show this flexibility by achieving the complete range of mean imbalances. This model also delivers very interesting behaviour for larger size trees. Figure 6.3.8 graphs the mean imbalance against the parameter $\zeta$ for size 6 trees. The minimum value that the

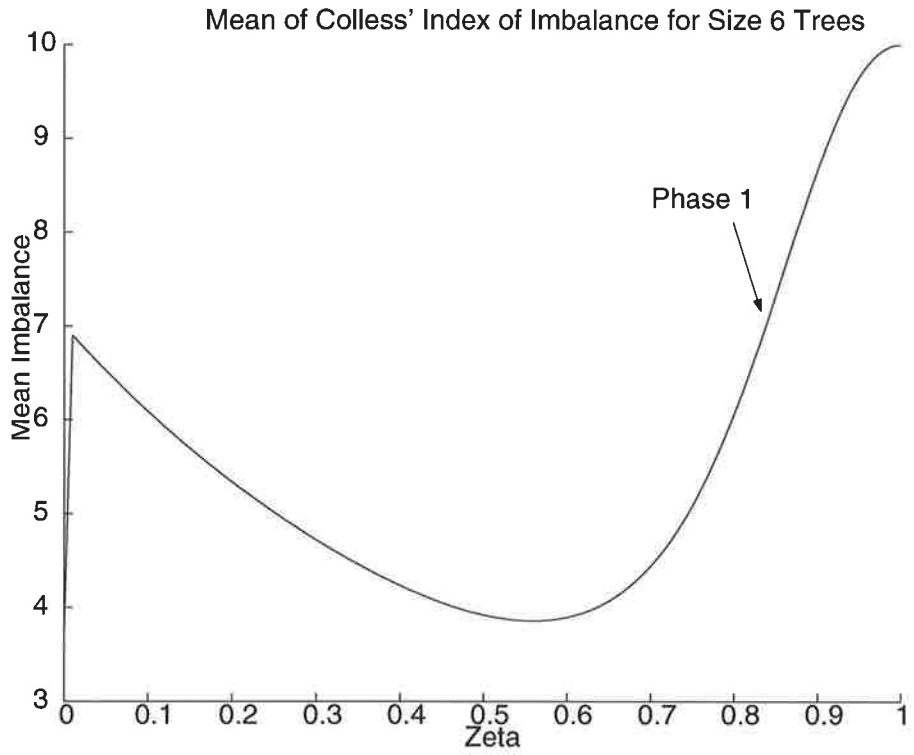


Figure 6.3.8: Mean Imbalance for Size 6 Trees

mean imbalance is allowed to take for size 6 trees is 2, when the $C_{2,6}$ imbalance class predominates, see Figure 6.2.1. However, as can be seen from the graph when $\zeta = 0$ the mean imbalance is approximately, 3.7. This implies that at $\zeta = 0$ the imbalance classes that predominate are $C_{2,6}$ and $C_{5,6}$. They predominate because these topologies only require two $3 \rightarrow 2$ hidden transitions, whereas all other topologies require more. The mean is slightly above 3.5 because $C_{5,6}$ is populated by more topologies than $C_{2,6}$.

The second interesting feature of the graph in Figure 6.3.8 is the spike which occurs for small $\zeta$. This spike occurs because as $\zeta$ increases from zero the $C_{7,6}$ class dominates. This imbalance class is generated from the $\mathbb{F}_{2,5}$ and $\mathbb{F}_{1,1}$ topologically isomorphic classes, and for reasons similar to those given above when discussing size 5 trees, $\mathbb{F}_{2,5}$ is the dominating topologically isomorphic class for small $\zeta$, hence, size 6 trees that are generated from $\mathbb{F}_{2,5}$ must therefore also predominate.

The final interesting feature of the graph is the gradual decrease in mean imbalance from near 7 to about 3.88 at around $\zeta = 0.61$. Since the mean imbalance dips below 4, this implies that $C_{2,6}$ is gradually acquiring a more prominent role. With $\zeta$ increasing, the dynamics amongst phases becomes more difficult to analyse, however, the number of low probability transitions that need to occur for topologies in $C_{2,6}$ are decreasing and as a result the likelihood of the occurrence of $C_{2,6}$ topologies increases and that of $C_{7,6}$ topologies decreases. This is due, in part, to an increase in the probability of $2 \to 1,1$ and $1 \to 3,3$ transitions; transitions that favour more balanced topologies. Consequently, a number of the topologies in $C_{2,6}$ can now be generated without any low probability transitions.

All the previously studied models, other than the crBD, have the common feature of predicting imbalances that are higher than those found in nature. In this section we have given an example of a one-parameter family of MBTs that has the flexibility of generating a very broad range of mean imbalances. This model predicts mean imbalances that encompass all the predictions of the models discussed in this thesis, and much more. The MBT model is therefore extremely flexible and because of its numerical tractability provides an excellent model of the macroevolutionary process. The final section in this chapter is devoted to determining the computational complexity of the imbalance algorithm we developed in Section 6.2.

## 6.4 The Complexity of the Imbalance Algorithm

In Section 6.2 we discussed the details of the imbalance algorithm. To calculate the probabilities of the imbalance classes of size $s$, we saw that we needed to first determine the probabilities of all the imbalance classes of sizes $t < s$. We reproduce the algorithm here, for convenience,

Set

$$\boldsymbol{p}[\mathbb{C}_{0,1}] = (-D_0)^{-1}\boldsymbol{d},$$

since it consists of only one branch. Then, loop through all tree sizes, $2 \leq t \leq s$.

- For each tree size, $t$, loop through all the imbalances, $0 \leq i \leq 1/2(t-1)(t-2)$.
  - Set $\boldsymbol{p}[\mathbb{C}_{i,t}] = 0$.
  - Loop through all possible daughter subtree sizes, $1 \leq l \leq t-1$.
    - Loop through all possible daughter subtree imbalances, $0 \leq l_i \leq 1/2(l-1)(l-2)$.
      - Test to see whether there exists a non-empty imbalance class $\mathcal{C}_{r_i,t-l}$ such that,

        $$r_i = i - l_i - \big|t - 2l\big|,$$

        with $0 \leq r_i \leq 1/2(t-l-1)(t-l-2)$ and if $\mathcal{C}_{l_i,l}$ is also non-empty then set

$$\begin{aligned}\boldsymbol{p}[\mathbb{C}_{i,t}] \;=\; & \boldsymbol{p}[\mathbb{C}_{i,t}] + (-D_0)^{-1}B\big(\boldsymbol{p}[\mathbb{C}_{l_i,l}] \otimes \boldsymbol{p}[\mathbb{C}_{r_i,t-l}] + \\ & I\{\mathbb{C}_{l_i,l} \neq \mathbb{C}_{r_i,t-l}\}\boldsymbol{p}[\mathbb{C}_{r_i,t-l}] \otimes \boldsymbol{p}[\mathbb{C}_{l_i,l}]\big).\end{aligned}$$

Let $G(n,s)$ be complexity of the algorithm for size $s$ trees with $n$-phase MAPs. $G(n,s)$ is then given by,

$$G(n,s) = \sum_{t=2}^{s} \sum_{i=0}^{\frac{1}{2}(t-1)(t-2)} \sum_{l=1}^{t-1} \sum_{l_i=0}^{\frac{1}{2}(l-1)(l-2)} g(n), \tag{6.4.1}$$

where $g(n)$, gives the number of calculations required for the matrix multiplications of the last step of the algorithm. This equation can be understood when one notices that the first summation results from the first loop over tree size, the second summation is over all the possible imbalances of trees of size $t$, the third summation is over all the possible daughter subtree sizes and the final summation is over the allowed imbalance values for the daughter tree. All other steps in the algorithm are $O(1)$, including the if statement. Now $G(n,s)$ is an overestimation of the actual complexity of the algorithm since, in actual fact, if a suitable triplet does not exist to meet the conditions of the if statement then the matrix multiplications are not performed.

The function $G(n,s)$ is tedious to calculate exactly, consequently, we shall instead determine only its leading term. For large $s$, then, we have

$$\begin{aligned} G(n,s) &\approx \sum_{t=2}^{s} \sum_{i=0}^{\frac{1}{2}(t-1)(t-2)} \sum_{l=1}^{t-1} \frac{1}{2} l^2 g(n) \\ &\approx \sum_{t=2}^{s} \sum_{i=0}^{\frac{1}{2}(t-1)(t-2)} \frac{1}{6} t^3 g(n) \\ &\approx \sum_{t=2}^{s} \frac{1}{12} t^5 g(n) \\ &\approx \frac{1}{72} s^6 g(n). \end{aligned} \tag{6.4.2}$$

Thus the algorithm has complexity $O(s^6 g(n))$, however as we have stated previously, the actual complexity is less than this due to the fact that many suitable triplets do not exist and so the matrix multiplications do not need to be performed. Figure 6.4.1 depicts the CPU running time for the imbalance algorithm for a three phase model for tree sizes ranging from 1 to 50. Together with the imbalance algorithm we have graphed the function

$$f(s) = \frac{1}{72} s^{4.2},$$

to demonstrate that, $O(s^6 g(n))$ is clearly an overestimation. We chose the factor 4.2 purely in order to demonstrate that $O(s^6 g(n))$ is greater than the actual com-

putational complexity of the algorithm.

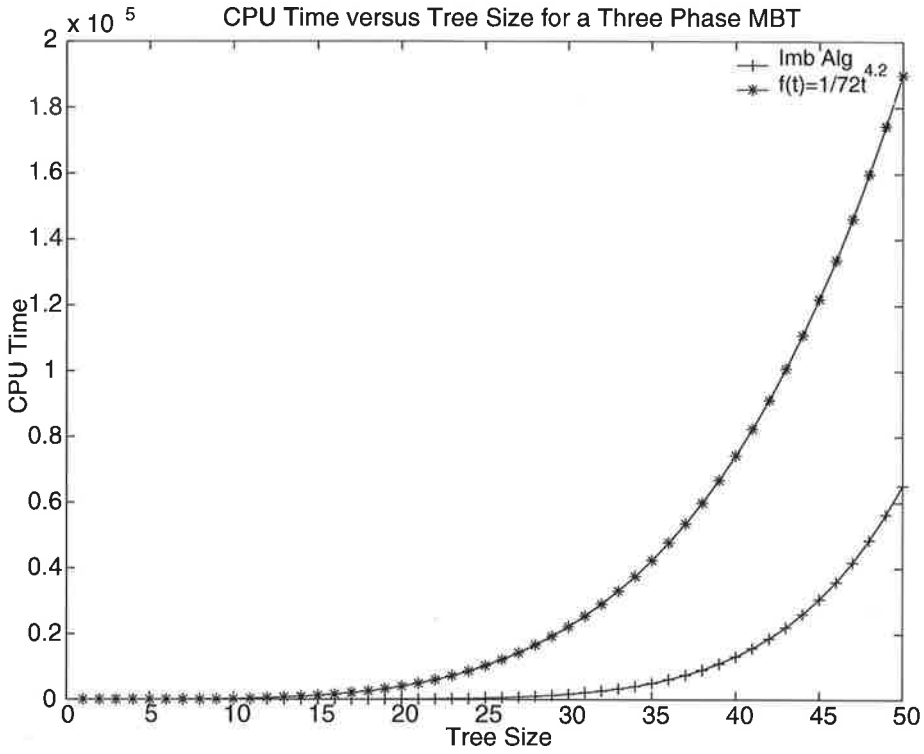


Figure 6.4.1: The computational complexity of the imbalance algorithm.

The MBT representation has allowed us to develop an algorithm that can calculate the distribution of imbalance given tree size and hence deduce the mean imbalance. This algorithm is of polynomial order, with exponent lower than 6 and probably quite significantly lower than 6.

The MBT model has two promising attributes that set it apart from the other models thus far proposed. The MBT is

- extremely flexible, as demonstrated in Section 6.3, and
- amenable to the use of algorithmic techniques.

In contrast, the MR model of Pinelis [26] suffers from an awkward representation that essentially closes the door to relatively efficient numerical analysis. What is also

promising is that by representing the binary-branch point ctMMTBP as an MBT a whole new world of possibilities in the domain of algorithmic analysis has opened up; see Chapter 7 for a further demonstration of the ability to develop algorithms for the MBT, in this case algorithms that determine the probability of eventual extinction. Chapter 8 presents a generalisation of these algorithms to the Markovian tree, the matrix analytic representation of the general ctMMTBP.

# Chapter 7

# Algorithmic Approaches for the MBT

## 7.1 Introduction

As already discussed in previous chapters, the MBT is a special case of the ctMMTBP, see for example, Chapters 3 and 5. The theoretical basis of the ctMMTBP is well established, [2, 21], however very little has been done in developing an algorithmic basis that can be used to determine measures that are of use in a modelling context, [5]. Dorman, Sinsheimer and Lange [5] provided a step towards rectifying some of this problem by providing an algorithmic approach to a ctMMTBP with Poissonian immigration. For example, they were able to calculate the probability of extinction of the process at any time. However, as they acknowledged, their algorithms failed in the important supercritical case as time gets large. The supercritical case is the important case because the probability of eventual extinction is always one for the sub-critical and critical cases.

In this chapter we provide two interesting algorithms that can determine the probability of eventual extinction of the MBT process in the supercritical case, and in Chapter 8 we generalise these algorithms to the general Markovian tree. The

theory of branching processes, [9], tells us that the probability of eventual extinction is the minimal non-negative solution to equation (5.2.13) of Chapter 5, which we reproduce here,

$$\mathbf{0} = \boldsymbol{d} + D_0 \boldsymbol{s} + B\left(\boldsymbol{s} \otimes \boldsymbol{s}\right). \tag{7.1.1}$$

Pre-multiplying equation (7.1.1) by $(-D_0)^{-1}$ and re-arranging we obtain,

$$\boldsymbol{s} = (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\left(\boldsymbol{s} \otimes \boldsymbol{s}\right). \tag{7.1.2}$$

Also note that the equation for the probability measure of the extinct space of trees for the MR model, as expressed in equation (5.4.19),

$$\boldsymbol{p}(\emptyset) = (-D)^{-1}\boldsymbol{d} + (-D)^{-1}B\left(\boldsymbol{p}(\emptyset) \otimes \boldsymbol{p}(\emptyset)\right), \tag{7.1.3}$$

has an identical structure to equation (7.1.2). Therefore, the algorithms developed here also find application in determining $\boldsymbol{p}(\emptyset)$ for the general MR model.

The remainder of this chapter is organised as follows. We discuss tree labelling in Section 7.2 followed by the Depth algorithm in Sections 7.3. The sample path classes of the Neuts algorithm are transformed to binary tree topologies in Section 7.3.1 and as a consequence a neat description of the sample path classes of each iteration of the Neuts algorithm is given in terms of binary tree topologies. The concept of the order of an MBT is defined in Section 7.4 followed by the Order algorithm in Section 7.5. Section 7.6 compares these two algorithms and shows that the Order algorithm converges at a faster rate. The final section, Section 7.7, discusses the logarithmic reduction algorithms and shows that, perhaps surprisingly, the Order algorithm is still the most efficient.

## 7.2 An aside: Tree Labelling and Representation

In Chapter 3 we discussed a node labelling system which we restate here. Suppose that we are at a node labelled $[\psi] = [0, i_1, \ldots, i_m]$, where $i_1, \ldots, i_m \in \{0, 1\}$, the

function $\alpha$ takes us from node $[\psi]$ to node,

$$\alpha(\psi) = [0, i_1, \ldots, i_{m-1}].$$

That is, $\alpha(\psi)$, moves from the current node $[\psi]$ up the tree to its parent node. Recall that we define each node with respect to node [0], consequently, the root node is labelled $[\alpha(0)]$. As a matter of nomenclature, when we are referring to the function $\alpha$ acting on a node $[\psi]$ we do not encase it in square brackets, however, if we are referring to the node $[\alpha(\psi)]$ then we do encase it in square brackets.

The portion of a branch that exists between the two nodes, $[\alpha(\psi)]$ and $[\psi]$, is the ordered pair, $([\alpha(\psi)], [\psi])$. We write $([\alpha(\psi)], [\psi])^{(i)}$ if this branch is an internal branch. We write $([\alpha(\psi)], [\psi])^{(e)}$ if this branch is an extinct branch, and finally, we write $([\alpha(\psi)], [\psi])^{(u)}$ if this branch is unevolved. If a superscipt is not specified then we just refer to that branch generically; its branch type is unimportant.

The function $\theta$ takes us from a node $[\psi]$ and moves us along the parental branch at $[\psi]$ to $[\psi, 1]$. In other words for nodes $[\psi] \neq [\alpha(0)]$,

$$\theta(\psi) = [\psi, 1],$$

and $\theta(\alpha(0)) = [0]$. Suppose there are at least $k$ nodes along the parental branch of a tree of topology $\mathcal{T}_{[\psi]}$, then $\theta^j(\psi)$ moves along the parental branch from $[\psi]$, $j$ parental subnodes for $j \leq k$. Clearly, $\theta^0(\psi) = [\psi]$. We define the number of internal nodes along the parent branch of a tree of topology, $\mathcal{T}$, to be $N(\mathcal{T})$. Commencing at $[\alpha(0)]$ and applying $\theta$, a total of $N(\mathcal{T}) + 1$ times takes us to the leaf node of $\mathcal{T}$'s parental branch. Now, if $[\psi] = [0, i_1, \ldots, i_m]$ is some node, then $\left|\psi\right| = m + 1$ is the depth of that node.

Suppose that the node $[\psi]$ is an internal node of a tree of topology $\mathcal{T}$. Then the topology of the tree based around node $[\psi]$ is the ordered set,

$$\mathcal{T}_{\psi} = \{([\alpha(\psi)], [\psi])^{(i)}, \mathcal{T}_{[\psi,0]}, \mathcal{T}_{[\psi,1]}\},$$

where $\mathcal{T}_{[\psi,0]}$ is the daughter subtree topology based around the daughter node $[\psi, 0]$, and $\mathcal{T}_{[\psi,1]}$ is the parental subtree topology based around the parental subnode $[\psi, 1]$.

Clearly, then, one can see that the probability of a random tree based around node $[\psi]$ eventually attaining a topology of $\mathcal{T}_{[\psi]}$ as $t \to \infty$ is given by,

$$\begin{aligned} \boldsymbol{p}(\mathcal{T}_{[\psi]}) &= P[\{([\alpha(\psi)],[\psi])^{(i)}\}]P[\mathcal{T}_{[\psi,0]}]P[\mathcal{T}_{[\psi,1]}] \\ &= (-D_0)^{-1}B\Big(\boldsymbol{p}(\mathcal{T}_{[\psi,0]}) \otimes \boldsymbol{p}(\mathcal{T}_{[\psi,1]})\Big), \end{aligned}$$

since each event is independent.

The set of internal nodes of an MBT with topology, $\mathcal{T}$, is denoted by $\mathbb{B}_{\mathcal{T}}$. The set of leaf nodes of that same topology is denoted by, $\mathbb{L}_{\mathcal{T}}$ and so the set of all nodes is,

$$\mathbb{N}_{\mathcal{T}} = \mathbb{B}_{\mathcal{T}} \cup \mathbb{L}_{\mathcal{T}}.$$

Note that we treat the root node, $[\alpha(0)]$ as being an element of $\mathbb{L}_{\mathcal{T}}$.

## 7.3 The Depth Algorithm

Let $q_i$ be the probability of eventual extinction of an MBT that commences its evolution in phase $i$ from node $\alpha(0)$. The vector, $\boldsymbol{q}$ is then the minimal non-negative solution of equation (7.1.2), [9], which we write as $\boldsymbol{s} = \boldsymbol{f}(\boldsymbol{s})$. Harris in his seminal work on branching processes, [9], exploited an iterative scheme to solve the equation for the probability of eventual extinction for the discrete-time multi-type branching process. In brief then, suppose that $\boldsymbol{q}_0$ is any vector in the unit cube of the appropriate dimension. Then,

$$\lim_{k\to\infty} \boldsymbol{f}_k(\boldsymbol{q}_0) = \boldsymbol{q}, \tag{7.3.1}$$

where $\boldsymbol{f}_{k+1}(\boldsymbol{s}) = \boldsymbol{f}\big(\boldsymbol{f}_k(\boldsymbol{s})\big)$ is the generating function of the offspring probability distribution for the $k+1$-st generation. A similar iterative approach was exploited by Neuts to solve for $G$ in the level-independent QBD environment. The algorithm of Neuts can be understood quite simply by a very neat physical interpretation as discussed in Chapter 4. At each step of the algorithm, the set of sample paths that

are measured are those that commence in $\mathcal{L}(m)$ and terminate in $\mathcal{L}(m-1)$ upon their first visit, such that

1. the maximum attainable level depends on how many iterations have been performed,
2. there is a restriction on the positions of the left and right transitions, and
3. there is a restriction on how many left transitions are allowed.

