Stochastic Models of Molecular Evolution: An Algebraical and Statistical Analysis

DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
des Fachbereichs Mathematik
der Ludwig-Maximilians-Universität München

vorgelegt von

Steffen Kläre

12. April 2005

1. Berichterstatter: Prof. Dr. Volkmar Liebscher
2. Berichterstatter: Prof. Dr. Arndt von Haeseler

Tag der mündlichen Prüfung: 08.07.2005

To Zakiya and Zillah.
May they live long and prosper.

Preface

First and foremost, I'd like to thank my supervisor Prof. Dr. Volkmar Liebscher. Volkmar has always been friendly and encouraging, maintaining the delicate balance between support and challenge. He is an inspiring mathematician and most importantly, an nice guy too.

This thesis was completed at the Institute of Biomathematics and Biometry of Prof. Dr. Rupert Lasser with the financial assistance of the GSF Research Center of Environment and Health in Neuherberg (January 2001-March 2004), a Marie Curie grant of the European Union (April 2003-July 2003), the Technical University Garching (April 2004-September 2004), the SFB 386 at the Ludwig-Maximilian-University Munich (October 2004-March 2005) and my parents.

There are quite a lot of people who have my gratitude for their influence on my life during the four years in Munich: all colleagues at IBB, especially Marie, who had the doubtful pleasure of sharing an office with me for most of the time; Prof. Dr. Gerhard Winkler who always had an interesting story to tell, and who invited me to the best wheat beer in the world; my flatmate and brother Michael; all the people from the board game group; and especially my friends from TSV Haar.

Special thanks to Maggie, whose help with the English language is invaluable to me, and whose hospitality during several occasions provided necessary time for relaxation and reflection.

Finally I want to thank my parents for supporting me in any possibly way, for being the best parents one could ask for, and for the ”tons” of Thüringer Rostbratwurst, they ”smuggled” into Bavarian territory to equip the Institute with something edible for some of its infamous barbecues.

Zusammenfassung

Die vorliegende Dissertationsschrift befasst sich mit den algebraischen Eigenschaften eines stochastischen Modells molekularer Evolution. Das betrachtete Modell ist Grundlage diverser phylogenetische Rekonstruktionsmethoden, die sich in erster Linie mit der Bestimmung von Verwandtschaftsverhältnissen zu einer gegebenen Menge von Angehörigen heute lebender Arten mittels den für diese Angehörigen ermittelten DNS-Sequenzen beschäftigen. Meist werden diese Verwandtschaftsverhältnisse durch sogenannte Abstammungsbäume dargestellt, obwohl in neuerer Zeit auch Netzwerke an Bedeutung gewonnen haben.

Aus stochastischer Sicht wird die Fragestellung wie folgt interpretiert: Gesucht ist ein Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ mit Knotenmenge $\mathcal{V}$ und Kantenmenge $\mathcal{E}$ und ein mit ihm assozierter stochastischer Prozess $\mathbf{X} : \mathcal{V} \to \mathcal{S}$, der jedem Knoten $\alpha \in \mathcal{V}$ einen Zustand x_α aus der genetischen Zustandsmenge $\mathcal{S}$ zuordnet, dessen Verteilung am besten zu den beobachteten DNS-Sequenzen an den Endknoten oder Blättern von $\mathcal{G}$ passt.

In der Regel werden folgende Anforderungen an die Modell-Parameter $\mathcal{G}$ und $\mathbf{X}$ gestellt: Die graphische Struktur wird als gewurzelter Baum $\mathcal{G}_\varrho = (V, E; \varrho)$ mit Wurzel ϱ angesehen. Der Prozess $\mathbf{X}$ genügt der *Markov-Eigenschaft*, d.h. er ist in einem Knoten α bedingt auf seinen direkten Vorfahren $\mathrm{pa}(\alpha)$ unabhängig von seinen Nichtnachkommen. Diese Bedingung ist gemäß Lauritzen [1996] äquivalent zur folgenden charakterisierenden Gleichung:

$$\mathbb{P}\Big(\bigcap_{\alpha\in V}\{X_\alpha = x_\alpha\}\Big) = \mathbb{P}(X_\varrho = x_\varrho)\prod_{(\gamma,\beta)\in E}\mathbb{P}(X_\beta = x_\beta \,|\, X_\gamma = x_\gamma). \tag{F}$$

Mit (F) läßt sich die obenstehende Aufgabenstellung wie folgt formulieren: Finde eine gewurzelte Baumstruktur $\mathcal{T}_\varrho$ mit Blattmenge $\mathcal{L} = \{\beta_1, \ldots, \beta_n\}$ und einen assozierten Markov-Prozess $\mathbf{X}$, so daß die durch

$$\mathbb{P}(X^{\mathcal{L}} = \underline{x}) = \sum_{\substack{\underline{y}\,\in\,\mathcal{S}^{n+m}\\ \underline{y}|_{\mathcal{L}}\,=\,\underline{x}}} \mathbb{P}(X_\varrho = y_\varrho)\prod_{(\alpha_1,\alpha_2)\in E}\mathbb{P}(X_{\alpha_2} = y_{\alpha_2} \,|\, X_{\alpha_1} = y_{\alpha_1}) \tag{\dagger}$$

gegebene Verteilung die in den Blättern erhobenen Daten am besten beschreibt. Die Kompliziertheit des betrachteten Problems hängt von der Anzahl der Blätter und

der Größe des Alphabets ab. Zum Beispiel hat das polynomielle System (†) für 10 Blätter und 20 Zustände genau 20^{10} Gleichungen. Aus rechentechnischen und Interpretationsgründen ist es daher von Vorteil, diese Zahlen passend zu veringern.

Die Zahl möglicher Zustände wird mittels der Annahme unabhängiger, identisch verteilter Sequenzpositionen reduziert. In diesem Fall stellt jede Position eine Stichprobe des gesuchten Prozesses auf den Blättern dar. Die Annahme wird nur ungern gemacht, da sie die Anwendbarkeit der entwickelten Methoden stark einschränkt.

Die Zahl möglicher Bäume zu einer gegebenen Menge wird durch die sogenannte *supertree*-Theorie reduziert. In dieser Theorie wird die Blattmenge $\mathcal{L}$ in sich überlappende Teilmengen aufgeteilt. Für die jeweilige Teilmenge wird die reduzierte Baumstruktur zu (†) generiert und anschließend wird aus den gewonnenen, reduzierten Baumstrukturen ein sogenannter Konsensbaum zusammengestellt.

In Chang [1996] wurde gezeigt, daß unter der Annahme, daß ein Markov-Prozess existiert, dieser durch die Dreierbäume, also Bäume mit genau drei Blättern rekonstruiert werden kann. Dies ist besonders vorteilhaft, weil es genau einen Baum mit drei Blättern gibt.

Diese Arbeit greift diesen Ansatz auf mit dem Ziel, Bedingungen an Blattverteilungen zu bestimmen, unter denen ein Markov-Prozess auf einem Dreierbaum existiert. Diese Bedingungen sind Polynome, deren gemeinsame Nullstellen eine algebraische Lösung für (†) besitzen. Der Ansatz wird auf drei Modell-Spezifikationen angewendet: das allgemeine Zwei-Zustands-Modell, das Neyman N_k Modell und das Kimura 2ST Modell. Für alle drei Modelle werden Polynome ermittelt, die obiger Beschreibung genügen. Außerdem wird für alle Modelle die algebraische Lösung von (†) explizit ausgerechnet und Bedingungen angegeben, unter denen die berechnete Lösung einen Markov-Prozess charakterisiert. Dabei sind die bestimmten Lösungen nur eindeutig bis auf Permutationen aufgrund der dem Gleichungssystem immanenten Symmetrien.

Zusätzlich werden für das Zwei-Zustands-Modell notwendige Bedingungen für die Existenz eines Markov-Prozesses auf einem Baum mit vier Blättern ermittelt. Diese Bedingungen werden aus den Lösungen für die Dreiersubbäume ermittelt. Die gefundenen Eigenschaften lassen vermuten, daß aus Markov-Prozessen auf Dreiersubbäumen keine hinreichenden Bedingungen für die Existenz eines passenden Markov-Prozesses auf dem Superbaum gewonnen werden können. Für das Neyman N_k und das Kimura 2ST Modell werden die Ergebnisse auf die zeitstetige Spezialisierung der Markov-Prozesse, das sogenannte Ratenmodell, übertragen.

Um letztendlich den Bogen zur Baumrekonstruktion zu schließen, wird ein Algorithmus präsentiert, der aus den Dreier-Blattverteilungen für eine gemeinsame Verteilung für n Sequenzen einen Baum mit n Blättern und einen Markov Prozess darauf generiert. Dieser Algorithmus erlaubt es, die Rekonstruktion aus Subbäumen genauer zu betrachten, auch wenn er nicht zeiteffizient ist.

Contents

List of Figures

List of Tables

Introduction

The theory of molecular evolution investigates how genes and genomes evolve. The subcategory of molecular phylogenetics develops methods for inferring evolutionary relationships among organisms, genes and proteins. Generally, this inference is done by deriving a tree from available sequence data. Usually, these data are n aligned DNA sequences of length N, where each sequence represents a species. Several methods were introduced which presented tree structures based on these data. The Maximum Likelihood approach first suggested by Felsenstein [1981] uses a class of Markov models of molecular evolution to derive a tree and a characterization of a stochastic process on the tree. This model class is the center point of this thesis.

A Markov model of the development of a particular species assumes that the future evolution of the species is independent of its *history* given the sequence data of its immediate predecessor. For example, given the properties of the forebears of the brontosaurs the development of the brontosaurs is considered independent of all other species that lived before and during the lifetime of brontosaurs. This property is also known as the *ordered directed Markov property.* If the evolutionary processes were governed by reproduction only, this assumption is quite reasonable.

A Markov process on a rooted tree is characterized by a joint distribution in the vertices over an appropriate state space which obeys the *factorization property.* The factorization property states that a joint distribution decomposes into a product of edge-related functions, i.e. for each edge a function is declared that depends on the states in the incident vertices only.

For a parametrization of these Markov processes on rooted trees the element of choice is a "transition kernel". Since transition kernels implicitly contain a direction their application to undirected trees appears somewhat pointless. However, an important observation is the relative irrelevance of the overall direction of the tree, i.e. a consistent change of the direction of the edges does not influence the characterization of the Markov process. Another consequence of this observation is the non-identifiability of a root from the data. Most of these statements are direct conclusions from results of Lauritzen [1996].

Usually, the chosen property on undirected trees is the local Markov property (e.g. Chang [1996]) where a vertex is conditionally independent of the remaining vertices given its immediate neighbor vertices. On rooted trees the ordered directed Markov property is preferred (e.g. Semple and Steel [2003]), since it appears weaker but more intuitive than the local Markov property.

An often, if reluctantly, made simplification of the model is the assumption that molecular evolution is only governed by homogeneous point mutations, i.e. changes across sites are regarded independent and identically distributed. Accordingly, the state space of the Markov process can be reduced from sequences to genetical alphabets, lowering the number of possible states considerably. Also, the simplification permits to consider aligned sequences of length N as a sample of size N. From the biological point of view such an assumption neglects certain established properties of molecular evolution. For instance, the *wobble effect* (e.g. Yap and Speed [2005]) proposes, that amino acid encoding in DNA yields different mutation rates for each of the three encoding sites. This contradicts the homogeneity of point mutations. Hence, the results of this work should only be considered in connection with DNA sequences from non-coding regions. Moreover, since the simplification does not allow recombination, it should only be applied to DNA sequences which are not subject to recombination, insertion or deletion. The best known examples are mitochondrial DNA and the DNA of the human Y chromosome.

The main objective of phylogenetic analysis is the identification of a tree and of the mutation process on the tree from the knowledge of the process at its leaves, i.e. the knowledge of the sequence data from present day species. In terms of the factorization property this is similar to solving the polynomial equation system which is derived from equating the observed leaf distribution to the theoretical leaf distribution expressed in terms of the transition kernels. This thesis examines the potential of this objective on three model specifications: the general two state model, the Neyman N_k model and the Kimura 2ST model.

The sample set obtained through the simplification gives the observed joint leaf distribution. For the inference of the tree the joint distributions suffice (e.g. Baake [1998]). However an inference of the Markov process demands more information than pairs of leaves can provide. For this inference joint distributions of triples of leaves are needed, as was proved by Chang [1996]. But joint distributions of triples of leaves only allow to reconstruct a Markov process if the full joint distribution on the leaves was subject to a Markov process. Yet in order to observe a Markov process on a tree with $n > 3$ leaves its restriction to all triple trees must necessarily be a Markov process. Thus it is useful to analyze the models at triple trees and then consider an extension to supertrees. Here, an extension from triple trees to quartet trees is attempted for the general two state model.

One indicator whether a joint leaf distribution comes from a Markov process are *phylogenetic invariants*. These are polynomials in the leaf distribution whose joint roots provide a solution for the polynomial equation system. These polynomials are the subject of various papers (see e.g. Hagedorn and Landweber [2000] or Allman and Rhodes [2003]), and are seen as important tools to understand Markov processes on trees. If the polynomial equation system has for a given leaf distribution a solution, then this is an algebraic solution, i.e. each variable has values in $\mathbb{C}$. In addition, if this algebraic solution is composed of transition kernels, it is stochastically admissible, thus characterizing a Markov process on a tree for the regarded joint leaf distribution.

For the general two state model on a triple tree only one phylogenetic invariant was observed, and every leaf distribution is a root of this invariant because it demands that the elements of input vector sum to one. Therefore, almost all joint distributions of three leaves over a two state alphabet provide an algebraic solution for the equation system. Except for the case of uncorrelated leaves and one unknown case, the number of possible solutions is finite, in particular it is exactly two. Generally, this work uses results from algebraic geometry to show that vectors with an infinite set of solutions are a zero set in the set of all solutions. Lazarfeld [1966] and Pearl and Tarsi [1986] also considered the same model in different contexts.

For the Neyman N_k model on a triple tree two phylogenetic invariants were observed, one of which again demands an element sum of one. For every input two solutions were observed. These solutions are functions of the pairwise leaf distributions which is consistent with the general perception that symmetric processes can be derived from pairwise leaf distributions (e.g. Baake [1998]). For every pairwise leaf distribution numerous triple leaf distributions exist. The two observed solutions are subject to two distinct triple leaf distributions, and only if the input leaf distribution is a root of the obtained invariants, it is one of them.

For the Kimura 2ST model on a triple tree 18 phylogenetic invariants were computed using the software package **Singular** (Greuel et al. [2001]), again including the summation polynomial. Analyzing the associated equation system returned four solutions, which again are functions of the pairwise leaf distributions. These solutions return two triple leaf distributions, and only if the observed leaf distribution is a root of the phylogenetic invariants, it is one of them.

Obviously, an algebraic solution does not characterize a Markov process. Hence, additional conditions are needed. Solutions of the two state model characterized a Markov process if the sign of the conditional correlation of two leaves given the third does not reverse the sign of the unconditional correlation of the two leaves, which is not zero since the leaves are not independent. For the Neyman N_k and the Kimura 2ST model the conditions of stochastic admissibility are concerned with the similarity of pairwise aligned sequences. The condition is regularly verified in real data.

The results for triple trees under the two state models were extended to quartet trees. This approach yielded some phylogenetic invariants, but also led to the conjecture, that the existence of a Markov process on a quartet tree cannot be guaranteed by just analyzing the compatibility of the parameters on the associated triple trees.

This conjecture is also strengthened by the phylogenetic invariants obtained for the Neyman N_k and the Kimura 2ST model on triple trees, and the Neyman N_2 model on quartet trees. The parameters for these models are derived from pairwise distributions but the phylogenetic invariants depend on the triple and quartet leaf distributions, respectively.

One popular model specification is the rate model, where evolution is assumed to be governed by a constance rate of change. Usually, the rate model is called time-continuous, whereas the models solely defined by transition kernels at the edges are called time-discrete. Both model classes are connected by declaring transition kernels for a particular edge under the time-continuous model as the matrix exponential of the rate matrix times the edge length. This connection is used to transfer the results related to the Neyman N_k and the Kimura 2ST model to the rate model. Unfortunately, the models did not provide sufficient equations to infer a set of rates as well as a set of edge lengths, but only a mixture of both. The positivity conditions directly linked to the rate model rejected all but one of the obtained solutions from the discrete model class. In other words, time-continuity destroyed the multiplicity of solutions of the model.

A couple of statistic tools to approximate a stochastically inadmissible solution by an admissible one are presented at the end of the thesis. The observations from the three considered models showed that even though algebraic solutions can almost always be obtained from a set of input sequences, finding an acceptable and stochastically admissible approximation is difficult. One estimator tackles the problem by manipulating the solutions, whereas the other estimator manipulates the observed leaf distribution. Simulations indicated that leaf distributions which assign at least one state with probability zero, provide an inadmissible solutions. Since input data almost ever yield leaf distributions with this property it is reasonable to try to erase this obstacle. Further simulations showed that positivity of the leaf distribution is still not sufficient to guarantee an admissible solution.

To measure the quality of such approximations, some kinds of confidence regions are presented. Most types are based on the information from Brown et al. [2001].

Finally, to relate the results from the triple tree discussions to the task of molecular phylogenetics, an algorithm is presented. This algorithm computes for an input set of aligned sequences for n leaves a tree and a characterization of a Markov process on the tree from all associated triples. It is not time-efficient, but permits to further investigate the reconstruction of trees from its triple trees.

Chapter 1 introduces the notion of trees in the graphical sense, names its elements, and provides some useful properties. The basic terminology is based on Lauritzen [1996], Lauritzen [2001] and Semple and Steel [2003]. Further, the notion of Markov processes on trees is introduced, and some of their useful properties are provided. The chapter includes a short overview of the considered models. It closes with the derivation of the polynomial equation system (LF), and presents some general properties.

Chapter 2 provides some notions from algebraic geometry. These notions include the so-called *elimination ideal*. A basis of this ideal contains all phylogenetic invariants needed to describe the space of leaf distributions with an algebraic solution of (LF). Most results are based on Shafarevich [1974] and Cox et al. [1997]. Also, using the Morse-Sard Theorem it is proved that vectors with an infinite number of solutions for (LF) form a zero set in the set of all vectors with a solution for (LF).

Chapter 3 examines the implications of the factorization property for a triple leaf distribution under the general two state model. The scenario was already discussed in Lazarfeld [1966] and Pearl and Tarsi [1986]. This thesis extends their results, most notably by considering the extension of the results to quartet trees.

Chapter 4 analyzes the implications of the factorization property for a triple leaf distribution under the two simplest symmetric models, the Neyman N_k and the Kimura 2ST model. The chapter presents phylogenetic invariants, explicit forms for the algebraic solutions, and conditions under which a solution characterizes a Markov process. In addition, the results are extended to the rate approach.

Chapter 5 introduces some statistical tools for estimating and evaluating an obtained Markov process characterization. These tools include two estimators, several simultaneous confidence regions and an algorithm to analyze the reconstruction of supertrees from their associated triple trees.

All chapters end with a separate section that contains the proofs to all results presented in the chapter. This was done for readability and consistency reasons.

Chapter 1

Basic Definitions and Properties

A goal of molecular phylogenetics is the visualization of relationships in a given set of species or individuals within a species. Usually, this is done by a graph, preferably a tree, where the leaves depict the considered species, and the inner structure illustrates the relationship between them. The introduction of the basic terminology and properties of graphs and trees is the subject of Section 1.1.

With the introduction of Maximum Likelihood methods for phylogenetic reconstruction in Felsenstein [1981], the theory of Markov processes on graphs got a boost. In addition, recent research tries to synthesize ancient proteins using insights obtained from characterizations of similar processes (e.g. Brooks et al. [2004]). The properties of Markov processes on trees are the subject of Section 1.2.

Section 1.3 relates the model to the task of phylogenetic reconstruction, and introduces some additional assumptions on the Markov model, which are commonly used though not very popular. Also, the model classes examined in Chapters 3 and 4 are presented. In particular, the general two state model, the Neyman N_k model and the Kimura 2ST model are regarded. These models were chosen because they are the simplest available but still complex enough to provide insights into the properties of the general Markov model.

The actual task of this work is the derivation of conditions on leaf distributions under which a Markov process exists on the underlying tree structure. Section 1.4 presents the mathematical formulation of the task, and some immediate consequences. One important observation is the irrelevance of the position of a root for the characterizing Markov distribution.

Section 1.5 contains all proofs for results presented in this chapter. For the reader's convenience this structure is retained throughout the work.

1.1 Graphs and Trees

This section will present the terms and properties of graphs and trees needed for this thesis. The notation is based on Lauritzen [1996, sect. 2.1] and its revised version Lauritzen [2001].

1.1.1 General Definitions

Here graphs and the parts relevant to introduce trees are defined. Later chapters will solely work on trees, and hence no additional terminology is needed.

Definition 1.1.1. *A* graph *$\mathcal{G}$ is a tuple $(\mathcal{V}, \mathcal{E})$ consisting of a finite* vertex *set $\mathcal{V}$ and a set $\mathcal{E}$ of* edges *connecting pairs of vertices in $\mathcal{V}$.*

If an edge $e \in \mathcal{E}$ connects two vertices $\alpha, \beta \in \mathcal{V}$, then α and β are called *adjacent*, and e is called *incident* to α and β, respectively. Edges can occur in two possible types, *directed* and *undirected.* If an edge e between α and β is undirected, α and β are called *neighbors* and the edge is denoted by $e := \langle\!\langle \alpha, \beta \rangle\!\rangle$. The neighbors of a vertex $\alpha \in \mathcal{V}$ are denoted by $\mathrm{ne}(\alpha)$. If e is directed from α to β, then α is called *parent* of β, β is called *child* of α, and the edge is denoted by $e := (\alpha, \beta)$. The children and parents for a vertex $\alpha \in \mathcal{V}$ are denoted by $\mathrm{ch}(\alpha)$ and $\mathrm{pa}(\alpha)$, respectively.

Vertices are usually classified according to their *degree.* A vertex $\alpha \in \mathcal{V}$ has a degree of $\deg(\alpha) = n \geq 0$ if the number of edges incident to α is exactly n. Using this definition three classes are presented: A vertex $\iota \in \mathcal{V}$ is called *isolated* if no edge is incident to ι, i.e. $\deg(\iota) = 0$. A vertex $\beta \in \mathcal{V}$ is called *terminal vertex*, *end point* or *leaf* if exactly one edge is incident to β, i.e. $\deg(\beta) = 1$. All other vertices are called *inner vertices*, i.e. at least two edges are incident to an inner vertex. The sets of all isolated and inner vertices, and leaves to a given graph $\mathcal{G}$ are denoted by $\mathcal{I}(\mathcal{G})$, $\mathcal{N}(\mathcal{G})$ and $\mathcal{L}(\mathcal{G})$, respectively. Usually, the graph is fixed and therefore, the sets are abbreviated by $\mathcal{I}$, $\mathcal{N}$ and $\mathcal{L}$ instead.

In terms of molecular phylogeny, a vertex denotes a species and edges describe relations between them. In particular, leaves represent recent species, inner vertices represent ancestral species, and isolated vertices indicate alien species and are usually not incorporated.

A basic feature of the notion of a graph is that it is a visual object. It is conveniently represented by a picture, where a *dot* is used for a vertex. Further, a *line* joining α and β represents the undirected edge $\langle\!\langle \alpha, \beta \rangle\!\rangle$, whereas an *arrow* from α pointing towards β is used for the directed edge $(\alpha, \beta) \in \mathcal{E}$. As an example consider Figure 1.1.

The direction of edges transfers to graphs in the following way: If all edges of a graph $\mathcal{G}$ are undirected, the graph is called *undirected.* Similarly, if all edges are directed, the graph is called *directed.* By replacing all directed edges $(\alpha, \beta) \in \mathcal{E}$ in a directed graph $\mathcal{G}$ by their undirected counterparts $(\!(\alpha, \beta)\!)$ one obtains its *undirected version* $\mathcal{G}^u$.

Next, a certain group of vertices is regarded. A *path* is an ordered set of distinct vertices $(\alpha_1, \ldots, \alpha_n)$ where α_i and α_{i+1}, $i = 1, \ldots, n-1$ are adjacent. Therefore, a path is also defined through a set of edges $(e_1, \ldots, e_{n-1})$, where e_i is the incident edge to α_i and α_{i+1}, $i = 1, \ldots, n-1$. If for a path all edges $e_1, \ldots, e_{n-1}$ are undirected, the path is *undirected.* If all edges are directed and point in the same direction, the path is called *directed.* The length of a path is given by the number of edges it runs through. For instance, the path $(\alpha_1, \ldots, \alpha_n)$ has length $n-1$. As an example, consider the paths in Figure 1.1. Here, $(\alpha_1, \ldots, \alpha_4)$ and $(\beta_1, \ldots, \beta_4)$ are undirected paths, whereas $(\gamma_1, \gamma_2, \gamma_3)$ is a directed path. When changing the direction of edge (γ_2, γ_3) the direction of the path is lost.

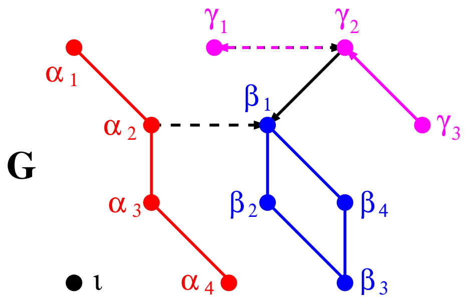

Figure 1.1: A graph and its components. The vertex ι is an isolated point, the vertex α_2 has neighbors α_1 and α_2 and a child in β_1. β_1 has two parents in α_2 and γ_2.

For some properties defined later, *separating* vertex sets are of interest. For vertex sets $A, B \subset \mathcal{V}$ with $A \cap B = \emptyset$, the set $S \subset \mathcal{V}$ is said to *separate* A from B if all paths between vertices $\alpha \in A$ and $\beta \in B$ lead through S, i.e. for every path $(\alpha_1, \ldots, \alpha_n)$ exist indices $k < l$ such that $\alpha_i \in A$ for $i < k$, $\alpha_j \in B$ for $j > l$ and $\alpha_k, \ldots, \alpha_l \in S$. For instance, the vertex δ in Figure 1.1 separates the set $\{\gamma_1, \gamma_2, \gamma_3\}$ from the remaining vertices of the graph.

The aim is to define trees in the generally accepted version. To do this, a few additional definitions are needed. Two vertices β_1 and β_2 are called *connected* if a path $(\alpha_1, \ldots, \alpha_n)$ exists with $\alpha_1 = \beta_1$ and $\alpha_n = \beta_2$. If every pair of vertices is connected, the graph is called *connected.* A path $(\alpha_1, \ldots, \alpha_n)$ is called *cycle* if $\alpha_1 = \alpha_n$ and $n > 2$. Figure 1.1 is connected if the vertex ι is deleted. The path $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_1)$ describes a cycle.

With these notions the definition of trees can be provided:

Definition 1.1.2. *A* tree $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ *is a connected, cycle-free graph.*

1.1.2 Properties of Trees

With this short introduction to graphs the underlying structure of the main objects of this thesis are defined: trees. This subsection will present some of the most important properties for the thesis.

Lemma 1.1.1. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote a tree. Then, the following statements are equivalent:*

1. *Between any two vertices β_1 and β_2 in $\mathcal{V}$ exists a unique path in $\mathcal{T}$.*
2. *If $\mathcal{E} \neq \emptyset$, then $\sharp(\mathcal{L}) \geq 2$.*
3. *If $\sharp(\mathcal{V}) = n$ then $\sharp(\mathcal{E}) = n - 1$.*
4. *Assume that for all $\alpha \in \mathcal{N}$ the degree is at least three. Then, $\sharp(\mathcal{L}) \geq \sharp(\mathcal{N})$.*

These are well-known properties. The unique path between vertices β_1 and β_2 is denoted by $\mathrm{p}(\beta_1, \beta_2)$. As proposed earlier, the length of the path is equal to the number of edges it runs through. Accordingly, the length of a path on a tree will be denoted by $\sharp(\mathrm{p}(\alpha, \beta))$.

Inner vertices of degree two will be called *non-furcating vertices*, because the edges incident to such a vertex can be joined uniquely without destroying the connectivity of the tree, i.e. such a vertex can be deleted without destroying the structure of the tree. Similarly, inner vertices of degree three will be called *furcating vertices*, because their deletion will destroy the connectivity of the graph. So far, trees are treated as undirected. For this work one class of directed trees is of interest: the rooted trees. For their introduction regard the relation $\prec_\varrho$, $\varrho \in \mathcal{V}$ defined by:

$$\alpha \prec_\varrho \beta \quad :\Leftrightarrow \quad \alpha \in \mathrm{p}(\varrho, \beta), \quad \alpha, \beta \in \mathcal{V}.$$

This relation gives a partial ordering of $\mathcal{V}$:

Lemma 1.1.2. *For every $\varrho \in \mathcal{V}$ the relation $\prec_\varrho$ is a partial ordering on $\mathcal{V}$ with minimal element ϱ.*

The partial ordering $\prec_\varrho$ induces a direction on an undirected tree $\mathcal{T}$ by choosing only those edges $(\alpha, \beta) \in \mathcal{E}$ with $\alpha \prec_\varrho \beta$. The resulting directed tree is called *rooted tree*, and is denoted by $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$. In $\mathcal{T}_\varrho$ all edges are directed away from ϱ. Note, that for every vertex $\alpha \in \mathcal{V}$ such a partial ordering can be defined.

Corollary 1.1.3. *For every undirected tree $\mathcal{T}$ exists the family of rooted trees $(\mathcal{T}_\alpha)_{\alpha \in \mathcal{V}}$.* □

Note that every rooted tree $\mathcal{T}_\varrho$ is defined on the same vertex set $\mathcal{V}$ as is $\mathcal{T}$, and the adjacency between pairs of vertices is transferred from $\mathcal{T}$ to $\mathcal{T}_\varrho$. A very useful property concerns the number of possible parents a vertex can have in a rooted tree.

Corollary 1.1.4. *Let $\mathcal{T}_\varrho$ denote a rooted tree. Then every vertex $\alpha \in \mathcal{V} \setminus \{\varrho\}$ has exactly one parent* $\mathrm{pa}(\alpha)$. *The root has no parents, i.e. it is orphaned.*

On a rooted tree $\mathcal{T}_\varrho$ and a vertex $\alpha \in \mathcal{V}$ the following vertex sets are defined:

$$\begin{aligned}
\mathrm{de}(\alpha) &:= \{\beta \in \mathcal{V} \setminus \{\alpha\} : \alpha \in \mathrm{p}(\varrho, \beta)\}, \\
\mathrm{nd}(\alpha) &:= \mathcal{V} \setminus (\mathrm{de}(\alpha) \cup \{\alpha\}), \\
\mathrm{an}(\alpha) &:= \{\beta \in \mathcal{V} \setminus \{\alpha\} : \beta \in \mathrm{p}(\varrho, \alpha)\}, \\
\mathrm{hi}(\alpha) &:= \{\beta \in \mathcal{V} \setminus \{\alpha\} : \sharp(\mathrm{p}(\varrho, \beta)) \leq \sharp(\mathrm{p}(\varrho, \alpha)).
\end{aligned}$$

The set $\mathrm{de}(\alpha)$ contains all vertices which are separated by α from the root ϱ, and the elements of $\mathrm{de}(\alpha)$ are called *descendants* of α. Accordingly, $\mathrm{nd}(\alpha)$ contains the *non-descendants* of α. The set $\mathrm{an}(\alpha)$ contains all vertices which lie on the path between the root ϱ and α. The elements of $\mathrm{an}(\alpha)$ will be called *ancestors* of α. Finally, $\mathrm{hi}(\alpha)$ contains all vertices whose path has at most the same length as the path between ϱ and α. The set $\mathrm{hi}(\alpha)$ will be called the *history* of α. In evolutionary terms, the history of a fixed species, represented by α, contains all species that had influence on the development of α, either through ancestry or environment. Obviously, one has $\mathrm{an}(\alpha) \subset \mathrm{hi}(\alpha) \subset \mathrm{nd}(\alpha)$.

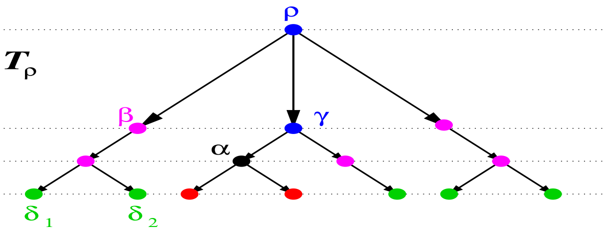

Figure 1.2: Vertex sets for a given vertex α on a rooted tree $\mathcal{T}_\rho$. The vertices colored red describe the descendants, the blue colored vertices are the ancestors, the magenta colored the remaining vertices of the history of α and the green colored vertices the remaining non-descendants of of α. γ is the unique parent of α.

The tree $\mathcal{T}_\rho$ presented in Figure 1.2 also visualizes a weakness of the definition of the ancestral set. If non-furcating vertices like β are deleted from the structure, the leaves δ_1 and δ_2 belong to the history $\mathrm{hi}(\alpha)$. Since the history should contain all species that lived before and with the species represented by α, this is an unwanted effect. However, at the moment nothing can be done about it.