Consequently, this space although well characterised, is not easily described. It is more difficult to give a physical interpretation for the algorithm of Harris, particularly since $\boldsymbol{q}_0$ can be any vector in the unit cube.

In this section we derive an algorithm that has similarities to both the Harris algorithm and the Neuts algorithm. We call the algorithm the Depth algorithm, in line with the physical interpretation of equation (7.1.2). The major difference between the algorithm presented here and the algorithm of Harris is that here we are dealing with the binary-branch point ctMMTBP, not the discrete-time process. The interpretation of the Depth algorithm is, as expected, different to that of the Neuts algorithm. There is however a one-to-one correspondence between the sample paths of the Neuts algorithm and the binary tree topologies of the Depth algorithm; we exploit this correspondence in Section 7.3.1 to give a better description of those sample paths.

In the MBT context, then, we shall implement the following recursion on equation (7.1.2),

$$\begin{aligned} \boldsymbol{s}(0) &= (-D_0)^{-1}\boldsymbol{d} \\ \boldsymbol{s}(l) &= (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\left(\boldsymbol{s}(l-1)\otimes\boldsymbol{s}(l-1)\right),\ l\geq 1. \end{aligned} \tag{7.3.2}$$

The similarity to the algorithm of Neuts can now be seen directly. However, a straight application of the interpretation of the Neuts algorithm is not valid. The simple left and right transition structure of the level-independent two-dimensional

QBD process is not sufficiently rich to physically account for the evolution of the MBT on the space of extinct trees. However, by discarding the level-based physical interpretation we can give the Depth algorithm a more natural interpretation.

**Definition 2** *The depth, $\delta$, of a Markovian binary tree of topology $\mathcal{T}$ is*

$$\delta(\mathcal{T}) = \max_{\psi \in \mathbb{B}_{\mathcal{T}}} \{|\psi|\}.$$

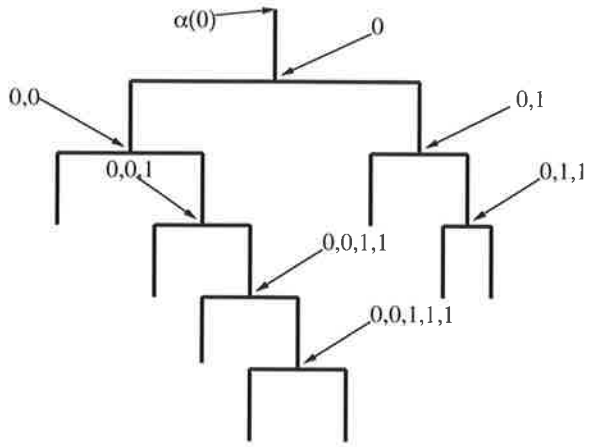


Figure 7.3.1: An example of an MBT of depth 5.

Figure 7.3.1 gives an example of an MBT whose root node and internal nodes have been labelled. Its depth, $\delta$ is therefore,

$$\delta(\mathcal{T}) = \max\{|0|, |0,0|, |0,1|, |0,0,1|, |0,1,1|, |0,0,1,1|, |0,0,1,1,1|\} = 5.$$

We now state and prove a lemma that is important for the correct physical interpretation of the Depth algorithm. Let $\mathcal{T}(t)$ be an evolving MBT, and, let $|\mathcal{T}(t)|$ be the total number of branches at time $t$. Then,

**Lemma 10** $\lim_{t\to\infty} |\mathcal{T}(t)| < \infty$, *almost surely, if and only if* $\lim_{t\to\infty} \delta(\mathcal{T}(t)) < \infty$, *almost surely.*

**Proof :** Since all states with a non-zero number of branches are transient, $\lim_{t\to\infty} |\mathcal{T}(t)| < \infty$, almost surely, if and only if the tree becomes extinct, almost

surely. Further, the tree is extinct, in the limit as $t \to \infty$, if and only if on every branch there are a finite number of nodes and so $\lim_{t\to\infty} \delta(\mathcal{T}(t)) < \infty$. ■

As a result of the above lemma, let us denote the space of extinct binary trees by $\mathbb{T}_{\delta<\infty}$.

If $[\psi]$ is a node, then let $\phi(\alpha(\psi))$ be the phase that the branch $([\alpha(\psi)], [\psi])$ was in immediately following node $[\alpha(\psi)]$. The probability of eventual extinction, $q_i$, of a topology $\mathcal{T}_{[\psi]}$ whose parent branch, $([\alpha(\psi)], [\psi])$, commenced in phase $\phi(\alpha(\psi)) = i$ has the following physical interpretation,

$$\begin{aligned} q_i &= P\big[|\mathcal{T}_{[\psi]}| < \infty \big| \phi(\alpha(\psi)) = i\big] \\ &= P\big[\delta(\mathcal{T}_{[\psi]}) < \infty \big| \phi(\alpha(\psi)) = i\big]. \end{aligned} \tag{7.3.3}$$

Thus the probability that an MBT will eventually become extinct is equivalent to the probability that it will have finite depth as $t \to \infty$ given that it began in some phase $\phi(\alpha(\psi))$. In order to simplify the above equation and those that are of a similar form we shall use an abuse of notation to avoid "storms of subscripts" [18], and so we write instead,

$$\boldsymbol{q} = P\big[\delta(\mathcal{T}_{[\psi]}) < \infty \big| \phi(\alpha(\psi))\big]. \tag{7.3.4}$$

So $\boldsymbol{q}$ is just the measure of the space of extinct trees $\mathbb{T}_{\delta<\infty}$.

Define the sequence $\boldsymbol{q}(l)$ for $l \geq 1$ to be the probability that a tree beginning with one branch eventually becomes extinct under the taboo that its depth is $\delta < l + 1$. Mathematically,

$$\boldsymbol{q}(l) = P\big[\delta(\mathcal{T}_{[\psi]}) < l + 1 \big| \phi(\alpha(\psi))\big], \tag{7.3.5}$$

for all $l \geq 0$. In other words $\boldsymbol{q}(l)$ is the probability that a tree will have depth less than $l + 1$ as $t \to \infty$, which is equivalent to saying that $\boldsymbol{q}(l)$ is the probability measure of all extinct trees with $\delta < l + 1$. Clearly $\boldsymbol{q}(0) = (-D_0)^{-1}\boldsymbol{d}$, because the root branch must undergo a catastrophe before a birth.

**Theorem 11** *The sequence* $\{\boldsymbol{q}(l)\}$ *for* $l \geq 0$ *defined by equation (7.3.5) is monotonically increasing and converges to the vector* $\boldsymbol{q}$. *The sequence* $\{\boldsymbol{q}(l)\}$ *also satisfies*

$$\begin{aligned} \boldsymbol{q}(0) &= (-D_0)^{-1}\boldsymbol{d} \\ \boldsymbol{q}(l) &= (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\left(\boldsymbol{s}(l-1) \otimes \boldsymbol{s}(l-1)\right), \; l \geq 1. \end{aligned} \tag{7.3.6}$$

**Proof :** The fact that $\{\boldsymbol{q}(l)\}$ is monotonically increasing is obvious. That it converges to $\boldsymbol{q}$ is also obvious since $\lim_{l\to\infty} \boldsymbol{q}(l) = \lim_{l\to\infty} P\left[\delta(\mathcal{T}_{[\psi]}) < l+1 \middle| \phi(\alpha(\psi))\right] = P\left[\delta(\mathcal{T}_{[\psi]}) < \infty \middle| \phi(\alpha(\psi))\right] = \boldsymbol{q}$.

The next step is to show that $\{\boldsymbol{q}(l)\}$ actually satisfies equation (7.3.6). A tree may eventually become extinct in one of two ways, either there is a catastrophe at $[\psi]$, or there is a branch point at $[\psi]$ and $([\alpha(\psi)], [\psi])$ becomes an internal branch. The two subtrees that are based around $[\psi, 0]$ and $[\psi, 1]$ must then both independently eventually become extinct.

The probability of the first scenario is just given by $(-D_0)^{-1}\boldsymbol{d}$. The second scenario is slightly more complicated. The probability that the root branch, $([\alpha(\psi)], [\psi])$, will eventually give birth to a daughter branch before it undergoes a catastrophe is given by $(-D_0)^{-1}B$. The daughter and parental subtrees that are based around $[\psi, 0]$ and $[\psi, 1]$ must both eventually become extinct under the taboo that each subtree has a depth of at most $l-1$, so that the entire tree has a depth of at most $l$. But the probability of eventual extinction of a tree with at most depth $l-1$ is just $\boldsymbol{q}(l-1)$. Since each of the two subtrees generated by the daughter and the parent branches are independent, we have that,

$$\boldsymbol{q}(l) = (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\left(\boldsymbol{q}(l-1) \otimes \boldsymbol{q}(l-1)\right), \tag{7.3.7}$$

and the proof is complete. ■

The Depth algorithm converges linearly with respect to depth since at each step, $l$, of the algorithm, $\boldsymbol{q}(l)$ is the probability measure of all topologies that have a depth, $\delta \leq l$. At the next step the space increases by all the tree topologies that

have a depth, $\delta = l + 1$. The number of extra topologies that are included at each step is therefore increased by only a finite number of trees.

### 7.3.1 A New Interpretation for the Sample Paths of the Neuts Algorithm

The space of sample paths that are included at each step of the Neuts algorithm cannot be easily characterized, because of the restrictions placed on the number and positions of the left transitions, see Chapter 4. On the other hand, the set of topologies that are included at each step of the Depth algorithm is easily described. For example, at the $l$-th step, the set of tree topologies consists of all those topologies of at most depth $l$.

The similarity between equations (7.3.6) for the MBT and

$$G(l) = (-A_1)^{-1}A_2 + (-A_1)^{-1}A_0G^2(l-1), \tag{7.3.8}$$

for the level-independent QBD process suggests that there exists some relationship between the sample paths of the Neuts algorithm and the tree topologies of the Depth algorithm. In fact, there is a very intimate relationship: the set of sample paths measured at each step of the Neuts algorithm can be transformed to the set of tree topologies that are measured at the equivalent step of the Depth algorithm; this transformation is one-to-one.

To show this to be true, we must understand the characteristics of the sample paths better. The matrices $G(l)$, given by equation (7.3.8) for all $l \geq 1$, record the probability that a sample path commencing in $\mathcal{L}(m)$ will visit $\mathcal{L}(m-1)$ under the taboo that it must remain below $\mathcal{L}(m+l+1)$, by undergoing at most $2^l$ left transitions. The sample paths that fit the above description also have one more restriction, the position of those left transitions is constrained. It is the fact that equation (7.3.8) is quadratic in $G$ that causes these constraints and it is these constraints that make it difficult to easily describe the sets of sample paths at each step

of the algorithm. For the remainder of this section we shall define and investigate these conclusions through consideration of the Depth algorithm.

We know for example, that there are an uncountably infinite number of trees that have the same topology. These trees differ only in the underlying phase process that has occured along each of their branches. We group these trees together because their topologies are identical. Similarly, there are an uncountably infinite number of sample paths that are related by the positions of their left and right transitions. These sample paths differ because the phase transitions for each sample path are different.

Let us denote left transitions by L and right transitions by R. As an example, consider the sample path that is defined by, RLRRLLL. There are an uncountably infinite number of sample paths that have the above characteristic; they differ only in their underlying phase process and hence in the length of time until each transition occurs. The probability measure of this set of sample paths is easily seen to be given by,

$$(-A_1)^{-1}A_0(-A_1)^{-1}A_2(-A_1)^{-1}A_0(-A_1)^{-1}A_0(-A_1)^{-1}A_2(-A_1)^{-1}A_2(-A_1)^{-1}A_2.$$

Let $\mathcal{S}_{m,m-1}$ denote a set of sample paths from level $m$ to level $m-1$ that differ only in their underlying phase transitions. Thus, if $\mathcal{S}_{m,m-1}$ = RLRRLLL, then the probability measure of this class is given by,

$$\begin{aligned} P(\mathcal{S}_{m,m-1}) &= (-A_1)^{-1}A_0(-A_1)^{-1}A_2(-A_1)^{-1}A_0(-A_1)^{-1}A_0 \\ &\quad \times(-A_1)^{-1}A_2(-A_1)^{-1}A_2(-A_1)^{-1}A_2. \end{aligned}$$

In other words the class, $\mathcal{S}_{m,m-1}$, ignores all the hidden transitions that may occur. These transitions are accounted for by the $(-A_1)^{-1}$ term.

Now, let $\mathbb{S}^{(l)}_{m,m-1}$ be the set of sample path classes that commence in $\mathcal{L}(m)$ and visit $\mathcal{L}(m-1)$ that are measured at step $l$ of the algorithm of Neuts. It easy to show that the number of sample path classes at this step is given by

$$N(l) = 1 + N^2(l-1),$$

with $N(0) = 1$. This follows directly from equation (7.3.8) and the fact that

$$G(0) = (-A_1)^{-1}A_2.$$

The spaces $\mathbb{S}^{(l)}_{m,m-1}$ for all $m$ are independent of the level $m$. We denote the sample path classes from $\mathbb{S}^{(l)}_{m,m-1}$ as $\mathcal{S}^l_{m,m-1,i}$, where $i \in \{0, 1, \ldots, N(l) - 1\}$, thus,

$$\mathbb{S}^{(l)}_{m,m-1} = \left\{\mathcal{S}^l_{m,m-1,0}, \mathcal{S}^l_{m,m-1,1}, \ldots, \mathcal{S}^l_{m,m-1,N(l)-1}\right\}, \tag{7.3.9}$$

where for each $l \geq 0$ we reserve, $\mathcal{S}^l_{m,m-1,0}$ to be the sample path class that consists of a single left transition from $\mathcal{L}(m)$ to $\mathcal{L}(m-1)$. Let $\mathcal{S}^1_{m,m+1}$ denote the sample path class that consists of a single right transition from $\mathcal{L}(m)$ to $\mathcal{L}(m+1)$. Every sample path class $\mathcal{S}^l_{m,m-1,i}$ for $i = 1, \ldots, N(l) - 1$ can be constructed by combining two sample path classes, one from $\mathbb{S}^{(l-1)}_{m+1,m}$, and the other from $\mathbb{S}^{(l)}_{m,m-1}$, so let $\mathcal{S}^{l-1}_{m+1,m,j} \in \mathbb{S}^{(l-1)}_{m+1,m}$ and $\mathcal{S}^{l-1}_{m,m-1,k} \in \mathbb{S}^{(l-1)}_{m,m-1}$, then we write,

$$\mathcal{S}^l_{m,m-1,i} = \mathcal{S}^1_{m,m+1}\mathcal{S}^{l-1}_{m+1,m,j}\mathcal{S}^{l-1}_{m,m-1,k}, \tag{7.3.10}$$

where $l \geq 1$, $j, k \in \{0, \ldots, N(l-1) - 1\}$ and $i = N(l-1)j + (k+1)$. This representation is clearly the natural representation when one studies equation (7.3.8). All of the sample path classes at step $l$, except the one sample path class that consists of one left transition, begin with an initial right transition that takes us from $\mathcal{L}(m)$ to $\mathcal{L}(m+1)$, to which we then append a sample path class that commences in $\mathcal{L}(m+1)$ and terminates in $\mathcal{L}(m)$. This sample path class has the constraints and taboos that are imposed at step $l-1$ of the algorithm. To this sample path class we finally append a sample path class that commences in $\mathcal{L}(m)$ and terminates in $\mathcal{L}(m-1)$ that also is under the constraints and taboos imposed upon it at step $l-1$ of the algorithm; this is exactly as in equation (7.3.10). We choose to order the sample path classes in such a way that when we combine the $j$-th and $k$-th sample path classes of the previous iteration we place the new sample path class in the position, $i = N(l-1)j + (k+1)$. What we wish to do, then, is to show that we can transform the set of these sample paths to the equivalent set of binary tree topologies whilst also maintaining that same ordering for the binary tree space.

To begin with, we explain the transformation from the set of sample path classes to the set of binary trees. There are two types of transitions in the Neuts algorithm, left and right transitions. There are also two types of transitions in the Depth algorithm, branch extinctions and internal branch point generation. It then seems natural to apply the following transformation,

- left transition $\rightarrow$ branch extinctions, and
- right transition $\rightarrow$ internal node.

More formally, if the left most unevolved branch is, $([\alpha(\psi)], [\psi])^{(u)}$, then the next right transition of the Neuts sample path class generates an internal node at $[\psi]$ by creating a new daughter branch. Immediately after the transition then, we have

$$[\psi] \rightarrow \left(([\alpha(\psi)], [\psi])^{(i)}, ([\psi], [\psi, 0])^{(u)}, ([\psi], [\psi, 1])^{(u)}\right),$$

where $([\psi], [\psi, 0])^{(u)}$ and $([\psi], [\psi, 1])^{(u)}$ are the unevolved daughter and parental branches. On the other hand if the next transition of the Neuts sample path class is a left transition, then the branch $([\alpha(\psi)], [\psi])^{(u)}$ is made extinct. In other words,

$$([\alpha(\psi)], [\psi])^{(u)} \rightarrow ([\alpha(\psi)], [\psi])^{(e)}.$$

Let $\Psi$ be the transformation that takes us from the space of sample path classes to the space of extinct binary trees, $\mathbb{T}_{\delta<\infty}$,

$$\Psi : \mathbb{S}_{m,m-1} \rightarrow \mathbb{T}_{\delta<\infty}.$$

To show how we transform from $\mathbb{S}_{m,m-1}$ to $\mathbb{T}_{\delta<\infty}$ consider the following sample path class, RRLLRLL. This sample path class consists of two right transitions, followed, by two left transitions another right transition, and then finally two left transitions. Let us perform the transformation to this sample path, so

$$\Psi(\text{RRLLRLL}) = \Psi(\text{R})\Psi(\text{R})\Psi(\text{L})\Psi(\text{L})\Psi(\text{R})\Psi(\text{L})\Psi(\text{L}),$$

because we apply the transformation to each transition from the sample path individually. Applying the transformation to the first right transition we obtain,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(u)},([0],[0,1])^{(u)}\Big)\Psi(\mathrm{R})\Psi(\mathrm{L})\Psi(\mathrm{L})\Psi(\mathrm{R})\Psi(\mathrm{L})\Psi(\mathrm{L}).$$

After the completion of the first transformation we are left with a tree that has two unevolved branches, $([0],[0,0])^{(u)}$ and $([0],[0,1])^{(u)}$. The left most unevolved branch is $([0],[0,0])^{(u)}$. The next transformation is $\Psi(\mathrm{R})$, and this therefore acts on branch $([0],[0,0])^{(u)}$. This right transition generates a branch point at $[0,0]$, and so we obtain,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(u)},([0,0],[0,0,1])^{(u)},([0],[0,1])^{(u)}\Big)$$
$$\Psi(\mathrm{L})\Psi(\mathrm{L})\Psi(\mathrm{R})\Psi(\mathrm{L})\Psi(\mathrm{L}).$$

The next transition is a left transition, and we apply this to the left-most unevolved branch which in this case is $([0,0],[0,0,0])^{(u)}$, and so we get,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(e)},([0,0],[0,0,1])^{(u)},([0],[0,1])^{(u)}\Big)$$
$$\Psi(\mathrm{L})\Psi(\mathrm{R})\Psi(\mathrm{L})\Psi(\mathrm{L}).$$

The next transition is a left transition again, and applying the transformation to the left-most unevolved branch $([0,0],[0,0,1])^{(u)}$, we obtain,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(e)},([0,0],[0,0,1])^{(e)},([0],[0,1])^{(u)}\Big)\Psi(\mathrm{R})\Psi(\mathrm{L})\Psi(\mathrm{L}).$$

At this point the left subtree based around node $[0,0]$ is terminated. The left most unevolved branch is now, $([0],[0,1])^{(u)}$. To this branch we then apply the transformation of the next right transition, to get

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(e)},([0,0],[0,0,1])^{(e)},([0],[0,1])^{(i)},$$
$$([0,1],[0,1,0])^{(u)},([0,1],[0,1,1])^{(u)}\Big)\Psi(\mathrm{L})\Psi(\mathrm{L}).$$

The next transition is also a left transition and we apply the transformation to the leftmost unevolved branch $([0,1],[0,1,0])^{(u)}$ to obtain,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(e)},([0,0],[0,0,1])^{(e)},([0],[0,1])^{(i)},$$
$$([0,1],[0,1,0])^{(e)},([0,1],[0,1,1])^{(u)}\Big)\Psi(\mathrm{L}).$$

The final transition is a left transition and we apply the transformation to the last remaining unevolved branch, $([0,1],[0,1,1])^{(u)}$, to obtain,

$$\Big(([\alpha(0)],[0])^{(i)},([0],[0,0])^{(i)},([0,0],[0,0,0])^{(e)},([0,0],[0,0,1])^{(e)},([0],[0,1])^{(i)},$$
$$([0,1],[0,1,0])^{(e)},([0,1],[0,1,1])^{(e)}\Big).$$

It is obvious that this transformation is indeed well defined and one-to-one. This is due to the simple mapping procedure that we apply; each transition has its unique position. We shall prove shortly that the space $\mathbb{S}^{(l)}_{m,m-1}$ is mapped into the space of binary tree topologies that have depth $\delta \leq l$, which we denote here by $\mathbb{T}_{\delta\leq l}$. This then implies that the space of sample path classes at each step of the Neuts algorithm is transformed into the well described space of binary tree topologies at the equivalent step of the Depth algorithm.

To begin with it is also easy to see that at each step, $l$, of the Depth algorithm there are

$$N(l) = N^2(l-1)+1,$$

topologies of depth $\delta \leq l$ with $N(0) = 1$, since equation (7.3.2) is so similar to equation (7.3.8). So we can label the topologies in $\mathbb{T}_{\delta\leq l}$ just as we did the sample path classes from $\mathbb{S}^{(l)}_{m,m-1}$, so

$$\mathbb{T}_{\delta\leq l} = \{\mathcal{T}_{0,l},\mathcal{T}_{1,l},\ldots,\mathcal{T}_{N(l)-1,l}\}, \tag{7.3.11}$$

where we reserve $\mathcal{T}_{0,l}$ to be the trivial single branch topology for all $l \geq 0$. We can represent every other topology in $\mathbb{T}_{\delta\leq l}$ by combining two topologies from $\mathbb{T}_{\delta\leq l-1}$ at node [0], so

$$\mathcal{T}_{i,l} = \{([\alpha(0)],[0])^{(i)},\mathcal{T}_{j,l-1,[0,0]},\mathcal{T}_{k,l-1,[0,1]}\}, \tag{7.3.12}$$

such that, $i = N(l-1)j+(k+1)$, and where $\mathcal{T}_{j,l-1,[0,0]}$ and $\mathcal{T}_{k,l-1,[0,1]}$ are the topologies of the two trees commencing at [0]. The similarities between, equations (7.3.9) and (7.3.11), and between equations (7.3.10) and (7.3.12) are striking. We next wish to prove the following theorem to formalise this relationship.

**Theorem 12** *If $\mathcal{S}^l_{m,m-1,i}$ is a sample path class of $\mathbb{S}^{(l)}_{m,m-1}$, then*

$$\Psi(\mathcal{S}^l_{m,m-1,i}) = \mathcal{T}_{i,l} \in \mathbb{T}_{\delta \leq l}.$$

**Proof :** To show that,

$$\Psi(\mathcal{S}^l_{m,m-1,i}) = \mathcal{T}_{i,l}$$

is true, we also use induction. Now, for $l = 0$ we have that $\mathbb{S}^{(0)}_{m,m-1} = \{\mathcal{S}^0_{m,m-1,0}\}$, therefore,

$$\Psi(\mathcal{S}^0_{m,m-1,0}) = \{([\alpha(0)], [0])^{(e)}\} = \mathcal{T}_{0,0},$$

and so the hypothesis is clearly true for $l = 0$. Suppose it is true for $l - 1$, that is,

$$\Psi(\mathcal{S}^{l-1}_{m,m-1,j}) = \mathcal{T}_{j,l-1},$$

for all $j = 0, \ldots, N(l-1) - 1$. Clearly, $\Psi(\mathcal{S}^l_{m,m-1,0}) = \mathcal{T}_{0,l}$ for all $l$ so we need to only show that,

$$\Psi(\mathcal{S}^l_{m,m-1,i}) = \mathcal{T}_{i,l},$$

for all $i = 1, \ldots, N(l) - 1$. So,

$$\begin{aligned}
\Psi(\mathcal{S}^l_{m,m-1,i}) &= \Psi(\mathcal{S}^1_{m,m+1}\mathcal{S}^{l-1}_{m+1,m,j}\mathcal{S}^{l-1}_{m,m-1,k}) && (7.3.13)\\
&= \Psi(\mathcal{S}^1_{m,m+1})\Psi(\mathcal{S}^{l-1}_{m+1,m,j})\Psi(\mathcal{S}^{l-1}_{m,m-1,k}) && (7.3.14)\\
&= \{([\alpha(0)], [0])^{(i)}, \mathcal{T}_{j,l-1,[0,0]}, \mathcal{T}_{k,l-1,[0,1]}\}, && (7.3.15)
\end{aligned}$$

where $i = N(l-1)j + (k+1)$ and in the second step we have used the induction hypothesis. We know that the position of a tree in $\mathbb{T}_{\delta \leq l}$ that is formed by combining the $j$-th with the $k$-th subtree from $\mathbb{T}_{\delta \leq l-1}$ is $N(l-1)j + (k+1)$ but this is just $i$, so,

$$\Psi(\mathcal{S}^l_{m,m-1,i}) = \{([\alpha(0)], [0])^{(i)}, \mathcal{T}_{j,l-1,[0,0]}, \mathcal{T}_{k,l-1,[0,1]}\} = \mathcal{T}_{i,l}, \quad (7.3.16)$$

and the theorem is proved. ∎

**Corollary 13**

$$\Psi(\mathbb{S}^{(l)}_{m,m-1}) = \mathbb{T}_{\delta \leq l}. \quad (7.3.17)$$

**Proof :** Follows immediately from Theorem 12. ■

We have now shown that there exists a one-to-one correspondence between the sample path classes of the Neuts algorithm and the binary tree topologies of the Depth algorithm. Further, at each step, $k$, we can map the $N(k)$ sample paths of that step to the $N(k)$ binary trees that are of depth less than or equal to $k$ and therefore give an easily identifiable description to the sample paths of the Neuts algorithm in terms of binary tree topology.

In Section 7.5, an algorithm that is analogous to the QBD algorithm $U$ is developed. We call this algorithm the Order algorithm. There is no counterpart to this algorithm in the branching process domain. We shall see that the Order algorithm converges at a faster rate than the Depth algorithm, in the same way that the algorithm $U$ converges at a faster rate than the algorithm of Neuts.

## 7.4 The Order of an MBT: Definition

**Definition 3** *The order of a single branch is 0. The order of an internal node, $\theta^k(\psi)$, on the parental branch of a topology $\mathcal{T}_{[\psi]}$, is denoted by $\mathbb{O}_n\big(\theta^k(\psi)\big)$ and is given by,*

$$\mathbb{O}_n\big(\theta^k(\psi)\big) = \mathbb{O}\big(\mathcal{T}_{[\theta^k(\psi),0]}\big),$$

*for $k = 0, 1, 2, \ldots, N(\mathcal{T}_\psi)$ . Finally, the order of $\mathcal{T}_\psi$ is given by,*

$$1 + \max_{k=0,1,2,\ldots,N(\mathcal{T}_\psi)} \mathbb{O}_n(\theta^k(\psi)).$$

From the definition it is clear that we need to calculate the order of an MBT recursively. We start with all the nodes of the parental branch and then work our way down the subtrees until we reach the leaf branches. The order of the tree is then 1 more than the highest order node along the parental branch. Consider the tree that is depicted in Figure 7.4.1. Let the topology of the entire tree be denoted by, $\mathcal{T}$. The parental branch of $\mathcal{T}$ has only two internal nodes along the parental

branch, and they are, [0] and [0, 1]. Let us calculate $\mathbb{O}_n(0,1)$ first,

$$\mathbb{O}_n(0,1) = \mathbb{O}(\mathcal{T}_{[0,1,0]}),$$

and

$$\mathbb{O}(\mathcal{T}_{[0,1,0]}) = 1 + \mathbb{O}_n(0,1,0) = 1 + \mathbb{O}(\mathcal{T}_{[0,1,0,0]}) = 1 + 0 = 1,$$

since $\mathcal{T}_{[0,1,0,0]}$ is a single branch and hence has order zero. So,

$$\mathbb{O}_n(0,1) = \mathbb{O}(\mathcal{T}_{[0,1,0]}) = 1.$$

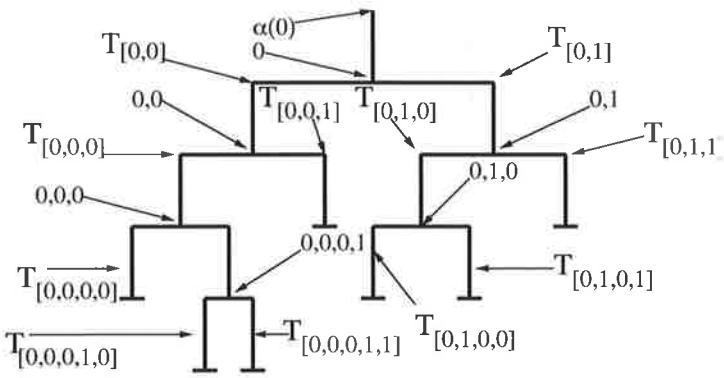


Figure 7.4.1: An example of an order calculation.

Let us now calculate $\mathbb{O}_n(0)$. This is a little more difficult due to the more complex nature of the subtree, $\mathcal{T}_{[0,0]}$. Let us go through the calculation. Now,

$$\mathbb{O}_n(0) = \mathbb{O}(\mathcal{T}_{[0,0]}),$$

and then,

$$\mathbb{O}(\mathcal{T}_{[0,0]}) = 1 + \mathbb{O}_n(0,0),$$

since $\mathcal{T}_{[0,0]}$ has only one internal parental node. However,

$$\mathbb{O}_n(0,0) = \mathbb{O}(\mathcal{T}_{[0,0,0]}),$$

and since $\mathcal{T}_{[0,0,0]}$ is a subtree with two internal parental nodes, we know that,

$$\mathbb{O}(\mathcal{T}_{[0,0,0]}) = 1 + \max\{\mathbb{O}_n(0,0,0), \mathbb{O}_n(0,0,0,1)\}.$$

It is clear that

$$\mathbb{O}_n(0,0,0,1) = \mathbb{O}(\mathcal{T}_{[0,0,0,1,0]}) = 0,$$

since this subtree consists of only a single branch. In addition, we have that,

$$\mathbb{O}_n(0,0,0) = \mathbb{O}(\mathcal{T}_{[0,0,0,0]}) = 0,$$

since this subtree also consists of only one branch. Therefore, we have that,

$$\mathbb{O}(\mathcal{T}_{[0,0,0]}) = 1 + \max\{\mathbb{O}_n(0,0,0), \mathbb{O}_n(0,0,0,1)\} = 1 + 0 = 1.$$

Consequently, then,

$$\mathbb{O}_n(0,0) = \mathbb{O}(\mathcal{T}_{[0,0,0]}) = 1,$$

and thus,

$$\mathbb{O}(\mathcal{T}_{[0,0]}) = 1 + \mathbb{O}_n(0,0) = 1 + 1 = 2,$$

resulting in,

$$\mathbb{O}_n(0) = \mathbb{O}(\mathcal{T}_{[0,0]}) = 2.$$

Finally then, we have that,

$$\mathbb{O}(\mathcal{T}) = 1 + \max\{\mathbb{O}_n(0), \mathbb{O}_n(0,1)\} = 1 + \max\{2,1\} = 1 + 2 = 3.$$

So the order of this particular tree is therefore three.

Note that trees from the same topologically isomorphic class can have different orders. By rotating uneven branch points, the nodes of the parental branch change and hence the daughter subtrees also change. Since order is calculated with respect to the parental branch and its internal nodes, the calculation of the order may therefore be different for topologically isomorphic trees.

The number of topologies that are of order $l$ for all $l \geq 1$ is infinite. One can see this, by understanding that if the orders of all the nodes of a tree of order $l$ are at

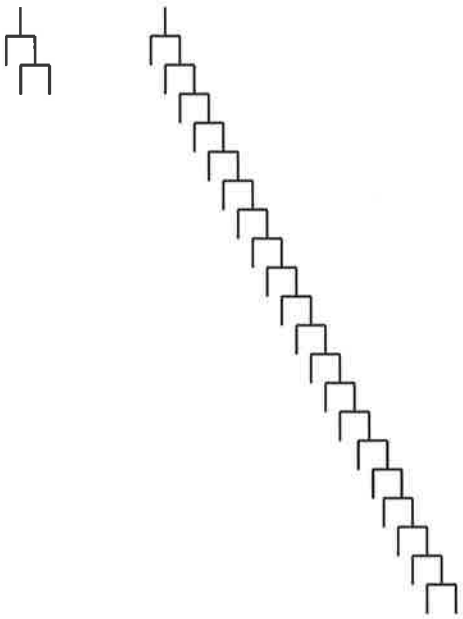

Figure 7.4.2: Two different trees of order one.

most $l-1$ then we can construct a tree from an infinite number of building blocks of order at most $l-1$. Figure 7.4.2 depicts two trees of order one. The second topology has twenty internal parental nodes each of order 0, whereas the first topology has only two.

**Lemma 14** $\lim_{t\to\infty} \left|\mathcal{T}(t)\right| < \infty$, *almost surely, if and only if* $\mathbb{O}(\mathcal{T}(t)) < \infty$, *almost surely.*

**Proof :** Since all the states with a non-zero number of branches is transient, we have that $\lim_{t\to\infty} \left|\mathcal{T}(t)\right| < \infty$, almost surely, if and only if the tree has become extinct, almost surely. Further, the tree is extinct if and only if on every branch there are a finite number of nodes and so, $\mathbb{O}(\mathcal{T}(t)) < \infty$ as $t \to \infty$, almost surely.

■

## 7.5 The Order Algorithm

We saw in the previous section that there are an infinite number of topologies that are of order $l \geq 1$, whereas for depth $l \geq 1$ there are only a finite number of topologies. As a result, if we could devise an algorithm that includes all the topologies of order $k \leq l$ at step $l$, then we would have a much more efficient algorithm than the Depth algorithm. The Order algorithm that we discuss in this section is just that algorithm.

The Order algorithm is analogous to algorithm $U$ from Chapter 4. We begin by re-writing equation (7.1.2) as,

$$\boldsymbol{s} = (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\big(\boldsymbol{s} \otimes I^{(1)}\big)\boldsymbol{s}. \tag{7.5.1}$$

If we substitute this equation into itself we obtain,

$$\begin{aligned} \boldsymbol{s} &= (-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\big(\boldsymbol{s} \otimes I^{(1)}\big)(-D_0)^{-1}\boldsymbol{d} + (-D_0)^{-1}B\big(\boldsymbol{s} \otimes I^{(1)}\big)^2\boldsymbol{s} \\ &= (-D_0)^{-1}\boldsymbol{d} + X(-D_0)^{-1}\boldsymbol{d} + X^2\boldsymbol{s}, \end{aligned}$$

where $X = (-D_0)^{-1}B\big(\boldsymbol{s} \otimes I^{(1)}\big)$. If we now repeat this substitution $l$ times we obtain,

$$\boldsymbol{s} = \sum_{k=0}^{l-1} X^k(-D_0)^{-1}\boldsymbol{d} + R_l(\boldsymbol{s}), \tag{7.5.2}$$

where $R_l(\boldsymbol{s}) = X^l\boldsymbol{s}$ is the remainder term. Now if we take the limit as $l \to \infty$ we obtain,

$$\boldsymbol{s} = \sum_{k=0}^{\infty} X^k(-D_0)^{-1}\boldsymbol{d} + R(\boldsymbol{s}), \tag{7.5.3}$$

where $R(\boldsymbol{s}) = \lim_{l\to\infty} R_l(\boldsymbol{s})$. The above expression is well defined because $\boldsymbol{s}$ is a probability. However, the remainder term may be non-zero. Consider now the minimal non-negative solution to equation (7.1.2), $\boldsymbol{q}$. Now, $\boldsymbol{q}$ must also be the minimal non-negative solution to equation (7.5.3), so we write,

$$\boldsymbol{q} = \sum_{k=0}^{\infty} U^k(-D_0)^{-1}\boldsymbol{d} + R(\boldsymbol{q}), \tag{7.5.4}$$

where $U = (-D_0)^{-1}B\big(\boldsymbol{q} \otimes I^{(1)}\big)$. It is a well known fact from branching process theory that $\boldsymbol{q}$ is the probability measure of all sample paths that eventually have zero living particles. Due to transience and regularity this probability is the same as the probability of all the topologies that consist of a finite number of branches as $t \to \infty$. The first term of equation (7.5.4) is the probability measure of these topologies, and consequently the second term is the probability measure of all those topologies that eventually become extinct after having an infinite number of branches. However, we know from branching process theory that this occurs only on a set of measure zero, and hence the second term, $R(\boldsymbol{s}) = \mathbf{0}$, for the physically significant solution, $\boldsymbol{s} = \boldsymbol{q}$.

We therefore consider the expression

$$\boldsymbol{s} = \sum_{k=0}^{\infty} \Big((-D_0)^{-1}B\big(\boldsymbol{s} \otimes I^{(1)}\big)\Big)^k (-D_0)^{-1}\boldsymbol{d}. \tag{7.5.5}$$

The form of equation (7.5.5) has a very interesting physical interpretation, an interpretation that allows us to develop the Order algorithm.

Let us begin by analyzing,

$$U = (-D_0)^{-1}B\big(\boldsymbol{q} \otimes I^{(1)}\big).$$

Suppose we are at a node, say $[\alpha(\psi)]$, and suppose that eventually an observable event occurs at node $[\psi]$, so that node $[\psi]$ becomes an internal node. The daughter branch at node $[\psi]$ evolves into a subtree that eventually becomes extinct with topology, $\mathcal{T}_{[\psi,0]}$. For the purposes of the above expression, whilst the daughter subtree is evolving towards extinction, the parental branch, $([\psi],[\psi,1])^{(u)}$ is unevolved until the daughter subtree has become extinct with that topology. We call the entity that is based around $[\psi]$, that has representation,

$$\{([\alpha(\psi)],[\psi])^{(i)}, \mathcal{T}_{[\psi,0]}, ([\psi],[\psi,1])^{(u)}\}$$

a $U$-unit. Note that $\mathbb{O}_n(\psi) < \infty$ since $\big|\mathcal{T}_{[\psi,0]}\big| < \infty$. The probability of a $U$-unit, $U$, is given by the product of the probability of the initial branch point, $(-D_0)^{-1}B$, the

eventual extinction of the daughter subtree, $\boldsymbol{q}$ and the suspension of the parental branch $I^{(1)}$, and since each event is independent we have,

$$U = (-D_0)^{-1}B\big(\boldsymbol{q} \otimes I^{(1)}\big).$$

Figure 7.5.1 depicts an example of a $U$-unit; the arrow on the parent branch im-

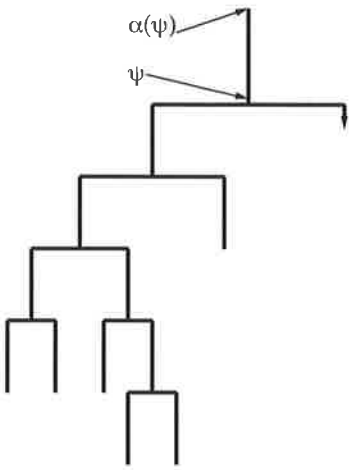


Figure 7.5.1: An example of a $U$-unit.

mediately following $[\psi]$, indicates that the parent branch is suspended, until the daughter subtree has become extinct.