Sometimes, literature distinguishes rooted trees with a root of degree two and of higher degree. E.g., Huelsenbeck and Bollback [2001] called a rooted tree $\mathcal{T}_\varrho$ with $\deg(\varrho) = 2$ *rooted*, and with $\deg(\varrho) \geq 3$ *unrooted*. However, this work won't apply this distinction and calls any tree directed by a partial ordering $\prec_\varrho$ a rooted tree. The reasons for this decision will become clear in Section 1.2.

Another vertex of interest is the so-called *most recent common ancestor* $\mathrm{mrca}(\alpha, \beta)$ of two vertices $\alpha, \beta \in \mathcal{V}$. The most recent common ancestor is defined by the following condition:

$$\mathrm{p}(\varrho, \mathrm{mrca}(\alpha, \beta)) = \mathrm{p}(\varrho, \alpha) \cap \mathrm{p}(\varrho, \beta),$$

i.e. $\mathrm{mrca}(\alpha, \beta)$ is the vertex at which the paths from the root to α and β split.

Usually, a special class of trees is preferred, the *binary* or *bifurcating trees.*

Definition 1.1.3. *A* binary tree *is a tree* $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ *where for all* $\alpha \in \mathcal{V}$

$$\deg(\alpha) = \begin{cases} 3, & \alpha \in \mathcal{N}, \\ 1, & \alpha \in \mathcal{L}. \end{cases}$$

This structure is preferred since it is seen as improbable that more than two new species are connected with a furcating vertex. The special structure of binary trees allows a computation of their number of edges and vertices from the number of leaves.

Lemma 1.1.5.(Prop. 14.1 in Waterman [1995]) *An undirected binary tree with n leaves has exactly $n - 2$ inner vertices and $n - 3$ inner edges. There are*

$$(2n-5)!! = \prod_{k=1}^{n-2} (2k-1) = \frac{(2n-5)!}{2^{n-3}(n-3)!}$$

distinct undirected binary trees with n leaves. □

Corollary 1.1.3 states that for an undirected tree a unique family of rooted trees $(\mathcal{T}_\varrho)_{\varrho \in \mathcal{V}}$ exists with the same vertex set and the same adjacency for pairs of vertices. Hence, the number of vertices and inner edges remains the same for rooted binary trees but the possible number of rooted binary trees with n leaves is the number of undirected binary trees with n leaves times the number of vertices, which in that case is $2(n - 1)$.

For any given number of leaves no other tree can have more furcating vertices than a binary tree. The other extreme are trees with only one inner vertex. Such trees are called *star trees.* Clearly, for any number of leaves these are unique. Therefore for fixed leaf number n, the number of possible trees is increasing in the number of inner vertices. Denote by $\mathfrak{T}_k$ the set of all trees with $k = 1, \ldots, n-2$ inner vertices.

The number of all possible trees with n leaves is bounded from above by:

$$\sum_{k=0}^{n-3} \sharp(\mathfrak{T}_k) \leq \sum_{k=1}^{n-3} \frac{(2n-5)!}{2^{n-3}(n-3)!} = (n-2)\frac{(2n-5)!}{2^{n-3}(n-3)!} =: K(n).$$

Hence, the total number of trees with n leaves lies between $\sharp(\mathfrak{T}_{n-2})$ and $K(n)$. Table 1.1 provides some numerical examples for the presented boundaries:

n	binary	$K(n)$
3	1	1
4	3	6
5	15	45
6	105	420
8	10,395	62,370
10	2,027,025	16,216,200
15	7,905,853,580,625	102,776,096,548,125

Table 1.1: Number of trees depending on the number of leaves.

These numbers suggest why it is preferred to restrict reconstruction methods to binary trees. But even then, the number of possible binary trees is much to high to employ optimization methods that run over all possible trees. This problem is addressed in the next subsection.

1.1.3 Supertrees and their Restrictions

Recent methods of phylogenetic reconstruction provide genealogic trees for large numbers of leaves. As Table 1.1 suggests, a selection from the overall set of possible trees is a rather hopeless approach. However, for a small number of leaves the amount of associated trees is reasonably small. Therefore, reconstructing large trees from a set of smaller trees is a better approach. To describe this approach, several structures must be introduced.

Definition 1.1.4. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote a tree, and $A \subset \mathcal{V}$ is an arbitrary vertex set. The* subgraph *of $\mathcal{T}$* generated by *A is the graph $\mathcal{G}_A := (A, \mathcal{E}_A)$ given by the edge set*

$$\mathcal{E}_A := \{(\alpha, \beta) \in \mathcal{E} : \alpha, \beta \in A\} \cup \{(\!(\alpha, \beta)\!) \in \mathcal{E} : \alpha, \beta \in A\}.$$

This definition has an immediate consequence.

Lemma 1.1.6. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote a tree, and $\mathcal{G}_A$ be a subgraph to a vertex set A. If $\mathcal{G}_A$ is connected, it is a tree. In that case $\mathcal{G}_A$ is called* subtree *of $\mathcal{T}$.*

For a rooted tree $\mathcal{T}_\varrho$ and the previously introduced vertex sets the following subtrees are defined:

- $\mathcal{T}_{\mathrm{de}(\alpha)} := \mathcal{G}_{\{\alpha\}\cup\mathrm{de}(\alpha)}$ denotes the *descendant tree* rooted at α,
- $\mathcal{T}_{\mathrm{nd}(\alpha)} := \mathcal{G}_{\{\alpha\}\cup\mathrm{nd}(\alpha)}$ denotes the *non-descendant tree* to α,
- $\mathcal{T}_{\mathrm{hi}(\alpha)} := \mathcal{G}_{\{\alpha\}\cup\mathrm{hi}(\alpha)}$ denotes the *historical tree* of α.

Figure 1.3 shows an example for the introduced subtrees for the tree $\mathcal{T}_\rho$ first presented in Figure 1.2.

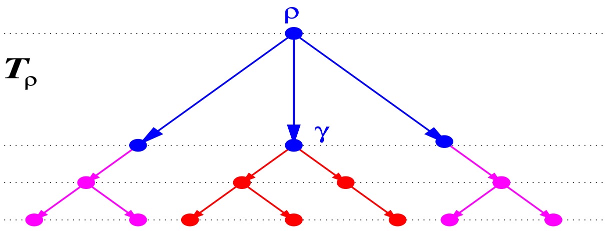

Figure 1.3: Kinds of subtrees. The red substructure together with γ describes the descendant tree of γ, the blue structure describes the historical tree of γ, and the blue structure together with the magenta colored structures presents the non-descendant tree to γ. The subgraph generated by the magenta colored structures is not connected, and hence it is no subtree.

The task of phylogenetic reconstruction is the derivation of the tree from its leaves. For this purpose another substructure of trees will be introduced, the so-called *restrictions*. For restrictions a certain vertex is of interest.

Lemma 1.1.7. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree and α, β, γ distinct vertices in $\mathcal{V}$. Then, a unique vertex $\varrho_{\alpha\beta\gamma} \in \mathcal{V}$ exists with $\mathrm{p}(\alpha,\beta)\cap\mathrm{p}(\alpha,\gamma)\cap\mathrm{p}(\beta,\gamma) = \{\varrho_{\alpha\beta\gamma}\}$. The vertex $\varrho_{\alpha\beta\gamma}$ is called the* trifurcating vertex *of $\alpha, \beta, \gamma \in \mathcal{V}$.*

With trifurcating vertices the definition of restrictions is quite straightforward. For this definition roots are of no concern. Therefore, without loss of generality restrictions will be introduced on undirected trees.

Definition 1.1.5. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree. A* restriction *to a vertex set $A \subset \mathcal{V}$ is a tree $\mathcal{T}_A$ composed of the vertex set $\mathcal{V}_A := A \cup \{\varrho_{\alpha\beta\gamma} : \alpha, \beta, \gamma \in A\}$ and the edge set:*

$$\mathcal{E}_A := \left\{(\!(\alpha,\beta)\!) : \alpha,\beta \in \mathcal{V} \text{ and } \mathrm{p}_{\mathcal{T}}(\alpha,\beta) \cap \mathcal{V}_A = \{\alpha,\beta\}\right\},$$

where $\mathrm{p}_{\mathcal{T}}(\alpha,\beta)$ denotes the path between α and β on $\mathcal{T}$. Appropriately, $\mathcal{T}$ is called supertree *of $\mathcal{T}_A$.*

If for n leaves a sufficient set of restrictions is available, then the supertree can be reconstructed. On the general properties and problems of a supertree reconstruction see e.g. Bininda-Emonds et al. [2002] or Semple and Steel [2003, chap. 6]. For this work two classes of restrictions are of particular interest.

Definition 1.1.6. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree with $n \geq 4$ leaves.*

1. *A* triple tree *is a restriction $\mathcal{T}_L$ of $\mathcal{T}$ to a set of three distinct leaves $L := \{\alpha, \beta, \gamma\} \subset \mathcal{L}$.*
2. *A* quartet tree *is a restriction $\mathcal{T}_L$ of $\mathcal{T}$ to a set of four distinct leaves $L := \{\alpha, \beta, \gamma, \delta\} \subset \mathcal{L}$.*

Quartet trees are a popular class for reconstruction since a quartet tree is the smallest tree that contains structural properties other than connectedness. They are used in the software-package TREE-PUZZLE(cf. Schmidt et al. [2002]), and their functionality is discussed in various works (see e.g. Strimmer and von Haeseler [1996] or Waterman [1995, chap. 14]). For any selection of four leaves $\alpha, \beta, \gamma, \delta \in \mathcal{V}$ three different quartet trees are possible, namely $(\alpha\beta)|(\gamma\delta)$, $(\alpha\gamma)|(\beta\delta)$ and $(\alpha\delta)|(\beta\gamma)$. This notation is also called *split notation*, the dash in the middle symbolizes the inner edge for the associated quartet tree. See Figure 1.4 for a good example.

Triple trees are structurally unique but only carry the information that the considered leaves are connected. Without additional properties associated with them no reconstruction can be attempted. In this thesis, the additional property is the parametrization of a Markov process on the vertices of the triple tree. Markov processes on trees will be introduced in Section 1.2, an algorithm for the tree construction will be presented in Section 5.4.

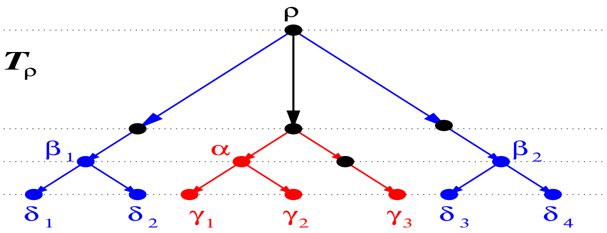

Figure 1.4: The figure shows the rooted tree $\mathcal{T}_\rho$ and some restrictions, in particular, the quartet tree for leaves $\{\delta_1, \ldots, \delta_4\}$ with inner vertices β_1, β_2, and the triple tree for leaves $\{\gamma_1, \gamma_2, \gamma_3\}$ with inner vertex α. The quartet tree has no root due to multiple choices, whereas the root for a triple tree will always be chosen as the inner vertex, here α. All black vertices cannot be observed in the restrictions.

The main focus of this work is on triple trees. Occasionally, quartet trees will be used to bring obtained results into a better view in terms of the reconstruction of trees. Generally, for the reconstruction of supertrees from triple trees the number of triple trees associated with a furcating vertex is of interest. If one deletes an inner vertex $\alpha \in \mathcal{N}$ together with the edges incident to it from a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, the tree decomposes into $\deg(\alpha)$ disjoint subtrees. Denote the set of this subtrees by $\mathfrak{G}_\alpha$. Using $\mathfrak{G}_\alpha$ the number of possible triple trees with α as trifurcating vertex is easily computed.

Lemma 1.1.8. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree and $\alpha \in \mathcal{N}(\mathcal{T})$ a furcating vertex. Then the number of triple trees with α as trifurcating vertex is given by:*

$$\mathrm{nt}_3(\alpha) := \sum_{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3 \in \mathfrak{G}_\alpha} \sharp(\mathcal{L}(\mathcal{T}_1))\sharp(\mathcal{L}(\mathcal{T}_2))\sharp(\mathcal{L}(\mathcal{T}_3)). \tag{1.1.1}$$

For a reconstruction method evaluating all triple trees, each inner vertex $\alpha \in \mathcal{N}$ is found in $\mathrm{nt}_3(\alpha)$ different triple trees. For example, the vertex ρ in Figure 1.4 is the trifurcating vertex for 12 different triple trees whereas α, β_1 and β_2 are the trifurcating vertices for five triple trees each.

Overall a reconstruction of a tree with n leaves can be achieved by regarding $\binom{n}{3}$ distinct triple trees with a weight function compared to $3\binom{n}{4}$ different quartet trees or $(2n-5)!!$ binary trees.

n	triples	quartets	binary
4	4	3	3
5	10	15	15
6	20	45	105
7	35	105	945
8	45	210	10.395
9	84	378	135.135
10	120	630	2.027.025
20	1.140	14.535	$2,2 \cdot 10^{20}$

Table 1.2: The number of possible binary trees for n leaves and the number of their restrictions with three and four leaves, respectively.

As Table 1.2 shows, for a large number of leaves the number of triple trees is much smaller than the number of quartet trees or binary trees. Hence together with a reasonable weight function the computational time could be reduced considerably. Markov processes on trees offer various possible weight functions.

1.2 Markov Models on Trees

This section introduces the notion of conditional independence, and applies this notion to define processes on trees with certain properties. These properties are first introduced on undirected trees, and then with slight variations on rooted trees. The notions are based on Lauritzen [1996, chap. 3] and its revised form Lauritzen [2001]. Most results are immediate consequences of results from these works. Throughout this thesis it is sufficient to restrict the considerations to discrete random variables.

1.2.1 Conditional Independence

Let X, Y, Z denote discrete random variables with joint probability distribution μ over a discrete space $\mathcal{X} := \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$. The following abbreviations are useful:

$$\mu_{XY}(x,y) := \mu(X=x, Y=y), \quad \mu_{X|Y}(x,y) := \mu(X=x|Y=y),$$
$$\mu_X(x) := \mu(X=x),$$

where the equations describe the pairwise joint probability, the conditional probability for $X = x$ given $Y = y$ and the marginal distribution in X, respectively. Accordingly, the notion for the joint distribution in all random variables is $\mu_{XYZ} = \mu$. The law of total probability (see also (1.5.2)) indicates that all those abbreviations can be described by μ, thus justifying the use of the letter μ in the abbreviations.

Definition 1.2.1. *Let X, Y, Z denote discrete random variables with joint probability distribution μ. X is called* conditionally independent of Y given Z under μ *and write $X \perp\!\!\!\perp Y \mid Z\ [\mu]$ if for μ-almost all $x \in \mathcal{X}_X$, $y \in \mathcal{X}_Y$, $z \in \mathcal{X}_Z$ the following equality holds*

$$\mu_{XY|Z}(x,y,z) = \mu_{X|Z}(x,z) \cdot \mu_{Y|Z}(y,z). \tag{1.2.1}$$

If Z is trivial X is said to be independent of Y*, and write $X \perp\!\!\!\perp Y$.*

Lemma 1.2.1. *Let X, Y, Z and W denote discrete random variables with joint distribution μ and let h denote an arbitrary measurable function on the sample space of X. Then, the ternary relation $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ has the following properties:*

(C1) *if $X \perp\!\!\!\perp Y \mid Z$, then $Y \perp\!\!\!\perp X \mid Z$;*

(C2) *if $X \perp\!\!\!\perp Y \mid Z$ and $U = h(X)$, then $U \perp\!\!\!\perp Y \mid Z$;*

(C3) *if $X \perp\!\!\!\perp Y \mid Z$ and $U = h(X)$, then $X \perp\!\!\!\perp Y \mid (Z, U)$;*

(C4) *if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, then $X \perp\!\!\!\perp (W, Y) \mid Z$.*

Note, that the converse to (C4) follows from (C2) and (C3). Another property of the conditional independence relation is often used:

(C5) If $X \perp\!\!\!\perp Y \,|\, Z$ and $X \perp\!\!\!\perp Z \,|\, Y,$ then $X \perp\!\!\!\perp (Y, Z)$.

However, this property does not hold universally, but only under additional conditions - essentially, that there are no non-trivial logical relationships between Y and Z. A trivial counterexample appears when $X = Y = Z$ with $\mu_X(1) = \mu_X(0) = 1/2$. One condition for the validity of (C5) is the strict positivity of the joint distribution μ.

Lemma 1.2.2.(Proposition 3.1 in Lauritzen [1996]) *Let X, Y, Z denote discrete random variables with a joint distribution μ. If μ is strictly positive, then (C5) is valid.* □

Quite a few works on modeling molecular evolution assume strict positivity (cf. Chang [1996], Baake [1998]). This work refrains from making this assumption since the input data usually available do not support it.

As already indicated, the relationship of families of random variables over a set of vertices is of great interest. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an arbitrary graph and $\mathbf{X} := (X_\alpha)_{\alpha \in \mathcal{V}}$ arrays of random variables X_α in finite discrete measurable spaces $\mathcal{X}_\alpha$. For $A \subseteq \mathcal{V}$ denote by $\mathcal{X}^A := \times_{\alpha \in A} \mathcal{X}_\alpha$ the cross product of the measurable space of the random field $X^A := (X_\alpha)_{\alpha \in A}$. Typical elements of $\mathcal{X}^A$ are denoted by $x^A := (x_\alpha)_{\alpha \in A}$.

For vertex sets $A, B, C \subseteq \mathcal{V}$ the following abbreviation for the conditional independence relation is introduced

$$A \perp\!\!\!\perp B \,|\, C \quad :\Leftrightarrow \quad X^A \perp\!\!\!\perp X^B \,|\, X^C.$$

With this abbreviation the properties (C1)-(C4) translate for vertex sets $A, B, C \subseteq \mathcal{V}$ to

(S1) if $A \perp\!\!\!\perp B \,|\, C,$ then $B \perp\!\!\!\perp A \,|\, C;$

(S2) if $A \perp\!\!\!\perp B \,|\, C$ and $D \subseteq A,$ then $D \perp\!\!\!\perp B \,|\, C;$

(S3) if $A \perp\!\!\!\perp B \,|\, C$ and $D \subseteq A,$ then $A \perp\!\!\!\perp B \,|\, (D \cup C);$

(S4) if $A \perp\!\!\!\perp B \,|\, C$ and $A \perp\!\!\!\perp D \,|\, (B \cup C),$ then $A \perp\!\!\!\perp (D \cup B) \,|\, C.$

Moreover, for disjoint subsets A, B, C and D of $\mathcal{V}$, (C5) transfers to

(S5) if $A \perp\!\!\!\perp B \,|\, (C \cup D)$ and $A \perp\!\!\!\perp C \,|\, (B \cup D),$ then $A \perp\!\!\!\perp (B \cup C) \,|\, D.$

Similar to (C5), property (S5) only holds for additional conditions, for example if the joint distribution μ is strictly positive.

For the abbreviations for conditional and marginal probabilities the convention will be extended such that the vertex sets are used as indices instead of the random variables, e.g. $\mu_{A|B} := \mu_{X^A|X^B}$ for $A, B \subseteq \mathcal{V}$. The joint probability over $\mathbf{X}$ will still be denoted by $\mu := \mu_{\mathcal{V}}$.

1.2.2 Markov Models on Undirected Trees

This subsection will introduce some properties for a process $\mathbf{X}$ on a vertex set $\mathcal{V}$ of an undirected tree $\mathcal{T}$ with values in a state space $\mathcal{X}$. It is well-known that such a process is characterized by a joint distribution μ over $\mathcal{X}^{\mathcal{V}}$. Of special interest in this work are joint distribution which *factorize*.

Definition 1.2.2. *A joint distribution μ over $\mathcal{X}^{\mathcal{V}}$ is said to* factorize *according to an undirected tree $\mathcal{T}$ if for all edges $⟅\alpha, \beta⟆ \in \mathcal{E}$ non-negative functions $\psi_{\alpha\beta} : \mathcal{X}_\alpha \times \mathcal{X}_\beta \to \mathbb{R}$ exist such that the probabilities for states $x^{\mathcal{V}} \in \mathcal{X}^{\mathcal{V}}$ can be written as:*

$$\mu(x^{\mathcal{V}}) = \prod_{⟅\alpha,\beta⟆ \in \mathcal{E}} \psi_{\alpha\beta}(x_\alpha, x_\beta). \tag{1.2.2}$$

If μ factorizes it is said to have property (F)*.*

The functions $\psi_{\alpha\beta}$, $⟅\alpha, \beta⟆ \in \mathcal{E}$ are not unique. For example, one could fix a leaf $\varrho \in \mathcal{L}$ with the partial ordering $\prec_\varrho$ and set

$$\psi_{\alpha\beta}(x_\alpha, x_\beta) = \left\{ \begin{array}{ll} \mathbb{P}(X_\beta = x_\beta \,|\, X_\alpha = x_\alpha), & \alpha \prec_\varrho \beta, \\ \mathbb{P}(X_{\mathrm{ch}(\varrho)} = x_{\mathrm{ch}(\varrho)}, X_\varrho = x_\varrho), & \alpha = \varrho \end{array} \right. . \tag{1.2.3}$$

This assignment for $\psi_{\alpha\beta}$ is unique because the degree of a leaf is one, and therefore $\varrho \in \mathcal{L}$ has only one child. This is the preferred assignment for $\psi_{\alpha\beta}$ (e.g. Chang [1996] or Huelsenbeck and Bollback [2001]). The assignment of the partial ordering also gives an idea about the factorization on directed trees. But this is the subject of Subsection 1.2.3.

For an introduction of Markov properties the factorization property lacks a proper interpretation. To accommodate for this, consider the following properties: A probability measure μ over $\mathcal{X}^{\mathcal{V}}$ is said to obey

(P) the *pairwise Markov property*, relative to $\mathcal{T}$, if for any pair $\alpha, \beta \in \mathcal{V}$ of non-adjacent vertices:

$$\alpha \perp\!\!\!\perp \beta \,|\, \mathcal{V} \setminus \{\alpha, \beta\};$$

(L) the *local Markov property*, relative to $\mathcal{T}$, if for any vertex $\alpha \in \mathcal{V}$:

$$\alpha \perp\!\!\!\perp \mathcal{V} \setminus (\{\alpha\} \cup \mathrm{ne}(\alpha)) \,|\, \mathrm{ne}(\alpha);$$

(G) the *global Markov property*, relative to $\mathcal{T}$, if for any triple (A, B, S) of disjoint subsets of $\mathcal{V}$ the set S separates A from B in $\mathcal{T}$:

$$A \perp\!\!\!\perp B \,|\, S.$$

The pairwise Markov property states, that the information available from β that is relevant for a prediction for α is already contained in the information from the remaining vertices. Analogously, the local Markov property states that the neighbors of α also contain the information from the non-adjacent vertices of α that is relevant for a prediction for α. In Chang [1996] this is the property assigned to the process. The global Markov property is the strongest property because the separation property states that any neighbor β of α already contains all available information from the subtrees separated from α by β. Figure 1.5 illustrates these properties on a tree.

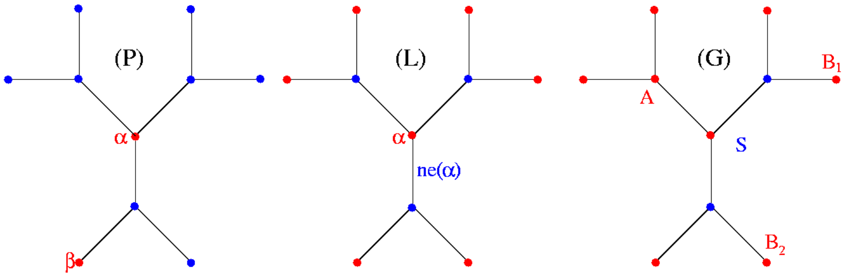

Figure 1.5: The three colorations of $\mathcal{T}$ demonstrate the properties (P), (L) and (G), respectively. The blue vertices are the condition for the independence of α from the other red vertices in (P) and (L), and S separates A from B_1 and B_2 in (G).

The relationship between the presented Markov properties is the object of the next considerations. Equivalence would be preferable, because the factorization property is the main object of the next chapters. The following properties can be observed:

Proposition 1.2.3. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree and μ a probability distribution on $\mathcal{X}^{\mathcal{V}}$.*

1. *The following implication rule holds:* (F) $\Rightarrow$ (G) $\Rightarrow$ (L) $\Rightarrow$(P)*;*

2. *The following equivalence holds:* (F) $\Leftrightarrow$ (G)*;*

3. *If μ is such that (S5) holds for disjoint sets $A, B, C, D \subseteq \mathcal{V}$, then*

$$\text{(F)} \quad \Leftrightarrow \quad \text{(G)} \quad \Leftrightarrow \quad \text{(L)} \quad \Leftrightarrow \quad \text{(P)}.$$

Hence, if a joint distribution μ over $\mathcal{X}^{\mathcal{V}}$ satisfies (S5) over a tree, equivalence is attained. Generally, on undirected graphs equivalence is only observed if μ is a strictly positive joint distribution. This statement was proved by several authors, but is usually attributed to Hammersley and Clifford [1971]. Strict positivity is also a consequence of the conditions on the functions $(\psi_{\alpha\beta})_{\langle\!\langle\alpha,\beta\rangle\!\rangle\in\mathcal{E}}$ in Chang [1996], guaranteeing the equivalence of local Markov property and factorization property.

In Matús [1992] classes of graphs are presented on which equivalences of properties is obtained with structural arguments. The results of this work have the following consequence for trees:

Lemma 1.2.4. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree. Then, the factorization property and the local Markov property are equivalent if and only if $\mathcal{E}$ contains at most one inner edge.*

In other words, on star trees and trees with one inner edge the joint distribution μ need not satisfy additional conditions to provide the equivalence of (F) and (L). Since, triple trees are star trees, and quartet trees are the only class of binary trees with one inner edge, they are subject to this equivalence. However, Lemma 1.2.4 also indicates that this equivalence does not translate to supertrees with more than one inner edge. The next statement is important for reconstruction methods:

Proposition 1.2.5. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an unrooted tree and μ a joint distribution over $\mathcal{X}^{\mathcal{V}}$. Further, A denotes a subset of $\mathcal{V}$, $\mathcal{T}_A$ is the associated restriction of $\mathcal{T}$, and $\mu_A = \mu|_A$ the constraint of μ to $\mathcal{T}_A$. If μ admits a factorization on $\mathcal{T}$ then μ_A also admits a factorization on $\mathcal{T}_A$. Generally, the converse does not hold.*

Therefore, if a factorizing distribution μ exists on the supertree $\mathcal{T}$, the constraints of μ to the restrictions of $\mathcal{T}$ factorize.

1.2.3 Markov Models on Rooted Trees

Usually, rooted trees are the preferred structure for phylogenetic reconstruction, since they suggest an evolutionary time system. This section will provide similar Markov properties to the properties presented for undirected trees and consider some implications. As before, the first property is a factorization property on rooted trees.

Definition 1.2.3. *A probability distribution μ allows a* recursive factorization *over a discrete state space $\mathcal{X}^{\mathcal{V}}$ on a rooted tree $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$, if a* root distribution *q^ϱ and a family of transition matrices $(P^{\alpha\beta})_{(\alpha,\beta)\in\mathcal{E}}$ exists such that μ can be written as:*

$$\mu(x^{\mathcal{V}}) = q^\varrho(x_\varrho) \prod_{(\alpha,\beta)\in\mathcal{E}} P^{\alpha\beta}(x_\beta, x_\alpha), \quad x^{\mathcal{V}} = (x_\alpha)_{\alpha\in\mathcal{V}} \in \mathcal{X}^{\mathcal{V}}. \tag{1.2.4}$$

Hence a factorizing distribution μ is characterized through a marginal distribution q^{ϱ} and family of transition matrices $(P^{\alpha\beta})_{(\alpha,\beta)\in\mathcal{E}}$. The recursive factorization will be denoted by (DF). Again, though the factorization property of a joint distribution μ is appropriate, it has no good interpretation in terms of conditional probabilities. Therefore, further Markov properties are introduced. A probability measure μ over $\mathcal{X}^{\mathcal{V}}$ on the rooted tree $\mathcal{T}_{\varrho}$ is said to obey

(DG) the *directed global Markov property*, relative to $\mathcal{T}_{\varrho}$, if for any triple (A, B, S) of disjoint subsets of $\mathcal{V}$ the set S separates A from B in $\mathcal{T}_{\varrho}$;

(DL) the *local directed Markov property*, relative to $\mathcal{T}_{\varrho}$, if any vertex is conditionally independent from its non-descendants given its parent vertex $\mathrm{pa}(\alpha)$:

$$\alpha \perp\!\!\!\perp \mathrm{nd}(\alpha) \setminus \{\mathrm{pa}(\alpha)\} \mid \mathrm{pa}(\alpha);$$

(DO) the *ordered directed Markov property*, relative to $\mathcal{T}_{\varrho}$, if any vertex is conditionally independent from its history given its parent vertex $\mathrm{pa}(\alpha)$:

$$\alpha \perp\!\!\!\perp \mathrm{hi}(\alpha) \setminus \{\mathrm{pa}(\alpha)\} \mid \mathrm{pa}(\alpha).$$

These properties have appropriate interpretations in terms of phylogenetic reconstruction. For instance, the ordered Markov property states that all the information available from the history of a certain species is contained in the information of the predecessor of the species. Also, (DO) is regularly used as the basis of model considerations in terms of phylogenetic reconstruction (e.g. Huelsenbeck and Bollback [2001] or Steel et al. [1998]). Figure 1.6 visualizes the presented properties.

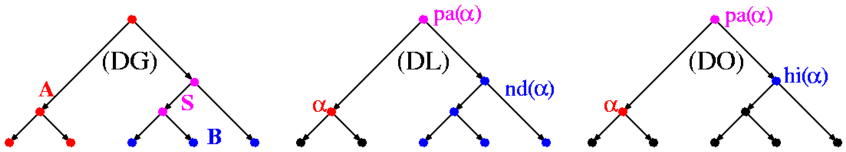

Figure 1.6: Markov properties on rooted trees. (DG) describes the global directed Markov property where A is conditionally independent of B given S. (DL) describes the local directed Markov property, and (DO) describes the ordered directed Markov property. In both properties α is given its parent conditionally independent of its non-descendants and its history, respectively.

The interesting fact about these properties is the fact that they are equivalent:

Theorem 1.2.6. *Let $\mathcal{T}_{\varrho} = (\mathcal{V}, \mathcal{E}; \varrho)$ denote a rooted tree. For a probability distribution μ over a discrete probability space $\mathcal{X}^{\mathcal{V}}$ the following equivalence is observed:*

$$\text{(DF)} \quad \Leftrightarrow \quad \text{(DG)} \quad \Leftrightarrow \quad \text{(DL)} \quad \Leftrightarrow \quad \text{(DO)}.$$

Due to this equivalence one just speaks of the *directed Markov property.* The relationship of Markov properties on undirected trees to the directed Markov property on rooted trees is explained in the following statement:

Proposition 1.2.7. *Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote an undirected tree and μ a joint distribution over $\mathcal{X}^{\mathcal{V}}$. μ obeys the global Markov property relative to $\mathcal{T}$, if and only if it obeys the directed Markov property on the rooted tree $\mathcal{T}_\alpha$ for all $\alpha \in \mathcal{V}$.*

With Proposition 1.2.3.2 one can equivalently state, that a factorizing joint distribution μ on a undirected tree $\mathcal{T}$ also factorizes on all rooted trees $\mathcal{T}_\alpha$, $\alpha \in \mathcal{V}$. Therefore, a particular choice of root does not alter the joint distribution μ. This root irrelevance is due to the commutativity of joint probabilities and their influence on the definition of conditional probabilities, see (1.5.1).

Note, that Proposition 1.2.5 is also valid on rooted trees.

Corollary 1.2.8. *Let $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$ denote a rooted tree, $A \subseteq \mathcal{V}$, and μ is a factorizing distribution over $\mathcal{X}^{\mathcal{V}}$. Then, the constraint μ_A of μ to A factorizes on the restriction $\mathcal{T}_A$.*

In particular, this result provides the opportunity to regard restrictions like triple or quartet trees in order to derive the characterization of a possible Markov process on the supertree. However, one should always keep in mind that the existence of a factorizing distribution on the restrictions is only necessary but not sufficient for the existence of a factorizing distribution on the supertree. Sufficiency conditions are regarded in Section 1.4.

1.3 Biological Background

Usually, a stochastic approach to *molecular evolution* is made by treating it as a Markov process X on a rooted tree $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$ over a genetically motivated state space $\mathcal{X}$. The structural elements of $\mathcal{T}_\varrho$ are interpreted in the following way. The leaf set $\mathcal{L}$ depicts a set of extant species and the inner vertices depict their respective ancestors up to ϱ which describes their *most recent common ancestor* (mrca($\mathcal{L}$)). In that notion a tree over all extant species should have the ancestor of all species (if such a species exists) as a root. The best known example of such a tree is the Haeckel-tree.