Any extinct tree can be constructed by combining a finite number of $U$-units together; connecting the unevolved parental subnode $[\psi, 1]^{(u)}$ of the previous $U$-unit to the parent root node of the next $U$-unit. Finally, the tree is terminated by a catastrophic transition on the unevolved parental sub-node of the parental branch of the entire topology. For example, $U^k(-D_0)^{-1}\boldsymbol{d}$ is the probability that a tree becomes extinct with a topology that is constructed from $k$ $U$-units combined together, followed by a catastrophe; hence the parental branch has $k$ internal nodes. Therefore,

$$\boldsymbol{q} = \sum_{k=1}^{\infty} U^k(-D_0)^{-1}\boldsymbol{d},$$

gives us the probability of the set of extinct trees consisting of a finite number of $U$-units, or a finite number of internal nodes along the parental branch. Figure 7.5.2

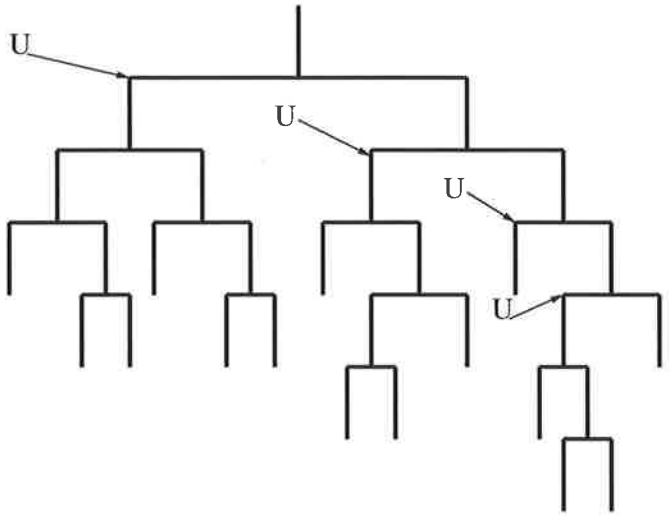


Figure 7.5.2: An example of an extinct tree built from four $U$-units

depicts a tree that is constructed from 4 $U$-units; a tree of this type is an element of the space with probability of $U^4(-D_0)^{-1}\boldsymbol{d}$.

Let $\mathcal{B}[\theta^k(\psi)]$ be the event that a parental branch commencing at node $\psi$ has undergone $k$ observable transitions since the node $[\psi]$. Let $\mathcal{A}[\theta^k(\psi)]$ be the event that the parental branch has undergone $k-1$ observable transitions since $[\psi]$ followed by a catastrophic transition at $[\theta^k(\psi)]$. Let $\phi_p(\psi)$ give the phase of the parental branch immediately after node $[\psi]$, in other words the initial phase of the branch $([\psi],[\psi,1])$. Thus, $\phi_p(\theta^k(\psi))$ is the phase of the parental branch immediately following node $[\theta^k(\psi)]$, in other words, the initial phase of the branch $([\theta^k(\psi)],[\theta^{k+1}(\psi)])$ for all $k$ such that the parental branch has at least $k$ parental subnodes.

**Definition 4** *The matrix $U=[U_{ij}]$ for $i,j=1,2,\ldots,n$ is defined to be*

$$U_{ij}=P[\mathcal{B}[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)<\infty\ \&\ \phi_p(\theta(\psi))=j\big|\phi_p(\psi)=i]. \tag{7.5.6}$$

We stress that $U$ is independent of the position of the initial node. However, the matrix $U$ is dependent upon the initial phase of the parent branch, $\phi_p(\psi)$, and on the phase of the parental branch immediately following $\theta(\psi)$, $\phi_p(\theta(\psi))$. Once again,

as an abuse of notation we shall also write $U$ as,

$$U = P[\mathcal{B}[\theta(\psi)],\, \mathbb{O}_n\big(\theta(\psi)\big) < \infty \;\&\; \phi_p(\theta(\psi))\big|\phi_p(\psi)], \tag{7.5.7}$$

in order to avoid the overuse of subscripts.

**Definition 5** *The vector* $\boldsymbol{q}$ *can be defined as,*

$$\boldsymbol{q} = P\big[\mathbb{O}(\mathcal{T}_{[\psi]}) < \infty\big|\phi\big(\alpha(\psi)\big)\big]. \tag{7.5.8}$$

The Order algorithm to determine the minimal non-negative solution to equation (7.5.5), and hence (7.1.2), is

$$\begin{aligned}
\boldsymbol{q}(0) &= (-D_0)^{-1}\boldsymbol{d}, && (7.5.9)\\
U(l) &= (-D_0)^{-1}B\big(\boldsymbol{q}(l-1)\otimes I^{(1)}\big),\ l \geq 1 && (7.5.10)\\
\boldsymbol{q}(l) &= \sum_{l=0}^{\infty}[U(l)]^l(-D_0)^{-1}\boldsymbol{d}, \qquad l \geq 1. && (7.5.11)
\end{aligned}$$

**Theorem 15** *The sequences* $\{U(l),\, l \geq 1\}$ *and* $\{\boldsymbol{q}(l),\, l \geq 0\}$ *defined by*

$$U(l) = P[\mathcal{B}[\theta(\psi)],\, \mathbb{O}_n(\theta(\psi)) < l \;\&\; \phi_p(\theta(\psi))\big|\phi_p(\psi)], \tag{7.5.12}$$

*and*

$$\boldsymbol{q}(l) = P[\mathbb{O}(\mathcal{T}_{[\psi]}) < l+1\big|\phi(\alpha(\psi))], \tag{7.5.13}$$

*satisfy (7.5.9)-(7.5.11). The two sequences are monotonically increasing and respectively converge to the matrix* $U$ *and the vector* $\boldsymbol{q}$*.*

**Proof :** We first show that the sequences $U(l)$ and $\boldsymbol{q}(l)$ defined by (7.5.12) and (7.5.13) monotonically increase and converge to $U$ and $\boldsymbol{q}$ respectively. Trivially, $\{U(l)\}$ and $\{\boldsymbol{q}(l)\}$ are monotonically increasing. Further,

$$\begin{aligned}
\lim_{l\to\infty} U(l) &= \lim_{l\to\infty} P[\mathcal{B}[\theta(\psi)],\, \mathbb{O}_n(\theta(\psi)) < l \;\&\; \phi_p(\theta(\psi))\big|\phi_p(\psi)]\\
&= P[\mathcal{B}[\theta(\psi)],\, \mathbb{O}_n(\theta(\psi)) < \infty \;\&\; \phi_p(\theta(\psi))\big|\phi_p(\psi)]\\
&= U,
\end{aligned}$$

and

$$\begin{aligned}\lim_{l\to\infty} \boldsymbol{q}(l) &= \lim_{l\to\infty} P[\mathbb{O}(\mathcal{T}) < l+1 \big| \phi(\alpha(0))] \\ &= P[\mathbb{O}(\mathcal{T}) < \infty \big| \phi(\alpha(0))] \\ &= \boldsymbol{q}.\end{aligned}$$

The matrix $U(l)$ gives the probability that beginning at some node, $[\psi]$ in phase $\phi_p(\psi)$, a branch point eventually occurs at the parental subnode, $[\theta(\psi)]$, the daughter subtree that is based around $[\theta(\psi), 0]$ has order at most $l-1$, so that, $\mathbb{O}_n(\theta(\psi)) < l$, and the parent branch at $[\theta(\psi)]$ is in phase $\phi_p(\theta(\psi))$. Therefore, we have,

$$U(l) = P\big[\mathcal{B}[\theta(\psi)],\, \mathbb{O}_n(\theta(\psi)) < l \;\&\; \phi_p(\theta(\psi)) \big| \phi_p(\psi)\big]. \tag{7.5.14}$$

However, the condition $\mathbb{O}_n(\theta(\psi)) < l$ is equivalent to saying that $\mathbb{O}(\mathcal{T}_{[\psi,0]}) < l$, so we can write,

$$U(l) = P\big[\mathcal{B}[\theta(\psi)],\, \mathbb{O}(\mathcal{T}_{[\psi,0]}) < l \;\&\; \phi_p(\theta(\psi)) \big| \phi_p(\psi)\big]. \tag{7.5.15}$$

Since each event occurs independently, we can re-write the above equation as,

$$\begin{aligned}U(l) \;=\; & \sum_{\phi_p(\theta(\psi))} \sum_{\phi(\alpha(\psi,0))} P\big[([\psi],[\theta(\psi)])^{(i)} \;\&\; \phi_{\mathrm{p}}\big(\theta(\psi)\big) \;\&\; \phi\big(\alpha([\psi,0])\big) \big| \phi_{\mathrm{p}}(\psi)\big] \\ & \times P\big[\mathbb{O}(\mathcal{T}_{[\psi,0]}) < l \big| \phi\big(\alpha\big([\psi,0]\big)\big)\big]\end{aligned}$$

The event $\mathcal{B}[\theta(\psi)]$ tells us that a branch point eventually occurs at node $[\theta(\psi)]^{(i)}$ with the daughter branch in phase $\phi(\alpha([\psi,0])$ and the parental branch in phase $[\phi_p(\theta(\psi))]$ immediately after the branch point, the probability of this is given by, $P\big[([\psi],[\theta(\psi)])^{(i)} \;\&\; \phi_{\mathrm{p}}\big(\theta(\psi)\big) \;\&\; \phi\big(\alpha([\psi,0])\big) \big| \phi_{\mathrm{p}}(\psi)\big]$ which is just $(-D_0)^{-1}B$. The term $P\big[\mathbb{O}(\mathcal{T}_{[\psi,0]}) < l \big| \phi\big(\alpha\big([\psi,0]\big)\big)\big]$ is just the probability of the tree based at $[\psi,0]$ becoming extinct with order at most $l-1$, which is given by $\boldsymbol{q}(l-1)$. Therefore we obtain

$$U(l) = (-D_0)^{-1}B\big(\boldsymbol{q}(l-1) \otimes I^{(1)}\big), \tag{7.5.16}$$

where the Kronecker product with $I^{(1)}$ represents that the present branch is frozen with probability 1.

Let $\mathcal{T}^k_{[\psi]}$ be a topology that is based around node $[\psi]$ whose parental branch has exactly $k$ internal nodes before terminating. The first internal parental branch subnode is of course $[\psi]$. Suppose that this topology has, $\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1$. The probability that a random tree will eventually have this property is just,

$$P[\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi(\alpha(\psi))]. \tag{7.5.17}$$

It is not difficult to see that the above expression is equivalent to,

$$P\big[\mathcal{A}\big[\theta^k(\psi)\big] \;\&\; \mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big], \tag{7.5.18}$$

since there are $k-1$ internal nodes from $[\psi]$, the first internal node. At the $k$-th parental subnode from the parent node $[\psi]$, a catastrophic event occurs. Now, the order of a tree is one more than the order of the node on the parental branch with the highest order, and so,

$$\begin{aligned} &P\big[\mathcal{A}\big[\theta^k(\psi)\big] \;\&\; \mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] = \\ &P\big[\mathcal{A}\big[\theta^k(\psi)\big] \;\&\; \max_{i=0,\ldots,k-1} \mathbb{O}_n\big(\theta^i(\psi)\big) < l \big| \phi\big(\alpha(\psi)\big)\big]. \end{aligned} \tag{7.5.19}$$

The expression in equation (7.5.19) is identical to saying that the orders of each of the nodes must all be less than $l$, so we have,

$$\begin{aligned} &P\big[\mathcal{A}\big[\theta^k(\psi)\big] \;\&\; \max_{i=0,\ldots,k-1} \mathbb{O}_n\big(\theta^i(\psi)\big) < l \big| \phi\big(\alpha(\psi)\big)\big] = \\ &P\big[\mathcal{A}\big[\theta^k(\psi)\big],\; \mathbb{O}_n(\psi) < l, \mathbb{O}_n\big(\theta(\psi)\big) < l, \ldots, \mathbb{O}_n\big(\theta^{k-1}(\psi)\big) < l \big| \phi\big(\alpha(\psi)\big)\big]. \end{aligned}$$

Now each node and hence daughter subtree evolves independently, so,

$$\begin{aligned} &P\big[\mathcal{A}\big[\theta^k(\psi)\big], \mathbb{O}_n(\psi)\big) < l, \mathbb{O}_n\big(\theta(\psi)\big) < l, \ldots, \mathbb{O}_n\big(\theta^{k-1}(\psi)\big) < l \big| \phi\big(\alpha(\psi)\big)\big] \\ = \;& \sum_{\phi_p(\psi)} \cdots \sum_{\phi_p(\theta^{k-1}(\psi))} P\big[\mathcal{B}\big[\psi\big], \mathbb{O}_n(\psi)\big) < l \;\&\; \phi_p(\psi)\big) \big| \phi(\alpha(\psi)))\big] \\ \times \;& P\big[\mathcal{B}\big[\theta(\psi)\big], \mathbb{O}_n\big(\theta(\psi)\big) < l, \;\&\; \phi_p\big(\theta(\psi)\big) \big| \phi_p(\psi)\big)\big] \\ &\vdots \\ \times \;& P\big[\mathcal{B}\big[\theta^{k-1}(\psi)\big], \mathbb{O}_n\big(\theta^{k-1}(\psi)\big) < l, \;\&\; \phi_p\big(\theta^{k-1}(\psi)\big) \big| \phi_p\big(\theta^{k-2}(\psi)\big)\big] \\ \times \;& P\big[\mathcal{A}\big[\theta^k(\psi)\big] \big| \phi_p\big(\theta^{k-1}(\psi)\big)\big]. \end{aligned}$$

The first $k$ terms are each just terms of the form of $U(l)$ and the last term is just equal to $(-D_0)^{-1}\boldsymbol{d}$, because after the $k$-th internal node of the parental branch, a catastrophe occurs. In other words,

$$P[\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi(\alpha(\psi))] = U^k(l)(-D_0)^{-1}\boldsymbol{d}. \tag{7.5.20}$$

However, to obtain $\boldsymbol{q}(l)$ we must sum over all $k$, that is, over all the possible number of internal branch points along the parental branch, so,

$$\begin{aligned} \boldsymbol{q}(l) &= \sum_{k=0}^{\infty} P[\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi(\alpha(\psi))] \\ &= \sum_{k=0}^{\infty} U^k(l)(-D_0)^{-1}\boldsymbol{d}. \end{aligned} \tag{7.5.21}$$

■

## 7.6 Comparing the Depth and Order Algorithms

In the level-independent QBD case, the algorithm $U$ converges linearly with respect to level and it converges at a faster rate than the Neuts algorithm. The reason for this is that all the sample paths included in each iteration $l$ of the Neuts algorithm are also included at the $l$-th iteration of the algorithm $U$. However, the algorithm $U$ does not place a constraint on the number or pattern of left transitions, unlike the Neuts algorithm and hence includes many more sample paths.

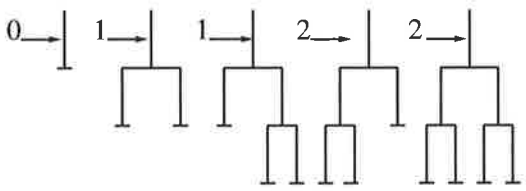


Figure 7.6.1: The space of trees included at the second iteration of the Depth algorithm, with their order also indicated

It is interesting to consider whether there is a similar relationship between the Depth and Order algorithms. Figure 7.6.1 illustrates all the topologies of the trees represented at the second iteration of the Depth algorithm. The order of each tree has also been indicated. The maximum order of the tree topologies at the second iteration is two. This illustrates a more general property.

**Theorem 16** *The highest order tree in the $l$-th iteration of the Depth algorithm is $l$.*

**Proof :** The proof is by induction. At the zeroth iteration of the Depth algorithm, only the single branch is included, and by definition the order of a single branch is zero. Hence the statement is true.

Suppose it is true for the $l$-th iteration, that is, the maximum order tree is $l$. At the $l+1$-st iteration of the Depth algorithm, the space of trees is constructed by combining two trees from the $l$-th step at a branch point; the daughter and parental subtrees. By the induction hypothesis, the daughter and parental subtrees are of order at most $l$. The order of the combined tree is 1 plus the order of the daughter subtree with the highest order, and this subtree is clearly the above daughter subtree. Therefore the order of the tree is at most $1+l$. Thus the maximum order at the $l+1$-th step is $l+1$. ■

The $l$-th iteration of the Order algorithm includes all trees of order at most $l$ without any restriction on the depth. Hence it includes all of the trees that are included at the $l$-th iteration of the Depth algorithm. Thus the Order algorithm converges more rapidly than the Depth algorithm. Figure 7.6.2 depicts a tree that appears at the first iteration of the Order algorithm, but not until the 20-th iteration of the Depth algorithm.

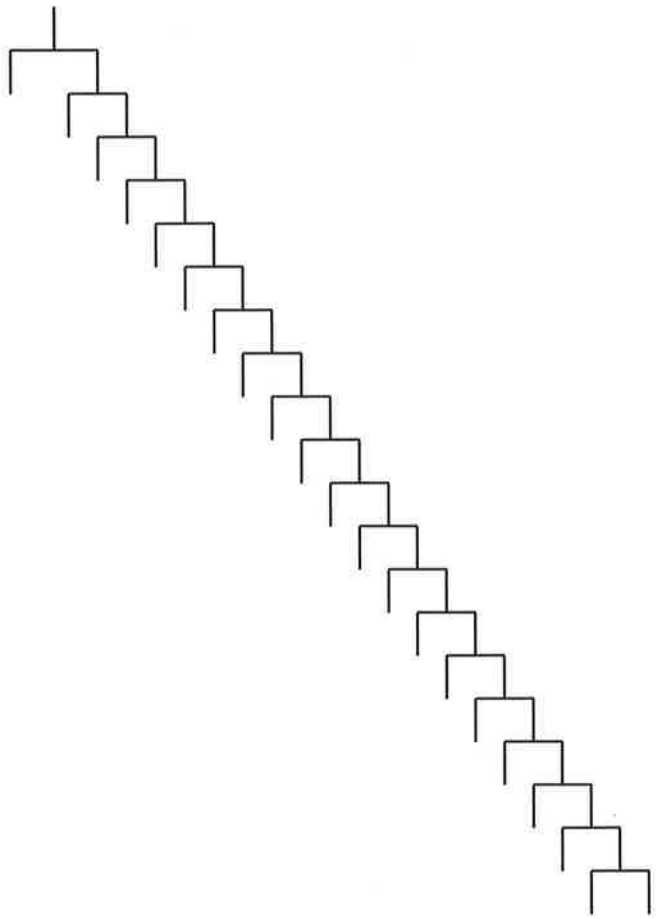

Figure 7.6.2: A tree of order 1 that only appears at the 20-th iteration of the Depth algorithm.

### 7.6.1 Numerical Comparison of the Depth and Order Algorithms

Consider an MBT with

$$D_0 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{7.6.1}$$

and

$$B = \begin{bmatrix} 1-\epsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}, \tag{7.6.2}$$

where $0 \leq \epsilon \leq 0.5$. Figure 7.6.3 compares the average CPU times for the Depth and Order algorithms using the above simple MBT. The algorithms were both ran one hundred times. As can be seen from the figure, the Order algorithm does indeed outperform the Depth algorithm, Notice how as $\epsilon$ approaches 0.5 both algorithms

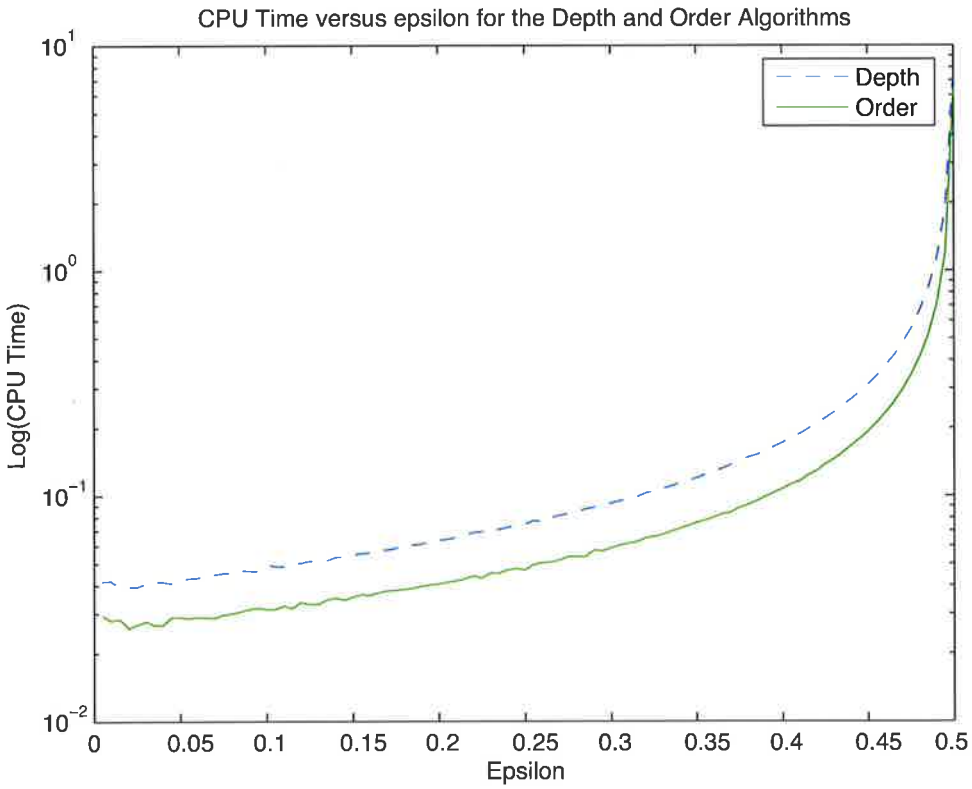


Figure 7.6.3: Comparison of the Depth and Order algorithms as $\epsilon$ varies from 0 to 0.5.

take considerably longer to determine the probability of eventual extinction, since the dominant eigenvalue of $D_0+BC$ approaches 0 and the process becomes critical.

## 7.7 Logarithmic Reduction Algorithms

The Neuts algorithm and the algorithm $U$ were developed to find $G$ in the level-independent QBD domain because each of the non-absorbing levels are identical. These algorithms can be interpreted physically by analyzing the levels of the QBD process and the sample paths in and between these levels. The MBT on the other hand, is a level-dependent QBD and it is rather striking that essentially level-independent algorithms can be developed. That these exist is a consequence of the special higher level transition structure of the MBT. This transition structure is based on the fact that there is no interaction between any of the living branches of an MBT; they evolve independently. This means that each branch of an MBT can be isolated and allowed to evolve whilst all the others are suspended. It is the independence of branch evolution that allows algorithms analogous to the Neuts algorithm and the algorithm $U$ to be applied successfully, despite the fact that they are essentially level-independent algorithms.

Since the Depth and Order algorithms converge linearly with respect to depth and order we wish to study whether there are algorithms that converge at a faster rate that can be implemented in the MBT regime. Algorithms for analyzing QBDs that converge quadratically with respect to level have already been suggested. For example, the logarithmic reduction algorithms [3, 17, 28] are in this class. It would be desirable to develop a quadratically convergent algorithm with respect to some quantity, such as level, or a more generalized order concept in the MBT regime. We shall investigate the level-independent and level-dependent logarithmic reduction algorithms each in turn.

The level-independent logarithmic reduction algorithm (LILRA) can be found in [17, 18] and we discussed it in Chapter 4. Define $\mathbb{C}(l)$ to be the complexity level of a tree for $l \in \{0, 1, 2, \ldots\}$. We call $\mathbb{C}$ the complexity level of the tree because we do not specify whether it is the level in a QBD sense, or another concept such as order or depth as defined earlier. Furthermore, define $\gamma(l)$ to be the first passage time

into complexity level $l$ for $l = 0, 1, 2, \ldots$, that is, $\gamma(l) = \inf\{t > 0 | X(t) \in \mathbb{C}(l)\}$.

The matrix $U^{[k]}$ has the form,

$$U^{[k]} = H^{[k]}L^{[k]} + L^{[k]}H^{[k]}, \tag{7.7.1}$$

where

$$H^{[k]} = P[\gamma(2^{k+1}) < \gamma(0), \ \& \ X(\gamma(2^{k+1}))|X(0) \in \mathbb{C}(2^k)], \tag{7.7.2}$$

and

$$L^{[k]} = P[\gamma(0) < \gamma(2^{k+1}), \ \& \ X(\gamma(0))|X(0) \in \mathbb{C}(2^k)]. \tag{7.7.3}$$

Suppose we are currently in complexity level, $\mathbb{C}(2^{k+1})$. From this complexity level, applying the operator $H^{[k]}$ takes us to complexity level $\mathbb{C}(2^{k+1}+2^k)$ and then applying the $L^{[k]}$ operator takes us down to complexity level $\mathbb{C}(2^{k+1})$. On the other hand, if we apply operator, $L^{[k]}$ first we find that we go down to $\mathbb{C}(2^k)$, and then applying operator $H^{[k]}$ takes us to level $\mathbb{C}(2^{k+1})$. By the mere complexity with which trees evolve, the manner in which the tree traverses from $\mathbb{C}(2^{k+1}+2^k)$ to $\mathbb{C}(2^{k+1})$ must be different from the way in which the tree traverses from $\mathbb{C}(2^k)$ to $\mathbb{C}(2^{k+1})$. In other words the two uses of the term $L^{[k]}$ require different expressions and similarly the two uses of $H^{[k]}$ also require different expressions. Consequently equation (7.7.1) is not valid for the MBT, and thus the level-independent algorithm is not applicable.

The level-dependent logarithmic reduction algorithm (LDLRA) [3, 28] can, of course, be applied to the QBD representation of the MBT. It is appealing since it is known to converge quadratically with respect to level. So whether the LDLRA should be applied depends on its overall efficiency when compared to the linearly convergent Order algorithm. Recall from Chapters 4 and 5 that the number of phases in the level-dependent QBD representation of the MBT at level $l$ is denoted by $M_l = n^l$, where there are $n$ distinct particle types.

At the $k$-th iteration, the Order algorithm has complexity $\mathcal{O}(n^3)$ whereas the LDLRA has complexity $\mathcal{O}\big(M_{2^{k-1}}M_{2^k}M_0\big)$. Now in the MBT case this means that the process has complexity,

$$\mathcal{O}\big(n^{2^{k-1}}n^{2^k}\big), \tag{7.7.4}$$

since there are $n^{2^{k-1}}$ phases in $\mathcal{L}(2^{k-1})$, $n^{2^k}$ phases in $\mathcal{L}(2^k)$ and one phase in $\mathcal{L}(0)$. This equation can be simplified to give,

$$\mathcal{O}\big(n^{3(2^{k-1})}\big). \tag{7.7.5}$$

Suppose then, that $n = 2$ and suppose that one thousand iterations of the Order algorithm are required to obtain an acceptable degree of accuracy, whereas only 10 of the LDLRA are required to obtain that same accuracy. Now the Order algorithm requires approximately $\mathcal{O}\big(2^3\big)$ calculations each iteration and there are one thousand iterations, so we have approximately eight thousand calculations. The LDLRA algorithm requires approximately, $\mathcal{O}\big((2^{3(2^9)})\big) = 2^{1536}$ calculations just for the tenth iteration. Clearly, what the LDLRA gains in convergence properties it loses in the number of calculations required at each iteration because of the massively increased size of the matrices involved in the calculation. Thus, despite the Order algorithm converging linearly, it is still more efficient than the LDLRA in most cases.

We could devise an algorithm based on the alternative state space representation given in Section 5.1.1, that is the representation where we count the number of branches in each of the phases. This representation has a smaller state space than the conventional representation we have employed throughout. Might not an algorithm based on this representation perform better than the Order algorithm? In this case, then, we have that,

$$\mathcal{O}\big(M_{2^{k-1}} M_{2^k} M_0\big) = \binom{2^{k-1}+n-1}{2^{k-1}} \binom{2^k+n-1}{2^k}.$$

Now suppose that $n = 2$ and say that we need ten iterations, then we require

$$\binom{2^{k-1}+n-1}{2^{k-1}} \binom{2^k+n-1}{2^k} = (2^9+1)(2^{10}+1) = 525825, \tag{7.7.6}$$

calculations just for the tenth iteration. Therefore the LDLRA based on this representation still seems to be more inefficient than the Order algorithm in most cases, although it would perform better than the LDLRA based on the original state space representation.

The algorithms that have been successfully developed for the MBT have been interpreted using specific measures of complexity, for example, the depth of the tree and the order of the tree. The LDLRA, which has a physical interpretation that is based on the level of the QBD, can be directly applied to the MBT. However such an approach yields matrices that become extremely large even at early stages of the algorithm and hence is not feasible. Another avenue for developing algorithms that converge faster than the Order algorithm is in generating different measures of complexity and finding algorithms with linear convergence with respect to the new measure. We have attempted such an approach, but we believe that these algorithms suffer from a flaw that is similar to the LDLRA flaw, namely, that they are plagued by matrices that at each iteration become progressively larger. The reason for this is that the subtree units, with which the trees are built at intermediate steps, will require multiple branches to be concurrently alive. Whilst we cannot rule out the possibility of the existence of such a measure of complexity, our experience points to no such measure, and thus no other algorithms that converge faster than the Order algorithm.

# Chapter 8

# The General Markovian Tree

## 8.1 Introduction

We have shown in the previous two chapters that the binary-branch point ctMMTBP when represented in the MBT format lends itself easily to algorithmic analysis. The next step in the process is to re-write the general ctMMTBP in terms of a structure similar to that of the MBT. This will confer to the ctMMTBP an excellent foundation from which to begin analysing the process algorithmically. The general Markovian tree, (MT), which forms the basis of this chapter, provides that alternative representation. This representation of the general ctMMTBP as a Markovian tree enables a field that is almost devoid of algorithmic approaches [5] to become subject to powerful matrix-analytic techniques. The alternative interpretation of hidden and observable transitions of the Markovian tree also provides a natural step towards developing a ctMMTBP structure that has correlations between the parent lifetime and the offspring distribution.

In Section 8.2 we define the MT representation. In Section 8.3 the equivalent ctMMTBP representation is stated. In Section 8.4 a general Markovian tree labelling system is introduced. In Section 8.5 the Depth algorithm is discussed. The Depth algorithm is equivalent to the Harris algorithm for the discrete-time multi-

type branching process [9], the difference being however, that we have given this algorithm a novel and interesting physical interpretation. We define the order of a Markovian tree in Section 8.6 and then finally, in Section 8.7 we discuss the Order algorithm for the MT.

## 8.2 The Markovian Tree: Definition

The Markovian tree is a level dependent process with states,

$$X(t) = \big(N(t), \phi_1(t), \ldots, \phi_{N(t)}(t)\big)$$

defined on the state space $\bigcup_{k=0}^{\infty} \big\{\{k\} \times \{1, \ldots, n\}^k\big\}$. The random variable, $N(t)$, denotes the number of living branches at time $t$ and the random variables, $\phi_k(t)$, for all $k \in \{1, \ldots, N(t)\}$ denote the phase of the $k$-th branch at time $t$. Let $\mathcal{N}^m$ be the $m$-fold Cartesian product of $\mathcal{N} = \{1, 2, \ldots, n\}$, for $m \geq 1$. The level of an MT is given by the number of branches that are alive; suppose that there are $m$ living branches, we denote the level by $\mathcal{L}(m)$ for $m \in \{0\} \cup \mathbb{Z}^+$,

$$\mathcal{L}(m) = \big\{(m, \phi_1, \ldots, \phi_m) \big| (\phi_1, \ldots, \phi_m) \in \mathcal{N}^m\big\}.$$

At $\mathcal{L}(m)$ there are $n^m$ possible states. The level, $\mathcal{L}(0)$, is populated by the state with zero branches, that is,

$$\mathcal{L}(0) = \{(0)\}.$$

The transition rate matrix for the process is defined to be

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ \boldsymbol{A}_{-1}^{(1)} & A_0^{(1)} & A_1^{(1)} & A_2^{(1)} & A_3^{(1)} & \ddots \\ \mathbf{0} & A_{-1}^{(2)} & A_0^{(2)} & A_1^{(2)} & A_2^{(2)} & \ddots \\ \mathbf{0} & 0 & A_{-1}^{(3)} & A_0^{(3)} & A_1^{(3)} & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{8.2.1}$$

Since the $Q$-matrix for the process is conservative we have that

$$\sum_{m=-1}^{\infty} A_m^{(k)} = 0,$$

for all $k \geq 1$. Note that we have changed the notation of the $A$ matrices from Chapter 5. The reason for this is that the $m$ in $A_m^{(k)}$ now refers to the number of new branches that are formed at a branch point, for all $m \in \{-1, 0, 1, \ldots\}$, whereas in Chapter 5 the standard QBD nomenclature was utilized. The reason for this change in nomenclature stems from the fact that we can now have more than one new branch emanating from a branch point and thus the process is no longer a QBD process.

The $n^k \times n^k$ matrices $A_0^{(k)}$ are given by,

$$A_0^{(k)} = A_0^{(k-1)} \oplus D_0, \text{ for } k \geq 1, \tag{8.2.2}$$

and $A_0^{(0)} = 0$. Where the matrix $D_0$ has the property that $(D_0)_{ii} < 0$ and $(D_0)_{ij} \geq 0$ for $1 \leq i \neq j \leq n$. The interpretation of $A_0^{(k)}$ is identical to that of equation (5.1.2) from Chapter 5. The matrices $A_{-1}^{(k)}$ are of dimension $n^k \times n^{k-1}$ and are given by

$$A_{-1}^{(k)} = \sum_{j=0}^{k-1} I^{(j)} \otimes \boldsymbol{d} \otimes I^{(k-1-j)}, \ k \geq 1 \tag{8.2.3}$$

where $I^{(0)} = 1$, and for $k \geq 1$ $I^{(k)}$ are the $n^k \times n^k$ identity matrices. The $n \times 1$ vector $\boldsymbol{d}$ has components $d_i \geq 0$ for $1 \leq 1 \leq n$ with at least one component being strictly greater than zero. The matrices $A_{-1}^{(k)}$ have interpretations identical to those of equation (5.1.1) of Chapter 5. The $A_m^{(k)}$ matrices are of dimension $n^k \times n^{k+m}$ for $m \geq 1$ and can be expressed as,

$$A_m^{(k)} = \sum_{j=0}^{k-1} I^{(j)} \otimes B_m \otimes I^{(k-1-j)}, \ k \geq 1. \tag{8.2.4}$$

The element $(B_m)_{i,i_0 i_1 \ldots i_{m-1} i_m}$ gives the rate at which a branch point occurs such that immediately after birth the parental branch will be in phase $i_m$ and the $m$ new daughter branches will be in phases, $i_0, i_1, \ldots, i_{m-1}$, given that the parent branch was in phase $i$ immediately before the birth. Therefore, $A_m^{(k)}$ gives the total rate at which a single branch from the $k$ possible branches will give rise to $m$ daughter branches in the one transition.

This alternative representation of the ctMMTBP in terms of an MT enables a different interpretation. This interpretation results from the distinction between observable (non-singular) transitions and hidden (singular) transitions. We are not concerned with particles but instead consider branches and the phase processes acting on these branches. The phase process generates the correlations that are possible between branch lifetimes, which now may be non-exponential, and the phases of the daughter branches at their birth.

At each branch point of the MBT, we designated the daughter branch to be the left branch and the parental branch to be the right branch. We saw the significance of this in Chapter 7 because this particular orientation allowed for a natural physical interpretation of the resulting algorithms. The purpose of this chapter is to provide numerical algorithms for the general Markovian tree, and as such a similar designation is required. In this case, the right most branch at each branch point is chosen to be the parental branch regardless of how many daughter branches have also been spawned.

Suppose that at time $t$ the process is in a state with $M$ branches and let branch $k \leq M$ be in phase $r$. Let the current state of the process therefore be,

$$\begin{array}{cccccccccc} (M, & a, & \ldots, & b, & r, & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & k+1 & \ldots & M \end{array},$$

where each branch is labelled by the number below that branch. The following transitions are then possible:

- A hidden transition to phase $j \neq r$, occurs with rate $(D_0)_{rj}$. This transition causes the state of the MT to become

$$\begin{array}{cccccccccc} (M, & a, & \ldots, & b, & j, & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & k+1 & \ldots & M \end{array},$$

- An observable transition that generates $m$ branches for $m \geq 1$. Such an observable transition occurs with rate $(B_m)_{r,i_0 i_1 \ldots i_{m-1} i_m}$. As stated above, we

orient the tree such that the right most branch is the parental branch. The new state of the MT is

$$\begin{array}{cccccccccccc} (M+m, & a, & \ldots, & b, & & i_0, & \ldots, & i_{m-1}, & & i_m, & c, & \ldots & d) \\ & 1 & \ldots & k-1 & k & \ldots & & k+m-1 & k+m & k+m+1 & \ldots & M+m \end{array},$$

where the daughter branches $i_0, i_1, \ldots, i_{m-1}$ are designated $k, \ldots, k+m-1$, the parental branch is now the $k+m$-th branch and the branches that were previously labelled $k+1, \ldots, M$ have been re-labelled to $k+m+1, \ldots, M+m$.

- Finally with rate $d_r$ a catastrophe occurs on branch $k$. This causes branch $k$ to cease to exist and the new state is

$$\begin{array}{cccccccc} (M-1, & a, & \ldots, & b, & c, & \ldots, & d) \\ & 1 & \ldots & k-1 & k & \ldots & M-1 \end{array}$$

The branches that were previously labelled $k+1, \ldots, M$ have been re-labelled to $k, \ldots, M-1$.

## 8.3 The Markovian Tree: ctMMTBP Representation

### 8.3.1 Definition

Let $\boldsymbol{f}(\boldsymbol{s}) = \left(f^{(1)}(\boldsymbol{s}), f^{(2)}(\boldsymbol{s}), \ldots, f^{(n)}(\boldsymbol{s})\right)$ be the generating function of the offspring probability distribution for the Markovian tree. Let $\boldsymbol{x}$ be an $n \times 1$ vector, and let $\boldsymbol{x}^{(m)}$ denote the $m$-fold Kronecker product of the vector $\boldsymbol{x}$. In other words, $\boldsymbol{x}^{(m)}$ is defined by,

$$\boldsymbol{x}^{(m)} = \boldsymbol{x}^{(m-1)} \otimes \boldsymbol{x}, \tag{8.3.1}$$

with $\boldsymbol{x}^{(1)} = \boldsymbol{x}$. We define the vector, $\widehat{\boldsymbol{d}}$, to be,

$$\hat{d}_i = \frac{d_i}{-(D_0)_{ii}}, \tag{8.3.2}$$

for all $i = 1, \ldots, n$, the matrix, $\widehat{D}_0$, to be

$$(\widehat{D}_0)_{ij} = \frac{(D_0)_{ij}}{-(D_0)_{ii}}, \tag{8.3.3}$$

for $1 \leq i \neq j \leq n$ with $(\widehat{D}_0)_{ii} = 0$, and the matrix $\widehat{B}_m$ for $m \geq 1$ and $i_0, i_1, \ldots, i_{m-1}, i_m \in \{1, \ldots, n\}$ to be

$$(\widehat{B}_m)_{i,i_0 \ldots i_{m-1} i_m} = \frac{(B_m)_{i,i_0 \ldots i_{m-1} i_m}}{-(D_0)_{ii}}. \tag{8.3.4}$$

The generating function of the offspring probability distribution can then be written in matrix form as

$$\boldsymbol{f}(\boldsymbol{s}) = \widehat{\boldsymbol{d}} + \widehat{D}_0 \boldsymbol{s} + \sum_{m=1}^{\infty} \widehat{B}_m \boldsymbol{s}^{(m+1)}. \tag{8.3.5}$$

### 8.3.2 Regularity and Mean Number of Branches

In order for the process to be regular (that is, non-explosive) [2] we require that

$$\frac{\partial f^{(i)}(\boldsymbol{s})}{\partial s_j} < \infty, \text{ for all } i, j = 1, \ldots, n. \tag{8.3.6}$$

Now, let $(C_m)_{i_0 \ldots i_{m-1} i_m, j}$ be the matrix that counts how many of the $m+1$ branches emanating from a node are in phase $j$ immediately after the creation of that node, that is,

$$(C_m)_{i_0 \ldots i_{m-1} i_m, j} = \sum_{k=0}^{m} I\{i_k = j\}. \tag{8.3.7}$$

The expected number of branches in phase $j$ given that the process began in phase $i$ can be calculated using [2]

$$M(t) = \exp(At), \tag{8.3.8}$$

where $A_{ij} = -(D_0)_{ii} b_{ij}$ and

$$b_{ij} = \left. \frac{\partial f^{(i)}(\boldsymbol{s})}{\partial s_j} \right|_{\boldsymbol{s}=\boldsymbol{e}} - \delta_{ij}. \tag{8.3.9}$$

In the case of the MT this is equal to

$$M(t) = \exp\left((D_0 + \sum_{m=1}^{\infty} B_m C_m)t\right). \tag{8.3.10}$$

The process is then

- subcritical if $\lambda_A < 0$,
- critical if $\lambda_A = 0$, and
- supercritical if $\lambda_A > 0$,

where $\lambda_A$ is the dominant eigenvalue of $A$.