Ideally, the state space $\mathcal{X}$ is the set of *sequences* or *words* of length m over a genetical alphabet. The most popular alphabets are:

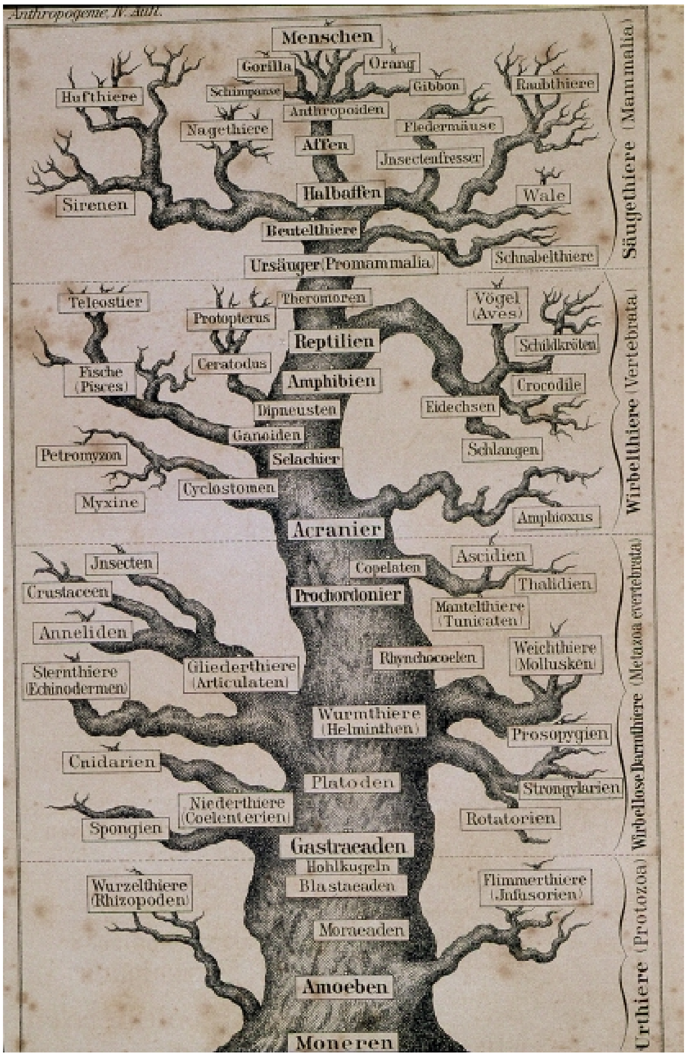

Figure 1.7: The Haeckel Tree

$$\mathcal{S}_2 := \{R, Y\}, \quad \mathcal{S}_4 := \{A, C, G, T\},$$
$$\mathcal{S}_{20} := \{w, m, y, q, f, i, g, v, h, e, l, p, s, c, a, r, n, d, t, k\},$$

depending whether one looks at purines vs. pyrimidines ($\mathcal{S}_2$), at nucleotides ($\mathcal{S}_4$) or at amino acids ($\mathcal{S}_{20}$).

An often, but reluctantly (e.g. Huelsenbeck and Bollback [2001]) made simplification concerns the evolution of sequences: It is assumed that all positions of a sequence evolved independently and identically distributed. In other words, the process of molecular evolution is assumed to be driven solely by point mutations, and that no recombination, insertions or deletions occurred. It is a very restricting condition, and methods developed under such a model should only be applied to sequences from either mitochondrial DNA or of the Y-chromosome. Actually, most insights concerning the relationship of species or races are based on comparison of such sequences (e.g. Sykes [2001]). Under this assumption the state set can the restricted to one the alphabets. Then a set of n aligned sequences of N sites provides a sample of N independent observations of the process $\mathbf{X}$, and hence statistical methods can be applied to estimate the process in the n vertices which represent the n sequences.

$\mathbf{X}$ is characterized through a joint distribution $\mu := (\mu_x)_{x \in \mathcal{X}^{\mathcal{V}}}$ which assigns to every joint state $x \in \mathcal{X}^{\mathcal{V}} = \mathcal{X} \times \cdots \times \mathcal{X}$ a probability of occurrence. A joint distribution μ over $\mathcal{X}^{\mathcal{V}}$ which characterizes a Markov process X will be called a *Markov distribution.* Since $\mathbf{X}$ is a Markov process its characterizing distribution μ is subject to equation (1.2.4) and hence is described by choosing transition matrices $(P^e)_{e \in \mathcal{E}}$ and a root distribution μ^ϱ from a parametric subfamily.

This thesis will consider three model specifications given by the special structure of their transition matrices, namely the general two state model, the Neyman N_k model and the Kimura 2ST model.

Example 1.3.1. The *general two state model* considers the state space $\mathcal{S}_2$ or equivalently $\{0, 1\}$ and transition matrices of type:

$$p^\alpha := \begin{pmatrix} 1 - p_{01}^\alpha & p_{01}^\alpha \\ p_{10}^\alpha & 1 - p_{10}^\alpha \end{pmatrix}, \quad \mu^\varrho := \begin{pmatrix} q_0^\varrho \\ 1 - q_0^\varrho \end{pmatrix}$$

for $\alpha \in \mathcal{V} \setminus \{\varrho\}$. It is the simplest non-symmetric model, i.e. the transition from class one to class two has a different probability of occurrence than staying in one class. Apparently, one can apply this model to DNA-data by distinguishing two classes of states. Two of the three possible selections actually have an interpretation. The selection $\{A, G\}$ vs. $\{C, T\}$ is the purine vs. pyrimidine approach. The selection $\{A, T\}$ vs. $\{C, G\}$ would give an idea about the possible development of the often discussed $\{G, C\}$-content (e.g. Meunier and Duret [2004]). According to the presented article the evolution of the $\{G, C\}$-content is driven by recombination. Since the homogeneity assumption does not permit recombination and under the stability assumption for the $\{G, C\}$-content (cf. Meunier and Duret [2004]), the change of the content should be small if at all observable. The third classification

$\{A, C\}$ vs. $\{G, T\}$ seems to be of no interest.

Example 1.3.2. The simplest way to incorporate a larger state space $\mathcal{X}$ is to assign a probability p^e for the overall probability of change along an edge e and then distributing it equally to all states. For instance, if $\mathcal{X} = \mathcal{S}_k := \{0, 1, \ldots, k-1\}$, the change from state $x \in \mathcal{S}$ to state $y \neq x$ has probability $p^e/(k-1)$. In addition, if the marginal distribution in the root ϱ is assumed to be stationary, i.e. $\mu^\varrho = (1/k, \ldots, 1/k)$, the resulting model is called the *Neyman N_k model* (eg. Semple and Steel [2003]). The transition matrix for an edge $e \in \mathcal{E}$ according to this model is described by:

$$P^e := \begin{pmatrix} 1-p_e & \frac{p_e}{k-1} & \cdots & \frac{p_e}{k-1} \\ \frac{p_e}{k-1} & 1-p_e & \cdots & \frac{p_e}{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_e}{k-1} & \frac{p_e}{k-1} & \cdots & 1-p_e \end{pmatrix}$$

Due to the symmetric structure of the transition matrices for all edges the stationarity of the marginal distributions translates to all vertices, i.e. $\mu^\alpha = \mu^\varrho$ for all $\alpha \in \mathcal{V}$. Hence, the model is characterized through one parameter per edge. The special case N_4 is better known as the *Jukes-Cantor-model.*

Example 1.3.3. Although the Neyman approach is easy and can be applied to any number of states, more complex models are preferred to accommodate certain observations in real data. One such observation is addressed by the *Kimura 2ST model.* Examining the classification of nucleotides into purines and pyrimidines showed that a change within a class is more probable than a change between classes. A change within a class is called TRANSITION, and a change between classes is called TRANSVERSION. The Kimura 2ST model is defined over the state space $\mathcal{S}_4$ or equivalently $\{0, 1, 2, 3\}$, and regards the states as stationarily distributed at the vertices, in this case $\mu^\alpha = (1/4, 1/4, 1/4, 1/4)$, $\alpha \in \mathcal{V}$. As already proposed, the states are divided into two classes, namely *purines* ($\{0, 1\} = \{A, G\}$) and *pyrimidines* ($\{2, 3\} = \{C, T\}$). The associated transition matrix for an edge $e \in \mathcal{E}$ is given by:

$$P^e := \begin{pmatrix} 1-p_e-2q_e & p_e & q_e & q_e \\ p_e & 1-p_e-2q_e & q_e & q_e \\ q_e & q_e & 1-p_e-2q_e & p_e \\ q_e & q_e & p_e & 1-p_e-2q_e \end{pmatrix}, \tag{1.3.1}$$

Here, p_e denotes the probability of a TRANSITION and $2q_e$ is the probability of a TRANSVERSION along edge $e \in \mathcal{E}$.

Example 1.3.4. There are two other rather popular specifications, namely the *rate model* and the *rate model with molecular clock.* The example will introduce those models only as far as they are considered in the thesis. For a more complete look at these model specifications see eg. Waterman [1995, chap. 15].

Behind the development of the *rate model* was the assumption of a continuous time model, where a rate matrix Q describes the rates of change across states, i.e.

$$Q = \begin{pmatrix} -\sum_{i=2}^{k} q_{1i} & q_{12} & \cdots & q_{1k} \\ q_{21} & -\sum_{i\neq 2}^{k} q_{2i} & \cdots & q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & \cdots & -\sum_{i=1}^{k-1} q_{ki} \end{pmatrix}.$$

The transition matrix after time $t \geq 0$ is given by

$$P(t) = \exp(t \cdot Q) = \sum_{m=0}^{\infty} \frac{t^m}{m!} Q^m.$$

Thus, transition matrices for a particular edge e are given by $\exp(t_e Q)$ where t_e denotes the time associated with the length of edge $e \in \mathcal{E}$. Moreover, for $t = 0$ this approach yields $P(0) = \mathbb{1}_k$, the identity matrix in $\mathbb{R}^{k \times k}$, i.e. if no time elapsed the probability of change is zero. Thus, artificial edges of zero length are endowed with the identity matrix as their transition matrix.

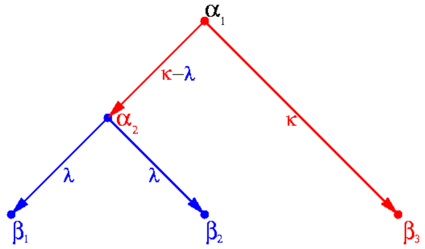

Figure 1.8: A rooted binary tree with molecular clock. The lengths of the edges (α_2, β_1) and (α_2, β_2) equals λ and the length of (α_1, α_2) is the length κ of (α_1, β_3) minus the λ.

Generally, rooted trees are preferred for their resemblance to a time line and therefore, the suggestion of a process running through time. However, as Proposition 1.4.1 will show, the Markov model without any further restriction does not prefer a particular root, i.e. any choice of inner vertex as root returns the same joint distribution to the Markov process.

One restriction providing a root is the *rate model with molecular clock*. It forces a root to a tree by demanding that paths between the root and leaves have equal lengths. This approach is called molecular clock since it is based on the assumption that for extant species the same evolutionary time elapsed since their mrca roamed the earth. As an example consider Figure 1.8. Here, the edges (α_2, β_1) and (α_2, β_2) have the same lengths and the length of edge (α_1, β_3) is equal to the length of the path $\mathrm{p}(\alpha_1, \beta_1)$. Methods using this approach provided good approximations of the real evolutionary time. Probably the best known result was the placing of the mrca of chimp and human three million years back which was at that time a much shorter period as was assumed by anthropologists (eg. Gribbin and Cherfas [2001]).

This concludes the introduction of models of molecular evolution considered in this work. Obviously, there are a lot more models each of which serves the visualization of certain aspects observed in data. However, their introduction is not subject of this thesis. For a satisfying overview Ewens and Grant [2001] is suggested.

1.4 The Task of Phylogenetic Reconstruction

Although the number of theories is astonishing, knowledge of evolutionary history is sparse and generally data are available only for species of recent times. Some fossil records provide approximations to evolutionary time, and thus for a molecular clock, and some insights into relationship and inheritance as well. In terms of the model this means that the inner structure of the tree is unknown.

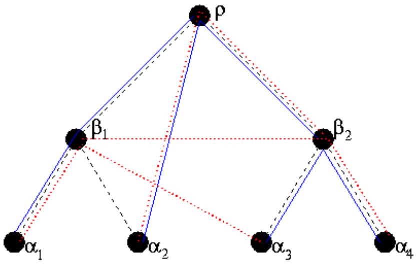

Figure 1.9: The leaves $\{\alpha_1, \ldots, \alpha_4\}$ are known, but the connection to the three inner vertices β_1, β_2 and ρ is unknown. The figure presents three possible rooted structures. The black dashed structure is rooted at ρ and splits α_1, α_2 from α_3, α_4. The blue solid structure presents a root change from ρ to β_1 and a reconnection of α_2 from β_1 to ρ. In this structure α_1 is called an *outgroup*. Finally, the red pointed structure puts the root back into ρ but splits α_1, α_3 from α_2, α_4 with outgroup α_2. Outgroups are usually used to place a root. Note, that the edge lengths in the picture are meaningless, they only visualize connectivity.

The knowledge of present species can be viewed as the knowledge of the Markov process on the leaves $\mathcal{L}$ of the sought tree $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$. Throughout the chapter the cardinality of $\mathcal{L}$ is $n \geq 3$. Thus in terms of the model, today's knowledge is given as a leaf distribution $\underline{m}$, which relates to the Markov distribution μ by

$$\underline{m}(\underline{x}) = \sum_{\underline{y} \in \mathcal{X}^{\sharp(\mathcal{N})}} \mu(\underline{x}, \underline{y}), \quad \underline{x} \in \mathcal{X}^n \tag{1.4.1}$$

The task of phylogenetic reconstruction is to find a Markov distribution μ which fits today's knowledge i.e. a given leaf distribution $\underline{m}$.

Definition 1.4.1. *Let $\mathcal{T}_\varrho$ denote a rooted tree and $\underline{m}$ a leaf distribution on $\mathcal{L}$. A Markov distribution μ on $\mathcal{V}$ satisfying (1.4.1) is called* Markov extension *of $\underline{m}$ over $\mathcal{T}_\varrho$.*

Since μ is a Markov distribution on a rooted tree, one applies (1.2.4) to (1.4.1) to get the following relationship between a leaf distribution and the transition matrices:

$$m(\underline{x}) = \sum_{\underline{x}^{\mathcal{N}} \in \mathcal{X}^{\mathcal{N}}} \mu_\varrho(x_\varrho) \prod_{(\alpha,\beta) \in \mathcal{E}} P^{\alpha\beta}_{x_\beta, x_\alpha}, \quad \underline{x} \in \mathcal{X}^{\mathcal{L}}. \tag{LF}$$

This equation is the basis of almost all following considerations. In terms of phylogenetic reconstruction the left hand side is known for all $\underline{x} \in \mathcal{X}^{\mathcal{L}}$ and the parameters of the associated right hand sides need to be retrieved.

Recall from Proposition 1.2.7, that for μ to be a Markov distribution the particular choice of ϱ is of no effect. For reconstruction methods the following properties must be regarded:

Proposition 1.4.1. *Let $\mathcal{T}_\varrho = (\mathcal{V}, \mathcal{E}; \varrho)$ denote a rooted tree and $\underline{m}$ an joint distribution on the leaves of $\mathcal{T}_\varrho$ with associated characterization $(P^{\alpha\beta})_{(\alpha,\beta)\in\mathcal{E}}$ and μ^ϱ for the associated Markov distribution μ. Then, the following properties are observable:*

1. *μ is also a Markov distribution on every rooted tree $\mathcal{T}_\alpha$, $\alpha \in \mathcal{V}$.*
2. *μ can be adapted to any tree obtained from $\mathcal{T}_\varrho$ by adding or deleting a vertex of degree two without violating the Markov property.*

Statement 2 implies, that vertices of degree two are not reconstructible from a leaf distribution. Statement 1 shows, that for computations the root can be placed at the best suited vertex without changing the Markov distribution. However, for the placement of a root, the structure of $\mathcal{T}$ must be known. Equation (1.1.5) provides a lower bound of possible structures with the number of binary trees to a given number of leaves. As Table 1.2 shows, for 20 leaves this number is with $2,2 \cdot 10^{20}$ already much too high to check all possible structures.

Thus alternative methods of reconstruction are sought. Table 1.2 suggests one, namely using sets of subtrees to reconstruct the supertree (e.g. Semple and Steel [2003, chap. 6]). A popular approach is using quartet trees (e.g. Strimmer and von

Haeseler [1996]), where only three possible ways of inner structures to four leaves are distinguishable. Hence, computing a sufficient set of quartet trees provides a way to reconstruct a supertree. A sufficient set of subtrees contains restriction trees for all leaves, and some overlap to connect them. Often, the quartet set contains incompatible quartets. For instance, the quartet splits $\alpha_1\alpha_2|\alpha_3\alpha_4$ and $\alpha_1\alpha_3|\alpha_2\alpha_5$ provide an incompatibility in the splitting of α_1, α_2 and α_3. Depending on the reconstruction method, such cases lead to information loss (supertree methods, eg. Bininda-Emonds et al. [2002]) or more structural complexity (phylogenetic networks eg. Bryant and Moulton [2002]).

Chang [1996] states, that if a factorizing joint distribution μ exists on the true (but unknown) tree $\mathcal{T}$, then the constraints of μ to the triple trees of $\mathcal{T}$ will return $\mathcal{T}$ and a characterization of μ. This is not completely true, because due to Proposition 1.4.1.2 such a reconstruction will not return non-furcating vertices of $\mathcal{T}$. However, this loss of information is acceptable because non-furcating vertices provide no additional information about the relationship of the leaves. Moreover, as Table 1.2 shows, the number of possible triple trees is another reduction of objects to consider.

1.5 Proofs

This section cumulates the proofs for all results of this chapter.

1.5.1 Proofs for Section 1.1

Section 1.1 contained the background from graph theory for trees.

Proof of Lemma 1.1.1. The presented statements follow from Theorem 1.2.1 and Proposition 1.2.5 in Semple and Steel [2003]. □

Proof of Lemma 1.1.2. In order to be a partial ordering, $\prec_\varrho$ needs to be reflexive, transitive and asymmetrical on $\mathcal{V}$. Let $\alpha, \beta, \gamma \in \mathcal{V}$. Reflexivity follows since $\alpha \in \mathrm{p}(\varrho, \alpha)$.

For transitivity let $\alpha \prec_\varrho \beta$ and $\beta \prec_\varrho \gamma$, i.e. $\alpha \in \mathrm{p}(\varrho, \beta)$ and $\beta \in \mathrm{p}(\varrho, \gamma)$. Since $\mathcal{T}$ is a tree, $\mathrm{p}(\varrho, \beta) \subseteq \mathrm{p}(\varrho, \gamma)$ and thus, $\alpha \in \mathrm{p}(\varrho, \gamma)$, i.e. $\alpha \prec_\varrho \gamma$.

For asymmetry assume $\alpha \prec_\varrho \beta$ and $\beta \prec_\varrho \alpha$. The path on trees is unique, thus the assumption is only fulfilled if $\alpha = \beta$.

ϱ is the minimal element because $\varrho \prec_\varrho \alpha$ for all $\alpha \in \mathcal{V}$ holds. Hence, all statements of the lemma are accounted for. □

Proof of Corollary 1.1.4. Assume, that $\alpha \in \mathcal{V} \setminus \{\varrho\}$ has two distinct parents β_1 and β_2. But then $\mathcal{T}_\varrho$ would allow two different paths $\mathrm{p}_1(\varrho, \alpha)$ running through β_1, and $\mathrm{p}_2(\varrho, \alpha)$ running through β_2 contrary to the path uniqueness proved in Lemma

1.1.1. Assume, ϱ has a parent vertex γ. Then $\mathcal{T}_\varrho$ contains an edge (γ, ϱ). But this implies $\gamma \prec_\varrho \varrho$ contrary to the fact that ϱ is the minimum of $\prec_\varrho$ on $\mathcal{V}$. Therefore, ϱ is parent-less and thus, the corollary is proved. □

Proof of Lemma 1.1.6. Assume $\mathcal{G}_A$ is connected. Due to the definition, $\mathcal{T}$ does not contain a cycle. But since $\mathcal{G}_A$ only inherits edges from $\mathcal{T}$ it also is cycle-free. Thus, $\mathcal{G}_A$ is cycle-free and connected, which is the definition of a tree. This completes the proof. □

Proof of Lemma 1.1.7. The intersection of paths must be non-empty since the vertices are connected. If, w.l.o.g., $\beta \in \mathrm{p}(\alpha, \gamma)$, the intersection of the three paths returns β as the only element.

Now assume that at least two distinct vertices ϱ_1, ϱ_2 are in the intersection of the paths. W.l.o.g., rewrite two paths as unions

$$\begin{aligned}\mathrm{p}(\alpha,\beta) &= \mathrm{p}(\alpha,\varrho_1) \cup \mathrm{p}(\varrho_1,\varrho_2) \cup \mathrm{p}(\varrho_2,\beta),\\ \mathrm{p}(\alpha,\gamma) &= \mathrm{p}(\alpha,\varrho_1) \cup \mathrm{p}(\varrho_1,\varrho_2) \cup \mathrm{p}(\varrho_2,\gamma).\end{aligned}$$

This yields for the third path:

$$\begin{aligned}\mathrm{p}(\beta,\gamma) &= \mathrm{p}(\beta,\varrho_2) \cup \mathrm{p}(\varrho_2,\varrho_1) \cup \mathrm{p}(\varrho_1,\gamma)\\ &= \mathrm{p}(\beta,\varrho_2) \cup \mathrm{p}(\varrho_2,\varrho_1) \cup \mathrm{p}(\varrho_1,\varrho_2) \cup \mathrm{p}(\varrho_2,\gamma).\end{aligned}$$

Apparently, the path between ϱ_1 and ϱ_2 occurs twice on the right hand side. Thus, $\mathrm{p}(\beta,\gamma) = \mathrm{p}(\beta,\varrho_2) \cup \mathrm{p}(\varrho_2,\gamma)$ and the intersection of the three paths contains only ϱ_2 contrary to the assumption and thus, the lemma is proved. □

Proof of Lemma 1.1.8. Selecting a triple begins with selecting three leaves. Since α is meant to be their trifurcating vertex, the leaves need to be separated by α. Thus, each has to come from a different subtree from $\mathfrak{G}_\alpha$. Select three distinct subtrees from $\mathfrak{G}_\alpha$. The number of possible triples generated from those three subtrees is the product of the number of their leaves. By summing over all possible selections of three distinct subtrees one finally gets (1.1.1) and the lemma is proved. □

1.5.2 Basic Notions from Probability Theory

The following statements are well-known in the field of probability theory. The order of results is copied from Semple and Steel [2003, p.4]. Let $\mathcal{X}$ denote a sample space. Provided $\mathbb{P}(B) > 0$ the *conditional probability* of an event A given B is given by:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)}. \tag{1.5.1}$$

For the proofs of Section 1.2 the following elementary results are useful.

(i) (*Law of total probability*) If $B_1, \dots, B_k$ partition $\mathcal{X}$ and A is an event in $\mathcal{X}$, then:

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \tag{1.5.2}$$

(ii) (*Bayes' rule*) If A and B are events in $\mathcal{X}$ and $\mathbb{P}(A), \mathbb{P}(B) > 0$, then

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}. \tag{1.5.3}$$

The value of $\mathbb{P}(A)$ on the right hand side of (1.5.3) is often evaluated by the law of total probability.

(iii) (*The product rule*) If $A_1, \dots, A_k$ are events in $\mathcal{X}$, then

$$\mathbb{P}(A_1, \dots, A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1, A_2)\dots\mathbb{P}(A_k|A_{k-1}, \dots, A_1). \tag{1.5.4}$$

1.5.3 Proofs for Section 1.2

Section 1.2 presented Markov properties on trees.

Proof of Lemma 1.2.1. See for instance Lemma 5.2 in Dawid [1980]. □

Proof of Proposition 1.2.3. For statement 1 see Proposition 3.8 in Lauritzen [1996].

With statement 1 only the direction (G)⇒(F) must be shown to prove statement 2. This is done by ordering the vertices placing the leaves last, preceded by their immediate neighbors and so on, then applying the product rule (1.5.4) and finally use the global Markov property on the ensuing equation. The resulting factorization is of the form presented in (1.2.3) and therefore μ factorizes on $\mathcal{T}$.

Finally, statement 3 follows from Theorem 2.7 in Lauritzen [1996] together with statement 2. This completes the proof. □

Proof of Lemma 1.2.4. According to Proposition 1 in Matús [1992], equivalence of global and local Markov property is given if and only if the considered graph $\mathcal{G}$ has no subgraphs with exactly four vertices and exactly two parallel edges. Any subgraph with four vertices of trees with at most two inner vertices is either connected or contains at least one isolated point. Further assume that a tree has the inner vertices $\alpha_1, \alpha_2, \alpha_3$ and edges $⟪\alpha_1, \alpha_2⟫$ and $⟪\alpha_2, \alpha_3⟫$, and the leaf β_1 is adjacent to α_1 whereas the leaf β_3 is adjacent to α_3. Then the subgraph with vertices $\alpha_1, \alpha_3, \beta_1, \beta_3$ has exactly two parallel edges $⟪\alpha_1, \beta_1⟫$ and $⟪\alpha_3, \beta_3⟫$. This completes the proof. □

Proof of Proposition 1.2.5. The statement is an extension of Proposition 3.22 in Lauritzen [1996]. Let $\mathcal{T}_A = (A, \mathcal{E}_A)$ denote a restriction of $\mathcal{T} = (\mathcal{V}, \mathcal{E})$, and μ_A is the constraint of μ to $\mathcal{T}_A$. W.l.o.g., $\varrho \in A$. Then similar to (1.4.1) one has with Proposition 1.2.7 and (1.2.4) for $\underline{x} \in \mathcal{X}^A$:

$$\begin{aligned}\mu_A(\underline{x}) &= \sum_{\underline{y}\in\mathcal{X}^{\mathcal{V}\setminus A}} \mu(\underline{x},\underline{y}) = q^{\varrho}_{x_\varrho} \prod_{(\alpha,\beta)\in\mathcal{E}_A} P^{\alpha\beta}_{x_\beta,x_\alpha} \sum_{\underline{y}\in\mathcal{X}^{\mathcal{V}\setminus A}} \prod_{(\gamma,\delta)\in\mathcal{E}\setminus\mathcal{E}_A} P^{\gamma\delta}_{x_\delta,x_\gamma} \\ &= q^{\varrho}_{x_\varrho} \prod_{(\alpha,\beta)\in\mathcal{E}_A} P^{\alpha\beta}_{x_\beta,x_\alpha},\end{aligned}$$

since vertices of degree two vanish from the equation with

$$\sum_{x\in\mathcal{X}} P^{\alpha\mathrm{ch}(\alpha)}_{zx} P^{\alpha\mathrm{pa}(\alpha)}_{xy} = \mu_{\mathrm{ch}(\alpha)|\mathrm{pa}(\alpha)}(y,z), \quad y,z\in\mathcal{X}, \tag{1.5.5}$$

and $\sum_{x\in\mathcal{X}} P^{\alpha\beta}_{xy} = 1$ for all $y \in \mathcal{X}$. This completes the proof. □

Proof of Theorem 1.2.6. The equivalence chain (DF)⇔(DG)⇔(DL) is the statement of Theorem 3.27 in Lauritzen [1996]. (DL)⇒(DO) follows with (S2) and $\mathrm{hi}(\alpha) \subseteq \mathrm{nd}(\alpha)$ for all $\alpha \in \mathcal{V}$. (DO)⇒(DF) is proved similarly to Lemma 1.2.4 using (1.5.4) and applying (DO) on the result. This completes the proof. □

Proof of Proposition 1.2.7. This follows immediately from Proposition 3.28 in Lauritzen [1996] and Proposition 1.2.3.2. □

1.5.4 Proofs for Section 1.4

This subsection will verify the statements made concerning the reconstruction task.

Proof of Proposition 1.4.1. Statement 2 follows from (1.5.5), and statement 1 follows from Proposition 1.2.5. □

Chapter 2

Algebraic Geometry

The previous section described some stochastic features of the model. This section will provide algebraic tools to answer the following questions concerning the recovery of a Markov process through equation (LF):

1. When does a leaf distribution $\underline{m}$ has a solution, i.e. when does it have a Markov-like extension (see Def. 2.1.3)?
2. When is the number of solutions finite then?
3. How many solutions do exist for a given leaf distribution $\underline{m}$?

The answers presented in the following section will be of a general kind. In particular, for question 3 only a lower and an upper bound are presented. Moreover, note that only conditions for Markov-like extensions can be established with algebraic geometry. The structure of the section follows the stated questions. The notation and results provided here are mostly taken from Cox et al. [1997]. Conditions for Markov extensions on triple trees for the general two state, Neyman N_k and Kimura 2ST model are presented in later chapters.

2.1 Rewriting the Questions

This section will provide the general language of polynomials and varieties. At the end of it, the above questions will be restated in the notion of varieties.

Definition 2.1.1.(Def.s 1.1.1-3 in Cox et al. [1997]) *A* monomial *in $t_1, \ldots, t_r$ is a product of the form*

$$t_1^{a_1} \cdot t_2^{a_2} \cdots t_r^{a_r},$$

where all of the exponents $a_1, \ldots, a_r$ are nonnegative integers. The total degree $|a|$ *of this monomial is the sum $\alpha_1 + \cdots + \alpha_r$.*

A polynomial f *in* $t_1, \dots, t_r$ *with coefficients in* $\mathbb{C}$ *is a finite linear combination (with coefficients in* $\mathbb{C}$*) of monomials. A polynomial* f *is written in the form*

$$f = \sum_a c_a t^a, \quad c_a \in \mathbb{C},$$

where the sum is over a finite number of $r-$ *tuples* $a = (a_1, \dots, a_r)$*. The set of all polynomials in* $t_1, \dots, t_r$ *with coefficients in* $\mathbb{C}$ *is denoted* $\mathbb{C}[t_1, \dots, t_r]$*.*
The total degree *of a polynomial* f *in* $\mathbb{C}[t_1, \dots, t_r]$*, denoted* $\deg(f)$*, is the maximum* $|a|$ *such that the coefficient* c_a *is nonzero.*

Usually, some characteristics of polynomials are given by their zero points or roots. Therefore, a system of polynomials can be described by its joint roots. The set of joint roots is called a *variety*:

Definition 2.1.2.(Def. 1.2.1 in Cox et al. [1997]) *Let* $f_1, \dots, f_s$ *be polynomials in* $\mathbb{C}[t_1, \dots, t_r]$*. The set* $\mathbf{V}(f_1, \dots, f_s)$ *defined through*

$$\mathbf{V}(f_1, \dots, f_s) := \{(a_1, \dots, a_r) \in \mathbb{C}^r : f_i(a_1, \dots, a_r) = 0 \textit{ for all } 1 \le i \le s\}$$

is called the affine variety *defined by* $f_1, \dots, f_s$*.*

Thus an affine variety $\mathbf{V}(f_1, \dots, f_s) \subset \mathbb{C}^r$ is the set of all solutions of the system of equations $f_1(t_1, \dots, t_r) = \cdots = f_s(t_1, \dots, t_r) = 0$.
The first task is to apply the above notation to the system (LF). For a proper application assign the integers s and r with their representant from the Markov model. h denotes the number of polynomials. System (LF) has as many equations as the joint leaf distribution $\underline{m}$ has elements, i.e. $\sharp(\mathcal{S})^{\sharp(\mathcal{L})}$. Set $k := \sharp(\mathcal{S})$ and $n = \sharp(\mathcal{L})$. Denote the parameters on the right hand side of (LF) by $(p_1, \dots, p_r)$ where the ordering should be chosen appropriately. Usually, $r = (k-1) + k(k-1)\sharp(\mathcal{E})$, where $\sharp(\mathcal{E})$ is the number of edges in the tree $\mathcal{T}$, and $s = k^n$ is the number of polynomials. Thus, a suitable ordering could be given by assigning the first $k(k-1)$ parameters to the first edge, the second $k(k-1)$ parameters to the second and so on and the final $k-1$ parameters would stand for the root parameters. Clearly, for this example an ordering of the edges needs to be included. Finally, to transfer the problem of finding a solution to the system (LF) to finding the roots of an associated system assign a suitable ordering to the elements of the leaf distribution $\underline{m}$, i.e. $\underline{m} := (m_i)_{i=1}^s$. For instance if $\mathcal{S} := \{0, 1, \dots, k-1\}$ the ordering could look like

$$i = x_1 k^{n-1} + x_2 k^{n-1} + \cdots + x_{n-1} k + x_n + 1, \quad x_j \in \mathcal{S},\ j = 1, \dots, n.$$

With these conventions rewrite (LF) by $m_i = f_i(p_1, \dots, p_r)$, $i = 1, \dots, s$ and define

$$g_i(m_1, \dots, m_s, p_1, \dots, p_r) := m_i - f_i(p_1, \dots, p_r), \quad i = 1, \dots, s. \tag{2.1.1}$$

Denote by $W := \mathbf{V}(g_1, \dots, g_s)$ the variety of system (2.1.1). Then finding a solution of (LF) w.r.t. to a given joint leaf distribution $\underline{m}$ is equivalent to computing the intersection:

$$\mathfrak{S}_{\underline{m}} := W \cap \{z \in \mathbb{C}^{s+r} : \ z_i = m_i, \ i = 1, \dots, s\}.$$

With these notions the questions asked at the beginning of the section are translated into:

1. When is $\mathfrak{S}_{\underline{m}}$ nonempty?
2. If $\mathfrak{S}_{\underline{m}}$ is nonempty, is it finite?
3. If $\mathfrak{S}_{\underline{m}}$ is finite, what is its cardinality?

For the language of the next sections the following definition is necessary.

Definition 2.1.3. *Let $\underline{m}$ denote a leaf distribution on a tree $\mathcal{T}$. A solution of (LF) w.r.t. $\underline{m}$ is called* Markov-like extension*. If a solution is stochastically admissible, i.e. if the solution describes a set of transition matrices and a root distribution, it is called* Markov extension*.*

2.2 Existence of a Solution

This section will answer the first question. For the identification of leaf distributions that have a Markov-like extension this is the most important question. The answer provided here is commonly accepted and also discussed on numerous occasions (eg. Allman and Rhodes [2003] or Pachter and Sturmfels [2004]). To start the section, *ideals* are introduced. These are polynomials that have a certain set of zero points in common. In a way, this notion already provides an idea into which direction the answer is headed.

Definition 2.2.1.(Def.s 1.4.1+2 in Cox et al. [1997]) *A subset $I \subset \mathbb{C}[t_1, \dots, t_r]$ is an* ideal *if it satisfies*

1. *$0 \in I$.*
2. *If $f, g \in I$, then $f + g \in I$.*
3. *If $f \in I$ and $q \in \mathbb{C}[t_1, \dots, t_r]$, then $qf \in I$.*

For polynomials $f_1, \dots, f_s$ in $\mathbb{C}[t_1, \dots, t_r]$ set

$$\langle f_1, \dots, f_s \rangle := \left\{ \sum_{i=1}^{s} q_i f_i : \ q_1, \dots, q_s \in \mathbb{C}[t_1, \dots, t_r] \right\}. \tag{2.2.2}$$

Consider $\underline{x} \in \mathbb{C}^r$ such that $f(\underline{x}) = g(\underline{x}) = 0$. Then also $(f+g)(\underline{x}) = f(\underline{x}) + g(\underline{x}) = 0$ and $(qf)(\underline{x}) = q(\underline{x})f(\underline{x}) = 0$. Therefore, $\underline{x}$ is a zero point of all polynomials in the ideal I. Accordingly a variety $\mathbf{V}$ defines a unique ideal $I(\mathbf{V})$ by

$$I = \{f : \mathbb{C}^r \to \mathbb{C},\ f(\underline{x}) = 0 \text{ for all } \underline{x} \in \mathbf{V}\}.$$

Coming back to (2.2.2) one observes, that $I_f := \langle f_1, \ldots, f_s \rangle$ defines an ideal (cf. Lemma 1.4.3 in Cox et al. [1997]). Such an ideal has an elegant interpretation in terms of polynomial equations. Given $f_1, \ldots, f_s \in \mathbb{C}[t_1, \ldots, t_r]$ one gets the system of equations $f_1 = 0, \ldots, f_s = 0$. From these equations, one can derive others using basic algebraic operations. For example, if one multiplies the first equation by $q_1 \in \mathbb{C}[t_1, \ldots, t_r]$, the second by $q_2 \in \mathbb{C}[t_1, \ldots, t_r]$ etc. and then adds the resulting equations one obtains:

$$q_1 f_1 + q_2 f_2 + \cdots + q_s f_s = 0,$$

which is a consequence of the original system. Note that the left hand side of this equation is exactly an element of the ideal $\langle f_1, \ldots f_s \rangle$. Thus, one can think of $\langle f_1, \ldots f_s \rangle$ as consisting of all "polynomial consequences" of the equations $f_1 = f_2 = \cdots = f_s = 0$. $(f_1, \ldots, f_s)$ is called the *basis* of I_f.

Coming back to the task at hand, the ideal $\langle g_1, \ldots, g_s \rangle$ with g_i defined through (2.1.1) contains all polynomials whose roots are in $\mathbf{V}(g_1, \ldots, g_s)$. To answer question 1 one has to derive from this ideal another ideal $\hat{I} \subset \mathbb{C}[t_1, \ldots, t_s]$. Such an ideal $\hat{I}$ contains all implications of a Markov process for its leaf distribution. Due to this observation the elements of $\hat{I}$ are called *phylogenetic invariants* (cf. Allman and Rhodes [2003]). The next task is to propose a way to compute $\hat{I}$.

For this way, consider an equation system in $x_1, \ldots, x_s$ which has an infinite number of solutions. To compute a characterization of the solution space solve the *polynomial parametrization*:

$$\begin{aligned} x_1 &= f_1(t_1, \ldots, t_r), \\ \vdots\ &\ \ \vdots \\ x_s &= f_s(t_1, \ldots, t_r), \end{aligned} \tag{2.2.3}$$

where $f_1, \ldots, f_s$ are polynomials in $\mathbb{C}[t_1, \ldots, t_r]$.

Under this circumstance, (LF) can be seen as a parametrization of the subset of $\mathbb{C}^s$, that also contains all leaf distributions with Markov- and Markov-like extension. But it should be noted that the subset will be larger than the set of leaf distributions as the following example shows:

Example 2.2.1. In Section 3.1 the two state three leaves case is discussed in detail. There, the system (LF) has the form:

$$\begin{aligned}
m_{000} &= (1-p_0^\alpha)(1-p_0^\beta)(1-p_0^\gamma)q^\varrho + (1-q^\varrho)p_1^\alpha p_1^\beta p_1^\gamma,\\
m_{001} &= (1-p_0^\alpha)(1-p_0^\beta)p_0^\gamma q^\varrho + (1-q^\varrho)p_1^\alpha p_1^\beta(1-p_1^\gamma),\\
m_{010} &= (1-p_0^\alpha)p_0^\beta(1-p_0^\gamma)q^\varrho + (1-q^\varrho)p_1^\alpha(1-p_1^\beta)p_1^\gamma,\\
m_{011} &= (1-p_0^\alpha)p_0^\beta p_0^\gamma q^\varrho + (1-q^\varrho)p_1^\alpha(1-p_1^\beta)(1-p_1^\gamma),\\
m_{100} &= p_0^\alpha(1-p_0^\beta)(1-p_0^\gamma)q^\varrho + (1-q^\varrho)(1-p_1^\alpha)p_1^\beta p_1^\gamma,\\
m_{101} &= p_0^\alpha(1-p_0^\beta)p_0^\gamma q^\varrho + (1-q^\varrho)(1-p_1^\alpha)p_1^\beta(1-p_1^\gamma),\\
m_{110} &= p_0^\alpha p_0^\beta(1-p_0^\gamma)q^\varrho + (1-q^\varrho)(1-p_1^\alpha)(1-p_1^\beta)p_1^\gamma,\\
m_{111} &= p_0^\alpha p_0^\beta p_0^\gamma q^\varrho + (1-q^\varrho)(1-p_1^\alpha)(1-p_1^\beta)(1-p_1^\gamma).
\end{aligned}$$

This is a parametrization for the equation

$$m_{000} + m_{001} + m_{010} + m_{011} + m_{100} + m_{101} + m_{110} + m_{111} = 1, \tag{2.2.4}$$

Obviously, the set of vectors in $\mathbb{C}^8$ satisfying (2.2.4) is not restricted to the cube $[0,1]^8$. To restrict the vector space of solutions to the associated leaf distributions, one has to add the inequality, $m_{xyz} \geq 0$ for all $x, y, z \in \{0, 1\}$.

The above example gave a glimpse of the way question 1 will be answered. Equation (2.2.4) provides the basis to the associated ideal $\hat{I}$ in the two state three leaves case. However, for model specifications with more than just two states or more than three leaves one polynomial won't be enough. The following notion help identifying $\hat{I}$:

Definition 2.2.2.(Def. 3.1.1 in Cox et al. [1997]) *Given $I = \langle f_1, \dots f_s\rangle \subset \mathbb{C}[t_1, \dots, t_r]$, the sth* elimination ideal *I_s is the ideal of $\mathbb{C}[t_{s+1}, \dots, t_r]$ defined by*

$$I_s := I \cap \mathbb{C}[t_{s+1}, \dots, t_r].$$

In this notation the sought ideal $\hat{I}$ is given by:

$$\hat{I} := I_s = \langle g_1, \dots, g_s\rangle \cap \mathbb{C}[m_1, \dots, m_s].$$

The following statement will provide a method to generate the basis $g_1, \dots, g_s$ to this ideal:

Proposition 2.2.1.(Thm. 3.3.1 in Cox et al. [1997]) *Let $G : \mathbb{C}^r \to \mathbb{C}^s$ be a function determined by the polynomial parametrization (2.2.3). Further, let $I = \langle x_1 - f_1, \dots, x_s - f_s\rangle \subset \mathbb{C}[t_1, \dots, t_r, x_1, \dots, x_s]$ and let $I_s = I \cap \mathbb{C}[x_1, \dots, x_s]$ be the sth elimination ideal. Then $\mathbf{V}(I_s)$ is the smallest affine variety in $\mathbb{C}^s$ containing $G(\mathbb{C}^r)$.* □

This method is more or less the reversal of the polynomial parametrization and is called *polynomial implicitization.* With the introduction of such a method question 1 has the following answer:

Theorem 2.2.2. *If the polynomial system (LF) has a solution w.r.t. a vector* $\underline{m}$*, then* $\underline{m}$ *is an element of the variety* $\mathbf{V}(\hat{I})$*.*

Hence, $\underline{m} \in \mathbf{V}(\hat{I})$ is necessary to have an non-empty set $\mathfrak{S}_{\underline{m}}$. Explicitly computing a basis for $\hat{I}$ is no easy task. Apart from the two state model where Example 2.2.1 implicitly shows the generation of the basis (2.2.4) one will find it exceedingly hard to do it by hand. Hence, computational support is needed. The author used two software packages, **Mathematica** (Wolfram [2003]) and **Singular** (Greuel et al. [2001]) for this purpose. The results will be discussed in Section 2.5.

2.3 Finitely Many Solutions

The next question asked for conditions on a leaf distribution $\underline{m}$ under which $\mathfrak{S}_{\underline{m}}$ has finitely many elements. Recall $r = (k-1)(k\sharp(\mathcal{E})+1)$ and $s = k^n$.

Theorem 2.3.1. *Let* $F := (f_1, \ldots, f_s) : \mathbb{C}^r \to \mathbb{C}^s$ *denote the polynomial equation system (LF). Then, a* $\underline{p}_0 \in \mathbb{C}^r$ *exists with* $\mathrm{rk}(\mathrm{D}F(\underline{p}_0)) = r$*.*
Moreover, let $\{i_1, \ldots, i_r\} \subset \{1, \ldots, s\}$ *denote a set of indices, such that* $(\mathrm{D}f_{i_j}(\underline{p}_0))_{j=1}^r$ *is a family of linearly independent vectors. The set*

$$M_0 = \big\{\underline{n} : \exists \underline{m} \in \mathbb{C}^s \text{ with } m_{i_j} = n_j,\ j = 1, \ldots, r \text{ and } \sharp(\{\underline{p} : F(\underline{p}) = \underline{m}\}) = \infty\big\}$$

is a Lebesgue zero set in $F(\mathbb{C}^r)$*.*

The complement of M_0 implies that the set $\{\underline{m} \in \mathbb{C}^s : \exists \underline{n} \in M_0^c :$ with $m_{i_j} = n_j,\ j = 1, \ldots, r\}$ contains all vectors $\underline{m} \in \mathbb{C}^s$ with $\sharp(\{\underline{p} \in \mathbb{C}^r : f(\underline{p}) = \underline{m}\}) < \infty$, and since M_0 is a Lebesgue zero set, almost all vectors $\underline{m} \in \mathbb{C}^s$ have a finite number of solutions.

With $\mathrm{rk}(\mathrm{D}F(\underline{p}_0)) = r$ another statement is connected, namely:

Lemma 2.3.2. *The number of phylogenetic invariants is bounded from below by* $s - r$*.*

This is obvious, since $f(\mathbb{C}^r) \subsetneq \mathbb{C}^s$ yields that the image space of f must be described by at least $s - r$ additional equations, the phylogenetic invariants. Table 2.1 offers a selected number of leaves and states.

Table 2.1 shows that even for the small system of four leaves and four states the minimal number of phylogenetic invariants is almost 200. Section 2.5 will provide

n	k	s	r	$s-r$
3	2	8	7	1
3	3	27	20	7
3	4	64	39	25
4	2	16	11	5
4	3	81	32	49
4	4	256	63	193

Table 2.1: The number of equations and variables for binary trees with three or four leaves in two, three or four states. The last column presents the lower bound for the number of phylogenetic invariants needed to identify a leaf distribution with Markov-like extension.

some more concern when it comes to the number of invariants.

2.4 The Number of Solutions

Finally, question 3 is considered, which asks for the cardinality of $\mathfrak{S}_{\underline{m}}$. Generally, this question cannot be answered exactly. However, there are some ways to obtain lower and upper bounds for the number of solutions. First, consider the following example:

Example 2.4.1. Consider the Markov model with four states and four leaves. Further, assume that the underlying tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ has the following structure:

$$\begin{aligned}\mathcal{V} &= \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \varrho_1, \varrho_2\},\\ \mathcal{E} &= \{(\varrho_1, \alpha_1), (\varrho_1, \alpha_2), (\varrho_1, \varrho_2), (\varrho_2, \alpha_3), (\varrho_2, \alpha_4)\}.\end{aligned}$$

The states in the leaves are distributed according to the quartet leaf distribution $\underline{m}$. Equation (LF) has the following form:

$$m(x_1, x_2, x_3, x_4) = \sum_{x_5=1}^{4} \mu_{\varrho_1}(x_5) P^{\alpha_1}_{x_5 x_1} P^{\alpha_2}_{x_5 x_2} \sum_{x_6=1}^{4} P^{\varrho_2}_{x_5 x_6} P^{\alpha_3}_{x_6 x_3} P^{\alpha_4}_{x_6 x_4}.$$

If π denotes a permutation mapping of the set $\{1, 2, 3, 4\}$ the following equation holds too:

$$(2.4.5)\qquad m(x_1, x_2, x_3, x_4) = \sum_{x_5=1}^{4} \mu_{\varrho_1}(\pi(x_5)) P^{\alpha_1}_{\pi(x_5) x_1} P^{\alpha_2}_{\pi(x_5) x_2} \sum_{x_6=1}^{4} P^{\varrho_2}_{\pi(x_5) x_6} P^{\alpha_3}_{x_6 x_3} P^{\alpha_4}_{x_6 x_4}.$$

The same manipulation works with a permutation of the states in ϱ_2. These permutations result in column and row permutations of the transition matrices thus providing alternative solutions. Hence, one solution generates $(k!)^2 = 576$ other solutions within this model, which are for generic $\underline{m}$ different from each other.

Generally, one can prove the following result:

Lemma 2.4.1. *Let μ be a Markov process on a rooted tree $\mathcal{T}_\varrho := (\mathcal{V}, \mathcal{E}; \varrho)$ with leaf set $\mathcal{L}$ and set $\mathcal{N}$ of inner vertices. Further, let $(P^\alpha)_{\alpha\in\mathcal{V}\setminus\{\varrho\}}$ and μ_ϱ denote a parametrization of μ and let $\Pi_\mathcal{S}$ denote the set of permutations of the states in $\mathcal{S}$. Then, modifying the parameters for an inner vertex $\alpha \in \mathcal{N}$, $\alpha \neq \varrho$, by*

$$\hat{P}^\alpha_{xy} = P^\alpha_{x\pi(y)}, \quad \hat{P}^\beta_{xy} = P^\beta_{\pi(x)y}, \quad \beta \in \mathrm{ch}(\alpha),\, x, y \in \mathcal{S}, \pi \in \Pi_\mathcal{S} \tag{2.4.6}$$

yields a parametrization for a Markov process with the same leaf distribution as μ.

For $\alpha = \varrho$ one can modify the parameters by

$$\hat{\mu}^\varrho_y = \mu^\varrho_{\pi(y)}, \quad \hat{P}^\beta_{xy} = P^\beta_{\pi(x)y}, \quad \beta \in \mathrm{ch}(\alpha),\, x, y \in \mathcal{S}, \pi \in \Pi_\mathcal{S} \tag{2.4.7}$$

to obtain a Markov process with the same leaf distribution as μ.

Thus there are $(k!)^{\sharp(\mathcal{N})}$ such alternative parameterizations for the leaf distributions μ.

Hence, a unique solution exists in general only *up to permutation* of inner states. The number $(k!)^{\sharp(\mathcal{N})}$ is also a lower bound to the possible number of solutions, i.e. a lower bound to the cardinality of $\mathfrak{S}_{\underline{m}}$, at least if there is one solution where all probabilities in μ^ϱ and all rows of the transition matrices P^e, $e \in \mathcal{E}$, are different. In Chang [1996] uniqueness was established by restricting the transition matrices to a reconstruction argument, i.e. by denying the choice of permutation matrices. This is a useful assumption for a couple of reasons. Firstly, the number of possible permutations becomes large the more states one assumes and secondly, phylogenetic inference usually assumes small step change, i.e. the considered transition matrices are diagonally dominant. To keep this observation, *reconstructible classes* are introduced. If a matrix A is in such an reconstructible class, depending on the particular definition certain permutations of A cannot be in this class. Next, a look at the three models considered in this thesis and the reason why the result was restricted to general Markov processes.

Example 2.4.2. According to Lemma 2.4.1 the general two state model on three leaves has at least two solutions. Section 3.1 will verify this observation.

The proposed permutations in (2.4.6) are translatable into row or column permutations for the transition matrices. For a symmetric transition matrix P^α this kind of permutation ends in a non-symmetric matrix $\hat{P}^\alpha$, i.e. the alternative solution leaves the model, although in terms of (LF) the same leaf distribution is recovered. Hence, the permutation approach leaves a symmetric models, if the state space $\mathcal{S}$ consists of more than two states. However, for the Kimura 2ST model one can observe a "freak" permutation. Looking at (1.3.1) one observes that swapping row one with row two and row three with row four yields a matrix that is also has the form of (1.3.1). Hence, the lower bound for solutions for the Kimura 2ST model is two. For a closer look at the subject of unique solutions for the particular models see Lemma 3.1.2, Proposition 4.1.6 and Proposition 4.2.4. Latter propositions will provide an additional kind of symmetry. However, these symmetries provide a solution that is subject to another leaf distribution $\widehat{\mathbf{m}}$.

Upper bounds are provided by *Bezout's Theorem* (Theorem 8.7.10 in Cox et al. [1997]). It bounds the number of possible solutions by the product of the total degrees of the associated equations of the system. The total degree of any polynomial in (LF) is $\sharp(\mathcal{E})+1$, since at least one monomial in such a polynomial contains one probability for each edge plus a probability for the root. Together with the number of equations this yields the upper bound $(\sharp(\mathcal{E})+1)^s$. This observation yields the following result:

Corollary 2.4.2. *Let $g_1, \dots, g_s$ denote the polynomials from (2.1.1). Then,* $\deg(g_i) = \sharp(\mathcal{E})+1$ *for all $i = 1, \dots, s$, and if a vector $\underline{m}$ generates a finite number of solutions for (LF), the number of different unique solutions up to permutation for reconstructible classes of transition matrices is bounded from above by:*

$$\frac{(\sharp(\mathcal{E})+1)^s}{(k!)^{\sharp(\mathcal{N})}}. \tag{2.4.8}$$

Theorem 2.3.1 states that with the exception of a zero set of vectors all vectors provide a finite number of solutions for (LF). $(\sharp(\mathcal{E})+1)^s$ contains all symmetric solutions, and hence by dividing this number by the number of symmetrical solutions, one gets an upper bound to the number of unique solutions. Unfortunately, when looking at the case $n = 3$, $k = 2$ one has $\sharp(\mathcal{E}) = 3$ and $\sharp(\mathcal{N}) = 1$, and thus this number is 32768. Clearly, this number is much to high and the only insight one can derive from this proposition is that the number of solutions is finite.

Bernstein's Theorem (Theorem 1 in Huber and Sturmfels [1997]) provides a complicated way of computing a (possibly better) upper bound to the number of solutions. However, this approach will not be discussed here.

2.5 Discussion

The previous section provided some very interesting theoretical results. However, their applicability should be considered. In particular, if one tries to obtain a basis to the proposed ideal $\hat{I}$ of phylogenetic invariants to the factorization property (LF). This section reports the hazards the author faced when tackling the problem.

Various algebraic softwares have incorporated an elimination algorithm that provides for the polynomial implicitization. Most are based on the computation of a so-called *Gröbner basis*, which basically is the preferred ideal basis in the field due to the following facts. Any polynomial $f \in \mathbb{C}[t_1, \dots, t_r]$ is divisible into

$$f = q_1 f_1 + \cdots + q_s f_s + u,$$

if $f_1, \dots, f_s$ is a basis to an ideal I, $q_1, \dots, q_s \in \mathbb{C}[t_1, \dots, t_r]$ and a remainder $u \in \mathbb{C}[t_1, \dots, t_r]$ which is not divisible by $f_1, \dots, f_s$. Usually this remainder is not unique depending on the order of the basis polynomials. But for Gröbner bases the remainder is unique (Prop. 2.6.1 in Cox et al. [1997]). Moreover, every ideal $I \neq \emptyset$ has a Gröbner basis (Coro. 2.5.6 in Cox et al. [1997]) and for a given ideal I with Gröbner basis G the set $G_s := G \cap \mathbb{C}[t_{s+1}, \dots, t_r]$ is a Gröbner basis of the sth elimination ideal I_s (Thm. 3.1.2 in Cox et al. [1997]).

One software package that provides the elimination ideal through Gröbner basis computation is the already mentioned **Mathematica** . However, as Cox et al. [1997, page 114] propose, this is not always a useful approach:

> ... In some cases (such as the implicitization problem to be studied in §3), we only want to eliminate certain variables, and we do not care about the others. In such a situation, it is a bit of overkill to compute a Groebner basis with lex order. This is especially true since lex order can lead to some very unpleasant Groebner bases...

To underline this statement consider the Kimura 2ST model. According to Theorem 2.3.1 at least four polynomials are needed for the generation of $\hat{I}$ in that case. **Mathematica** computed several hours and produced 24 polynomials that filled more than 200 A4 pages of output. Clearly, the usefulness of such a result is disputable.

Singular on the other hand is a software package solely made for the purpose of algebraic geometry. Its function for deriving a basis of an elimination ideal is much faster and provides more suitable results. For the Kimura 2ST model **Singular** produced within minutes 18 polynomials with about 24 A4 pages of output. Hence, this result is much better in quantitative and interpretational terms, although the overall benefit is still in doubt. Who likes to leaf through 24 pages of polynomials?

Also, w.r.t. Table 2.1 one has to realize that the derivation of a proper polynomial basis for the ideal of phylogenetic invariants will become more difficult and their interpretation even more questionable. Hence, an identification of phylogenetic invariants with a meaningful interpretation (as suggested by Allman and Rhodes [2003]) could be much more beneficial than the knowledge of the whole basis without an interpretation. In Section 3.3 some phylogenetic invariants for the extension of a

Markov process from the triple trees to quartet trees under the two state model are presented.

2.6 Proofs

Finally, the proofs to the presented results are provided. The chapter revolved around three questions. The first question asked for conditions for the existence of an algebraic solution of (LF), and was answered by Theorem 2.2.2. For the proof of this result consider the following helpful statements:

Proposition 2.6.1.(Prop. 1.4.8 in Cox et al. [1997]) *Let V and W be affine varieties in $\mathbb{C}^n$. Then:*

1. *$V \subset W$ if and only if $I(V) \supset I(W)$.*
2. *$V = W$ if and only if $I(V) = I(W)$.*

□

In other words, decreasing the cardinality of a variety provides more polynomials with the same roots. The next result is very helpful when trying to identify a certain variety. If an established basis is unsuitable for deriving certain properties a base change is possible.

Lemma 2.6.2.(Prop. 1.4.4 in Cox et al. [1997]) *If $f_1, \ldots, f_s$ and $g_1, \ldots, g_s$ are bases of the same ideal in $\mathbb{C}[t_1, \ldots, t_r]$, so that $\langle f_1, \ldots f_m\rangle = \langle g_1, \ldots, g_s\rangle$, then $\mathbf{V}(f_1, \ldots, f_s) = \mathbf{V}(g_1, \ldots, g_s)$.* □

With these results the proof for the answer to question 1 is straight forward:

Proof of Theorem 2.2.2. The statement follows immediately from the previously made statements and from Proposition 2.2.1. □

So far for the first question. For the next question asked for conditions for a finite number of solutions. Varieties are called *irreducible* if they cannot be decomposed into subvarieties. For instance, points, lines and planes are irreducible varieties (see e.g. §.5 in Cox et al. [1997]). For the dimension of a variety refer to Chapter 9 in Cox et al. [1997]. A mapping $f : X \to Y$ of irreducible varieties is called regular, if for $\underline{x} \in X$ polynomials $f_1, \ldots, f_s$, $s = \dim(Y)$ exist with $f(\underline{x}) = (f_1(\underline{x}), \ldots, f_s(\underline{x}))$. For the proof of the answer present in Theorem 2.3.1 consider the following results:

Proposition 2.6.3.(Theorem I.6.7 in Shafarevich [1974]) *If $f : X \to Y$ is a regular mapping of irreducible varieties, $f(X) = Y$, $\dim(X) = r$, $\dim(Y) = s$, then $s \leq n$ and*

1. $\dim f^{-1}(\underline{y}) \geq n - s$ *for every point* $\underline{y} \in Y$*;*
2. *in* Y *exists a non-empty open set* U *such that* $\dim f^{-1}(\underline{y}) = r - s$ *for* $\underline{y} \in U$*.*

□

The next result provides a way to compute the exact dimension of an irreducible variety.

Proposition 2.6.4.(Theorem II.1.3 in Shafarevich [1974]) *The dimension of the tangent space at a single point* $\underline{x} \in \mathbb{Q}^n$ *is equal to the dimension of the (irreducible) variety.* □

In principle the selection of the rational point includes that the dimension is minimal. Other points with higher dimension could exist but these form a sparse set. With this information consider Theorem 2.3.1.

Proof of Theorem 2.3.1. The first statement follows when evaluating the functional matrix DF at a rational point $x \in \mathbb{C}^r$. With Proposition 2.6.3.2 and Proposition 2.6.4 such a point exists.

F is a polynomial mapping and therefore an infinitely differentiable. Then according to the Morse-Sard Theorem (see e.g. Thm. 1.3 in Hirsch [1976]) the set M_0 is a Lebesgue zero set in $F(\mathbb{C}^8)$. This completes the proof. □

Proof of Lemma 2.3.2. This is a straight forward statement. With Proposition 2.6.3.1 the dimension of the vector space of all vectors $\underline{m}$ with a solution for (LF) is at most r. Since $\underline{m} \in \mathbb{C}^s$ one needs at least $s - r$ conditions to reduce the space. Here, these conditions are the phylogenetic invariants. □

So far for question 2. It remains to consider the statements made in connection with question 3.

Proof of Lemma 2.4.1. The first statement declares that a permutation of state probabilities in any inner vertex, including the root, does not alter the Markov process. Equation (2.4.5) already is the proof of this statement because such a permutation results only in a permutation of summands which leaves the left hand side invariant.

The second statement on the number of possible permutations is easily computed. Apparently there are $k!$ possible permutations per inner vertex and $\sharp(\mathcal{N})$ different inner vertices. Since a state permutation in one vertex is independent of the state in another vertex, the overall number of permutations equals the product, i.e. $(k!)^{\sharp(\mathcal{N})}$ and thus the lemma is proved. □

Proof of Corollary 2.4.2. According to Bezout's Lemma the number of possible solutions of a polynomial equation system is bounded from above by the product of the total degree of the equations. Recall (LF):

$$m_{\underline{x}} = \sum_{\underline{y}, \underline{y}|_{\mathcal{L}} = \underline{x}} q^{\varrho}_{y_\varrho} \prod_{(\alpha,\beta)\in\mathcal{E}} P^{\alpha\beta}_{x_\beta x_\alpha}.$$

Clearly, for every $\underline{x}$ one finds at least one monomial of degree $\sharp(\mathcal{E})+1$, i.e. every polynomial in (LF) has total degree of $\sharp(\mathcal{E})+1$. The number of polynomials equals $s = k^n$, and therefore the number of solutions of (LF) is bounded from above by $(\sharp(\mathcal{E})+1)^s$. When looking at reconstructible classes one has to deny all solution that can be obtained by permutation. Therefore, with Lemma 2.4.1 the number of solutions in an reconstructible class is bounded from above by $(\sharp(\mathcal{E})+1)^s/(k!)^{\sharp(\mathcal{N})}$. This completes the proof. □

Chapter 3

Stochastic Models of Molecular Evolution in Two States

This chapter examines the extendability of a leaf distribution to a Markov distribution on a triple tree under the general two state model. The molecular significance of the model is sparse but existent. As already mentioned in Example 1.3.1, the examination of the evolution of sequences in two states can be equally interesting as the analysis of the evolution of nucleotide sequences.

The model also has its applications in other fields. Most notably, Lazarfeld [1966] used it to produce and interpret decision or correlation trees for psychological tests, and Pearl and Tarsi [1986] applied it to certain features in the field of artificial intelligence. Those papers proved that in the generic case the model equations have a unique algebraic solution using the parametrization approach (cf. Cox et al. [1997, § 1.3]) but didn't present a closed form of this solution.

Chapter 2 introduced phylogenetic invariants as an tool to test if a leaf distribution $\underline{m}$ has an algebraic solution for (LF), i.e. if $\underline{m}$ has a Markov-like extension to the underlying tree. As Example 2.2.1 indicates, the two state model on triple tree only has the invariant (2.2.4), which demands that the elements of the vector $\underline{m}$ sum to one.

This chapter will use brute force to compute a complete solution of the system. From these computations conditions for existence and uniqueness are derived. In addition, the degenerate cases are considered, and more importantly, conditions for the existence of a stochastically admissible solution are established.

Following this analysis the results are extended to quartet trees. To achieve that, one quartet tree is fixed and the specifications of its inferred triple trees are compared. This attempt returns some phylogenetic invariants.

For a closer insight into the results their implications on the symmetrical model are discussed. Note, that symmetrical in that case means that the transition matrix for every edge w.r.t. this model is symmetrical.

3.1 The General Two State Case on Triple Trees

This section presents a complete analysis of the two state three leaves specification of the Markov model of molecular evolution. For this purpose let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denote a triple tree with $\mathcal{V} := \{\alpha, \beta, \gamma, \varrho\}$ and $\mathcal{E} := \{(\varrho, \alpha), (\varrho, \beta), (\varrho, \gamma)\}$. A joint distribution on the leaf set $\mathcal{L} := \{\alpha, \beta, \gamma\}$ is denoted by $\underline{m} := (m_{xyz})_{x,y,z \in \{0,1\}}$. As said the task is to identify a Markov extension $\underline{\mu} := (\mu_{\mathcal{V}}(u, x, y, z))_{u,x,y,z \in \{0,1\}}$ to a given leaf distribution $\underline{m} := \underline{\mu_{\mathcal{L}}}$ with $m_{xyz} = \mu|_{\mathcal{L}}(x, y, z)$ for $x, y, z \in \{0, 1\}$. The section will present conditions under which such an extension exists and give an explicit characterization of it. This is done by first solving the induced system (3.1.1) algebraically and then computing conditions under which the generated terms are probabilities. To allow for an algebraic solution, $\underline{m}$ has to fulfil certain conditions. These conditions are presented and the case of their violation is discussed.

3.1.1 Basic Model Properties

According to equation (1.2.4) a Markov distribution on the triple tree $\mathcal{T}$ is characterized by a root distribution $(q^{\varrho}_{w})_{w \in \{0,1\}} := (\mu_{\varrho}(w))_{w \in \{0,1\}}$ and a family of transition kernels $(p^{\delta})_{\delta \in \mathcal{L}}$ with $p^{\delta}_{uw} = \mu(X_{\delta} = u | X_{\varrho} = w)$ for $u, w \in \{0, 1\}$. Equation (LF) provides the starting point of this chapter, namely the equation system

$$(3.1.1) \qquad m_{xyz} = \mu_{0xyz} + \mu_{1xyz} = q^{\varrho}_{0} p^{\alpha}_{0x} p^{\beta}_{0y} p^{\gamma}_{0z} + q^{\varrho}_{1} p^{\alpha}_{1x} p^{\beta}_{1y} p^{\gamma}_{1z}, \quad x, y, z \in \{0, 1\}.$$

Equation (3.1.1) yields eight equations in seven variables. With Proposition 2.6.3 at least one phylogenetic invariant is needed. This invariant is given through

Lemma 3.1.1. *Let $\underline{m}$ denote a leaf distribution. If system (3.1.1) has an algebraic solution w.r.t. $\underline{m}$, then*

$$(3.1.2) \qquad \sum_{x,y,z \in \{0,1\}} m_{xyz} = 1$$

is satisfied.

Apparently, (3.1.2) is a defining property for any distribution in eight states. Hence, this invariant does not restrict the set leaf distributions with an algebraic extension. This provides the possibility to compute a solution by considering only seven of the eight equations. The following definition will give a more precise description of solutions.

Definition 3.1.1. *An* algebraic solution *of (3.1.1) is composed of a vector* $q^\varrho \in \mathbb{C}^2$ *and a set of matrices* $p^\delta \in \mathbb{C}^{2\times 2}$, $\delta \in \mathcal{L}$ *with the constraints*

$$(3.1.3) \qquad q_0^\varrho + q_1^\varrho = 1 \quad \textit{and} \quad p_{w0}^\delta + p_{w1}^\delta = 1 \quad \textit{for } \delta \in \mathcal{L},\ w \in \{0,1\}.$$

A stochastically admissible solution *is an algebraic solution with* $q^\varrho \in [0,1]^2$ *and* $p^\delta \in [0,1]^{2\times 2}$, $\delta \in \mathcal{L}$.