### 8.3.3 Probability of Eventual Extinction

The final property we wish to discuss in this section is the probability of eventual extinction, which we once again denote by $\boldsymbol{q}$. It is well known from branching process theory [2] that if $\lambda_A \leq 0$ the process will eventually become extinct almost surely and if $\lambda_A > 0$ then $\boldsymbol{q} < \boldsymbol{e}$ component-wise. It is this final case that interests us the most. Recall that the probability of eventual extinction of a continuous-time Markovian multi-type branching process is the minimal non-negative solution [2] to

$$\boldsymbol{u}(\boldsymbol{s}) = \boldsymbol{0}, \tag{8.3.11}$$

and for the MT $\boldsymbol{u}(s)$ can easily be shown to be

$$\boldsymbol{u}(\boldsymbol{s}) = \boldsymbol{d} + D_0\boldsymbol{s} + \sum_{m=1}^{\infty} B_m \boldsymbol{s}^{(m+1)}, \tag{8.3.12}$$

using equation (8.3.5). We therefore have that $\boldsymbol{q}$ is the minimal non-negative solution to

$$\boldsymbol{u}(\boldsymbol{s}) = \boldsymbol{d} + D_0\boldsymbol{s} + \sum_{m=1}^{\infty} B_m \boldsymbol{s}^{(m+1)} = \boldsymbol{0}. \tag{8.3.13}$$

We multiply this equation by $(-D_0)^{-1}$ and re-arrange to obtain,

$$\boldsymbol{s} = (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty} (-D_0)^{-1} B_m \boldsymbol{s}^{(m+1)}, \tag{8.3.14}$$

which is the form that is most useful for the discussion of the Depth and Order algorithms to follow.

## 8.4 An Aside: Labelling the Nodes of an MT

Recall that in Chapter 3 we defined a node labelling system for binary trees. In this system, each node was uniquely specified by a binary sequence. Suppose we are at node $[\psi] = [0, i_1, \ldots, i_l]$, where $i_1, \ldots, i_l \in \{0, 1\}$, and if this node is an internal node, then

- the node that is at a depth of $l + 2$ and to the left of $[\psi]$ (which we call the daughter node) is denoted by $[\psi, 0] = [0, i_1, \ldots, i_l, 0]$, and
- the node that is at a depth of $l + 2$ and to the right of $[\psi]$ (which we call the parental subnode) is denoted by $[\psi, 1] = [0, i_1, \ldots, i_l, 1]$.

This node labelling system can by generalized to the Markovian tree. We begin labelling the first non-root node of the MT by [0]. Now, consider the node that has label,

$$[0, i_1, i_2, \ldots, i_l]$$

where $i_1, \ldots, i_l$ are non-negative integers, and suppose that $m+1$ branches emanate from this node; $m$ of these being the daughter branches and one of these being the parental branch. We label the nodes at the tips of the unevolved daughter branches as

$$[0, i_1, i_2, \ldots, i_l, 0]\ [0, i_1, i_2, \ldots, i_l, 1] \quad \ldots \quad [0, i_1, i_2, \ldots, i_l, m-1]$$

and we label the tip of the unevolved parental branch, called the parental subnode, by

$$[0, i_1, i_2, \ldots, i_l, m].$$

Let $\psi$ denote a sequence of integers, such that the first is always a zero. The node labelled by $[\psi]$ has $|\psi|$ indices. In addition, the number of indices of $[\psi]$, $|\psi|$, gives the depth of that node. Let $\alpha(\psi)$ be the mapping that moves us up the tree from $[\psi] = [0, i_1, \ldots, i_{l-1}, i_l]$ to node, $[0, i_1, \ldots, i_{l-1}]$. We call $[0, i_1, \ldots, i_{l-1}]$ the parent node of $[\psi]$. We define the root node of the tree to be $[\alpha(0)]$. This labelling system is depicted in Figure 8.4.1 for a Markovian tree.

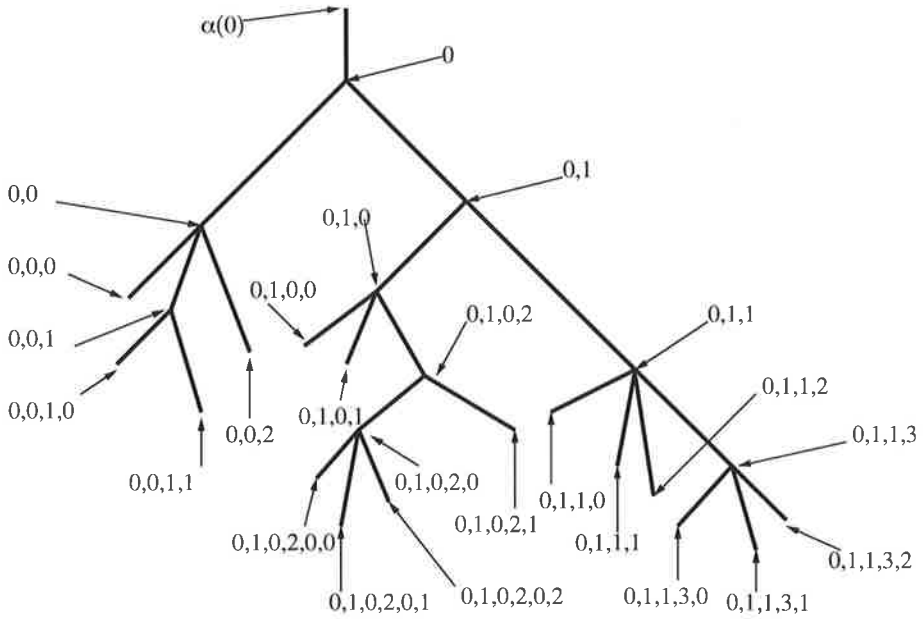


Figure 8.4.1: Labelling nodes in an MT

Just as in the MBT, that portion of a branch between the nodes $[\alpha(\psi)]$ and $[\psi]$ is the ordered pair, $([\alpha(\psi)], [\psi])$. We write $([\alpha(\psi)], [\psi])^{(i)}$ if this branch is an internal branch. We write $([\alpha(\psi)], [\psi])^{(e)}$ if this branch is extinct, and finally we write $([\alpha(\psi)], [\psi])^{(u)}$ if this branch is unevolved. If a superscript is not specified then we just refer to that branch generically; its branch type is unimportant.

Now, suppose that there are $m+1$ subtrees that emanate from node $[\psi]$; $m$ of these being the daughter branches and one of these being the parental branch. We represent the tree that commences from $[\alpha(\psi)]$ by the ordered set,

$$\{([\alpha(\psi)], [\psi])^{(i)}, \mathcal{T}_{[\psi,0]}, \mathcal{T}_{[\psi,1]}, \ldots, \mathcal{T}_{[\psi,m-1]}, \mathcal{T}_{[\psi,m]}\},$$

where $\mathcal{T}_{[\psi,j]}$ is the topology of the $j$-th daughter subtree that is based around $[\psi, j]$, for all $j = 0, 1, 2, \ldots, m-1$ and $\mathcal{T}_{[\psi,m]}$ is the topology of the parental subtree based around the parental subnode $[\psi, m]$.

Let the set of branch points of an MT of topology $\mathcal{T}$ be denoted by $\mathbb{B}_{\mathcal{T}}$, let the set of leaf nodes of an MT of that same topology be denoted by $\mathbb{L}_{\mathcal{T}}$, we then have

that,

$$\mathbb{N}_{\mathcal{T}} = \mathbb{B}_{\mathcal{T}} \cup \mathbb{L}_{\mathcal{T}},$$

is the set of nodes of an MT of topology $\mathcal{T}$.

Since the number of daughter branches that are generated at each internal node, $[\psi]$, is finite but unbounded, we let,

$$\sigma(\psi) = \max\{j : [\psi, j] \in \mathbb{N}_{\mathcal{T}}\}, \tag{8.4.1}$$

be the total number of branches that emanate from $[\psi]$. The parental branch is always that branch that is created from nodes, $[\psi]$ and $[\psi, \sigma(\psi)]$, that is

$$([\psi], [\psi, \sigma(\psi)]).$$

Suppose that $[\psi]$ is either the root node or an internal node, then let the function, $\theta$ be defined by,

$$\theta(\alpha(0)) = [0],$$

and for $[\psi] \neq [\alpha(0)]$,

$$\theta(\psi) = [\psi, \sigma(\psi)].$$

The function $\theta$ is well defined, and maps a node, $[\psi]$ of depth $|\psi|$ to the parental subnode, $[\psi, \sigma(\psi)]$ that emanates from $\psi$ and which is at a depth of $|\psi|+1$. Therefore $\theta^k(\psi)$ traces the pathway of the parental branch that commences from node $\psi$, provided that the parental branch is of at least length $k$ from node $[\psi]$. Clearly, $\theta^0(\psi) = [\psi]$. Finally, if $[\psi]$ is a node, then, $\phi(\alpha(\psi))$ denotes the phase of the branch $([\alpha(\psi)], [\psi])$ immediately after $[\alpha(\psi)]$. The phase of the parental branch, $([\psi], [\theta(\psi)])$ immediately after the node $[\psi]$ is denoted by $\phi_p(\psi)$.

## 8.5 The Depth Algorithm

The Depth algorithm in the context of the MT is the continuous-time analogue of the algorithm of Harris [9]. The Depth algorithm has a very interesting physical

interpretation, similar to that in the MBT, namely, that the maximum depth the Markovian trees can reach at each step of the algorithm can increase by one only. Consequently, we follow a procedure that is similar to Chapter 7.

The Depth algorithm is the recursion on the following set of equations,

$$\boldsymbol{s}(0) = (-D_0)^{-1}\boldsymbol{d} \tag{8.5.1}$$

$$\boldsymbol{s}(l) = (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\boldsymbol{s}^{(m+1)}(l-1), \text{ for, } l \geq 1. \tag{8.5.2}$$

**Definition 6** *The depth, $\delta(\mathcal{T})$, of an MT of topology $\mathcal{T}$ is defined to be,*

$$\delta(\mathcal{T}) = \max_{\psi \in \mathbb{B}_{\mathcal{T}a}} \{|\psi|\}.$$

The MT, $\mathcal{T}$ depicted in Figure 8.5.1 has,

$$\begin{aligned}\delta(\mathcal{T}) &= \max_{\psi \in \mathbb{B}_{\mathcal{T}a}} \{|\psi|\} \\ &= \{|0|, |0,1|, |0,2|, |0,3|, |0,1,0|, |0,1,3|, |0,3,1|, |0,3,1,1|, |0,3,1,1,0|\} \\ &= 5.\end{aligned}$$

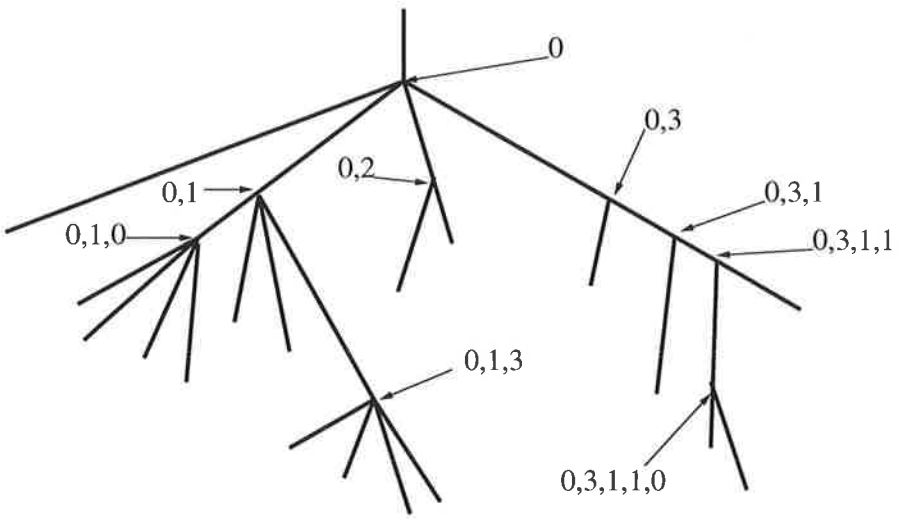


Figure 8.5.1: An MT of depth 5.

**Lemma 17** $\lim_{t\to\infty}\left|\mathcal{T}(t)\right| < \infty$. *almost surely, if and only if* $\lim_{t\to\infty}\delta(\mathcal{T}(t)) < \infty$, *almost surely.*

**Proof :** The proof is similar to Lemma 10 from Chapter 7. ∎

Since trees that eventually becomes extinct are of finite depth, almost surely, the probability of eventual extinction of an MT is

$$\begin{aligned} \boldsymbol{q} &= P\big[\left|\mathcal{T}_{[\psi]}\right| < \infty \big| \phi(\alpha(\psi))\big] \\ &= P\big[\delta(\mathcal{T}_{[\psi]}) < \infty \big| \phi(\alpha(\psi))\big]. \end{aligned} \tag{8.5.3}$$

Let $\boldsymbol{q}(l)$ be the probability that a tree, $\mathcal{T}_{[\psi]}$ commencing with one branch will eventually become extinct under the taboo that $\delta(\mathcal{T}_{[\psi]}) < l+1$, for all $l \geq 0$. That is,

$$\boldsymbol{q}(l) = P\big[\delta(\mathcal{T}_{[\psi]}) < l+1 \big| \phi(\alpha(\psi))\big]. \tag{8.5.4}$$

Notice that

$$\boldsymbol{q}(0) = (-D_0)^{-1}\boldsymbol{d}, \tag{8.5.5}$$

because an extinct tree of zero depth cannot undergo any observable transitions.

**Theorem 18** *The sequence* $\{\boldsymbol{q}(l)\}$, *for* $l \geq 0$, *defined by equations (8.5.4) and (8.5.5) is monotonically increasing and converges to the vector* $\boldsymbol{q}$. *The sequence* $\{\boldsymbol{q}(l)\}$ *also satisfies,*

$$\boldsymbol{s}(0) = (-D_0)^{-1}\boldsymbol{d} \tag{8.5.6}$$

$$\boldsymbol{s}(l) = (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\boldsymbol{s}^{m+1}(l-1),\ l \geq 1. \tag{8.5.7}$$

**Proof :** Once again the proof of this theorem follows a very similar format to Theorem 11 from Chapter 7. The fact that $\{\boldsymbol{q}(l)\}$ is monotonically increasing is obvious. That it converges to $\boldsymbol{q}$ is also obvious since $\lim_{l\to\infty}\boldsymbol{q}(l) = \lim_{l\to\infty} P\big[\delta(\mathcal{T}_{[\psi]}) < l+1\big|\phi(\alpha(\psi))\big] = P\big[\delta(\mathcal{T}_{[\psi]}) < \infty\big|\phi(\alpha(\psi))\big] = \boldsymbol{q}$.

To show that $\boldsymbol{q}(l)$ satisfies equation (8.5.7) let us understand the physical evolution of the process. There are only two pathways with which a tree of depth,

$\delta(\mathcal{T}_{[\psi]}) < l+1$, can eventually become extinct. The first is a direct extinction, where the parent branch undergoes a catastrophic transition before any births. The probability of this scenario is just $(-D_0)^{-1}\boldsymbol{d}$. In the second pathway the parent undergoes an observable transition at node $[\psi]$ spawning a finite but unbounded number of daughters with probability $\sum_{m=1}^{\infty}(-D_0)^{-1}B_m$. Clearly, in order for the tree to eventually become extinct with depth $\delta < l+1$, all the daughter subtrees and the parental subtree must each independently become extinct with depths $\delta < l$. The probability of this second pathway is given by $\sum_{m=1}^{\infty}(-D_0)^{-1}B_m\boldsymbol{q}^{m+1}(l-1)$. Hence we have that,

$$\boldsymbol{q}(l) = (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\boldsymbol{q}^{m+1}(l-1). \tag{8.5.8}$$

and the proof is complete. ■

At the $l$-th step of the algorithm the space of extinct trees that are measured includes all those trees from step $l-1$, that is, those trees that have depths, $\delta < l$, plus all those trees of depth $\delta = l$. Consequently, at each step only a finite number of extra trees are included. In Section 8.7 we discuss the Order algorithm which includes infinitely many extra trees at each step and therefore converges at a faster rate than the Depth algorithm but before we can do this we define the order of an MT.

## 8.6 The Order of an MT: Definition

Similarly to Section 7.4 we define the order of an MT. Now, let $N(\mathcal{T}_{[\psi]})$ denote the number of internal nodes along the parental branch of $\mathcal{T}_{[\psi]}$.

**Definition 7** *The order of a single branch is* 0. *The order of a node,* $\theta^k(\psi)$, *of a topology* $\mathcal{T}_{[\psi]}$, *denoted by* $\mathbb{O}_n\big(\theta^k(\psi)\big)$ *is given by,*

$$\mathbb{O}_n(\theta^k(\psi)) = \max_{j=1,2,\ldots,\sigma(\theta^k(\psi))-1}\{\mathbb{O}(\mathcal{T}_{[\theta^k(\psi),j]})\},$$

*for* $k = 0, 1, 2, \ldots, N(\mathcal{T}_{[\psi]})$. *Finally, the order of* $\mathcal{T}_{[\psi]}$ *is given by,*

$$1 + \max_{k=0,1,2,\ldots,N(\mathcal{T}_{[\psi]})} \{\mathbb{O}(\theta^k(\psi))\}.$$

As in Chapter 7 it is clear that we need to calculate the order of an MT recursively. The procedure is similar to that described in that chapter. To illustrate the procedure in the MT case we calculate the order of the topology depicted in Figure 8.6.1. The order of the tree, $\mathcal{T}$ depicted in the figure is,

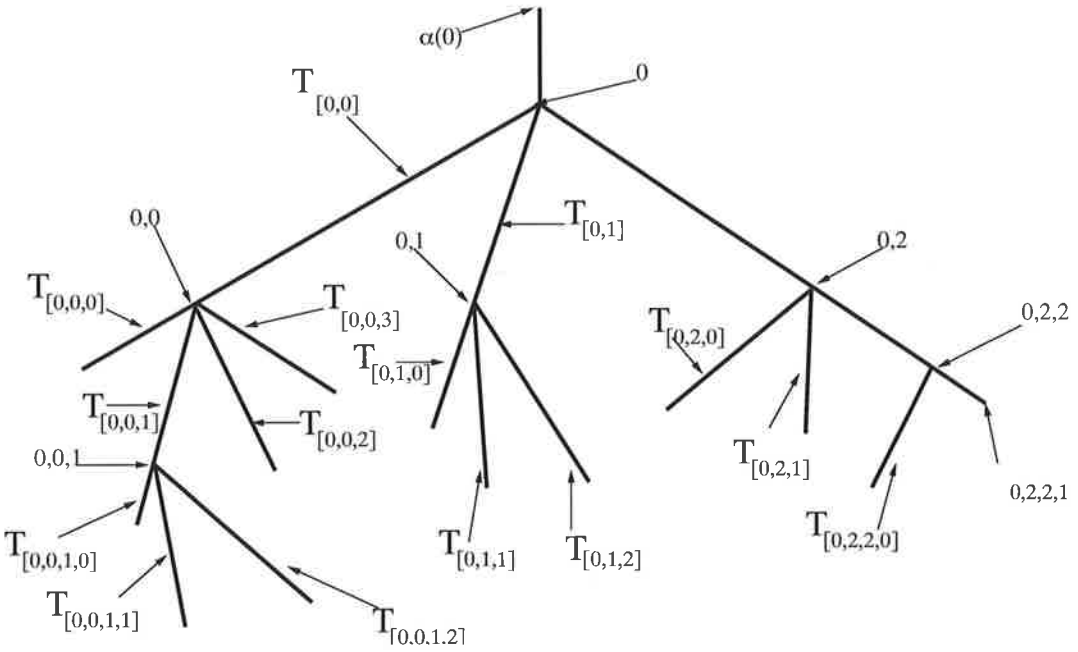


Figure 8.6.1: An example of an order calculation

$$\mathbb{O}(\mathcal{T}) = 1 + \max\{\mathbb{O}_n(0), \mathbb{O}_n(0,2), \mathbb{O}_n(0,2,2)\}.$$

Now,

$$\mathbb{O}_n(0,2,2) = \mathbb{O}(\mathcal{T}_{[0,2,2,0]}) = 0,$$

since $\mathcal{T}_{[0,2,2,0]}$ consists of a single branch. Similarly,

$$\mathbb{O}_n(0,2) = \max\{\mathbb{O}(\mathcal{T}_{[0,2,0]}), \mathbb{O}(\mathcal{T}_{[0,2,1]})\} = 0,$$

since both daughter subtrees, $\mathcal{T}_{[0,2,0]}$ and $\mathcal{T}_{[0,2,1]}$ each consist of only a single branch.