Clearly, the existence of a Markov extension is equivalent to the existence of a stochastically admissible solution to (3.1.1). Due to (3.1.3) a solution is determined by the root probability q_0^ϱ and one column of the transition matrices for each leaf, namely p_{00}^δ, p_{10}^δ, $\delta \in \mathcal{L}$.

Equation (3.1.1) contains certain symmetries which have influence on the uniqueness of solutions.

Lemma 3.1.2. *Let* $\underline{m}$ *denote a leaf distribution and let* q_0^ϱ, p_{00}^δ, p_{10}^δ, $\delta \in \mathcal{L}$ *identify an algebraic solution of (3.1.1) w.r.t.* $\underline{m}$.

1. *Let* $\pi : \mathcal{L} \to \mathcal{L}$ *denote a permutation mapping on the leaves. Then, the parameters* q_0^ϱ, $\hat{p}_{00}^{\pi(\delta)}$, $\hat{p}_{10}^{\pi(\delta)}$, $\delta \in \mathcal{L}$ *with* $\hat{p}_{w0}^{\pi(\delta)} = p_{w0}^\delta$, $w \in \{0,1\}$ *identify an algebraic solution of (3.1.1) w.r.t.* $\pi(\underline{m})$, *i.e. the permutation of the vector elements of* $\underline{m}$ *consistent with* π.
2. *The parameters* $\hat{q}_0^\varrho$, $\hat{p}_{00}^\delta$, $\hat{p}_{10}^\delta$, $\delta \in \mathcal{L}$ *with* $\hat{q}_0^\varrho = q_1^\varrho$, $\hat{p}_{w0}^\delta = p_{(1-w)0}^\delta$, $w \in \{0,1\}, \delta \in \mathcal{L}$ *identify a solution of (3.1.1) w.r.t.* $\underline{m}$.

The first statement shows that a permutation of the leaf labels results in a permutation of the state probabilities in $\underline{m}$ but retains the structure of the process. This observation is valid for all star trees. The second statement claims that a permutation of the root state probabilities implies a permutation of the rows of the transition matrices but preserves the structure of the leaf distribution $\underline{m}$. Hence, a solution of (3.1.1) w.r.t. $\underline{m}$ always identifies a number of alternative solutions. Therefore, if a solution exists, it can be *unique up to symmetry* only. This should be kept in mind when encountering the phrase *unique* in this chapter.

3.1.2 The Algebraic Solution

This section presents the algebraic solution to system (3.1.1), the conditions for its existence and its closed form. The observations of this section establish that almost always a unique solution up to symmetry exists and that the obtained conditions are quite intuitive.

To describe the solution some further abbreviations are needed. First, applying (3.1.3) yields pairwise leaf probabilities and root leaf probabilities ($x, y \in \{0, 1\}$):

$$m_{xy\Sigma} := m_{xy0} + m_{xy1} = q_0^\varrho\, p_{0x}^\alpha p_{0y}^\beta + q_1^\varrho\, p_{1x}^\alpha p_{1y}^\beta, \tag{3.1.4}$$

$$m_{x\Sigma\Sigma} := m_{x00} + m_{x01} + m_{x10} + m_{x11} = q_0^\varrho\, p_{0x}^\alpha + q_1^\varrho\, p_{1x}^\alpha. \tag{3.1.5}$$

The probabilities $m_{x\Sigma z}$, $m_{\Sigma yz}$, $m_{\Sigma y\Sigma}$ and $m_{\Sigma\Sigma z}$ are computed accordingly. Further abbreviations are needed (again $x, y, z \in \{0, 1\}$)

$$\begin{aligned} r_{xyz}^\alpha &:= m_{xy\Sigma} m_{\Sigma\Sigma z} + m_{x\Sigma z} m_{\Sigma y\Sigma} - m_{\Sigma yz} m_{x\Sigma\Sigma} - m_{xyz}, \\ r_{xyz}^\beta &:= m_{xy\Sigma} m_{\Sigma\Sigma z} + m_{\Sigma yz} m_{x\Sigma\Sigma} - m_{x\Sigma z} m_{\Sigma y\Sigma} - m_{xyz}, \\ r_{xyz}^\gamma &:= m_{x\Sigma z} m_{\Sigma y\Sigma} + m_{\Sigma yz} m_{x\Sigma\Sigma} - m_{xy\Sigma} m_{\Sigma\Sigma z} - m_{xyz}. \end{aligned} \tag{3.1.6}$$

and

$$\begin{aligned} s_{xyz}^{\beta\gamma} &:= m_{xyz} m_{x\Sigma\Sigma} - m_{xy\Sigma} m_{x\Sigma z}, & t_{yz}^{\beta\gamma} &:= m_{\Sigma yz} - m_{\Sigma y\Sigma} m_{\Sigma\Sigma z}, \\ s_{xyz}^{\alpha\gamma} &:= m_{xyz} m_{\Sigma y\Sigma} - m_{xy\Sigma} m_{\Sigma yz}, & t_{xz}^{\alpha\gamma} &:= m_{x\Sigma z} - m_{x\Sigma\Sigma} m_{\Sigma\Sigma z}, \\ s_{xyz}^{\alpha\beta} &:= m_{xyz} m_{\Sigma\Sigma z} - m_{x\Sigma z} m_{\Sigma yz}, & t_{xy}^{\alpha\beta} &:= m_{xy\Sigma} - m_{x\Sigma\Sigma} m_{\Sigma y\Sigma}. \end{aligned} \tag{3.1.7}$$

$t_{11}^{\alpha\beta}$ is equivalent to the covariance between X_α and X_β, whereas $s_{11z}^{\alpha\beta}$ is the conditional covariance between X_α and X_β given X_γ has value $z \in \{0, 1\}$ up to a scalar. The other terms have a similar relevance. The product $t_{xy}^{\alpha\beta} t_{xz}^{\alpha\gamma} t_{yz}^{\beta\gamma}$ is of particular interest.

Lemma 3.1.3. *Let $\underline{m}$ denote a joint leaf distribution on $\mathcal{T}$. Suppose for some $x, y, z \in \{0, 1\}$ that $t_{xy}^{\alpha\beta} t_{xz}^{\alpha\gamma} t_{yz}^{\beta\gamma} = 0$. Then also*

$$t_{(1-x)y}^{\alpha\beta} t_{(1-x)z}^{\alpha\gamma} t_{yz}^{\beta\gamma} = 0, \quad t_{x(1-y)}^{\alpha\beta} t_{xz}^{\alpha\gamma} t_{(1-y)z}^{\beta\gamma} = 0, \quad t_{xy}^{\alpha\beta} t_{x(1-z)}^{\alpha\gamma} t_{y(1-z)}^{\beta\gamma} = 0. \tag{3.1.8}$$

According to Lemma 3.1.3, if one product vanishes all products of type (3.1.8) vanish. This property simplifies some proofs. To state the conditions for uniqueness define $\chi_{xyz} := r_{xyz}^\alpha - 4 s_{xyz}^{\beta\gamma} t_{yz}^{\beta\gamma}$.

Theorem 3.1.4. *Let $\underline{m} = (m_{xyz})_{x,y,z\in\{0,1\}}$ denote a joint leaf distribution on the triple tree $\mathcal{T}$. Assume further*

$$t_{00}^{\alpha\beta} t_{00}^{\alpha\gamma} t_{00}^{\beta\gamma} \neq 0 \textit{ and } \chi_{000} \neq 0. \tag{3.1.9}$$

Then the system (3.1.1) has a unique algebraic solution up to symmetry. The fol-

lowing expressions describe this solution:

$$q_0^{\varrho} = \frac{1}{2} + \frac{r_{000}^{\alpha} + 2m_{0\Sigma\Sigma}t_{00}^{\beta\gamma}}{2\sqrt{\chi_{000}}}, \tag{3.1.10}$$

$$p_{00}^{\alpha} = -\frac{r_{000}^{\alpha} - \sqrt{\chi_{000}}}{2t_{00}^{\beta\gamma}}, \quad p_{00}^{\beta} = -\frac{r_{000}^{\beta} - \sqrt{\chi_{000}}}{2t_{00}^{\alpha\gamma}}, \quad p_{00}^{\gamma} = -\frac{r_{000}^{\gamma} - \sqrt{\chi_{000}}}{2t_{00}^{\alpha\beta}}, \tag{3.1.11}$$

$$p_{10}^{\alpha} = -\frac{r_{000}^{\alpha} + \sqrt{\chi_{000}}}{2t_{00}^{\beta\gamma}}, \quad p_{10}^{\beta} = -\frac{r_{000}^{\beta} + \sqrt{\chi_{000}}}{2t_{00}^{\alpha\gamma}}, \quad p_{10}^{\gamma} = -\frac{r_{000}^{\gamma} + \sqrt{\chi_{000}}}{2t_{00}^{\alpha\beta}}. \tag{3.1.12}$$

Recalling for $\delta_1, \delta_2 \in \mathcal{L}$, that $t^{\delta_1\delta_2}$ denotes the covariance between X_{δ_1} and X_{δ_2}, condition (3.1.9) demands that there is stochastic dependence between the leaves. This is quite intuitive because independence would suggest that there is no connection between the leaves, i.e. there is no tree to the given species.

3.1.3 Characterization of Markov Extensions under the Two State Model

After establishing an algebraic solution it is useful to check whether the solution is consistent with the model. A stochastic model of molecular evolution is clearly described in terms of probabilities instead of general complex numbers. The solution presented in Theorem 3.1.4 is not necessarily stochastically admissible, as the following example shows.

Example 3.1.1. Consider the following artificial leaf distribution for the two state model:

$$\underline{m} = (164, 189, 41, 151, 165, 25, 141, 124)/1000$$

This vector satisfies condition (3.1.9) since

$$t_{00}^{\alpha\beta}t_{00}^{\alpha\gamma}t_{00}^{\beta\gamma} = -0.000216104.$$

However, further computations show that

$$\chi_{000} = -0.000843645,$$

such that the radical terms become complex, for instance:

$$q_0^{\varrho} \approx 0.5 - 0.078453098\mathbf{i}.$$

This example demonstrates that the conditions given in Theorem 3.1.4 are not sufficient to get a stochastically admissible solution. Finding conditions, under which a leaf distribution is stochastically admissible is the purpose of this section.

The following sets are the starting point for the conditions.

$$S_{xy}^{\alpha\beta} :=\{t_{xy}^{\alpha\beta},\ s_{xy0}^{\alpha\beta},\ s_{xy1}^{\alpha\beta}\},\quad S_{xz}^{\alpha\gamma} := \quad \{t_{xz}^{\alpha\gamma},\ s_{x0z}^{\alpha\gamma},\ s_{x1z}^{\alpha\gamma}\},\quad S_{yz}^{\beta\gamma} :=\{t_{yz}^{\beta\gamma},\ s_{0yz}^{\beta\gamma},\ s_{1yz}^{\beta\gamma}\}.$$

The sets will be called *covariance sets*, since they contain values which up to a scalar denote the unconditional covariance between the indexed leaves and the conditional covariances between those leaves given the remaining leaf has a certain state. Such a set *has a sign* if all contained terms have the same sign or the conditional covariances are zero. The unconditional probabilities cannot be zero owing to condition (3.1.9). The stochastic admissibility of a solution depends on the signs of the sets.

Theorem 3.1.5. *Let $x, y, z \in \{0, 1\}$ and let $S_{xy}^{\alpha\beta}, S_{xz}^{\alpha\gamma}$ and $S_{yz}^{\beta\gamma}$ have a sign. The number of sets with a negative sign is even if and only if the unique solution up to symmetry of (3.1.1) given by Theorem 3.1.4 is stochastically admissible.*

Even though pairs of leaves are not allowed to be independent it is feasible for them to be conditionally independent w.r.t. the third leaf. Another implication is that at least two leaves must be positively correlated and if pairs (X_α, X_β) and (X_α, X_γ) are positively correlated then (X_β, X_γ) must be positively correlated as well.

Example 3.1.2. Recall the leaf distribution presented in Example 3.1.1. Computing the covariance sets for this distribution yields:

$$S_{00}^{\alpha\beta} = \{0.057065, 0.016359, 0.019661\},\quad S_{00}^{\alpha\gamma} = \{-0.073495, -0.027085, -0.016207\},$$
$$S_{00}^{\beta\gamma} = \{0.051527, 0.017015, 0.016935\}.$$

Clearly, all sets have a sign, but an odd number of sets has negative sign. Therefore, inadmissibility is shown.

The example shows that a strictly positive leaf distribution is not necessarily Markov-extendable. Chapter 5 presents a heuristic approach to manipulate data for stochastic admissibility.

3.1.4 Degenerate Cases

To complete the analysis of solutions for (3.1.1) a look at the degenerate cases must be included. Three possible cases may occur:

$$t_{00}^{\alpha\beta}t_{00}^{\alpha\gamma}t_{00}^{\beta\gamma} \neq 0, \chi_{000} = 0,\quad t_{00}^{\alpha\beta}t_{00}^{\alpha\gamma}t_{00}^{\beta\gamma} = 0, \chi_{000} \neq 0,\quad t_{00}^{\alpha\beta}t_{00}^{\alpha\gamma}t_{00}^{\beta\gamma} = 0, \chi_{000} = 0.$$

The first case needs further consideration. However, the second and third case can be reduced to $t_{xy}^{\alpha\beta}t_{xz}^{\alpha\gamma}t_{yz}^{\beta\gamma} = 0$. This condition is equivalent to

$$t_{xy}^{\alpha\beta} = 0 \quad \text{or} \quad t_{xz}^{\alpha\gamma} = 0 \quad \text{or} \quad t_{yz}^{\beta\gamma} = 0.$$

In probabilistic words, one looks at the cases where for two leaves δ_1, δ_2 the random variables X_{δ_1} and X_{δ_2} are uncorrelated. For random variables with only two possible states this implies stochastic independence. A first observation is the following

Lemma 3.1.6. *Let $x, y, z \in \{0,1\}$ and $t_{xy}^{\delta_1\delta_2} = 0$. Then, at least another covariance is zero.*

Probabilistically, if X_{δ_1} and X_{δ_2} are independent of each other, then X_{δ_3} is also independent of at least one of them. The implication of these cases is recorded in the next theorem.

Theorem 3.1.7. *Let $t_{xy}^{\alpha\beta} = 0$. Then Lemma 3.1.6 holds and the following line-ups are possible up to symmetry:*

(i) $t_{xz}^{\alpha\gamma} = 0$ *and* $t_{yz}^{\beta\gamma} \neq 0$*. Then, for $u, y, z \in \{0,1\}$ this gives* $p_{0y}^{\beta} \neq p_{1y}^{\beta}$, $p_{0z}^{\gamma} \neq p_{1z}^{\gamma}$, $p_{0x}^{\alpha} = p_{1x}^{\alpha} = m_{x\Sigma\Sigma}$, $0 < q_0^{\varrho} < 1$, *and*

$$q_0^{\varrho} = \frac{m_{\Sigma\Sigma z} - p_{1z}^{\gamma}}{p_{0z}^{\gamma} - p_{1z}^{\gamma}}, \quad p_{uz}^{\beta} = \frac{m_{\Sigma yz} - m_{\Sigma y\Sigma} p_{(1-u)z}^{\gamma}}{m_{\Sigma\Sigma z} - p_{(1-u)z}^{\gamma}},$$

with free parameters p_{0z}^{γ} and p_{1z}^{γ}. Lemma 3.1.2 provides analogue results for the remaining two cases where one covariance is not zero.

(ii) $t_{xz}^{\alpha\gamma} = t_{yz}^{\beta\gamma} = 0$*. Then,*

(a) $p_{0x}^{\alpha} = p_{1x}^{\alpha} = m_{x\Sigma\Sigma}$, $p_{0y}^{\beta} = p_{1y}^{\beta} = m_{\Sigma y\Sigma}$, $p_{0z}^{\gamma} \neq p_{1z}^{\gamma}$ *and*

$$q_0^{\varrho} = \frac{m_{\Sigma\Sigma z} - p_{1z}^{\gamma}}{p_{0z}^{\gamma} - p_{1z}^{\gamma}}$$

with free parameters p_{0z}^{γ} and p_{1z}^{γ}. The cases

$$\begin{aligned} &p_{0x}^{\alpha} = p_{1x}^{\alpha} = m_{x\Sigma\Sigma}, \quad p_{0y}^{\beta} \neq p_{1y}^{\beta}, \quad p_{0z}^{\gamma} = p_{1z}^{\gamma} = m_{\Sigma\Sigma z} \\ &p_{0x}^{\alpha} \neq p_{1x}^{\alpha}, \quad p_{0y}^{\beta} = p_{1y}^{\beta} = m_{\Sigma y\Sigma}, \quad p_{0z}^{\gamma} = p_{1z}^{\gamma} = m_{\Sigma\Sigma z} \end{aligned}$$

return resembling notions.

(b) $p_{0x}^{\alpha} = p_{1x}^{\alpha} = m_{x\Sigma\Sigma}$, $p_{0y}^{\beta} = p_{1y}^{\beta} = m_{\Sigma y\Sigma}$, $p_{0z}^{\gamma} = p_{1z}^{\gamma} = m_{\Sigma\Sigma z}$ *with free parameter q_0^{ϱ}.*

(c) $q_0^{\varrho} = 0$ *and* $p_{1x}^{\alpha} = m_{x\Sigma\Sigma}$, $p_{1y}^{\beta} = m_{\Sigma y\Sigma}$, $p_{1z}^{\gamma} = m_{\Sigma\Sigma z}$ *and the remaining three parameters,* p_{0x}^{α}, p_{0y}^{β}, p_{0z}^{γ}*, are arbitrary. Similarly, the case $q_0^{\varrho} = 1$ yields* $p_{0x}^{\alpha} = m_{x\Sigma\Sigma}$, $p_{0y}^{\beta} = m_{\Sigma y\Sigma}$, $p_{0z}^{\gamma} = m_{\Sigma\Sigma z}$ *and arbitrary parameters* p_{1x}^{α}, p_{1y}^{β}, p_{1z}^{γ}.

A structural interpretation of Theorem 3.1.7 is given in Figure 3.1. Note, that in the two state case independence is equivalent to non-correlativeness. Generally, independence suggests non-connectivity resulting in the rejection of a tree model. Independence in terms of phylogeny means that the common ancestor of two representees is so far back in the evolutionary time scale that the relatedness of sequences is barely observable. However, this kind of independence should also be observable when trying to align such sequences.

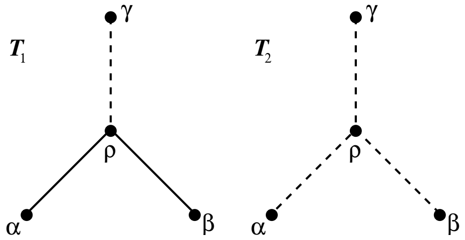

Figure 3.1: Degenerate cases. $\mathcal{T}_1$ shows a tree where the leaf γ is independent from the other two leaves. $\mathcal{T}_2$ shows a structure, where all leaves are pairwise independent. Here, a dashed line indicates that the leaf vertex of this edge is independent from the rest of the structure and in principle even isolated, but still "connected" through a transition matrix as given in Theorem 3.1.7.

Another implication of Theorem 3.1.7 is the insight that for all vectors with eight entries that sum to one a solution to (3.1.1) exists even though uniqueness cannot be guaranteed. In other words, vectors satisfying (3.1.9) have a finite set of interrelated solutions and vectors subject to Lemma 3.1.6 return fields of solutions.

For the remaining degenerate case

$$t_{00}^{\alpha\beta}t_{00}^{\alpha\gamma}t_{00}^{\beta\gamma} \neq 0, \chi_{000} = 0, \tag{3.1.13}$$

the following statement can be presented.

Lemma 3.1.8. *If a leaf distribution $\underline{m}$ obeys (3.1.13), then the equation system has no solution. The set of all leaf distributions obeying (3.1.13) is a Lebesgue zero set in the set of all possible leaf distributions on triple trees.*

The statement of the lemma indicates that the variety $\mathbf{V}_2$ of all leaf distributions for which (3.1.1) has a solution, is not affine. This follows from Proposition 2.2.1 which states that the elimination ideal with basis (3.1.2) provides the smallest affine variety containing the variety $\mathbf{V}_2$. However, it is not known whether or not a leaf distribution obeying (3.1.13) exists.

3.2 Special Case: The Symmetrical Two State Model

The previous section observed the implications of (LF) for a leaf distribution under the general two state model on a triple tree. This section translates these results for the symmetrical two state model on a triple tree. This model is also known as the Neyman N_2 model (cf. Semple and Steel [2003, chap. 8.5] and Example 1.3.2). Regard the tree introduced at the beginning of Section 3.1, i.e. $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with

$$\mathcal{V} = \{\alpha, \beta, \gamma, \varrho\} \quad \text{and} \quad E = \{(\varrho, \alpha), (\varrho, \beta), (\varrho, \gamma)\}$$

and leaf set $\mathcal{L} := \{\alpha, \beta, \gamma\}$.

The N_2 model is characterized by

$$q_0^\varrho = 1/2 = q_1^\varrho, \quad \text{and} \quad p_{01}^\delta = p_\delta = p_{10}^\delta, \quad \delta \in \mathcal{L},$$

i.e. the root distribution is stationary and the transition matrix is symmetrical, which immediately implies that the single leaf distributions are stationary as well. These observations reduce (3.1.1) to the following system

$$\begin{aligned} 2m_{000} &= (1-p_\alpha)(1-p_\beta)(1-p_\gamma) + p_\alpha p_\beta p_\gamma = 2m_{111}, \\ 2m_{001} &= (1-p_\alpha)(1-p_\beta)p_\gamma + p_\alpha p_\beta (1-p_\gamma) = 2m_{110}, \\ 2m_{010} &= (1-p_\alpha)p_\beta(1-p_\gamma) + p_\alpha(1-p_\beta)p_\gamma = 2m_{101}, \\ 2m_{011} &= (1-p_\alpha)p_\beta p_\gamma + p_\alpha(1-p_\beta)(1-p_\gamma) = 2m_{100}. \end{aligned} \tag{3.2.1}$$

Hence, condition (3.1.2) simplifies to

$$2(m_{000} + m_{001} + m_{010} + m_{100}) = 1. \tag{3.2.2}$$

This is the phylogenetic invariant for the N_2 model on a triple tree. With this equation a solution of (3.2.1) can be obtained by selecting three of the four equations.

An expression commonly used in the field of phylogenetic reconstruction is the so called *Hamming distance* between two sequences. Usually, it signifies the number of element-wise differences of the sequences. In probabilistic terms it can be defined as the probability that $X_\beta \neq X_\alpha$ e.g.,

$$d_{\alpha\beta} = m_{01\Sigma} + m_{10\Sigma} = m_{010} + m_{011} + m_{100} + m_{101} = 2m_{010} + 2m_{100}. \tag{3.2.3}$$

The distances $d_{\alpha\gamma}$ and $d_{\beta\gamma}$ are defined similarly.

The goal of the remainder of this section is to apply the results from Section 3.1.2 to this specification and to give some interpretations. Inserting the assumptions of the N_2 model into Theorem 3.1.4 yields:

Corollary 3.2.1. *If all Hamming distances are different from* $1/2$*, the solution for (3.1.1) under the* N_2 *model is unique up to symmetry. It is given as*

$$p_\alpha = \frac{1}{2} - \frac{\Delta}{2(1-2d_{\beta\gamma})},\; p_\beta = \frac{1}{2} - \frac{\Delta}{2(1-2d_{\alpha\gamma})},\; p_\gamma = \frac{1}{2} - \frac{\Delta}{2(1-2d_{\alpha\beta})},$$

where

$$\Delta = \sqrt{(1-2d_{\alpha\beta})(1-2d_{\alpha\gamma})(1-2d_{\beta\gamma})}.$$

The term Δ is the equivalent to χ_{xyz}, $(x, y, z \in \{0, 1\})$ from the general model. Apparently, the non-zero condition (3.1.9) is satisfied if no Hamming distance is one half.

Next, the conditions for a stochastic solution will be considered. The sign sets $S_{01}^{\alpha\beta}$ etc. translate to

$$\begin{aligned} S^{\alpha\beta} &= \{1 - 2d_{\alpha\beta}, m_{001}m_{000} - m_{010}m_{100}\}, \\ S^{\alpha\gamma} &= \{1 - 2d_{\alpha\gamma}, m_{010}m_{000} - m_{001}m_{100}\}, \\ S^{\beta\gamma} &= \{1 - 2d_{\beta\gamma}, m_{100}m_{000} - m_{001}m_{010}\} \end{aligned}$$

independent of the choice of a reduced state. Again, the sign of such a set exists when all elements have the same sign, which in turn is the sign of the set.

Corollary 3.2.2. *A solution for the* (N_2) *model is stochastically admissible if the sets* $S^{\alpha\beta}$, $S^{\alpha\gamma}$ *and* $S^{\beta\gamma}$ *have a sign and the number of negative signs is even.*

Phylogenetic data usually indicate Hamming distances smaller than 1/2 under the N_2 model (eg. Lake [1997]). The remaining cases only occur, if the sequences of two considered species would differ in more than 50% of the sites. Such an alignment is almost impossible to obtain (e.g. Waterman [1995]).

This section closes with a look at the case where the Hamming distances are exactly 1/2. Obviously, these results are special cases of Theorem 3.1.7. Solutions will be presented as vectors $(p_\alpha, p_\beta, p_\gamma) \in \mathbb{R}^3$

Corollary 3.2.3. *Let* $d_{\alpha\beta} = 1/2$. *Then at least another Hamming distance is* 1/2 *and the following line-ups are possible:*

(i) $d_{\alpha\gamma} = 1/2$ *and* $d_{\beta\gamma} = 1/2$. *Then, the leaf distribution* $\underline{m}$ *is the uniform distribution on* $\{0, 1\}^3$ *and the set of solutions is presented by*

$$\{(t, 1/2, 1/2) : t \in \mathbb{R}\} \cup \{(1/2, t, 1/2) : t \in \mathbb{R}\} \cup \{(1/2, 1/2, t) : t \in \mathbb{R}\}.$$

(ii) $d_{\alpha\gamma} = 1/2$ *and* $d_{\beta\gamma} \neq 1/2$. *Then the set of solutions is presented by*

$$\{(1/2, t, f(t, d_{\beta\gamma})) : t \in \mathbb{R}\} \cup \{(1/2, f(t, d_{\beta\gamma}), t) : t \in \mathbb{R}\},$$

where

$$f(t,y) = \frac{y-t}{1-2t}, \quad t \in \mathbb{R},\ y > 0.$$

The cases $d_{\alpha\beta} = d_{\beta\gamma} = 1/2 \neq d_{\alpha\gamma}$ *and* $d_{\alpha\beta} \neq 1/2 = d_{\alpha\gamma} = d_{\beta\gamma}$ *yield similar results.*

Similar to Theorem 3.1.7 the implications for the possible structure are visualized by Figure 3.1, i.e. for case (i) one vertex could be treated as isolated or identical to the inner vertex and for (ii) two isolated vertices can be assumed. The case (3.1.13) does not occur for the symmetrical model since here $\chi_{000} = t_{00}^{\alpha\beta} t_{00}^{\alpha\gamma} t_{00}^{\beta\gamma}$.

3.3 Extending the Results to Quartet Trees

So far, the main observation of this chapter is that an algebraic solution of (LF) w.r.t. a leaf distribution under the general two state model on a triple tree can almost always be found. However, as Section 1.4 stated, the goal of phylogenetic reconstruction is to obtain a tree with a possibly large number of leaves. This section examines the possibilities of extending the results from triple trees to quartet trees by examining some straightforward conditions for an extension. With Lemma 2.3.2 and Table 2.1 at least five phylogenetic invariants must be obtained from such an examination in order to guarantee an algebraic solution of (LF) for a quartet leaf distribution.

Regard the quartet tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with:

$$\begin{aligned} \mathcal{V} &= \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \varrho_1, \varrho_2\}, \\ \mathcal{E} &= \{(\varrho_1,\alpha_1), (\varrho_1,\alpha_2), (\varrho_1,\varrho_2), (\varrho_2,\alpha_3), (\varrho_2,\alpha_4)\} \end{aligned} \tag{3.3.1}$$

and leaf set $\mathcal{L} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. The associated triple trees are denoted by $\mathcal{T}^i = (\mathcal{V}^i, \mathcal{E}^i)$, $i = 1,2,3,4$, with

$$\begin{aligned} \mathcal{V}^1 &= \{\alpha_1, \alpha_2, \alpha_3, \varrho_1\}, & \mathcal{E}^1 &= \{(\varrho_1,\alpha_1), (\varrho_1,\alpha_2), (\varrho_1,\alpha_3)\}, \\ \mathcal{V}^2 &= \{\alpha_1, \alpha_2, \alpha_4, \varrho_1\}, & \mathcal{E}^2 &= \{(\varrho_1,\alpha_1), (\varrho_1,\alpha_2), (\varrho_1,\alpha_4)\}, \\ \mathcal{V}^3 &= \{\alpha_1, \alpha_3, \alpha_4, \varrho_2\}, & \mathcal{E}^3 &= \{(\varrho_2,\alpha_1), (\varrho_2,\alpha_3), (\varrho_2,\alpha_4)\}, \\ \mathcal{V}^4 &= \{\alpha_2, \alpha_3, \alpha_4, \varrho_2\}, & \mathcal{E}^4 &= \{(\varrho_2,\alpha_2), (\varrho_2,\alpha_3), (\varrho_2,\alpha_4)\}. \end{aligned} \tag{3.3.2}$$

The relationship of $\mathcal{T}$ and $\{\mathcal{T}^i\}_{i=1}^4$ is visualized in Figure 3.2.

An initial quartet distribution $\underline{m}$ on $\mathcal{L}$ provides triple leaf distributions $\underline{m}^i$ to each triple trees $\mathcal{T}^i$ by the following computation (set $a,b,c,d \in \{0,1\}$):

$$\begin{aligned} m^1_{abc} &= m_{abc0} + m_{abc1}, & m^2_{abd} &= m_{ab0d} + m_{ab1d}, \\ m^3_{acd} &= m_{a0cd} + m_{a1cd}, & m^4_{bcd} &= m_{0bcd} + m_{1bcd}. \end{aligned} \tag{3.3.3}$$

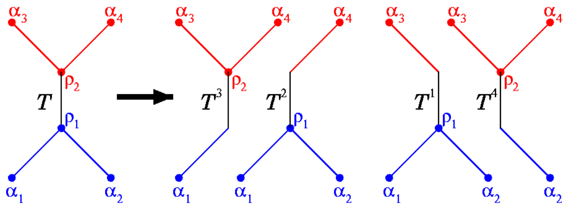

Figure 3.2: The picture shows the four triples to the quartet tree $\alpha\beta|\gamma\delta$. The colored inner vertices describe the associated furcating vertex.

For those distributions Theorem 3.1.4 returns transition parameters, which in turn are indexed by the triple identifier, for instance $p^{\alpha,1}_{ra}$ for the transition probability from ϱ_1 to α in triple $\mathcal{T}^1$. Covariances $t_{\alpha\beta}$ do not need an index since they are defined using pairwise and marginal leaf probabilities.

If $\underline{m}$ is subject to a Markov distribution μ, its restrictions $\underline{m}^i$ to the triple trees $\mathcal{T}^i$, $i = 1, 2, 3, 4$ are sufficient for a reconstruction of μ. This observation is verified in Chang [1996] under some conditions posed on the transition probabilities of the process. In that case, the transition parameters for edges $(\varrho_i, \alpha_j) \in \mathcal{E}$ must be equal on all triple trees containing ϱ_i and α_j. This equality condition will be called *compatibility*

This section looks for conditions on $\underline{m}$ for the existence of μ. A first step is applying Theorem 3.1.4 to μ^i, $i = 1, 2, 3, 4$, and analyzing the obtained parameters on $\mathcal{T}$:

Theorem 3.3.1. *Let $\mathcal{T}$ denote the quartet tree given in (3.3.1) with its associated triple trees $\mathcal{T}^i$, $i = 1, 2, 3, 4$ and let $\underline{m}$ denote a leaf distribution on $\mathcal{L}$. A Markov-like extension of $\underline{m}$ w.r.t. the system (LF) poses the following necessary conditions on $\underline{m}$ $(a, b, c, d \in \{0, 1\})$:*

$$
\begin{aligned}
0 =& r^{\alpha_1,1}_{abc} t^{\alpha_2\alpha_4}_{ad} - r^{\alpha_1,2}_{abd} t^{\alpha_2\alpha_3}_{ac}, \quad 0 = r^{\alpha_2,1}_{abc} t^{\alpha_1\alpha_4}_{bd} - r^{\alpha_2,2}_{abd} t^{\alpha_1\alpha_3}_{bc},\\
0 =& r^{\alpha_3,3}_{acd} t^{\alpha_2\alpha_4}_{bc} - r^{\alpha_3,4}_{bcd} t^{\alpha_1\alpha_4}_{ac}, \quad 0 = r^{\alpha_4,3}_{acd} t^{\alpha_1\alpha_3}_{bd} - r^{\alpha_4,4}_{bcd} t^{\alpha_2\alpha_3}_{bc},\\
0 =& t^{\alpha_1\alpha_4}_{ad} t^{\alpha_2\alpha_3}_{bc} - t^{\alpha_1\alpha_3}_{ac} t^{\alpha_2\alpha_4}_{bd}.
\end{aligned} \tag{3.3.4}
$$

Further, assume (3.1.9) to hold.