Node [0] is slightly more complicated, because the daughter subtrees are more complex than single branch trees, but still

$$\mathbb{O}_n(0) = \max\{\mathbb{O}(\mathcal{T}_{[0,0]}), \mathbb{O}(\mathcal{T}_{[0,1]})\}.$$

Now

$$\mathbb{O}(\mathcal{T}_{[0,0]}) = 1 + \mathbb{O}_n(0,0),$$

since there is only one branch point on the parental branch of $\mathcal{T}_{[0,0]}$. The order of this branch point is,

$$\mathbb{O}_n(0,0) = \max\{\mathbb{O}(\mathcal{T}_{[0,0,0]}), \mathbb{O}(\mathcal{T}_{[0,0,1]}), \mathbb{O}(\mathcal{T}_{[0,0,2]})\}.$$

The orders of $\mathcal{T}_{[0,0,0]}$ and $\mathcal{T}_{[0,0,2]}$ are both zero, because these daughter subtrees each consist of only one branch. On the other hand, $\mathcal{T}_{[0,0,1]}$ has order,

$$\mathbb{O}(\mathcal{T}_{[0,0,1]}) = 1 + \mathbb{O}_n(0,0,1),$$

because $\mathcal{T}_{[0,0,1]}$ has only one node along its parental branch. Now

$$\mathbb{O}_n(0,0,1) = \max\{\mathbb{O}(\mathcal{T}_{[0,0,1,0]}), \mathbb{O}(\mathcal{T}_{[0,0,1,1]})\} = 0,$$

since both subtrees are only of single branches. Hence,

$$\mathbb{O}(\mathcal{T}_{[0,0,1]}) = 1 + \mathbb{O}_n(0,0,1) = 1 + 0 = 1,$$

and therefore,

$$\mathbb{O}_n(0,0) = \max\{\mathbb{O}(\mathcal{T}_{[0,0,0]}), \mathbb{O}(\mathcal{T}_{[0,0,1]}), \mathbb{O}(\mathcal{T}_{[0,0,2]})\} = \max\{1,0,0\} = 1,$$

so that,

$$\mathbb{O}(\mathcal{T}_{[0,0]}) = 1 + \mathbb{O}_n(0,0) = 1 + 1 = 2.$$

The order of the subtree $\mathcal{T}_{[0,1]}$ can be found in a similar manner and is $\mathbb{O}(\mathcal{T}_{[0,1]}) = 1$. So,

$$\mathbb{O}_n(0) = \max\{\mathbb{O}(\mathcal{T}_{[0,0]}), \mathbb{O}(\mathcal{T}_{[0,1]})\} = \max\{2,1\} = 2,$$

and as a result the order of $\mathcal{T}$ is,

$$\mathbb{O}(\mathcal{T}) = 1 + \max\{\mathbb{O}_n(0), \mathbb{O}_n(0,2), \mathbb{O}_n(0,2,2)\} = 1 + \max\{2,0,0\} = 1 + 2 = 3.$$

Trees that are topologically isomorphic can have different orders. Rotating the nodes changes the parental branch and the daughter subtrees at each node along the parental branch and this changes the calculation of the order.

We stated in the case of the MBT that the number of topologies that exist at any order $l \geq 1$ are infinite, and we saw this because a tree of order $l$ can be constructed such that at each node of the parental branch the daughter subtree has order at most $l-1$, and there can be any number of parental branch subnodes, so the possible number of topologies is infinite. This generalises in a similar manner to the MT; at each and every internal node of the parental branch if the order of the node is at most $l-1$ then the order of the tree is $l$. Parental branches with any number of internal nodes can be created and so the number of topologies with order $l$ is clearly infinite.

**Lemma 19** $\lim_{t\to\infty} \left|\mathcal{T}\right| < \infty$ *if and only if* $\mathbb{O}(\mathcal{T}) < \infty$, *almost surely.*

**Proof :** Once again the proof is similar to Lemma 14 from Chapter 7. ■

## 8.7 The Order Algorithm

The Order algorithm, as we shall see, is a significant improvement on the Depth algorithm. This algorithm reduces to the MBT Order algorithm developed in Chapter 7 if we restrict the process to spawn only one daughter branch at each branch point. Let us re-write equation (8.3.14) as

$$\boldsymbol{s} = (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\Big(\boldsymbol{s}^m \otimes I^{(1)}\Big)\boldsymbol{s}, \tag{8.7.1}$$

where $I^{(1)}$ is the $n \times n$ identity matrix. If we substitute this equation into the right hand side we obtain,

$$\begin{aligned} \boldsymbol{s} &= (-D_0)^{-1}\boldsymbol{d} + \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\left(\boldsymbol{s}^m \otimes I^{(1)}\right)(-D_0)^{-1}\boldsymbol{d} \\ &+ \left(\sum_{m=1}^{\infty}(-D_0)^{-1}B_m\left(\boldsymbol{s}^m \otimes I^{(1)}\right)\right)\left(\sum_{m=1}^{\infty}(-D_0)^{-1}B_m\left(\boldsymbol{s}^m \otimes I^{(1)}\right)\right)\boldsymbol{s} \\ &= (-D_0)^{-1}\boldsymbol{d} + X(-D_0)^{-1}\boldsymbol{d} + X^2\boldsymbol{s}. \end{aligned}$$

where $X = \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\left(\boldsymbol{s}^m \otimes I^{(1)}\right)$. Now if we do this $l$ times, we obtain,

$$\boldsymbol{s} = \sum_{k=0}^{l} X^k(-D_0)^{-1}\boldsymbol{d} + R_l(\boldsymbol{s}), \tag{8.7.2}$$

where $R_l(\boldsymbol{s}) = X^{l+1}\boldsymbol{s}$ is the remainder term. Now if we take the limit as $l \to \infty$ we obtain,

$$\boldsymbol{s} = \sum_{k=0}^{\infty} X^k(-D_0)^{-1}\boldsymbol{d} + R(\boldsymbol{s}), \tag{8.7.3}$$

where $R(\boldsymbol{s}) = \lim_{l\to\infty} R_l(\boldsymbol{s})$. The above expression is well defined because the left hand side, $\boldsymbol{s} \leq \boldsymbol{e}$ componentwise. However, in general the remainder term does not need to be zero, but we can show that for $\boldsymbol{q}$, the minimal non-negative solution of equation (8.7.1), the remainder term $R(\boldsymbol{s})$ is zero on physical grounds, as follows.

Substituting $\boldsymbol{q}$ into equation (8.7.3) we obtain,

$$\boldsymbol{q} = \sum_{k=0}^{\infty} X^k(-D_0)^{-1}\boldsymbol{d} + R(\boldsymbol{q}), \tag{8.7.4}$$

and we know that since $\boldsymbol{q}$ is the minimal non-negative solution of (8.7.1) it is also the minimal non-negative solution to the above equation. Now we also know from branching process theory that $\boldsymbol{q}$ is the probability measure of all the sample paths that eventually have zero living particles. Due to the transience and regularity of the ctMMTBP these sample paths are equivalent, almost surely, to the space of extinct trees that have a finite number of extinct leaf branches. The first term of equation (8.7.4) is the probability measure of all those extinct topologies with a

finite number of branches. On the other hand the second term can be interpreted as being the probability measure of those topologies that have an infinite number of leaf branches that eventually become extinct. Once again using the transience and regularity of the process this space has probability measure zero and so $R(\boldsymbol{q}) = 0$ for $\boldsymbol{q}$. As a result we consider only,

$$\boldsymbol{s} = \sum_{k=0}^{\infty} X^k(-D_0)^{-1}\boldsymbol{d}. \tag{8.7.5}$$

Let

$$U = \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\big(\boldsymbol{q}^m \otimes I^{(1)}\big), \tag{8.7.6}$$

and set $\boldsymbol{s} = \boldsymbol{q}$, the probability of eventual extinction, in equation (8.7.5). Now $\boldsymbol{q}$ and $U$ have very interesting physical interpretations based on their respective equations above. Let us first interpret the matrix $U$. Consider, the $m$-th term from the summation, that is,

$$U_m = (-D_0)^{-1}B_m\big(\boldsymbol{q}^m \otimes I^{(1)}\big). \tag{8.7.7}$$

This term gives the probability that there are $m$ daughters spawned at a branch point each of which generates a subtree that eventually becomes extinct. For the purposes of this expression the parental branch remains alive. We call such a structure a $U_m$-unit; if we do not specify how many daughter branches there are we call the structure a $U$-unit. Figure 8.7.1 represents a $U_4$-unit. The parent branch gives birth to four daughter branches. These four daughter branches generate subtrees that eventually become extinct, whilst the arrow on the parent branch indicates that its evolution has been suspended, that is, it is an unevolved branch. In general, the suspension of the evolution of the parental branch is made manifest in equation (8.7.7) by seeing that the $m$ daughter branches are made extinct by the $\boldsymbol{q}^m$ term, whilst the evolution of the parental branch is governed by the identity matrix, $I^{(1)}$, and hence does not evolve. Allowing the parental branch to remain idle while its daughters all become extinct is possible because of the independent evolution of each branch subsequent

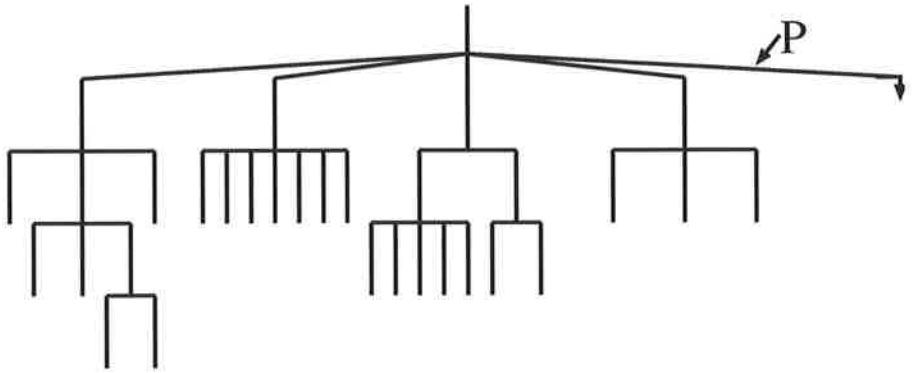


Figure 8.7.1: An example of a $U_4$-unit.

to its birth. To evaluate the matrix $U$ we sum over all $m$, since there is no restriction on the number of births.

To construct an extinct tree, we connect the parental subnode of the previous $U$-unit to the root node of a new $U$-unit. An extinct tree can only be of finite size, almost surely, so only a finite number of $U$-units can be connected. Following the connection of the final $U$-unit, the parent branch must undergo a catastrophe before any other observable transition. As an example, Figure 8.7.2 depicts an extinct tree

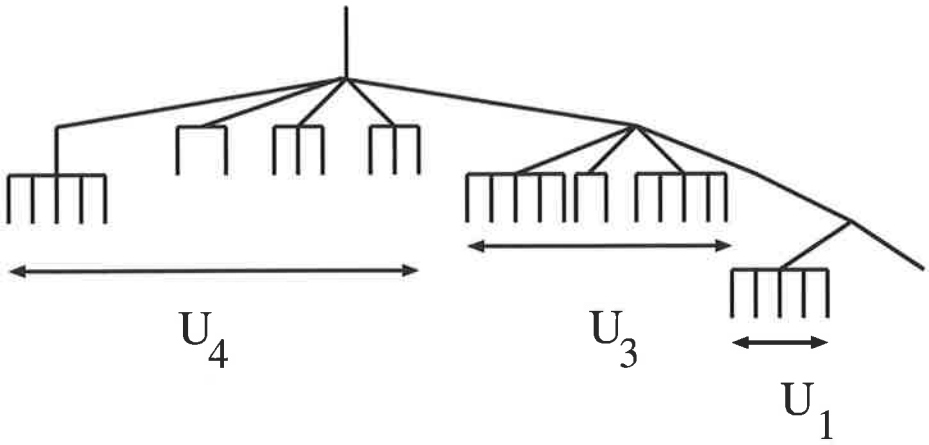


Figure 8.7.2: An example of a tree with three $U_k$-units.

that is constructed from a $U_4$-unit a $U_3$-unit and a $U_1$-unit before final extinction. The probability of obtaining a tree with this description is easily deduced to be,

$U_4U_3U_1(-D_0)^{-1}\boldsymbol{d}$. More generally,

$$U^k(-D_0)^{-1}\boldsymbol{d}, \tag{8.7.8}$$

gives the probability of generating an extinct tree from $k$ $U$-units. In other words, the above expression is the probability measure of the space of extinct trees that are constructed from any combination of $k$ $U$-units, with the parent branch of the $k$-th unit undergoing a catastrophe, with probability $(-D_0)^{-1}\boldsymbol{d}$, to render the tree extinct. Therefore,

$$\sum_{k=0}^{\infty} U^k(-D_0)^{-1}\boldsymbol{d},$$

is the probability measure of the space of all possible extinct trees.

To recap, the structural subunits that can be used to generate extinct trees are the $U$-units. These units are connected using their parental branches as we have shown above. Consider a node, $[\psi] \in \mathbb{B}_{\mathcal{T}}$, we can represent a $U$-unit whose daughter subtrees are spawned at $[\psi]$, as the following ordered set of branches and subtrees,

$$\{([\alpha(\psi)],[\psi])^{(i)}, \mathcal{T}_{[\psi,0]}, \mathcal{T}_{[\psi,1]}, \ldots, \mathcal{T}_{[\psi,\sigma(\psi)-1]}, ([\psi],[\theta(\psi)])^{(u)}\},$$

where the parental branch is denoted by $([\psi],[\theta(\psi)])^{(u)}$ to emphasize the fact that this branch is unevolved.

Now, let $\mathcal{B}_m[\theta^k(\psi)]$ be the event that a parental branch which commences from node $[\psi]$ has undergone $k$ observable transitions since $[\psi]$ such that at $\theta^k(\psi)$ $m$ daughter branches are spawned. If we do not specify the number of daughter branches we write, $\mathcal{B}[\theta^k(\psi)]$ for the event that a parental branch which commences from node $[\psi]$ has undergone $k$ observable transitions since $[\psi]$. Let $\mathcal{A}[\theta^k(\psi)]$ be the event that the parental branch has undergone $k-1$ observable transitions since node $[\psi]$ followed by a catastrophic transition at node $[\theta^k(\psi)]$. As before, let $\phi_p\big(\theta^k(\psi)\big)$ be the phase that the parental branch was in immediately after the $[\theta^k(\psi)]$-th branch point. The initial phase of the parent branch is $\phi_p(\psi)$. The initial phase of a tree of topology $\mathcal{T}_{[\psi]}$ is denoted by $\phi(\alpha(\psi))$.

**Definition 8** *The matrix $U$ is defined to be*

$$U = P\big[\mathcal{B}[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)\big) < \infty\ \&\ \phi_p\big(\theta(\psi)\big)\big|\phi_p(\psi)\big], \tag{8.7.9}$$

*for all $[\psi]$.*

**Definition 9** *The vector $\boldsymbol{q}$ is defined as,*

$$\boldsymbol{q} = P\big[\mathbb{O}(\mathcal{T}_{[\psi]}) < \infty\big|\phi\big(\alpha(\psi)\big)\big], \tag{8.7.10}$$

*for all $[\psi] \neq \alpha(0)$.*

Expressed in this way, we see $U$ as being the probability that beginning at some node, $[\psi]$ in phase $\phi_p(\psi)$, a branch point eventually occurs at node $[\theta(\psi)]$, the orders of each of the daughter subtrees are finite, so, $\mathbb{O}_n\big(\theta(\psi)\big)\big) < \infty$, and the parent branch is suspended at node $[\theta(\psi)]$ in phase $\phi_p\big(\theta(\psi)\big)$. The matrix, $U$, is actually independent of the position of the initial node because the subsequent evolution of any branch that is spawned from that node is independent of the rest of the tree immediately after its birth. The probability of eventual extinction of the tree $\mathcal{T}_{[\psi]}$, $\boldsymbol{q}$, is the probability that $\mathcal{T}_{[\psi]}$ has finite order as $t \to \infty$, given that it commenced from node $[\alpha(\psi)]$ in phase, $\phi\big(\alpha(\psi)\big)$.

The Order algorithm to determine the minimal non-negative solution to equation (8.7.5) is

$$\boldsymbol{s}(0) = (-D_0)^{-1}\boldsymbol{d}, \tag{8.7.11}$$

$$X(l) = \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\big(\boldsymbol{s}^m(l-1)\otimes I^{(1)}\big), \tag{8.7.12}$$

$$\boldsymbol{s}(l) = \sum_{k=0}^{\infty}X^k(l)(-D_0)^{-1}\boldsymbol{d}, \tag{8.7.13}$$

for $l \geq 1$.

**Theorem 20** *The sequences $\{U(l)\big|l \geq 1\}$ and $\{\boldsymbol{q}(l)\big|l \geq 0\}$ defined by*

$$U(l) = P\big[\mathcal{B}[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)\big) < l\ \&\ \phi_p\big(\theta(\psi)\big)|\phi_p(\psi)\big]. \tag{8.7.14}$$

*and*

$$\boldsymbol{q}(l) = P\big[\mathbb{O}(\mathcal{T}_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big], \tag{8.7.15}$$

*satisfy equations (8.7.11)-(8.7.13). The two sequences are monotonically increasing and respectively converge to the matrix $U$ and the vector $\boldsymbol{q}$.*

**Proof :** We first show that the sequences $\{U(l)\}$ and $\{\boldsymbol{q}(l)\}$ defined by (8.7.14) and (8.7.15) monotonically increase and converge to $U$ and $\boldsymbol{q}$ respectively. Trivially, $\{U(l)\}$ and $\{\boldsymbol{q}(l)\}$ are monotonically increasing. Further,

$$\begin{aligned}
\lim_{l\to\infty} U(l) &= \lim_{l\to\infty} P\big[\mathcal{B}[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)\big) < l \ \&\ \phi_p\big(\theta(\psi)\big) \big| \phi_p(\psi)\big] \\
&= P\big[\mathcal{B}[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)\big) < \infty \ \&\ \phi_p\big(\theta(\psi)\big) \big| \phi_p(\psi)\big]\big] \\
&= U,
\end{aligned}$$

and,

$$\begin{aligned}
\lim_{l\to\infty} \boldsymbol{q}(l) &= \lim_{l\to\infty} P\big[\mathbb{O}(\mathcal{T}_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] \\
&= P\big[\mathbb{O}(\mathcal{T}_{[\psi]}) < \infty \big| \phi\big(\alpha(\psi)\big)\big] \\
&= \boldsymbol{q}.
\end{aligned}$$

The matrix, $U_m(l)$ is the probability that beginning at some node $[\psi]$ and phase $\phi_p(\psi)$, a branch point eventually occurs at node $[\theta(\psi)]$, generating $m$ daughter subtrees such that the orders of each of the daughter subtrees is less than $l$, so, $\mathbb{O}_n\big(\theta(\psi)\big)\big) < l$, and the parental branch is suspended at node $[\theta(\psi)]$ in phase $\phi_p\big(\theta(\psi)\big)\}$. Therefore we have,

$$\begin{aligned}
U_m(l) &= P\big[\mathcal{B}_m[\theta(\psi)],\ \mathbb{O}_n\big(\theta(\psi)\big)\big) < l \ \&\ \phi_p\big(\theta(\psi)\big) | \phi_p(\psi)\big] \\
&= P\big[\mathcal{B}_m[\theta(\psi)],\ \max_{k\in\{0,1,\ldots,m-1\}} \{\mathbb{O}\big(\mathcal{T}_{[\psi,k]}\big)\} < l,\ \&\ \phi_p\big(\theta(\psi)\big) | \phi_p(\psi)\big] \\
&= P\big[\mathcal{B}_m[\theta(\psi)],\ \mathbb{O}\big(\mathcal{T}_{[\psi,0]}\big) < l, \ldots, \mathbb{O}\big(\mathcal{T}_{[\psi,m-1]}\big) < l,\ \&\ \phi_p\big(\theta(\psi)\big) | \phi_p(\psi)\big]
\end{aligned}$$

where in the second step we have used

$$\mathbb{O}_n(\theta(\psi)) = \max_{0,1,\ldots,m-1} \{\mathbb{O}\big(\mathcal{T}_{[\psi,k]}\big)\}$$

and in the third step we have used the fact that

$$\max_{0,1,\ldots,m-1}\{\mathbb{O}\big(\mathcal{T}_{[\theta(\psi),k]}\big)\} < l,$$

is equivalent to $\mathbb{O}(\mathcal{T}_{[\psi,k]}) < l$ for all $k = 0, 1, \ldots, m-1$. Thus, each and every daughter subtree must be extinct with order less than $l$. Since each event occurs independently we can write,

$$\begin{aligned} U_m(l) &= P\big[\mathcal{B}_m[\theta(\psi)],\ \mathbb{O}\big(\mathcal{T}_{[\psi,0]}\big) < l, \ldots, \mathbb{O}\big(\mathcal{T}_{[\psi,m-1]}\big) < l \ \&\ \phi_p\big(\theta(\psi)\big)|\phi_p(\psi)\big] \\ &= \sum_{\phi(\alpha(\psi,0)} \cdots \sum_{\phi(\alpha(\psi,m-1))} P\big[([\psi],[\theta(\psi)])^{(i)},\ \phi(\alpha(\psi,0)), \ldots,\ \phi(\alpha(\psi,m-1)), \\ &\qquad\qquad \& \ \phi_{\mathrm{p}}(\theta(\psi))\big|\phi_{\mathrm{p}}(\psi)\big] \\ &\times\ P\big[\mathbb{O}\big(\mathcal{T}_{[\psi,0]}\big) < l|\phi(\alpha(\psi,0))\big] \ldots P\big[\mathbb{O}\big(\mathcal{T}_{[\psi,m-1]}\big) < l|\phi(\alpha(\psi,m-1))\big]. \end{aligned} \tag{8.7.16}$$