Then the transition parameters for the edge (ϱ_1, ϱ_2) are determined through

$$
p^{\varrho_1\varrho_2}_{00} = \frac{1}{2} - \frac{r^{\alpha_3,1}_{abc} t^{\alpha_1\alpha_4}_{ad} - r^{\alpha_3,3}_{acd} t^{\alpha_1\alpha_2}_{ab}}{2t^{\alpha_1\alpha_2}_{ab}\sqrt{\chi^3_{acd}}} + \frac{t^{\alpha_1\alpha_4}_{ad}\sqrt{\chi^1_{abc}}}{2t^{\alpha_1\alpha_2}_{ab}\sqrt{\chi^3_{acd}}}, \tag{3.3.5}
$$

$$
p^{\varrho_1\varrho_2}_{10} = \frac{1}{2} - \frac{r^{\alpha_3,1}_{abc} t^{\alpha_1\alpha_4}_{ad} - r^{\alpha_3,3}_{acd} t^{\alpha_1\alpha_2}_{ab}}{2t^{\alpha_1\alpha_2}_{ab}\sqrt{\chi^3_{acd}}} - \frac{t^{\alpha_1\alpha_4}_{ad}\sqrt{\chi^1_{abc}}}{2t^{\alpha_1\alpha_2}_{ab}\sqrt{\chi^3_{acd}}} \tag{3.3.6}
$$

provided that both $t_{ab}^{\alpha_1\alpha_2}$ and χ_{acd}^{3} do not vanish.

When looking for compatibility conditions one finds that the conditions (3.3.4) are already observed when checking compatibility for terminal edges. Therefore, compatibility for the inner edge follows immediately from compatibility of the terminal edges.

(3.3.4) provides five phylogenetic invariants, therefore the necessary condition of Lemma 2.3.2 is satisfied. However, as an indication that these invariants are not sufficient consider the symmetrical N_2 model. Since the N_2 model is a special case, the above conditions must hold here, too.

In order to be subject to the N_2 model on a quartet tree, a quartet leaf distribution must satisfy:

$$m_{abcd} = m_{(1-a)(1-b)(1-c)(1-d)}, \quad a, b, c, d \in \{0, 1\}. \tag{3.3.7}$$

With this notion and the insights from Section 3.2 the statements of Theorem 3.3.1 turns into:

Corollary 3.3.2. *Let $\mathcal{T}$ denote the quartet tree given by (3.3.1) and $\underline{m}$ be a quartet leaf distribution on $\mathcal{T}$ satisfying (3.3.7). If (LF) has an algebraic solution w.r.t. $\underline{m}$ under the N_2 model, then the associated Hamming distances satisfy:*

$$d_{\alpha_1\alpha_3} d_{\alpha_2\alpha_4} = d_{\alpha_1\alpha_4} d_{\alpha_2\alpha_3}. \tag{3.3.8}$$

The parameter for the inner edge is then given by:

$$p_{\varrho_2} = \frac{1}{2}\left(1 - \sqrt{\frac{(1-2d_{\alpha_1\alpha_3})(1-2d_{\alpha_2\alpha_4})}{(1-2d_{\alpha_1\alpha_2})(1-2d_{\alpha_3\alpha_4})}}\right).$$

Obviously, the five invariants from (3.3.4) became just one invariant. This is due to the simple fact, that for the N_2 model the following equivalence is observable:

$$r_{abc}^{\alpha_1,1} = -t_{bc}^{\alpha_2\alpha_3} = \frac{1}{4}(1 - 2d_{\alpha_2\alpha_3}).$$

This equality is derived as (3.4.42) in the proof section. However, when looking at the lower bound for the number of necessary phylogenetic invariants proposed by Lemma 2.3.2 it becomes apparent that at least three invariants are needed for the N_2 model. This suggests that the polynomials given in (3.3.4) are not sufficient for the existence of an algebraic solution of (LF) w.r.t. to a given quartet leaf distribution. The following example shows that an extension from triple trees to quartet trees needs more than just compatible parameters.

Example 3.3.1. Consider the following vector $\underline{m}$ satisfying (3.3.7):

$$\underline{m} = (320, 50, 20, 30, 10, 15, 10, 45)/1000.$$

This vector yields Hamming distances:

$$\begin{aligned} d_{\alpha_1\alpha_2} &= 4/25, & d_{\alpha_1\alpha_3} &= 21/100, & d_{\alpha_1\alpha_4} &= 7/25, \\ d_{\alpha_2\alpha_3} &= 3/20, & d_{\alpha_2\alpha_4} &= 1/5, & d_{\alpha_3\alpha_4} &= 19/100 \end{aligned}$$

These Hamming distances satisfy (3.3.8). If this polynomial would be sufficient for compatibility, then $p^1_{\alpha_1}$ and $p^2_{\alpha_1}$ should be equal. Instead, one gets:

$$p^1_{\alpha_1} = 0.124691, \quad p^2_{\alpha_1} = 0.146918,$$

and obviously these values are far from equal.

A set of sufficient phylogenetic invariants for the N_2 model can be computed with **Singular** as well as **Mathematica** . Both softwares return four polynomials, one of which is the usual summation condition. The following statement presents the invariants generated by **Mathematica** :

Lemma 3.3.3. *Let $\mathcal{T}$ denote the quartet tree given by (3.3.1). $\underline{m}$ is a quartet leaf distribution on $\mathcal{T}$ satisfying (3.3.7). Then, (LF) has an algebraic solution w.r.t. $\underline{m}$ under the N_2 model only if $\underline{m}$ is a root of the following polynomials:*

$$\begin{aligned} f_1(x_1,\dots,x_8) &= 1 - 2(x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8), \\ f_2(x_1,\dots,x_8) &= 2(x_2x_5+x_3x_8-x_4x_7-x_1x_6), \\ f_3(x_1,\dots,x_8) &= 2(x_1x_7+x_4x_6-x_2x_8-x_3x_5), \\ f_4(x_1,\dots,x_8) &= 2x_4(x_6^2-x_7^2) - 2x_3(x_5^2-x_8^2) + 2(x_1+x_2)(x_5x_7-x_6x_8) \\ &\quad - 2(x_3-x_4)(x_5x_6-x_7x_8). \end{aligned}$$

The order of the entries of $\underline{m}$ are chosen in the already introduced fashion: $\underline{m} := (m_{0000}, m_{0001}, \dots, m_{0111})$. The polynomial f_1 is the summation condition for a quartet leaf distribution under the N_2 model. For an interpretation of f_2 and f_3 note, that $m_{0000}, m_{0011}, m_{0101}$ and m_{0110} denote the probabilities where both states are attained by an even number of leaves whereas $m_{0001}, m_{0010}, m_{0100}$ and m_{0111} denote the probabilities where both states are attained by an odd number of leaves. This suggests the conclusion that a distribution which is root of f_2 and f_3 has a evenly matched distribution between the stated cases.

Looking at the overall performance of the attempts of this section so far one can say, that searching for compatibility by looking at the inferred triple transition parameters is not sufficient to deduce the existence of an algebraic solution of (LF) on a quartet tree. One attempt for an explanation relates the invariants of Lemma 3.3.3 to the invariant in (3.3.8). The latter invariant only relates the pairwise leaf probabilities whereas the former set relates the quartet leaf probabilities in an irreducible fashion. Consequently, one can conjecture that looking at the triple constraints of a quartet leaf distribution $\underline{m}$ is not sufficient to find all needed conditions for the existence of an algebraic solution of (LF) w.r.t. $\underline{m}$ under any model. Further attempts on finding conditions won't be tested here.

However, presenting conditions for stochastic admissibility of a computed solution is still a valid task. One finds that Theorem 3.1.5 still holds, and the remaining conditions are obtained from the results of Theorem 3.3.1:

Theorem 3.3.4. *An extension of a quartet leaf distribution $\underline{m}$ on $\mathcal{T}$ is stochastically admissible, if condition (3.3.4) is satisfied, and the transition parameters for the associated triple trees satisfy Theorem 3.1.5, and:*

$$\begin{aligned} p_{00}^{\alpha_1,3}, p_{10}^{\alpha_1,3} &\in [\min\{p_{00}^{\alpha_1,1}, p_{10}^{\alpha_1,1}\}, \max\{p_{00}^{\alpha_1,1}, p_{10}^{\alpha_1,1}\}],\\ p_{00}^{\alpha_2,4}, p_{10}^{\alpha_2,4} &\in [\min\{p_{00}^{\alpha_2,1}, p_{10}^{\alpha_2,1}\}, \max\{p_{00}^{\alpha_2,1}, p_{10}^{\alpha_2,1}\}],\\ p_{00}^{\alpha_3,1}, p_{10}^{\alpha_3,1} &\in [\min\{p_{00}^{\alpha_3,3}, p_{10}^{\alpha_3,3}\}, \max\{p_{00}^{\alpha_3,3}, p_{10}^{\alpha_3,3}\}],\\ p_{00}^{\alpha_4,2}, p_{10}^{\alpha_4,2} &\in [\min\{p_{00}^{\alpha_4,3}, p_{10}^{\alpha_4,3}\}, \max\{p_{00}^{\alpha_4,3}, p_{10}^{\alpha_4,3}\}]. \end{aligned} \tag{3.3.9}$$

The following example presents an intuitive interpretation of (3.3.9).

Example 3.3.2. Consider the first equation. The term $p^{\alpha_1,3}$ denotes the transition matrix for edge (ϱ_2, α_1) and the term $p^{\alpha_1,1}$ denotes the transition matrix for edge (ϱ_1, α_1). Assume $p_{00}^{\alpha_1,1} > p_{10}^{\alpha_1,1}$. Then, condition (3.3.9) states that the probability of preserving state zero along edge (ϱ_2, α_1) must be smaller than the probability of preserving state zero along edge (ϱ_1, α_1) and further, the probability for a change from state one to state zero along edge (ϱ_2, α_1) must be larger than the same state change along edge (ϱ_1, α_1).
Since $p_{0a}^{\alpha_1,i} = 1 - p_{0(1-a)}^{\alpha_1,i}, i = 1, 2, 3$ this observation translates to all possible transitions. Apparently, this observation corresponds to the assumption that the probability of change along longer edges is larger than on shorter edges or in other words, the longer the considered time interval the more likely mutation occurred. If $p_{00}^{\alpha_1,1} < p_{10}^{\alpha_1,1}$ is observed, the implications need to be reversed, i.e. the probability of change is higher on shorter edges than on longer edges, which appears to be a rather dissatisfying occurrence.

The sign condition from Theorem 3.1.5 has its own implication on extension considerations. Recall that the sign of the covariance terms $t_{ab}^{\delta_1\delta_2}$ are the signs of the covariance sets, provided the covariance sets have a sign. Moreover, since the covariance terms are not subject to a particular triple structure, they can be considered independent of any particular structure. Keeping these observations in mind one can give the following statement:

Lemma 3.3.5. *Let $\mathcal{T}$ denote a tree with $n > 3$ leaves, $\mathcal{L}$ its leaf set and $\widehat{\underline{m}}$ a joint distribution on $\mathcal{L}$. If $\widehat{\underline{m}}$ has a Markov extension, $\mathcal{L}$ can be divided into two disjoint not necessary nonempty sets $\mathcal{L}_1$ and $\mathcal{L}_2$ with $\mathcal{L}_1 \cup \mathcal{L}_2 = \mathcal{L}$ and the covariance terms satisfy:*

$$\begin{aligned} t_{ab}^{\delta_1\delta_2} &> 0, \quad a,b \in \{0,1\},\ \delta_1,\delta_2 \in \mathcal{L}_i,\ i \in \{1,2\}, \\ t_{ac}^{\delta_1\delta_3} &< 0, \quad a,c \in \{0,1\},\ \delta_1 \in \mathcal{L}_i,\ \delta_2 \in \mathcal{L}_{3-i},\ i \in \{1,2\}. \end{aligned} \tag{3.3.10}$$

This property gives a good first test for the admissibility of a given leaf distribution. If only two sets satisfying (3.3.10) are observed, a first obstacle is taken. In that case, it is quite appropriate to place a root for the derived tree on the edge that connects both sets. This convention agrees with the concept of outgroups. Usually, one adds to the set of considered species another species , which from general understanding is not as close a relative as the other species are to each other. Unfortunately, for the approach suggested here, such an outgroup species must have a sequence which is different in at least 50% of all sites, and such alignments are not observed (e.g. Waterman [1995]).

3.4 Proofs

The section starts with some helpful properties of the terms defined in (3.1.6) and (3.1.7).

Lemma 3.4.1. *Let $x, y, z \in \{0,1\}$. The following equalities hold:*

$$\chi_{xyz} = (r_{xyz}^{\alpha})^2 - 4s_{xyz}^{\beta\gamma}t_{yz}^{\beta\gamma} = (r_{xyz}^{\beta})^2 - 4s_{xyz}^{\alpha\gamma}t_{xz}^{\alpha\gamma} = (r_{xyz}^{\gamma})^2 - 4s_{xyz}^{\alpha\beta}t_{xy}^{\alpha\beta}, \tag{3.4.1}$$

$$-t_{xy}^{\alpha\beta}t_{xz}^{\alpha\gamma} = m_{x\Sigma\Sigma}^2\, t_{yz}^{\beta\gamma} + m_{x\Sigma\Sigma}\, r_{xyz}^{\alpha} + s_{xyz}^{\beta\gamma}, \tag{3.4.2}$$

$$r_{xyz}^{\alpha} = -r_{x(1-y)z}^{\alpha} = -r_{xy(1-z)}^{\alpha} = -(r_{(1-x)yz}^{\alpha} + t_{yz}^{\beta\gamma}), \tag{3.4.3}$$

$$s_{xyz}^{\alpha\beta} = -s_{(1-x)yz}^{\alpha\beta} = -s_{x(1-y)z}^{\alpha\beta} = r_{xy(1-z)}^{\gamma} + s_{xy(1-z)}^{\alpha\beta} + t_{xy}^{\alpha\beta}, \tag{3.4.4}$$

$$t_{xy}^{\alpha\beta} = -t_{(1-x)y}^{\alpha\beta} = -t_{x(1-y)}^{\alpha\beta} = t_{(1-x)(1-y)}^{\alpha\beta}, \tag{3.4.5}$$

$$\chi_{xyz} = \chi_{(1-x)yz} = \chi_{x(1-y)z} = \chi_{xy(1-z)}, \tag{3.4.6}$$

$$\chi_{xyz} = (r_{xyz}^{\alpha} + 2m_{x\Sigma\Sigma}t_{yz}^{\beta\gamma})^2 + 4t_{xy}^{\alpha\beta}t_{xz}^{\alpha\gamma}t_{yz}^{\beta\gamma}. \tag{3.4.7}$$

Proof. These are simple computations. Showing one equality in each line is sufficient because the remaining equalities are computed similarly. Starting with (3.4.1)

$$\begin{aligned}
&(r^{\alpha}_{xyz})^2 - 4s^{\beta\gamma}_{xyz}t^{\beta\gamma}_{yz}\\
&= ((m_{xy\Sigma}m_{\Sigma\Sigma z} + m_{x\Sigma z}m_{\Sigma y\Sigma}) - (m_{\Sigma yz}m_{x\Sigma\Sigma} + m_{xyz})^2\\
&\quad - 4(m_{xyz}m_{x\Sigma\Sigma} - m_{xy\Sigma}m_{x\Sigma z})(m_{\Sigma yz} - m_{\Sigma y\Sigma}m_{\Sigma\Sigma z})
\end{aligned}$$

$$\begin{aligned}
&= m^2_{\Sigma\Sigma z}m^2_{xy\Sigma} + m^2_{\Sigma y\Sigma}m^2_{x\Sigma z} + m^2_{x\Sigma\Sigma}m^2_{\Sigma yz} + m^2_{xyz} + 2m_{x\Sigma\Sigma}m_{\Sigma yz}m_{xyz}\\
&\quad + 2m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}m_{xy\Sigma}m_{x\Sigma z} - 2m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{x\Sigma z}m_{\Sigma yz} - 2m_{x\Sigma\Sigma}m_{\Sigma\Sigma z}m_{xy\Sigma}m_{\Sigma yz}\\
&\quad - 2m_{\Sigma y\Sigma}m_{x\Sigma z}m_{xyz} - 2m_{\Sigma\Sigma z}m_{xy\Sigma}m_{xyz} - 4m_{x\Sigma\Sigma}m_{\Sigma yz}m_{xyz} + 4m_{xy\Sigma}m_{x\Sigma z}m_{\Sigma yz}\\
&\quad - 4m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}m_{xy\Sigma}m_{x\Sigma z} + 4m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}m_{xyz}\\
&= m^2_{\Sigma\Sigma z}m^2_{xy\Sigma} + m^2_{\Sigma y\Sigma}m^2_{x\Sigma z} + m^2_{x\Sigma\Sigma}m^2_{\Sigma yz} + m^2_{xyz} - 2m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{x\Sigma z}m_{\Sigma yz}\\
&\quad - 2m_{\Sigma y\Sigma}m_{x\Sigma z}m_{xyz} - 2m_{x\Sigma\Sigma}m_{\Sigma\Sigma z}m_{xy\Sigma}m_{\Sigma yz} - 2m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}m_{x\Sigma z}m_{\Sigma yz}\\
&\quad - 2m_{\Sigma\Sigma z}m_{xy\Sigma}m_{xyz} - 2m_{x\Sigma\Sigma}m_{\Sigma yz}m_{xyz} + 4m_{xy\Sigma}m_{x\Sigma z}m_{\Sigma yz} + 4m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}m_{xyz}\\
&= ((m_{xy\Sigma}m_{\Sigma\Sigma z} + m_{\Sigma yz}m_{x\Sigma\Sigma}) - (m_{x\Sigma z}m_{\Sigma y\Sigma} + m_{xyz}))^2\\
&\quad - 4(m_{\Sigma y\Sigma}m_{xyz} - m_{xy\Sigma}m_{\Sigma yz})(m_{x\Sigma z} - m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}) = (r^{\beta}_{xyz})^2 - 4s^{\alpha\gamma}_{xyz}t^{\alpha\gamma}_{xz}.
\end{aligned}$$

Equation (3.4.2) is verified by the following computations:

$$\begin{aligned}
&m^2_{x\Sigma\Sigma}t^{\beta\gamma}_{yz} + m_{x\Sigma\Sigma}r^{\alpha}_{xyz} + s^{\beta\gamma}_{xyz}\\
&= m^2_{x\Sigma\Sigma}(m_{\Sigma yz} - m_{\Sigma y\Sigma}m_{\Sigma\Sigma z}) + (m_{x\Sigma\Sigma}m_{xyz} - m_{xy\Sigma}m_{x\Sigma z})\\
&\quad + m_{x\Sigma\Sigma}(m_{xy\Sigma}m_{\Sigma\Sigma z} + m_{x\Sigma z}m_{\Sigma y\Sigma} - m_{\Sigma yz}m_{x\Sigma\Sigma} - m_{xyz})\\
&= -m^2_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{\Sigma\Sigma z} - m_{xy\Sigma}m_{x\Sigma z} + m_{x\Sigma\Sigma}m_{\Sigma y\Sigma}m_{x\Sigma z} + m_{x\Sigma\Sigma}m_{\Sigma\Sigma z}m_{xy\Sigma}\\
&= -(m_{xy\Sigma} - m_{x\Sigma\Sigma}m_{\Sigma y\Sigma})(m_{x\Sigma z} - m_{x\Sigma\Sigma}m_{\Sigma\Sigma z}) = -t^{\alpha\beta}_{xy}t^{\alpha\gamma}_{xz}.
\end{aligned}$$

Equation (3.4.7) immediately follows when (3.4.2) is inserted in the initial definition $\chi_{xyz} = (r^{\alpha}_{xyz})^2 - 4s^{\beta\gamma}_{xyz}t^{\beta\gamma}_{yz}$.

The remaining properties are proved applying (3.1.4) and (3.1.5). Start with replacing state x by state $1-x$ in (3.4.3):

$$\begin{aligned}
r^{\alpha}_{xyz} &= m_{xy\Sigma}m_{\Sigma\Sigma z} + m_{x\Sigma z}m_{\Sigma y\Sigma} - m_{\Sigma yz}m_{x\Sigma\Sigma} - m_{xyz}\\
&= (m_{\Sigma y\Sigma} - m_{(1-x)y\Sigma})m_{\Sigma\Sigma z} + (m_{\Sigma\Sigma z} - m_{(1-x)\Sigma z})m_{\Sigma y\Sigma}\\
&\quad - m_{\Sigma yz}(1 - m_{(1-x)\Sigma\Sigma}) - (m_{\Sigma yz} - m_{(1-x)yz})\\
&= m_{(1-x)yz} + m_{\Sigma yz}m_{(1-x)\Sigma\Sigma} - m_{(1-x)y\Sigma}m_{\Sigma\Sigma z} - m_{(1-x)\Sigma z}m_{\Sigma y\Sigma}\\
&\quad + 2m_{\Sigma y\Sigma}m_{\Sigma\Sigma z} - 2m_{\Sigma yz} = -(r^{\alpha}_{(1-x)yz} + 2t^{\beta\gamma}_{yz}).
\end{aligned}$$

Now for the state change from y to $1-y$

$$\begin{aligned}
r^{\alpha}_{xyz} &= m_{xy\Sigma}m_{\Sigma\Sigma z} + m_{x\Sigma z}m_{\Sigma y\Sigma} - m_{\Sigma yz}m_{x\Sigma\Sigma} - m_{xyz}\\
&= (m_{x\Sigma\Sigma} - m_{x(1-y)\Sigma})m_{\Sigma\Sigma z} + m_{x\Sigma z}(1 - m_{\Sigma(1-y)\Sigma})\\
&\quad - (m_{\Sigma\Sigma z} - m_{\Sigma(1-y)z})m_{x\Sigma\Sigma} - (m_{x\Sigma z} - m_{x(1-y)z})\\
&= m_{x(1-y)z} + m_{\Sigma(1-y)}m_{x\Sigma\Sigma} - m_{x\Sigma z}m_{\Sigma(1-y)\Sigma} - m_{x(1-y)\Sigma}m_{\Sigma\Sigma z} = -r^{\alpha}_{x(1-y)z}
\end{aligned}$$

Follow with (3.4.4), replace first z with $1-z$:

$$\begin{aligned} s^{\alpha\beta}_{xyz} &= m_{\Sigma\Sigma z}m_{xyz} - m_{x\Sigma z}m_{\Sigma yz} \\ &= (m_{xy\Sigma} - m_{xy(1-z)})(1 - m_{\Sigma\Sigma(1-z)}) - (m_{x\Sigma\Sigma} - m_{x\Sigma(1-z)})(m_{\Sigma y\Sigma} - m_{\Sigma y(1-z)}) \\ &= m_{x\Sigma\Sigma}m_{\Sigma y(1-z)} + m_{\Sigma y\Sigma}m_{x\Sigma(1-z)} - m_{\Sigma\Sigma(1-z)}m_{xy\Sigma} - m_{xy(1-z)} \\ &\quad + m_{xy\Sigma} - m_{x\Sigma\Sigma}m_{\Sigma y\Sigma} + m_{\Sigma\Sigma(1-z)}m_{xy(1-z)} - m_{x\Sigma(1-z)}m_{\Sigma y(1-z)} \\ &= r^{\gamma}_{xy(1-z)} + s^{\alpha\beta}_{xy(1-z)} + t^{\alpha\beta}_{xy}. \end{aligned}$$

Now replace y with $1-y$:

$$\begin{aligned} s^{\alpha\beta}_{xyz} &= m_{\Sigma\Sigma z}m_{xyz} - m_{x\Sigma z}m_{\Sigma yz} = m_{\Sigma\Sigma z}(m_{x\Sigma z} - m_{x(1-y)z}) - m_{x\Sigma z}(m_{\Sigma\Sigma z} - m_{\Sigma(1-y)z}) \\ &= m_{x\Sigma z}m_{\Sigma(1-y)z} - m_{\Sigma\Sigma z}m_{x(1-y)z} = -s^{\alpha\beta}_{x(1-y)z} \end{aligned}$$

Next for (3.4.5), replace x with $1-x$:

$$t^{\alpha\beta}_{xy} = m_{xy\Sigma} - m_{x\Sigma\Sigma}m_{\Sigma y\Sigma} = (m_{\Sigma y\Sigma} - m_{(1-x)y\Sigma}) - (1 - m_{(1-x)\Sigma\Sigma})m_{\Sigma y\Sigma} = -t^{\alpha\beta}_{(1-x)y}$$

Apply (3.4.3), (3.4.4) and (3.4.5) to get (3.4.6). First, replace x with $1-x$:

$$\begin{aligned} \chi_{xyz} &= (r^{\alpha}_{xyz})^2 - 4s^{\beta\gamma}_{xyz}t^{\beta\gamma}_{yz} = (r^{\alpha}_{(1-x)yz} + 2t^{\beta\gamma}_{yz})^2 - 4(r^{\alpha}_{(1-x)yz} + s^{\beta\gamma}_{(1-x)yz} + t^{\beta\gamma}_{yz})t^{\beta\gamma}_{yz} \\ &= (r^{\alpha}_{(1-x)yz})^2 - 4s^{\beta\gamma}_{(1-x)yz}t^{\beta\gamma}_{yz} = \chi_{(1-x)yz}. \end{aligned}$$

Replace y with $1-y$:

$$\chi_{xyz} = (r^{\alpha}_{xyz})^2 - 4s^{\beta\gamma}_{xyz}t^{\beta\gamma}_{yz} = (-r^{\alpha}_{x(1-y)z})^2 - 4(-s^{\beta\gamma}_{x(1-y)z})(-t^{\beta\gamma}_{(1-y)z}) = \chi_{x(1-y)z}.$$

The symmetry arguments from Lemma 3.1.2 transfers the results to the remaining combinations. Thus, all properties are verified. □

Remark 3.4.1. Equation (3.4.5) provides a standard property of a correlation mapping. Application of Lemma 3.4.1 to the formula from Theorem 3.1.4 giving of p^{α}_{00} yields:

$$p^{\alpha}_{00} = \frac{\sqrt{\chi_{000}}}{2t^{\beta\gamma}_{00}} - \frac{r^{\alpha}_{000}}{2t^{\beta\gamma}_{00}} = -\frac{\sqrt{\chi_{001}}}{2t^{\beta\gamma}_{01}} - \frac{r^{\alpha}_{001}}{2t^{\beta\gamma}_{01}} = \hat{p}^{\alpha}_{10},$$

i.e. the statement of Lemma 3.1.2.2 is observed in the given solution . Thus, it is sufficient to prove the statements for one state, say 000. This property gives the opportunity to use shorter abbreviations for the following proofs, namely $\chi := \chi_{000}$ and

$$\begin{aligned} r_\alpha &:= r^{\alpha}_{000}, & s_\alpha &:= s^{\beta\gamma}_{000}, & t_\alpha &:= t^{\beta\gamma}_{00}, \\ r_\beta &:= r^{\beta}_{000}, & s_\beta &:= s^{\alpha\gamma}_{000}, & t_\beta &:= t^{\alpha\gamma}_{00}, \\ r_\gamma &:= r^{\gamma}_{000}, & s_\gamma &:= s^{\alpha\beta}_{000}, & t_\gamma &:= t^{\alpha\beta}_{00}. \end{aligned}$$

3.4.1 Proofs for Section 3.1

Here, the general two state model on triple trees was examined.

Proof of Lemma 3.1.1. The summation of the right hand sides of (3.1.1) for all states gives one. Hence, the sum of all left hand sides needs to be one. This concludes the proof. □

Proof of Lemma 3.1.2. The proof is quite straightforward.

1. Apply the permutation π to equation (3.1.1) to get for $x, y, z \in \{0, 1\}$:
$$m_{xyz} = q_0^\varrho p_{0x}^\alpha p_{0y}^\beta p_{0z}^\gamma + q_1^\varrho p_{1x}^\alpha p_{1y}^\beta p^\gamma 1z = q_0^\varrho \hat{p}_{0x}^{\pi(\alpha)} \hat{p}_{0y}^{\pi(\beta)} \hat{p}_{0z}^{\pi(\gamma)} + q_1^\varrho \hat{p}_{1x}^{\pi(\alpha)} \hat{p}_{1y}^{\pi(\beta)} \hat{p}_{1z}^{\pi(\gamma)}.$$
2. Follows from commutativity of addition, since for $x, y, z \in \{0, 1\}$ one computes
$$\hat{q}_0^\varrho \hat{p}_{0x}^\alpha \hat{p}_{0y}^\beta \hat{p}_{0z}^\gamma + \hat{q}_1^\varrho \hat{p}_{1x}^\alpha \hat{p}_{1y}^\beta \hat{p}_{1z}^\gamma = q_1^\varrho p_{1x}^\alpha p_{1y}^\beta p_{1z}^\gamma + q_0^\varrho p_{0x}^\alpha p_{0y}^\beta p_{0z}^\gamma = m_{xyz}.$$

This completes the proof. □

Proof of Lemma 3.1.3. If $t_{xy}^{\alpha\beta} t_{xz}^{\alpha\gamma} t_{yz}^{\beta\gamma} = 0$, then at least one factor must be zero, w.l.o.g, $t_{xy}^{\alpha\beta} = 0$. Now, due to Lemma 3.4.1
$$t_{(1-x)y}^{\alpha\beta} = -t_{xy}^{\alpha\beta} = 0$$
and thus, also $t_{(1-x)y}^{\alpha\beta} t_{(1-x)z}^{\alpha\gamma} t_{yz}^{\beta\gamma} = 0$. Lemma 3.4.1 also guarantees the validity of the remaining equalities. □

Proof of Theorem 3.1.4. The proposed terms and conditions are derived by solving the following, equivalent system derived through (3.1.4) and (3.1.5) by adding suitable equations:

$$m_{000} = q_0^\varrho p_{00}^\alpha p_{00}^\beta p_{00}^\gamma + q_1^\varrho p_{10}^\alpha p_{10}^\beta p_{10}^\gamma, \tag{3.4.8}$$
$$m_{00\Sigma} = q_0^\varrho p_{00}^\alpha p_{00}^\beta + q_1^\varrho p_{10}^\alpha p_{10}^\beta, \tag{3.4.9}$$
$$m_{0\Sigma 0} = q_0^\varrho p_{00}^\alpha p_{00}^\gamma + q_1^\varrho p_{10}^\alpha p_{10}^\gamma, \tag{3.4.10}$$
$$m_{\Sigma 00} = q_0^\varrho p_{00}^\beta p_{00}^\gamma + q_1^\varrho p_{10}^\beta p_{10}^\beta, \tag{3.4.11}$$
$$m_{0\Sigma\Sigma} = q_0^\varrho p_{00}^\alpha + q_1^\varrho p_{10}^\alpha, \tag{3.4.12}$$
$$m_{\Sigma 0\Sigma} = q_0^\varrho p_{00}^\beta + q_1^\varrho p_{10}^\beta, \tag{3.4.13}$$
$$m_{\Sigma\Sigma 0} = q_0^\varrho p_{00}^\gamma + q_1^\varrho p_{10}^\gamma. \tag{3.4.14}$$

The goal of the following computations is to retrieve an equation which only depends on one variable, say p_{00}^α. Due to Lemma 3.1.2.2 computing for the other variables yields equivalent expressions. Equations (3.4.12)-(3.4.14) yield

$$q_1^\varrho p_{10}^\alpha = m_{0\Sigma\Sigma} - q_0^\varrho p_{00}^\alpha, \quad q_1^\varrho p_{10}^\beta = m_{\Sigma 0\Sigma} - q_0^\varrho p_{00}^\beta, \quad q_1^\varrho p_{10}^\gamma = m_{\Sigma\Sigma 0} - q_0^\varrho p_{10}^\gamma, \tag{3.4.15}$$

Insert (3.4.15) into (3.4.9) and apply $q_0^\varrho + q_1^\varrho = 1$ to get

$$\begin{aligned} m_{00\Sigma}q_1^\varrho &= q_0^\varrho q_1^\varrho p_{00}^\alpha p_{00}^\beta + (m_{0\Sigma\Sigma} - q_0^\varrho p_{00}^\alpha)(m_{\Sigma0\Sigma} - q_0^\varrho p_{00}^\beta) \\ &= q_0^\varrho p_{00}^\beta(q_1^\varrho p_{00}^\alpha - m_{0\Sigma\Sigma} + q_0^\varrho p_{00}^\alpha) + m_{\Sigma0\Sigma}(m_{0\Sigma\Sigma} - q_0^\varrho p_{00}^\alpha). \end{aligned}$$

Conduct similar computations for p_{00}^γ, and summarize the terms in dependence of p_{00}^α and q_0^ϱ:

$$q_0^\varrho p_{00}^\beta(p_{00}^\alpha - m_{0\Sigma\Sigma}) = t_\gamma + q_0^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma}), \tag{3.4.16}$$

$$q_0^\varrho p_{00}^\gamma(p_{00}^\alpha - m_{0\Sigma\Sigma}) = t_\beta + q_0^\varrho(m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0}). \tag{3.4.17}$$