Now the second step can be understood when one notices that $\mathcal{B}_m[\theta(\psi)]$ is the event that a branch point occurs to make $([\psi],[\theta(\psi)])$ an internal branch with $m$ daughter branches, and the probability of this is just $(-D_0)^{-1}B_m$; the terms $P\big[\mathbb{O}\big(\mathcal{T}_{[\psi,k]}\big) < l|\phi(\psi,k)\big]$ are just the probability that the subtree $\mathcal{T}_{[\psi,k]}$ becomes extinct with order at most $l-1$, in other words, $\boldsymbol{q}(l-1)$, for all $k = 0, 1, \ldots, m-1$ . We therefore have that equation (8.7.16) is

$$U_m(l) = (-D_0)^{-1}B_m\Big(\boldsymbol{q}^m(l-1) \otimes I^{(1)}\Big), \tag{8.7.17}$$

where the Kronecker product with $I^{(1)}$ represents the fact that the parental branch is frozen with probability 1. Now since a $U$-unit may be constructed by any finite number of daughter branches we sum over all possibilities and therefore we have that,

$$U(l) = \sum_{m=1}^{\infty}(-D_0)^{-1}B_m\Big(\boldsymbol{q}^m(l-1) \otimes I^{(1)}\Big). \tag{8.7.18}$$

Let $\mathcal{T}^k_{[\psi]}$ be a topology that is based around node $[\psi]$ whose parental branch has undergone $k$ branch points before undergoing a catastrophic transition. The first of these internal branch points is $[\psi]$. The probability that this tree has order,

$\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1$ is given by, $P[\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)]$ and it is not hard to see that this is equivalent, to,

$$P\big[\mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] = P\big[\mathcal{A}[\theta^k(\psi)], \mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big], \qquad (8.7.19)$$

because at the $k+1$ parental subnode, $[\theta^k(\psi)]$, from $[\alpha(\psi)]$, a catastrophic transition occurs rendering the parental branch extinct. Now, recall that the order of a tree is the maximum of the orders of all the nodes along the parental pathway, therefore, we have that,

$$\begin{aligned} &P\big[\mathcal{A}[\theta^k(\psi)], \mathbb{O}(\mathcal{T}^k_{[\psi]}) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] = \\ &P\big[\mathcal{A}[\theta^k\big(\psi\big)], \max_{i=0,\ldots,k-1} \mathbb{O}_n(\theta^i(\psi)) < l \big| \phi\big(\alpha(\psi)\big)\big]. \end{aligned}$$

However, the above expression is equivalent to each of the individual nodes having order less than l, so

$$\begin{aligned} &P\big[\mathcal{A}[\theta^k(\psi)], \max_{i=0,\ldots,k-1} \mathbb{O}_n\big(\theta^i\big(\psi\big)\big) < l \big| \phi\big(\alpha(\psi)\big)\big] = \\ &P\big[\mathcal{A}[\theta^k(\psi)],\ \mathbb{O}_n\big(\psi\big) < l,\ \mathbb{O}_n\big(\theta\big(\psi\big)\big) < l, \ldots,\ \mathbb{O}_n\big(\theta^{k-1}\big(\psi\big)\big) < l \big| \phi\big(\alpha(\psi)\big)\big] \end{aligned}$$

But since each of the nodes and their subtrees evolve independently, we can instead write,

$$\begin{aligned} &P\big[\mathcal{A}[\theta^k(\psi)], \max_{i=0,\ldots,k-1} \mathbb{O}(\theta^i(\psi)) < l \big| \phi(\alpha(\psi))\big] = \\ &\sum_{\phi_p(\psi)} \cdots \sum_{\phi_p(\theta^{k-1}(\psi))} P\big[\mathcal{B}_{\sigma(\psi)-1}[\psi],\ \mathbb{O}_n\big(\psi\big) < l,\ \Phi(\psi)\ \ \&\ \phi_p\big(\psi\big) \big| \phi\big(\alpha(\psi)\big)\big] \\ &\times P\big[\mathcal{B}_{\sigma(\theta(\psi))-1}\big[\theta\big(\psi\big)\big],\ \mathbb{O}_n\big(\theta(\psi)\big) < l,\ \Phi(\theta(\psi))\ \&\ \phi_p\big(\theta(\psi)\big) \big| \phi_p\big(\psi\big)\big] \ldots \\ &\times P\big[\mathcal{B}_{\sigma(\theta^{k-1}(\psi))-1}[\theta^{k-1}(\psi)],\ \mathbb{O}_n\big(\theta^{k-1}(\psi)\big) < l,\ \Phi(\theta^{k-1}(\psi))\ \& \\ &\qquad\qquad \phi_p\big(\theta^{k-1}(\psi)\big) \big| \phi_p\big(\theta^{k-2}(\psi)\big)\big] \\ &\times P\big[\mathcal{A}[\theta^k(\psi)] \big| \phi_p\big(\theta^{k-1}(\psi)\big)\big], \end{aligned} \qquad (8.7.20)$$

where $\Phi(\psi)$ is just shorthand for $\phi(\psi,0),\ldots,\phi(\psi,\sigma(\psi)-1)$, the phases of each daughter branch emanating from an internal node $[\psi]$. The first $k$ terms of equation (8.7.20) are each just the definition of $U(l)$, and the last term is just equal to

$(-D_0)^{-1}\boldsymbol{d}$, because after the $k$-th branch point of the parental branch, a catastrophe must occur. Hence we can write,

$$P\big[\mathbb{O}(\mathcal{T}_\psi^k) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] = U^k(l)(-D_0)^{-1}\boldsymbol{d}. \tag{8.7.21}$$

However, to obtain $\boldsymbol{q}(l)$ we must sum over all the possible number of branch points of the parental branch, so,

$$\begin{aligned} \boldsymbol{q}(l) &= \sum_{k=0}^{\infty} P\big[\mathbb{O}(\mathcal{T}_\psi^k) < l+1 \big| \phi\big(\alpha(\psi)\big)\big] \\ &= \sum_{k=0}^{\infty} U^k(l)(-D_0)^{-1}\boldsymbol{d}. \end{aligned} \tag{8.7.22}$$

■

The number of topologies that have order $l \geq 1$ is infinite. We can see this because such trees can be constructed by combining $U$-units together, whose individual orders must be less than $l-1$. We can combine any number of these units together. To re-construct the topologies of order $l$ we can connect any number of $U$-units from the pool of units of orders $l-1$. At each iteration of the Order algorithm, this is exactly what we do. Thus at each step, $k$, we recombine all the topologies of order $k-1$ and below to obtain all the topologies of order $k$ and below; there are an infinite number of ways of combining the topologies of order less than k. Hence, at each step we are including infinitely more topologies. In contrast, the Depth algorithm, which is similar to the Harris algorithm in the discrete-time multi-type branching process, only adds a finite number of new topologies at each step, in fact, if there were $N(k)$ topologies at the $k$-th step, an extra $N^2(k) - N(k) + 1$ are included at the $k+1$-th step, clearly a finite number for all finite $k$. As a result, the Order algorithm is a much more efficient algorithm to determine the probability of eventual extinction of a multi-type branching process.

The ease with which algorithms can be developed for the MBT carries across to the MT. In this chapter we have developed the Depth and Order algorithms for the MT. The Order algorithm is a novel way of calculating the extinction probability

for a ctMMTBP, an algorithm which is a significant improvement on the algorithm of Harris [9]. The MT representation of the ctMMTBP should provide the basis for better algorithmic analysis of the ctMMTBP and as a result, allow it to play a more prominent role in modelling physical phenomena.

# Chapter 9

# Conclusions and Further Research

## 9.1 Conclusions

It is widely believed that the rates of evolution, at the genetic level and at the macroevolutionary level [6, 22], have not been constant throughout all time. At the macroevolutionary level, these rate changes alter the overall topologies of phylogenetic trees thus, changes in the speciation and extinction rates of species have the effect of altering the imbalance of trees as they evolve. The simple models, such as the crBD and PDA models, models that do not account for rate variation and have clearly been shown to be inadequate in generating topologies that agree with observation [11, 22, 30]. It has become increasingly evident [1, 22] that more complex models of macroevolution are required in order to aid in the inference of phylogenetic tree topology. One step towards developing more sophisticated models is to allow for rate variation. Consequently, an excellent candidate for a macroevolutionary model is the ctMMTBP, [1, 22, 26], but the distinct lack of adequate numerical methods has hampered the use of the ctMMTBP as a modelling device. In this thesis we have addressed:

- the need for a reasonable macroevolutionary model based on the ctMMTBP, and

- the need for some algorithmic approaches to the ctMMTBP.

We have developed a model of the macroevolutionary process based on the binary-branch point ctMMTBP, which we have called the Markovian binary tree. To transform from the ctMMTBP branching structure to the MBT transition structure, singular branch points (ctMMTBP) were interpreted as hidden transitions (MBT) and binary branch points (ctMMTBP) were interpreted as being observable transitions. Observable transitions were regarded as corresponding to nodes in the phylogenetic trees, while hidden ones were not.

With this subtle change of interpretation, we were able to describe the dynamics of each branch of an MBT with a Markovian arrival process (MAP). As a result, the time until an observable event or to the extinction of a branch need not be exponentially distributed and there exist correlations between the offspring distribution and the lifetime of a branch. Using this MAP interpretation and the fact that the tree topologies are binary we represented the ctMMTBP as a level-dependent quasi-birth-death process. The states of the process were given by the number of branches and the phase that each branch was in. This representation allowed us to unambiguously keep track of the daughter and parental branches and thus reconstruct tree topologies.

In Chapter 5 we demonstrated that the MBT subsumed all of the simple macroevolutionary models such as the crBD, the PDA and the sPDA. What was of most significance however, was that we showed that the multi-rate (MR) model of Pinelis [26] was also subsumed by the MBT. The multi-rate model (MR) is also based on the binary-branch point ctMMTBP. In the MR model however, Pinelis [26] introduced the concept of a quasi-stable branch, that is, a branch that cannot ever become extinct or give birth to any daughter branches. The inclusion of this branch type and the necessary pruning of extinct branches from the tree topologies had the effect of complicating any analysis performed with the MR model. In fact, even showing that the MR subsumes the simple models of macroevolution is not a trivial task. We were able to prove in Chapter 5 however, that the MBT model did indeed subsume

the MR model and that there was no need for quasi-stable branches and the pruning of extinct branches could in fact be discarded. Thus the MBT not only subsumes the MR but is also amenable to simpler analysis.

Having demonstrated that the MBT is the most general model, we wished to test it in a macroevolutionary environment. As stated on numerous occasions during this thesis, the level of imbalance of phylogenetic trees is important because it has the potential to show how the rates of evolution have changed over time. Therefore any reasonable model of the macroevolutionary process must have the flexibility to account for these imbalances. The imbalance algorithm was developed in Chapter 6. The imbalance algorithm determined the mean imbalance of an MBT model conditional on tree size. This algorithm therefore,

1. extended the work of Heard [11] and Rogers [30] who did similar things for the crBD and PDA models, and

2. showed that the MBT is sufficiently flexible that any mean imbalance is possible conditional on tree size!

The combination of flexibility and algorithmic tractability make the MBT a formidable model in the macroevolutionary domain.

The subtle change of the interpretation of the branching structure of the ct-MMTBP to that of the MBT allowed us to develop two further algorithms in Chapter 7. These algorithms determine the probability of eventual extinction of the process in the interesting super-critical domain. The algorithms that we developed to solve for the probability of eventual extinction were called the Depth and Order algorithms. Surprisingly, both these algorithms did not require use of the level-dependent QBD representation of the MBT. This can be attributed to the independent evolution of each branch of an MBT. Further, we demonstrated that the Order algorithm was more efficient than the level-dependent logarithmic reduction algorithm. Furthermore, it was also demonstrated that the sample paths of the

Neuts algorithm at each step could be transformed to the set of the binary trees measured at the equivalent step of the Depth algorithm.

The ease with which the algorithms could be developed and implemented in the special case of the binary-branch point ctMMTBP was transferred to the general ctMMTBP. That is, we reinterpreted the transition branching structure of the general ctMMTBP in exactly the same way as we did in the MBT case: singular transitions corresponded to hidden transitions, and non-singular transitions corresponded to observable transitions. Using this interpretation we once again developed the alternative representation of the general ctMMTBP which we called the general Markovian tree (MT) in Chapter 8. It was then relatively straightforward to develop the Depth and Order algorithms for the MT in a similar fashion to the corresponding algorithms in the MBT environment.

## 9.2 Future work

The essential groundwork for the MBT as a macroevolutionary model has been laid out here. It possesses all the attributes that make it a good model for macroevolution. It is flexible, as witnessed by the fact that a one parameter four phase model was sufficient to span the entire range of mean imbalances for trees of size 5 and it is readily analyzable as witnessed by the fact that we developed a number of algorithms that were easy to implement.

What remains to be done is to apply the MBT to macroevolution. More specifically, to determine how the MBT can be fitted to the data so as to generate an actual model. Once this has been achieved, suitable tests of the models performance need to be devised and implemented. The statistically fitted MBT model can be applied to well known phylogenies to see what results are obtained. In other words, what topologies are generated by the MBT with highest likelihood? How do they differ from those of other studies? Does the MBT predict topologies that more closely represent the true topology?

Finally, representing the ctMMTBP as an MT gives to the ctMMTBP a richer transition structure. A transition structure that has correlations between branch lifetime and branch offspring distributions. It is our belief that since the MT is more amenable to algorithmic analysis, the MT representation may serve as a good starting point to develop other algorithms for the ctMMTBP. Algorithms that will enable the ctMMTBP to be used in a wide variety of modelling contexts.

# Bibliography


[1] D. J. Aldous. Stochastic Models and Descriptive Statistics for Phylogenetic Trees, from Yule to Today. *Statistical Science*, 16:23–34, 2001.

[2] K.B. Athreya and P.E. Ney. *Branching Processes.* Springer-Verlag, 1971.

[3] L. Bright and P.G. Taylor. Calculating the Equilibrium Distribution in Level-Dependent Quasi-Birth-and-Death Processes. *Communications in Statistics. Stochastic Models*, 11:497–525, 1995.

[4] D. H. Colless. Review of Phylogenetics: the Theory and Practice of Phylogenetic Systematics. *Systematic Zoology*, 31:100–104, 1982.

[5] K.S. Dorman, J.S. Sinsheimer, and K. Lange. In the Garden of Branching Processes. *SIAM Review*, 46:202–229, 2004.

[6] J.H. Gillespie. *The Causes of Molecular Evolution.* Oxford University Press, 1991.

[7] L. Gün. Experimental Results on Matrix Analytical Solution Techniques-Extensions and Comparisons. *Communications in Statistics. Stochastic Models.*, 5:669–682, 1989.

[8] E. F. Harding. The Probabilities of Rooted Tree-Shapes Generated by Random Bifurcations. *Advances in Applied Probability*, 3:44–77, 1971.

[9] T.E Harris. *The Theory of Branching Processes.* Springer-Verlag, 1963.

[10] S. B. Heard. Patterns in Tree Balance among Cladistic, Phenetic, and Randomly Generated Phylogenetic Trees. *Evolution*, 46:1818–1826, 1992.

[11] S. B. Heard. Patterns in Phylogenetic Tree Balance With Variable and Evolving Speciation Rates. *Evolution*, 50:2141–2148, 1996.

[12] S. B. Heard and A. O. Mooers. Signatures of Random and Selective Mass Extinctions in Phylogenetic Tree Balance. *Systematic Biology*, 51:889–897, 2002.

[13] D.G. Kendall. On the Generalized "Birth-and-Death" Process. *Annals of Mathematical Statistics*, 19:1–15, 1948.

[14] M. Kimmel and D.E. Axelrod. *Branching Processes in Biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer, 2002.

[15] M. Kirkpatrick and M. Slatkin. Searching for Evolutionary Patterns in the Shape of a Phylogenetic Tree. *Evolution*, 47:1171–1181, 1993.

[16] G. Latouche. Newton's Iteration for Non-Linear Equations in Markov Chains. *IMA Journal of Numerical Analysis*, 14:583–598, 1994.

[17] G. Latouche and V. Ramaswami. A Logarithmic Reduction Algorithm for Quasi-Birth-and-Death Processes. *Journal of Applied Probability*, 30:650–674, 1993.

[18] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modelling*. ASA SIAM, 1999.

[19] G. Latouche, A. Remiche, and P. G. Taylor. Transient (Markov) Arrival Processes. *Annals of Applied Probability*, 13:628–640, 2003.

[20] D.M. Lucantoni. New Results on the Single Server Queue with a Batch Markovian Arrival Process. *Communications in Statistics. Stochastic Models*, 7:1–46, 1991.

[21] C.J. Mode. *Multitype Branching Processes: Theory and Applications.* Number 34 in Modern Analytic and Computational Methods in Science and Mathematics. American Elsevier Pub. Co., 1971.

[22] A. O. Mooers and S. B. Heard. Inferring Evolutionary Processes from Phylogenetic Tree Shape. *The Quarterly Review of Biology*, 71:31–54, 1997.

[23] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models.* The John Hopkins University Press, 1981.

[24] M.F. Neuts. Moment Formulas for the Markov Renewal Branching Process. *Advances in Applied Probability*, 8:690–711, 1976.

[25] M.F. Neuts. A Versatile Markovian Point Process. *Journal of Applied Probability*, 16:764–779, 1979.

[26] I. Pinelis. Evolutionary models of phylogenetic trees. *Proceedings of the Royal Society Statistical Society of London*, 203:1425–1431, 2003. Appendix A available online at `http://journals.royalsoc.ac.uk/app/home/contribution.asp?wasp=a8dec894974947a5b93cf7579dc72f2e&referrer-parent&backto=issue,15,15;journal,60,206;homemainpublications,6,7;`.

[27] V. Ramaswami. Non-Linear Matrix Equations in Applied Probability-Solution Techniques and Open Problems. *SIAM Review*, 30:256–263, 1988.

[28] V. Ramaswami and P. G. Taylor. Some Properties of the Rate Operators in Level Dependent Quasi-Birth-and-Death Processes with a Countable Number of Phases. *Communications in Statistics. Stochastic Models*, 12:143–164, 1996.

[29] J. Riordan. *An Introduction to Combinatorial Analysis.* John Wiley and Sons, 1958.

[30] J. S. Rogers. Central Moments and Probability Distributions of Colless's Coefficient of Tree Imbalance. *Evolution*, 48:2026–2036, 1994.

[31] E. Seneta. *Non-negative Matrices and Markov Chains.* Springer-Verlag, 1981.

[32] B.A. Sevastyanov. The Theory of Branching Random Processes. *Uspehi Matematicheskikh Nauk*, 6:47–99, 1951.

[33] J. B. Slowinski. Probabilities of $n$-Trees Under Two Models: A Demonstration that Asymmetrical Interior Nodes are not Improbable. *Systematic Zoology*, 39:89–94, 1990.

[34] M. Steel and A. McKenzie. Properties of Phylogenetic Trees Generated by Yule-type Speciation Models. *Mathematical Biosciences*, 170:91–112, 2001.