Insert (3.4.15) into (3.4.16) to get:

$$\begin{aligned} q_1^\varrho p_{10}^\beta(p_{00}^\alpha - m_{0\Sigma\Sigma}) &= m_{\Sigma0\Sigma}(p_{00}^\alpha - m_{0\Sigma\Sigma}) - t_\gamma - q_0^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma}) \\ &= q_1^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma}). \end{aligned} \tag{3.4.18}$$

It would be convenient to erase q_1^ϱ from this equality. To do this without any kind of violation the case $q_1^\varrho = 0$ needs further consideration. Under the assumption $q_1^\varrho = 0$ equations (3.4.9), (3.4.12) and (3.4.13) produce $m_{00\Sigma} = p_{00}^\alpha p_{00}^\beta$, $m_{0\Sigma\Sigma} = p_{00}^\alpha$ and $m_{\Sigma0\Sigma} = p_{00}^\beta$ respectively. Thus, even when omitting q_1^ϱ from the notation of (3.4.18) the equality is preserved. The notion for p_{10}^γ dependent on p_{00}^α is derived in analogous manner:

$$p_{10}^\gamma(p_{00}^\alpha - m_{0\Sigma\Sigma}) = (m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0}). \tag{3.4.19}$$

Applying (3.4.16)-(3.4.19) to (3.4.11) yields

$$\begin{aligned} m_{\Sigma00}q_0^\varrho(p_{00}^\alpha - m_{0\Sigma\Sigma})^2 &= (q_0^\varrho)^2(p_{00}^\alpha - m_{0\Sigma\Sigma})^2 p_{00}^\beta p_{00}^\gamma + q_0^\varrho q_1^\varrho(p_{00}^\alpha - m_{0\Sigma\Sigma})^2 p_{10}^\beta p_{10}^\gamma \\ &= (t_\gamma + q_0^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma}))(t_\beta + q_0^\varrho(m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0})) \\ &\quad + q_0^\varrho q_1^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma})(m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0}) \\ &= t_\beta t_\gamma + q_0^\varrho(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma})(m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0}) + q_0^\varrho t_\beta(m_{\Sigma0\Sigma}p_{00}^\alpha - m_{00\Sigma}) \\ &\quad + q_0^\varrho t_\gamma(m_{\Sigma\Sigma0}p_{00}^\alpha - m_{0\Sigma0}) \\ &= t_\beta t_\gamma + q_0^\varrho(m_{00\Sigma}m_{0\Sigma0} - m_{00\Sigma}t_\beta - m_{0\Sigma0}t_\gamma) \\ &\quad + q_0^\varrho p_{00}^\alpha(p_{00}^\alpha m_{\Sigma0\Sigma}m_{\Sigma\Sigma0} + m_{\Sigma0\Sigma}t_\beta + m_{\Sigma\Sigma0}t_\gamma - m_{00\Sigma}m_{\Sigma\Sigma0} - m_{0\Sigma0}m_{\Sigma0\Sigma}) \\ &= t_\beta t_\gamma + q_0^\varrho(m_{0\Sigma\Sigma}m_{\Sigma\Sigma0}m_{00\Sigma} + m_{0\Sigma\Sigma}m_{\Sigma0\Sigma}m_{0\Sigma0} - m_{00\Sigma}m_{0\Sigma0}) \\ &\quad + q_0^\varrho p_{00}^\alpha(p_{00}^\alpha m_{\Sigma0\Sigma}m_{\Sigma\Sigma0} - 2m_{0\Sigma\Sigma}m_{\Sigma0\Sigma}m_{\Sigma\Sigma0}) \end{aligned}$$

Restructuring the terms gives the following relationship:

$$\begin{aligned} 0 &= t_\beta t_\gamma + q_0^\varrho(m_{0\Sigma\Sigma}m_{\Sigma\Sigma0}m_{00\Sigma} + m_{0\Sigma\Sigma}m_{\Sigma0\Sigma}m_{0\Sigma0} - m_{00\Sigma}m_{0\Sigma0} - m_{0\Sigma\Sigma}^2 m_{\Sigma00}) \\ &\quad + q_0^\varrho p_{00}^\alpha(p_{00}^\alpha(m_{\Sigma0\Sigma}m_{\Sigma\Sigma0} - m_{\Sigma00}) + 2(m_{0\Sigma\Sigma}m_{\Sigma00} - m_{0\Sigma\Sigma}m_{\Sigma0\Sigma}m_{\Sigma\Sigma0})) \\ &= q_1^\varrho t_\beta t_\gamma - q_0^\varrho m_{0\Sigma\Sigma}^2 t_\alpha - q_0^\varrho(p_{00}^\alpha)^2 t_\alpha + 2q_0^\varrho p_{00}^\alpha m_{0\Sigma\Sigma}t_\alpha \\ &= q_1^\varrho t_\beta t_\gamma - q_0^\varrho t_\alpha(p_{00}^\alpha - m_{0\Sigma\Sigma})^2. \end{aligned}$$

Using $q_0^\varrho + q_1^\varrho = 1$ provides a description of q_0^ϱ dependent of p_{00}^α, namely

$$q_0^\varrho(t_\alpha(p_{00}^\alpha - m_{0\Sigma\Sigma})^2 + t_\beta t_\gamma) = t_\beta t_\gamma. \tag{3.4.20}$$

Applying (3.4.15) to (3.4.8) results in:

$$\begin{aligned} m_{000} &= q_0^\varrho p_{00}^\alpha p_{00}^\beta p_{00}^\gamma + (m_{0\Sigma\Sigma} - q_0^\varrho p_{00}^\alpha) p_{10}^\beta p_{10}^\gamma \\ &= p_{00}^\alpha (q_0^\varrho p_{00}^\beta p_{00}^\gamma + q_1^\varrho p_{10}^\beta p_{10}^\gamma - p_{10}^\beta p_{10}^\gamma) + m_{0\Sigma\Sigma} k_{10}^\beta k_{10}^\gamma, \\ m_{000} - m_{\Sigma 00} p_{00}^\alpha &= (m_{0\Sigma\Sigma} - p_{00}^\alpha) p_{10}^\beta p_{10}^\gamma. \end{aligned} \tag{3.4.21}$$

Inserting (3.4.18) and (3.4.19) into equation (3.4.21) yields

$$(m_{000} - m_{\Sigma 00} p_{00}^\alpha)(m_{0\Sigma\Sigma} - p_{00}^\alpha) = (m_{\Sigma 0\Sigma} - m_{00\Sigma} p_{00}^\alpha)(m_{\Sigma\Sigma 0} - m_{0\Sigma 0} p_{00}^\alpha).$$

Minor reordering steps surrender the following quadratic equation

$$\begin{aligned} 0 &= (p_{00}^\alpha)^2 (m_{\Sigma 00} - m_{\Sigma 0\Sigma} m_{\Sigma\Sigma 0}) + (m_{0\Sigma\Sigma} m_{000} - m_{00\Sigma} m_{0\Sigma 0}) \\ &\quad + p_{00}^\alpha (m_{\Sigma\Sigma 0} m_{00\Sigma} + m_{\Sigma 0\Sigma} m_{0\Sigma 0} - m_{0\Sigma\Sigma} m_{\Sigma 00} - m_{000}) \\ 0 &= t_\alpha (p_{00}^\alpha)^2 + r_\alpha p_{00}^\alpha + s_\alpha. \end{aligned} \tag{3.4.22}$$

To generate a solution of (3.4.22) in p_{00}^α the condition $t_\alpha \neq 0$ must be satisfied. With Lemma 3.1.2.1 this observation transfers to all t_δ, $\delta \in \mathcal{L}$, thus the necessity of the first condition of (3.1.9) is verified. For the explicit description of p_{00}^α apply the well-known equation for solving quadratic equations to get:

$$(p_{00}^\alpha)^\pm = -\frac{r_\alpha \pm \sqrt{r_\alpha^2 - 4 s_\alpha t_\alpha}}{2t_\alpha}. \tag{3.4.23}$$

To compute the formula for q_0^ϱ insert (3.4.23) into (3.4.20). The computation will use the terms for $(p_{00}^\alpha)^-$:

$$\begin{aligned} q_0^\varrho \left(t_\alpha \big(-\frac{r_\alpha - \sqrt{\chi}}{2t_\alpha} - m_{0\Sigma\Sigma} \big)^2 + t_\beta t_\gamma \right) &= t_\beta t_\gamma, \\ q_0^\varrho \big(((r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha) - \sqrt{\chi})^2 + 4 t_\alpha t_\beta t_\gamma \big) &= 4 t_\alpha t_\beta t_\gamma. \end{aligned}$$

(3.4.7) implies $4t_\alpha t_\beta t_\gamma = \chi - (r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha)^2$. Hence, one gets

$$\begin{aligned} &2\sqrt{\chi}(\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha)) q_0^\varrho \\ &= (\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha))^2 + 2(r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha)(\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha)). \end{aligned}$$

A division by the factor for q_0^ϱ is admissible under the conditions (3.1.9), thus demanding condition $\chi \neq 0$ and

$$q_0^\varrho = \frac{1}{2} + \frac{r_\alpha + 2m_{0\Sigma\Sigma} t_\alpha}{2\sqrt{\chi}},$$

i.e. the desired form from (3.1.10). The application of $(p^\alpha_{00})^+$ will yield $1/2 - q^\varrho_0$. From this insights compute p^α_{10} by inserting $(p^\alpha_{00})^-$ and (q^ϱ_0) into (3.4.15):

$$\begin{aligned}
(q^\varrho_0)^- p^\alpha_{10} &= m_{0\Sigma\Sigma} - (q^\varrho_0)^+ (p^\alpha_{00})^+,\\
p^\alpha_{10} &= \frac{2\sqrt{\chi}}{\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha)} \left[m_{0\Sigma\Sigma} + \frac{(r_\alpha - \sqrt{\chi})(\sqrt{\chi} + (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))}{4t_\alpha\sqrt{\chi}} \right]\\
p^\alpha_{10} &= \frac{4m_{0\Sigma\Sigma}t_\alpha\sqrt{\chi} + (r_\alpha - \sqrt{\chi})(\sqrt{\chi} + (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))}{2t_\alpha(\chi - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))}\\
&= \frac{\sqrt{\chi}(2m_{0\Sigma\Sigma}t_\alpha - \sqrt{\chi}) + r_\alpha(r_\alpha + 2m_{0\Sigma\Sigma}) - r_\alpha\sqrt{\chi} + r_\alpha\sqrt{\chi}}{2t_\alpha(\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))}\\
&= \frac{(-r_\alpha - \sqrt{\chi})(\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))}{2t_\alpha(\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))} = -\frac{r_\alpha + \sqrt{\chi}}{2t_\alpha} = (p^\alpha_{00})^-,
\end{aligned}$$

To finish the proof insert formulas (3.1.10)-(3.1.12) into equation (3.4.8) while heeding condition (3.1.9):

$$\begin{aligned}
m_{000} &= p^\alpha_{00}p^\beta_{00}p^\gamma_{00}q^\varrho_0 + p^\alpha_{10}p^\beta_{10}p^\gamma_{10}(1 - q^\varrho_0),\\
&= -\frac{r_\alpha - \sqrt{\chi}}{2t_\alpha} \cdot \frac{r_\beta - \sqrt{\chi}}{2t_\beta} \cdot \frac{r_\gamma - \sqrt{\chi}}{2t_\gamma} \cdot \frac{\sqrt{\chi} + r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha}{2\sqrt{\chi}}\\
&\quad - \frac{r_\alpha + \sqrt{\chi}}{2t_\alpha} \cdot \frac{r_\beta + \sqrt{\chi}}{2t_\beta} \cdot \frac{r_\gamma + \sqrt{\chi}}{2t_\gamma} \cdot \frac{\sqrt{\chi} - r_\alpha - 2m_{0\Sigma\Sigma}t_\alpha}{2\sqrt{\chi}}.
\end{aligned}$$

Multiply by $-16t_\alpha t_\beta t_\gamma\sqrt{\chi}$:

$$\begin{aligned}
-16m_{000}t_\alpha t_\beta t_\gamma\sqrt{\chi} &= (\sqrt{\chi} + (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))(r_\alpha - \sqrt{\chi})(r_\beta - \sqrt{\chi})(r_\gamma - \sqrt{\chi})\\
&\quad + (\sqrt{\chi} - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha))(r_\alpha + \sqrt{\chi})(r_\beta + \sqrt{\chi})(r_\gamma + \sqrt{\chi})\\
&= 2\sqrt{\chi}\big(r_\alpha r_\beta r_\gamma + \chi(r_\alpha + r_\beta + r_\gamma) - (r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha)(\chi + r_\alpha(r_\beta + r_\gamma) + r_\beta r_\gamma)\big)\\
&= 2\sqrt{\chi}\big((\chi - r_\alpha^2)(r_\beta + r_\gamma) - 2m_{0\Sigma\Sigma}t_\alpha(\chi + r_\alpha(r_\beta + r_\gamma) + r_\beta r_\gamma)\big)\\
&= -4t_\alpha\sqrt{\chi}\big(2s_\alpha(r_\beta + r_\gamma) + m_{0\Sigma\Sigma}(\chi + r_\alpha(r_\beta + r_\gamma) + r_\beta r_\gamma)\big)
\end{aligned}$$

Divide by $-4t_\alpha\sqrt{\chi}$ and use $\chi = r_\beta - 4s_\beta t_\beta$ (according to (3.4.1))

$$\begin{aligned}
4m_{000}t_\beta t_\gamma &= 2s_\alpha(r_\beta + r_\gamma) + m_{0\Sigma\Sigma}(\chi + r_\alpha(r_\beta + r_\gamma) + r_\beta r_\gamma)\\
&= (2s_\alpha + m_{0\Sigma\Sigma}(r_\alpha + r_\beta))(r_\beta + r_\gamma) - 4m_{0\Sigma\Sigma}s_\beta t_\beta.
\end{aligned}$$

Note, that $r_\alpha + r_\beta = 2(m_{00\Sigma}m_{\Sigma\Sigma 0} - m_{000})$ and $r_\beta + r_\gamma = 2(m_{\Sigma 00}m_{0\Sigma\Sigma} - m_{000})$. Thus, divide by 4 to get

$$\begin{aligned}
m_{000}t_\beta t_\gamma &= (m_{000}m_{0\Sigma\Sigma} - m_{00\Sigma}m_{0\Sigma 0} + m_{0\Sigma\Sigma}(m_{00\Sigma}m_{\Sigma\Sigma 0} - m_{000}))(m_{\Sigma 00}m_{0\Sigma\Sigma} - m_{000})\\
&\quad - m_{0\Sigma\Sigma}(m_{000}m_{\Sigma 0\Sigma} - m_{00\Sigma}m_{\Sigma 00})(m_{0\Sigma 0} - m_{0\Sigma\Sigma}m_{\Sigma\Sigma 0})\\
&= -t_\beta\big(m_{00\Sigma}(m_{\Sigma 00}m_{0\Sigma\Sigma} - m_{000}) + m_{0\Sigma\Sigma}(m_{000}m_{\Sigma 0\Sigma} - m_{00\Sigma}m_{\Sigma 00}\big) = m_{000}t_\beta t_\gamma,
\end{aligned}$$

thus leading to a true statement. Therefore, given the assumptions the given expressions yield a solution. This completes the proof of the theorem. □

Proof of Theorem 3.1.5. This theorem gave conditions under which (3.1.1) has a stochastically admissible solution w.r.t. a given leaf distribution $\underline{m}$. The derivation of these conditions is accomplished by bounding the terms established in Theorem 3.1.4 between zero and one. For readability the notation from Remark 3.4.1 is inherited. The condition $\chi > 0$ is necessary for real valued parameters. This implies:

$$4t_\alpha t_\beta t_\gamma > -(r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha)^2. \tag{3.4.24}$$

Consider $0 \leq q_0^\varrho \leq 1$. Inserting formula (3.1.10) yields ($\sqrt{\chi}$ may denote the positive root of χ)

$$-\frac{1}{2} \leq \pm\frac{r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha}{2\sqrt{\chi}} \leq \frac{1}{2},$$
$$-\sqrt{\chi} \leq \pm(r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha) \leq \sqrt{\chi}.$$

This is equivalent to

$$(r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha)^2 \leq \chi.$$

Thus $\chi \geq 0$ and with (3.4.7) also $t_\alpha t_\beta t_\gamma \geq 0$. With (3.1.9) strict positivity is demanded. Also, $t_\alpha t_\beta t_\gamma > 0$ indicates the sign condition given for the covariance sets, since positivity is attained if either all covariance terms are positive or one is positive and the others are negative.

The remaining conditions are obtained by looking at the transition parameters. Due to Lemma 3.1.2, it is sufficient to consider the implications to one parameter set, p_{w0}^α, $w \in \{0, 1\}$ say. Since p_{00}^α and p_{10}^α must be admissible, start the considerations with

$$0 \leq -\frac{r_\alpha \pm \sqrt{\chi}}{2t_\alpha} \leq 1$$

Assume $t_\alpha > 0$. Then $0 \leq -r_\alpha \pm \sqrt{\chi} \leq 2t_\alpha$, and thus:

$$r_\alpha \leq \pm\sqrt{\chi} \leq r_\alpha + 2t_\alpha.$$

Therefore, $t_\alpha > 0$ implies $r_\alpha \leq 0$. Consider each bound separately and start with the lower bound

$$r_\alpha \leq \pm\sqrt{\chi} \quad \text{or} \quad r_\alpha^2 \geq r_\alpha^2 - 4s_\alpha t_\alpha$$

and thus, $s_\alpha \geq 0$. Look at the upper bound: With $\pm\sqrt{\chi} \leq r_\alpha + 2t_\alpha$, one finds

$$-4s_\alpha t_\alpha \leq 4r_\alpha t_\alpha + 4t_\alpha^2, \quad \text{and} \quad 0 \leq r_\alpha + s_\alpha + t_\alpha.$$

Insert the notation from (3.1.6) and (3.1.7) to get:

$$\begin{aligned}0 &\leq m_{\Sigma\Sigma 0}m_{00\Sigma} + m_{\Sigma 0\Sigma}m_{0\Sigma 0} - m_{0\Sigma\Sigma}m_{\Sigma 00} - m_{000} + m_{0\Sigma\Sigma}m_{000} - m_{00\Sigma}m_{0\Sigma 0}\\ &\quad + m_{\Sigma 00} - m_{\Sigma 0\Sigma}m_{\Sigma\Sigma 0}\\ &= (m_{\Sigma 00} - m_{000})(1 - m_{0\Sigma\Sigma}) - (m_{\Sigma 0\Sigma} - m_{00\Sigma})(m_{\Sigma\Sigma 0} - m_{0\Sigma 0})\\ &= m_{1\Sigma\Sigma}m_{100} - m_{10\Sigma}m_{1\Sigma 0}.\end{aligned}$$

This is the third and final term of $S^{\beta\gamma}_{00}$ and admissibility was obtained for $t_\alpha > 0$, $s_\alpha \geq 0$ and $m_{1\Sigma\Sigma}m_{100} - m_{10\Sigma}m_{1\Sigma 0} \geq 0$, i.e. for a positive sign of $S^{\beta\gamma}_{00}$. Analogue computations for $t_\alpha < 0$ provide similar results for a negative sign of $S^{\beta\gamma}_{00}$. With the statements of Lemma 3.1.2 the Theorem is proven. □

Proof of Lemma 3.1.6. Due to (3.4.5) setting $t^{\alpha\beta}_{00} = 0$ implies $t^{\alpha\beta}_{xy} = 0$ for all $x, y \in \{0, 1\}$. This completes the proof of the lemma. □

Proof of Theorem 3.1.7. This Theorem treats the implications for possible processes that yield a leaf distribution $\underline{m}$ subject to Lemma 3.1.6. Observe, that

$$\begin{aligned}t^{\alpha\beta}_{00} &= m_{00\Sigma} - m_{0\Sigma\Sigma}m_{\Sigma 0\Sigma} = p^\alpha_{00}p^\beta_{00}q^\varrho_0 + p^\alpha_{10}p^\beta_{10}q^\varrho_1 - (p^\alpha_{00}q^\varrho_0 + p^\alpha_{10}q^\varrho_1)(p^\beta_{00}q^\varrho_0 + p^\beta_{10}q^\varrho_1)\\ &= q^\varrho_0 q^\varrho_1(p^\alpha_{00}p^\beta_{00} + p^\alpha_{10}p^\beta_{10} - p^\alpha_{00}p^\beta_{10} - p^\alpha_{10}p^\beta_{00}).\end{aligned}$$

Doing this computations similarly for $t^{\alpha\gamma}_{00}$ and $t^{\beta\gamma}_{00}$ gives:

$$t^{\alpha\beta}_{00} = q^\varrho_0 q^\varrho_1(p^\alpha_{00} - p^\alpha_{10})(p^\beta_{00} - p^\beta_{10}), \tag{3.4.25}$$

$$t^{\alpha\gamma}_{00} = q^\varrho_0 q^\varrho_1(p^\alpha_{00} - p^\alpha_{10})(p^\gamma_{00} - p^\gamma_{10}), \tag{3.4.26}$$

$$t^{\beta\gamma}_{00} = q^\varrho_0 q^\varrho_1(p^\beta_{00} - p^\beta_{10})(p^\gamma_{00} - p^\gamma_{10}). \tag{3.4.27}$$

Thus, $t^{\alpha\beta}_{00} = 0$ yields $q^\varrho_0 = 0$, $q^\varrho_0 = 1$, $p^\alpha_{00} = p^\alpha_{10}$ or $p^\beta_{00} = p^\beta_{10}$.
First consider $q^\varrho_0 = 0$. Then $q^\varrho_1 = 1$, all covariances are zero due to (3.4.25), (3.4.26) and (3.4.27). If all covariances are zero one obtains:

$$s_\alpha = m_{000}m_{0\Sigma\Sigma} - m_{00\Sigma}m_{0\Sigma 0} = m_{0\Sigma\Sigma}(m_{000} - m_{0\Sigma\Sigma}m_{\Sigma 0\Sigma}m_{\Sigma\Sigma 0}). \tag{3.4.28}$$

Moreover, with (3.1.1), (3.1.4) and (3.1.5) one gets for $x, y, z \in \{0, 1\}$

$$m_{xyz} = p^\alpha_{1x}p^\beta_{1y}p^\gamma_{1z}, \quad m_{xy\Sigma} = p^\alpha_{1x}p^\beta_{1y}, \quad m_{x\Sigma\Sigma} = p^\alpha_{1x}, \tag{3.4.29}$$

$$s_\alpha = p^\alpha_{10}(p^\alpha_{10}p^\beta_{10}p^\gamma_{10} - p^\alpha_{10}p^\beta_{10}p^\gamma_{10}) = 0. \tag{3.4.30}$$

(3.4.29) yields through analogous computations $p^\alpha_{1x} = m_{x\Sigma\Sigma}$, $p^\beta_{1y} = m_{\Sigma y\Sigma}$, $p^\gamma_{1z} = m_{\Sigma\Sigma z}$ and parameters p^δ_{0u}, $u \in \{0, 1\}$, $\delta \in \mathcal{L}$ are free. Since $m_{0\Sigma\Sigma} = m_{000} + m_{001} + m_{010} + m_{011}$ the property $m_{0\Sigma\Sigma} = 0$ implies $m_{000} = 0$. Thus, applying equality (3.4.30) to (3.4.28) yields

$$m_{000} = m_{0\Sigma\Sigma}m_{\Sigma 0\Sigma}m_{\Sigma\Sigma 0} = p^\alpha_{10}p^\beta_{10}p^\gamma_{10},$$

i.e. the proposed parameters are a solution under the given assumptions and case *(ii.c)* is verified.

Next consider

$$p^\alpha_{00} = p^\alpha_{10}, \quad p^\beta_{00} = p^\beta_{10}, \quad p^\gamma_{00} \neq p^\gamma_{10}. \tag{3.4.31}$$

Then, again all covariances are zero due to (3.4.25) and (3.4.26). Further, (3.4.28) and (3.4.29) are retained while

$$s_\alpha = p^\alpha_{00}(p^\alpha_{00}p^\beta_{00}(q^\varrho_0 p^\gamma_{00} + q^\varrho_1 p^\gamma_{10}) - p^\alpha_{00}p^\beta_{00}(q^\varrho_0 p^\gamma_{00} + q^\varrho_1 p^\gamma_{10})) = 0. \tag{3.4.32}$$

With (3.4.29) the following assignments $p^\alpha_{ux} = m_{x\Sigma\Sigma}$, $p^\beta_{uy} = m_{\Sigma y\Sigma}$ are fixed while for free $p^\gamma_{00} \neq p^\gamma_{10}$ one gets from (3.4.14)

$$q^\varrho_0 = \frac{m_{\Sigma\Sigma z} - p^\gamma_{10}}{p^\gamma_{00} - p^\gamma_{10}}.$$

(3.4.29) and (3.4.32) finally show, that the proposed terms are a solution to (3.1.1) under (3.4.31). The remaining cases

$$\begin{aligned} p^\alpha_{00} = p^\alpha_{10}, \quad p^\beta_{00} \neq p^\beta_{10}, \quad p^\gamma_{00} = p^\gamma_{10}, \\ p^\alpha_{00} \neq p^\alpha_{10}, \quad p^\beta_{00} = p^\beta_{10}, \quad p^\gamma_{00} = p^\gamma_{10}. \end{aligned}$$

are handled similarly. The case

$$p^\alpha_{00} = p^\alpha_{10}, \quad p^\beta_{00} = p^\beta_{10}, \quad p^\gamma_{00} = p^\gamma_{10}$$

gives $p^\alpha_{0x} = m_{x\Sigma\Sigma}$, $p^\beta_{0y} = m_{\Sigma y\Sigma}$, $p^\gamma_{0z} = m_{\Sigma\Sigma z}$ and the free parameter q^ϱ_0 through analogous computations. Thus, cases *(ii.a)* and *(ii.b)* are treated.

Finally, consider

$$t^{\alpha\beta}_{xy} = t^{\alpha\gamma}_{xz} = 0, \quad t^{\beta\gamma}_{yz} \neq 0, \quad x, y, z \in \{0, 1\}.$$

Then, with (3.4.25)-(3.4.27) and (3.4.12) one gets for $x, y, z \in \{0, 1\}$:

$$p^\alpha_{0x} = p^\alpha_{1x} = m_{x\Sigma\Sigma}, \quad p^\beta_{0y} \neq p^\beta_{1y}, \quad p^\gamma_{0z} \neq p^\gamma_{1z}, \quad 0 < q^\varrho_0 < 1. \tag{3.4.33}$$

Inserting these properties into (3.1.1) returns

$$m_{xyz} = m_{x\Sigma\Sigma}(q^\varrho_0 p^\beta_{0y} p^\gamma_{0z} + q^\varrho_1 p^\beta_{1y} p^\gamma_{1z}) = m_{x\Sigma\Sigma} m_{\Sigma yz}, \quad x, y, z \in \{0, 1\}.$$

Thus, it remains to establish the solution to the system

$$m_{\Sigma 00} = q^\varrho_0 p^\beta_{00} p^\gamma_{00} + (1 - q^\varrho_0) p^\beta_{10} p^\gamma_{10}, \tag{3.4.34}$$

$$m_{\Sigma 0\Sigma} = q^\varrho_0 p^\beta_{00} + (1 - q^\varrho_0) p^\beta_{10}, \tag{3.4.35}$$

$$m_{\Sigma\Sigma 0} = q^\varrho_0 p^\gamma_{00} + (1 - q^\varrho_0) p^\gamma_{10}. \tag{3.4.36}$$

(3.4.35) and (3.4.36) yield:

$$q_0^\varrho = \frac{m_{\Sigma 0\Sigma} - p_{10}^\beta}{p_{00}^\beta - p_{10}^\beta} = \frac{m_{\Sigma\Sigma 0} - p_{10}^\gamma}{p_{00}^\gamma - p_{10}^\gamma}. \tag{3.4.37}$$

From the latter equality one establishes:

$$p_{00}^\beta = \frac{p_{10}^\beta(m_{\Sigma\Sigma 0} - p_{00}^\gamma) + m_{\Sigma 0\Sigma}(p_{00}^\gamma - p_{10}^\gamma)}{m_{\Sigma\Sigma 0} - p_{10}^\gamma}. \tag{3.4.38}$$

Note that the case $p_{10}^\gamma = m_{\Sigma\Sigma 0}$ was fully considered in the previous cases, thus the numerator stays valid. Now, insert (3.4.37) into (3.4.34) to get:

$$m_{\Sigma 00}(p_{00}^\gamma - p_{10}^\gamma) + p_{10}^\beta p_{10}^\gamma(m_{\Sigma\Sigma 0} - p_{00}^\gamma) = p_{00}^\beta p_{00}^\gamma(m_{\Sigma\Sigma 0} - p_{10}^\gamma).$$

Applying (3.4.38) yields:

$$\begin{aligned}
&m_{\Sigma 00}(p_{00}^\gamma - p_{10}^\gamma) + p_{10}^\beta p_{10}^\gamma(m_{\Sigma\Sigma 0} - p_{00}^\gamma) = p_{00}^\gamma(p_{10}^\beta(m_{\Sigma\Sigma 0} - p_{00}^\gamma) + m_{\Sigma 0\Sigma}(p_{00}^\gamma - p_{10}^\gamma)),\\
&m_{\Sigma 00}(p_{00}^\gamma - p_{10}^\gamma) - p_{10}^\beta(m_{\Sigma\Sigma 0} - p_{00}^\gamma)(p_{00}^\gamma - p_{10}^\gamma) = p_{00}^\gamma m_{\Sigma 0\Sigma}(p_{00}^\gamma - p_{10}^\gamma),
\end{aligned}$$

and thus,

$$p_{10}^\beta = \frac{m_{\Sigma 00} - m_{\Sigma 0\Sigma}p_{00}^\gamma}{m_{\Sigma\Sigma 0} - p_{00}^\gamma}. \tag{3.4.39}$$

Reinserting into (3.4.38) finally returns

$$p_{00}^\beta = \frac{m_{\Sigma 00} - m_{\Sigma 0\Sigma}p_{10}^\gamma}{m_{\Sigma\Sigma 0} - p_{10}^\gamma}. \tag{3.4.40}$$

Derive from (3.4.37) the equality $q_1^\varrho = -(m_{\Sigma\Sigma 0} - p_{00}^\gamma)/(p_{00}^\gamma - p_{10}^\gamma)$. Insert the computed terms into (3.4.34) to get

$$\begin{aligned}
m_{\Sigma 00} &= \frac{m_{\Sigma\Sigma 0} - p_{10}^\gamma}{p_{00}^\gamma - p_{10}^\gamma} p_{00}^\gamma \frac{m_{\Sigma 00} - m_{\Sigma 0\Sigma}p_{10}^\gamma}{m_{\Sigma\Sigma 0} - p_{10}^\gamma} - \frac{m_{\Sigma\Sigma 0} - p_{00}^\gamma}{p_{00}^\gamma - p_{10}^\gamma} p_{10}^\gamma \frac{m_{\Sigma 00} - m_{\Sigma 0\Sigma}p_{00}^\gamma}{m_{\Sigma\Sigma 0} - p_{00}^\gamma}\\
&= m_{\Sigma 00}\frac{p_{00}^\gamma - p_{10}^\gamma}{p_{00}^\gamma - p_{10}^\gamma},
\end{aligned}$$

i.e. the proposed parameters are indeed a solution under (3.4.33). The remaining cases for $x, y, z \in \{0, 1\}$

$$\begin{aligned}
&p_{0x}^\alpha \neq p_{1x}^\alpha, \quad p_{0y}^\beta = p_{1y}^\beta = m_{\Sigma y\Sigma}, \quad p_{0z}^\gamma \neq p_{1z}^\gamma, \quad 0 < q_0^\varrho < 1,\\
&p_{0x}^\alpha \neq p_{1x}^\alpha, \quad p_{0y}^\beta \neq p_{1y}^\beta, \quad p_{0z}^\gamma = p_{1z}^\gamma = m_{\Sigma\Sigma z}, \quad 0 < q_0^\varrho < 1
\end{aligned}$$

return similar results. This completes the proof. □

Proof of Lemma 3.1.8. The equation (3.4.23) was derived without any restriction to χ. When inserting (3.4.23) into (3.4.20) one obtains:

$$q_0^\varrho\big(((r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha) - \sqrt{\chi})^2 + 4t_\alpha t_\beta t_\gamma\big) = 4t_\alpha t_\beta t_\gamma. \tag{3.4.41}$$

With $\chi = 0$ and (3.4.7) one gets for (3.4.41):

$$q_0^\varrho\big(\underbrace{(r_\alpha + 2m_{0\Sigma\Sigma}t_\alpha)^2 + 4t_\alpha t_\beta t_\gamma}_{=0}\big) = 4t_\alpha t_\beta t_\gamma,$$

i.e. $t_\alpha t_\beta t_\gamma = 0$, and hence no leaf distribution with a solution for (3.1.1) obeys (3.1.13).

The function $\chi : \mathbb{C}^8 \to \mathbb{C}$ is a polynomial mapping and hence is infinitely differentiable. Thus with the Morse-Sard Theorem (see e.g. Thm. 1.3 in Hirsch [1976]) the set $\{\underline{m} \in \mathbb{C}^8 : \chi(\underline{m}) = 0\}$ is a Lebesgue zero set. This completes the proof. □

3.4.2 Proofs for Section 3.2

Here, the special case of the symmetrical two state model on triple trees was related to the observations from the general model.

Proof of Corollary 3.2.1. First, insert the model restrictions into (3.1.4)-(3.1.7). Develop for state 001 and start with the pairwise probabilities

$$\begin{aligned} m_{00\Sigma} &= m_{000} + m_{001} = \frac{1}{2} - m_{010} - m_{100} = \frac{1}{2}(1 - d_{\alpha\beta}),\\ m_{0\Sigma1} &= m_{001} + m_{011} = m_{001} + m_{100} = \frac{d_{\alpha\gamma}}{2} \end{aligned}$$

and similarly, $m_{\Sigma01} = d_{\beta\gamma}/2$. Now for the other expression

$$\begin{aligned} r^\alpha_{001} &= m_{00\Sigma}m_{\Sigma\Sigma1} + m_{0\Sigma1}m_{\Sigma0\Sigma} - m_{\Sigma01}m_{0\Sigma\Sigma} - m_{001}\\ &= \frac{1}{4}\left(1 - d_{\alpha\beta} + d_{\alpha\gamma} - d_{\beta\gamma} - 4m_{001}\right)\\ &= \frac{1}{4}(1 - 2m_{010} - 2m_{100} + 2m_{100} + 2m_{001} - 2m_{001} - 2m_{010} - 4m_{001})\\ &= \frac{1}{4}(1 - 4m_{010} - 4m_{001}) = \frac{1}{4}(1 - 2d_{\beta\gamma}),\\ t^{\beta\gamma}_{01} &= m_{\Sigma01} - m_{\Sigma\Sigma1}m_{\Sigma0\Sigma} = m_{001} + m_{010} - \frac{1}{4} = -\frac{1}{4}(1 - 2d_{\beta\gamma}) = -r^\alpha_{001}, \end{aligned}$$

and further

$$\begin{aligned}
s^{\beta\gamma}_{001} &= m_{001}m_{0\Sigma\Sigma} - m_{00\Sigma}m_{0\Sigma 1} = \frac{m_{001}}{2} - (\frac{1}{2} - m_{010} - m_{100})(m_{001} + m_{100}) \\
&= \frac{m_{001}}{2} - \frac{m_{001}}{2} - \frac{m_{100}}{2} + (m_{001} + m_{100})(m_{010} + m_{100}) = \frac{1}{4}(d_{\alpha\beta}d_{\alpha\gamma} - 2m_{100}), \\
\chi_{001} &= (r^{\alpha}_{001})^2 - 4s^{\beta\gamma}_{001}t^{\beta\gamma}_{01} = r^{\alpha}_{001}(r^{\alpha}_{001} + 4s^{\beta\gamma}_{001}) = \frac{1}{4}r^{\alpha}_{001}(1 - 2d_{\beta\gamma} + 4d_{\alpha\beta}d_{\alpha\gamma} - 8m_{100}) \\
&= \frac{1}{4}r^{\alpha}_{001}(1 - 4m_{001} - 4m_{010} - 8m_{100} + 4d_{\alpha\beta}d_{\alpha\gamma}) = \frac{1}{4}r^{\alpha}_{001}(1 - 2d_{\alpha\beta})(1 - 2d_{\alpha\gamma}) \\
&= \frac{1}{16}(1 - 2d_{\alpha\beta})(1 - 2d_{\alpha\gamma})(1 - 2d_{\beta\gamma}) = \frac{\Delta^2}{16}.
\end{aligned}$$

The equality

$$4t^{\beta\gamma}_{01} = -(1 - 2d_{\beta\gamma}) \tag{3.4.42}$$

shows that $t^{\beta\gamma}_{01} \neq 0$ is equivalent to $d_{\beta\gamma} \neq 1/2$ under the symmetrical model. In addition, $d_{\delta_1\delta_2} \neq 0$, $\delta_1 \neq \delta_2 \in \mathcal{L}$ implies $\chi_{001} \neq 0$, thus providing the conditions to transfer the results of Theorem 3.1.4 to the N_2 model. Use above conditions to establish the proposed solution. First, verify that q^{ϱ}_0 is indeed 1/2 by

$$q^{\varrho}_0 = \frac{1}{2} + \frac{r^{\alpha}_{001} + 2m_{0\Sigma\Sigma}t^{\beta\gamma}_{01}}{2\sqrt{\chi_{001}}} = \frac{1}{2} + \frac{r^{\alpha}_{001}(1 - 2m_{0\Sigma\Sigma})}{2\sqrt{\chi_{001}}} = \frac{1}{2},$$

since $m_{0\Sigma\Sigma} = 1/2$. Now for $p_\alpha = p^{\alpha}_{10}$

$$p_\alpha = -\frac{r^{\alpha}_{001}}{2t^{\beta\gamma}_{01}} + \frac{\sqrt{\chi_{001}}}{2t^{\beta\gamma}_{01}} = \frac{1}{2} - \frac{\Delta}{2(1 - 2d_{\beta\gamma})}.$$

The computations for p_β and p_γ are similar. With (3.1.10)-(3.1.12) and the relationship of systems (3.2.1) and (3.1.1) shows, that the presented terms form a solution. This completes the proof. □

Proof of Corollary 3.2.2. First, observe with (3.4.42) and (3.4.5) that for $\delta_1 \neq \delta_2 \in \mathcal{L}$ the property $t_{\delta_1\delta_2} > 0$ implies $d_{\delta_1\delta_2} < 1/2$ and equivalently, $t_{\delta_1\delta_2} < 0$ implies $d_{\delta_1\delta_2} > 1/2$. Thus, under the symmetrical model $t_{\alpha\beta}t_{\alpha\gamma}t_{\beta\gamma} > 0$ has the following implications:

$$d_{\alpha\beta} < 1/2, \quad d_{\alpha\gamma} < 1/2, \quad d_{\beta\gamma} < 1/2, \tag{3.4.43}$$

$$d_{\delta_1\delta_2} < 1/2, \quad d_{\delta_1\delta_3} > 1/2, \quad d_{\delta_2\delta_3} > 1/2, \quad \delta_1 \neq \delta_2 \neq \delta_3 \tag{3.4.44}$$

Next, similarly to the derivation of the conditions in Theorem 3.1.5 one bounds the obtained parameters between zero and one. Again, citing Lemma 3.1.2 it is sufficient to look at the implications for p_α to derive all conditions. Note, that due to (3.2.1) the following relationship between the joint leaf probabilities holds under the N_2 model:

$$m_{000} + m_{001} + m_{010} + m_{100} = \frac{1}{2}.$$

Now for the computations:

$$\begin{aligned}
0 &\leq p_\alpha \leq 1,\\
0 &\leq \frac{1}{2}\left(1 \pm \sqrt{\frac{(1-2d_{\alpha\beta})(1-2d_{\alpha\gamma})}{1-2d_{\beta\gamma}}}\right) \leq 1,\\
-1 &\leq \pm\sqrt{\frac{(1-2d_{\alpha\beta})(1-2d_{\alpha\gamma})}{1-2d_{\beta\gamma}}} \leq 1,\\
0 &\leq \frac{(1-2d_{\alpha\beta})(1-2d_{\alpha\gamma})}{1-2d_{\beta\gamma}} \leq 1.
\end{aligned}$$

The lower bound is maintained if (3.4.43) or (3.4.44) is satisfied. For the upper bound one has to distinguish the cases $d_{\beta\gamma} > 1/2$ and $d_{\beta\gamma} < 1/2$. Consider the first case and apply the definitions of the Hamming distances

$$\begin{aligned}
&(1-2d_{\alpha\beta})(1-2d_{\alpha\gamma}) \leq 1 - 2d_{\beta\gamma},\\
&d_{\alpha\beta} + d_{\alpha\gamma} - d_{\beta\gamma} - 2d_{\alpha\beta}d_{\alpha\gamma} \geq 0,\\
&m_{100} - 2(m_{001}m_{010} + m_{001}m_{100} + m_{010}m_{100} + m_{100}^2) \geq 0,\\
&m_{000}m_{100} - m_{001}m_{010} \geq 0.
\end{aligned}$$

Similarly, for $d_{\beta\gamma} < 1/2$ one gets

$$m_{000}m_{100} - m_{001}m_{010} \leq 0.$$

Transferring these results to the covariance conditions presented (3.4.43) and (3.4.44) yields the following cases:

$$\begin{aligned}
&m_{000}m_{001} \geq m_{010}m_{100}, \quad &&m_{000}m_{010} \geq m_{001}m_{100}, \quad &&m_{000}m_{100} \geq m_{001}m_{010},\\
&m_{000}m_{001} \leq m_{010}m_{100}, \quad &&m_{000}m_{010} \leq m_{001}m_{100}, \quad &&m_{000}m_{100} \geq m_{001}m_{010},\\
&m_{000}m_{001} \leq m_{010}m_{100}, \quad &&m_{000}m_{010} \geq m_{001}m_{100}, \quad &&m_{000}m_{100} \leq m_{001}m_{010},\\
&m_{000}m_{001} \geq m_{010}m_{100}, \quad &&m_{000}m_{010} \leq m_{001}m_{100}, \quad &&m_{000}m_{100} \leq m_{001}m_{010},
\end{aligned}$$

dependent on the signs of the Hamming distances. In all cases one probability is always on the larger side of the inequalities whereas the other three probabilities are twice on the smaller side. Thus, two probabilities are allowed to be zero and positivity is not necessary for admissibility of a solution. This completes the proof. □

Proof of Corollary 3.2.3. The implications concerning the Hamming distances immediately follow from Theorem 3.1.7 with (3.4.42). For the case study consider (3.2.1) first in the following updated form:

$$\begin{aligned} 2m_{001} &= p_\gamma + p_\alpha p_\beta - p_\alpha p_\gamma - p_\beta p_\gamma, \\ 2m_{010} &= p_\beta + p_\alpha p_\gamma - p_\alpha p_\beta - p_\beta p_\gamma, \\ 2m_{100} &= p_\alpha + p_\beta p_\gamma - p_\alpha p_\beta - p_\alpha p_\gamma. \end{aligned} \tag{3.4.45}$$

To start the verification of the cases start with the assumption $d_{\alpha\beta} = d_{\alpha\gamma} = 1/2$ and $d_{\beta\gamma} \neq 1/2$. These assumptions return $m_{001} = m_{010}$ and thus, $d_{\beta\gamma} = 4m_{001}$. Applying (3.2.3) to (3.4.45) yields the system:

$$0 = (1-2p_\alpha)(1-2p_\beta), \quad 0 = (1-2p_\alpha)(1-2p_\gamma), \quad 1 - 8m_{001} = (1-2p_\beta)(1-2p_\gamma).$$

Consider the possible cases: If $p_\alpha = 1/2$ and $p_\beta \neq 1/2$, $p_\gamma \neq 1/2$ the latter parameters have the following relationship which is derived from the equality $1 - 8m_{001} = (1-2p\beta)(1-2p_\gamma)$:

$$p_\beta = \frac{4m_{001} - p_\gamma}{1-2p_\gamma} = \frac{d_{\beta\gamma} - p_\gamma}{1-2p_\gamma}.$$

Defining for $t \in \mathbb{C} \setminus \{1/2\}$ and $y > 0$ the function

$$f(t,y) := \frac{y-t}{1-2t}$$

the subspace of solution vectors $(p_\alpha, p_\beta, p_\gamma)$ for (3.4.45) under the given assumptions $d_{\alpha\beta} = d_{\alpha\gamma} = 1/2$ and $d_{\beta\gamma} \neq 1/2$ is given by

$$\big\{(1/2, t, f(d_{\beta\gamma}, t)) : t \in \mathbb{C} \setminus \{1/2\}\big\} \cup \big\{(1/2, f(d_{\beta\gamma}, t), t) : t \in \mathbb{C} \setminus \{1/2\}\big\}.$$

For the similar cases $d_{\alpha\beta} = d_{\beta\gamma} = 1/2$, $d_{\alpha\gamma} \neq 1/2$ and $d_{\alpha\gamma} = d_{\beta\gamma} = 1/2$, $d_{\alpha\beta} \neq 1/2$ one gets analogue results

If, in addition to p_α, also $p_\beta = 1/2$, then the equation $1 - 8m_{001} = (1-2p\beta)(1-2p_\gamma)$ yields $d_{\beta\gamma} = 1/2$ and thus, $m_{xyz} = 1/8$ for all $x, y, z \in \{0,1\}$, i.e. a uniform leaf distribution is observed. p_γ remains a free parameter.

Conversely assume, $\underline{m}$ is a uniform leaf distribution. Then, all inferred Hamming distances are $1/2$ and with the system

$$0 = (1-2p_\alpha)(1-2p_\beta), \quad 0 = (1-2p_\alpha)(1-2p_\gamma), \quad 0 = (1-2p_\beta)(1-2p_\gamma),$$

inferred from (3.4.45) the associated space of solutions is given by

$$\{(t, 1/2, 1/2) : t \in \mathbb{C}\} \cup \{(1/2, t, 1/2) : t \in \mathbb{C}\} \cup \{(1/2, 1/2, t) : t \in \mathbb{C}\}.$$

This wraps up the proof of the corollary. □

3.4.3 Proofs for Section 3.3

Here, the extension of the results to quartet trees was analyzed.

Proof of Theorem 3.3.1. Consider the quartet tree $\mathcal{T} = (V, E)$ with (3.3.1) and inherit the notation from (3.3.2) and (3.3.3), i.e. consider the triple trees $\mathcal{T}^i = (V^i, E^i)$, and associated triple leaf distributions $\underline{m}^i$, $i = 1, 2, 3, 4$. The inferred parameters established via Theorem 3.1.4 will be indexed in that fashion.

To get compatible parameters one of the following two scenarios must be satisfied

$$p_{00}^{\delta,i} = p_{00}^{\delta,j}, \quad q_0^{\rho,i} = q_0^{\rho,j}, \tag{3.4.46}$$

$$p_{00}^{\delta,i} = p_{10}^{\delta,j}, \quad q_0^{\rho,i} = q_1^{\rho,j} \tag{3.4.47}$$

for $\delta \in \{\alpha_1, \alpha_2\}, \rho = \varrho_1, i = 1, j = 2$ and $\delta \in \{\alpha_3, \alpha_4\}, \rho = \varrho_2, i = 3, j = 4$. For $\delta \in L$, $i = 1, 2, 3, 4$ the following holds:

$$2p_{00}^{\delta,i} = (p_{00}^{\delta,i} + p_{10}^{\delta,i}) + (p_{00}^{\delta,i} - p_{10}^{\delta,i}), \quad 2p_{10}^{\delta,i} = (p_{00}^{\delta,i} + p_{10}^{\delta,i}) - (p_{00}^{\delta,i} - p_{10}^{\delta,i}), \tag{3.4.48}$$

i.e. looking at cases (3.4.46) or (3.4.47) is similar to looking at the differences $p_{00}^{\delta,i} - p_{10}^{\delta,i}$ and sums $p_{00}^{\delta,i} + p_{10}^{\delta,i}, \delta \in L$, $i = 1, 2, 3, 4$. When looking at the structure of (3.1.11) and (3.1.12) one observes for, α_1 say:

$$p_{00}^{\alpha_1,1} + p_{10}^{\alpha_1,1} = -\frac{r_{\alpha_1}^1}{t_{\beta\gamma}}, \quad p_{00}^{\alpha_1,1} - p_{10}^{\alpha_1,1} = \frac{\sqrt{\chi^1}}{t_{\beta\gamma}}.$$

Consider the differences and sums in the light of compatibility. Then, (3.4.46) can be written as:

$$\begin{aligned}
(p_{00}^{\alpha_1,1} + p_{10}^{\alpha_1,1}) + (p_{00}^{\alpha_1,1} - p_{10}^{\alpha_1,1}) &= (p_{00}^{\alpha_1,2} + p_{10}^{\alpha_1,2}) + (p_{00}^{\alpha_1,2} - p_{10}^{\alpha_1,2}),\\
(p_{00}^{\alpha_1,1} + p_{10}^{\alpha_1,1}) - (p_{00}^{\alpha_1,1} - p_{10}^{\alpha_1,1}) &= (p_{00}^{\alpha_1,2} + p_{10}^{\alpha_1,2}) - (p_{00}^{\alpha_1,2} - p_{10}^{\alpha_1,2}),
\end{aligned}$$

and (3.4.47) as:

$$\begin{aligned}
(p_{00}^{\alpha_1,1} + p_{10}^{\alpha_1,1}) + (p_{00}^{\alpha_1,1} - p_{10}^{\alpha_1,1}) &= (p_{00}^{\alpha_1,2} + p_{10}^{\alpha_1,2}) - (p_{00}^{\alpha_1,2} - p_{10}^{\alpha_1,2}),\\
(p_{00}^{\alpha_1,1} + p_{10}^{\alpha_1,1}) - (p_{00}^{\alpha_1,1} - p_{10}^{\alpha_1,1}) &= (p_{00}^{\alpha_1,2} + p_{10}^{\alpha_1,2}) + (p_{00}^{\alpha_1,2} - p_{10}^{\alpha_1,2}).
\end{aligned}$$

Hence, finding conditions for (3.4.46) and (3.4.47) is equivalent to finding conditions for

$$p_{00}^{\delta,i} + p_{10}^{\delta,i} = p_{00}^{\delta,j} + p_{10}^{\delta,j}, \quad |p_{00}^{\delta,i} - p_{10}^{\delta,i}| = |p_{00}^{\delta,j} - p_{10}^{\delta,j}| \tag{3.4.49}$$

for $\delta \in \{\alpha_1, \alpha_2\}, i = 1, j = 2$ and $\delta \in \{\alpha_3, \alpha_4\}, i = 3, j = 4$ and equivalently

$$\frac{r_{\alpha_1}^1}{t_{\alpha_2\alpha_3}} = \frac{r_{\alpha_1}^2}{t_{\alpha_2\alpha_4}}, \quad \frac{r_{\alpha_2}^1}{t_{\alpha_1\alpha_3}} = \frac{r_{\alpha_2}^2}{t_{\alpha_1\alpha_4}}, \quad \left|\frac{\sqrt{\chi_1}}{t_{\alpha_2\alpha_3}}\right| = \left|\frac{\sqrt{\chi_2}}{t_{\alpha_2\alpha_4}}\right|, \quad \left|\frac{\sqrt{\chi_1}}{t_{\alpha_1\alpha_3}}\right| = \left|\frac{\sqrt{\chi_2}}{t_{\alpha_1\alpha_4}}\right|, \tag{3.4.50}$$

$$\frac{r_{\alpha_3}^3}{t_{\alpha_1\alpha_4}} = \frac{r_{\alpha_3}^4}{t_{\alpha_2\alpha_4}}, \quad \frac{r_{\alpha_4}^3}{t_{\alpha_1\alpha_3}} = \frac{r_{\alpha_4}^4}{t_{\alpha_2\alpha_3}}, \quad \left|\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_4}}\right| = \left|\frac{\sqrt{\chi_4}}{t_{\alpha_2\alpha_4}}\right|, \quad \left|\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_3}}\right| = \left|\frac{\sqrt{\chi_4}}{t_{\alpha_2\alpha_3}}\right|. \tag{3.4.51}$$

Recall from (3.1.9) that $t_{\alpha_i\alpha_j} \neq 0$, $i \neq j$ and $\sqrt{\chi_i} \neq 0$ are necessary for the existence of a unique solution. Hence, division or multiplication with these terms does not pose problems to the equalities above.

Look at the following relations:

$$\left|\frac{\sqrt{\chi_1}}{t_{\alpha_2\alpha_3}}\right| = \left|\frac{\sqrt{\chi_2}}{t_{\alpha_2\alpha_4}}\right|, \quad \left|\frac{\sqrt{\chi_1}}{t_{\alpha_1\alpha_3}}\right| = \left|\frac{\sqrt{\chi_2}}{t_{\alpha_1\alpha_4}}\right|, \tag{3.4.52}$$

$$\left|\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_4}}\right| = \left|\frac{\sqrt{\chi_4}}{t_{\alpha_2\alpha_4}}\right|, \quad \left|\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_3}}\right| = \left|\frac{\sqrt{\chi_4}}{t_{\alpha_2\alpha_3}}\right|. \tag{3.4.53}$$

The absolute value implies two possibilities, both sides are positive or a sign change occurred. In particular, for (3.4.52) only the following cases can be observed:

$$\begin{gathered}
\frac{\sqrt{\chi_1}}{t_{\alpha_2\alpha_3}} = \frac{\sqrt{\chi_2}}{t_{\alpha_2\alpha_4}}, \quad \frac{\sqrt{\chi_1}}{t_{\alpha_1\alpha_3}} = \frac{\sqrt{\chi_2}}{t_{\alpha_1\alpha_4}}, \\
\frac{\sqrt{\chi_1}}{t_{\alpha_2\alpha_3}} = -\frac{\sqrt{\chi_2}}{t_{\alpha_2\alpha_4}}, \quad \frac{\sqrt{\chi_1}}{t_{\alpha_1\alpha_3}} = -\frac{\sqrt{\chi_2}}{t_{\alpha_1\alpha_4}}.
\end{gathered}$$

Restructuring yields the following equivalences:

$$\begin{gathered}
\frac{\sqrt{\chi_1}}{\sqrt{\chi_2}} = \frac{t_{\alpha_2\alpha_3}}{t_{\alpha_2\alpha_4}} = \frac{t_{\alpha_1\alpha_3}}{t_{\alpha_1\alpha_4}}, \quad -\frac{\sqrt{\chi_1}}{\sqrt{\chi_2}} = \frac{t_{\alpha_2\alpha_3}}{t_{\alpha_2\alpha_4}} = \frac{t_{\alpha_1\alpha_3}}{t_{\alpha_1\alpha_4}}, \\
\frac{\sqrt{\chi_3}}{\sqrt{\chi_4}} = \frac{t_{\alpha_1\alpha_3}}{t_{\alpha_2\alpha_3}} = \frac{t_{\alpha_1\alpha_4}}{t_{\alpha_2\alpha_4}}, \quad -\frac{\sqrt{\chi_3}}{\sqrt{\chi_4}} = \frac{t_{\alpha_1\alpha_3}}{t_{\alpha_2\alpha_3}} = \frac{t_{\alpha_1\alpha_4}}{t_{\alpha_2\alpha_4}}
\end{gathered}$$

and therefore,

$$t_{\alpha_2\alpha_3} t_{\alpha_1\alpha_4} = t_{\alpha_2\alpha_4} t_{\alpha_1\alpha_3}. \tag{3.4.54}$$

Next, consider the following equality and apply the conditions from (3.4.50) and (3.4.54):

$$\begin{aligned}
\frac{\chi_1}{t^2_{\alpha_2\alpha_3}} &= \frac{1}{t^2_{\alpha_2\alpha_3}}(r^1_{\alpha_1} + 2m_{0\Sigma\Sigma\Sigma} t_{\alpha_2\alpha_3})^2 + 4t_{\alpha_1\alpha_2}\frac{t_{\alpha_1\alpha_3}}{t_{\alpha_2\alpha_3}} \\
&= \left(\frac{r^1_{\alpha_1}}{t_{\alpha_2\alpha_3}} + 2m_{0\Sigma\Sigma\Sigma}\right)^2 + 4t_{\alpha_1\alpha_2}\frac{t_{\alpha_1\alpha_3}}{t_{\alpha_2\alpha_3}} \\
&= \left(\frac{r^2_{\alpha_1}}{t_{\alpha_2\alpha_4}} + 2m_{0\Sigma\Sigma\Sigma}\frac{t_{\alpha_2\alpha_4}}{t_{\alpha_2\alpha_4}}\right)^2 + 4t_{\alpha_1\alpha_2}\frac{t_{\alpha_1\alpha_4}}{t_{\alpha_2\alpha_4}} = \frac{\chi_2}{t^2_{\alpha_2\alpha_4}}.
\end{aligned}$$

Hence, the conditions (3.4.54) and

$$\frac{r^1_{\alpha_1}}{t_{\alpha_2\alpha_3}} = \frac{r^2_{\alpha_1}}{t_{\alpha_2\alpha_4}}, \quad \frac{r^1_{\alpha_2}}{t_{\alpha_1\alpha_3}} = \frac{r^2_{\alpha_2}}{t_{\alpha_1\alpha_4}}, \tag{3.4.55}$$

already contain the quadratic notion of (3.4.52). Analogously, equations

$$\frac{r^3_{\alpha_3}}{t_{\alpha_1\alpha_4}} = \frac{r^4_{\alpha_3}}{t_{\alpha_2\alpha_4}}, \quad \frac{r^3_{\alpha_4}}{t_{\alpha_1\alpha_3}} = \frac{r^4_{\alpha_4}}{t_{\alpha_2\alpha_3}} \tag{3.4.56}$$

contain the quadratic notion of (3.4.53), and hence (3.4.52) and (3.4.53) are redundant.

For a satisfying answer also the root distribution must be considered. First, assume $q_0^{\varrho_1,1} = q_0^{\varrho_1,2}$. Then $t_{\alpha_2\alpha_3}\sqrt{\chi_2} = t_{\alpha_2\alpha_4}\sqrt{\chi_1}$ and one computes:

$$\begin{aligned} q_0^{\varrho_1,1} - q_0^{\varrho_1,2} &= \frac{1}{2} - \frac{r^1_{\alpha_1} + 2m_{0\Sigma\Sigma\Sigma}t_{\alpha_2\alpha_3}}{2\sqrt{\chi_1}} - \frac{1}{2} + \frac{r^2_{\alpha_1} + 2m_{0\Sigma\Sigma\Sigma}t_{\alpha_2\alpha_4}}{2\sqrt{\chi_2}} \\ &= \frac{r^2_{\alpha_1}}{2\sqrt{\chi_2}} - \frac{r^1_{\alpha_1}}{2\sqrt{\chi_1}} + m_{0\Sigma\Sigma\Sigma}\underbrace{\left(\frac{t_{\alpha_2\alpha_4}}{\sqrt{\chi_2}} - \frac{t_{\alpha_2\alpha_3}}{\sqrt{\chi_1}}\right)}_{=0} \\ &= \frac{r^2_{\alpha_1}}{t_{\alpha_2\alpha_4}}\frac{t_{\alpha_2\alpha_4}}{2\sqrt{\chi_2}} - \frac{r^1_{\alpha_1}}{2\sqrt{\chi_1}} = \frac{r^1_{\alpha_1}}{t_{\alpha_2\alpha_3}}\frac{t_{\alpha_2\alpha_3}}{2\sqrt{\chi_1}} - \frac{r^1_{\alpha_1}}{2\sqrt{\chi_1}} = 0. \end{aligned}$$

Now assume $q_0^{\varrho_1,1} = q_1^{\varrho_1,2}$. Then $t_{\alpha_2\alpha_3}\sqrt{\chi_2} = -t_{\alpha_2\alpha_4}\sqrt{\chi_1}$ and one computes:

$$\begin{aligned} q_0^{\varrho_1,1} - q_1^{\varrho_1,2} &= \frac{1}{2} - \frac{r^1_{\alpha_1} + 2m_{0\Sigma\Sigma\Sigma}t_{\alpha_2\alpha_3}}{2\sqrt{\chi_1}} - \frac{1}{2} - \frac{r^2_{\alpha_1} + 2m_{0\Sigma\Sigma\Sigma}t_{\alpha_2\alpha_4}}{2\sqrt{\chi_2}} \\ &= -\left(\frac{r^1_{\alpha_1}}{2\sqrt{\chi_1}} + \frac{r^2_{\alpha_1}}{2\sqrt{\chi_2}}\right) - m_{0\Sigma\Sigma\Sigma}\underbrace{\left(\frac{t_{\alpha_2\alpha_3}}{\sqrt{\chi_1}} + \frac{t_{\alpha_2\alpha_4}}{\sqrt{\chi_2}}\right)}_{=0} \\ &= -\left(\frac{r^1_{\alpha_1}}{t_{\alpha_2\alpha_3}}\frac{t_{\alpha_2\alpha_3}}{2\sqrt{\chi_1}} + \frac{r^2_{\alpha_1}}{2\sqrt{\chi_2}}\right) = -\left(-\frac{r^2_{\alpha_1}}{t_{\alpha_2\alpha_4}}\frac{t_{\alpha_2\alpha_4}}{2\sqrt{\chi_2}} + \frac{r^2_{\alpha_1}}{2\sqrt{\chi_2}}\right) = 0. \end{aligned}$$

Since analogue computations for $q^{\varrho_2,3}$ and $q^{\varrho_2,4}$ yield analogue results, conditions (3.4.54) and (3.4.55) also imply the equality of the root distributions.

Finally, the transition parameters for edge (ϱ_1, ϱ_2) need to be derived. Other equivalences will arise during these calculations. The following equivalences must be observed:

$$\begin{aligned} p_{00}^{\alpha_3,1} &= p_{00}^{\varrho_1\varrho_2}p_{00}^{\alpha_3,3} + (1 - p_{00}^{\varrho_1\varrho_2})p_{10}^{\alpha_3,3}, \\ p_{10}^{\alpha_3,1} &= p_{10}^{\varrho_1\varrho_2}p_{00}^{\alpha_3,3} + (1 - p_{10}^{\varrho_1\varrho_2})p_{10}^{\alpha_3,3}. \end{aligned}$$

Equivalently, the parameters must satisfy:

$$\begin{aligned} p_{00}^{\alpha_4,2} &= p_{00}^{\varrho_1\varrho_2}p_{00}^{\alpha_4,3} + (1 - p_{00}^{\varrho_1\varrho_2})p_{10}^{\alpha_4,3}, \\ p_{10}^{\alpha_4,2} &= p_{10}^{\varrho_1\varrho_2}p_{00}^{\alpha_4,3} + (1 - p_{10}^{\varrho_1\varrho_2})p_{10}^{\alpha_4,3}. \end{aligned}$$

Hence, the following equivalences must be satisfied:

$$p_{00}^{\varrho_1\varrho_2} = \frac{p_{00}^{\alpha_3,1} - p_{10}^{\alpha_3,3}}{p_{00}^{\alpha_3,3} - p_{10}^{\alpha_3,3}} = \frac{p_{00}^{\alpha_4,2} - p_{10}^{\alpha_4,3}}{p_{00}^{\alpha_4,3} - p_{10}^{\alpha_4,3}},$$
$$p_{10}^{\varrho_1\varrho_2} = \frac{p_{10}^{\alpha_3,1} - p_{10}^{\alpha_3,3}}{p_{00}^{\alpha_3,3} - p_{10}^{\alpha_3,3}} = \frac{p_{10}^{\alpha_4,2} - p_{10}^{\alpha_4,3}}{p_{00}^{\alpha_4,3} - p_{10}^{\alpha_4,3}}.$$

To verify the equality, insert the representations from (3.1.11) and (3.1.12) and regard the difference of the terms depending on α_3 and α_4, respectively:

$$\begin{aligned}
&\frac{\frac{r^3_{\alpha_3}+\sqrt{\chi_3}}{2t_{\alpha_1\alpha_4}} - \frac{r^1_{\alpha_3}-\sqrt{\chi_1}}{2t_{\alpha_1\alpha_2}}}{\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_4}}} - \frac{\frac{r^3_{\alpha_4}+\sqrt{\chi_3}}{2t_{\alpha_1\alpha_3}} - \frac{r^2_{\alpha_4}-\sqrt{\chi_2}}{2t_{\alpha_1\alpha_2}}}{\frac{\sqrt{\chi_3}}{t_{\alpha_1\alpha_3}}}\\
&= \frac{r^3_{\alpha_3} - r^3_{\alpha_4}}{2\sqrt{\chi_3}} + \frac{t_{\alpha_1\alpha_3}(r^2_{\alpha_4} - \sqrt{\chi_2}) - t_{\alpha_1\alpha_4}(r^1_{\alpha_3} - \sqrt{\chi_1})}{2t_{\alpha_1\alpha_2}\sqrt{\chi_3}}\\
&= \frac{t_{\alpha_1\alpha_2}(r^3_{\alpha_3} - r^3_{\alpha_4}) - (t_{\alpha_1\alpha_4}r^1_{\alpha_3} - t_{\alpha_1\alpha_3}r^2_{\alpha_4})}{2\sqrt{\chi_3}} \overset{!}{=} 0.
\end{aligned}$$

Using the notions from (3.1.11) and (3.1.12) yields the following equality:

$$t_{\alpha_1\alpha_2}(r^3_{\alpha_3} - r^3_{\alpha_4}) - t_{\alpha_1\alpha_4}r^1_{\alpha_3} + t_{\alpha_1\alpha_3}r^2_{\alpha_4} = t_{\alpha_1\alpha_3}r^2_{\alpha_2} - t_{\alpha_1\alpha_4}r^1_{\alpha_2}.$$

But the right hand side of this equation is zero with (3.4.55), and thus this condition already guarantees the compatibility of the parameter for the inner edge. This completes the proof. □

Proof of Corollary 3.3.2. The proof contains two steps. Step one is the consideration of the invariants from (3.3.4) under the N_2 model and step two is the derivation of the parameter for the inner edge.

For step one, recall from (3.4.42) the following equalities:

$$r^1_{\alpha_1} = -t_{\alpha_2\alpha_3}, \quad r^2_{\alpha_1} = -t_{\alpha_2\alpha_4},$$

and analogue equalities for α_2, α_3 and α_4. These equalities yield:

$$r^1_{\alpha_1}t_{\alpha_2\alpha_4} - r^2_{\alpha_1}t_{\alpha_2\alpha_3} = 0,$$

i.e. the first four invariants are zero due to the model properties. The fifth invariant yields:

(3.4.57)
$$\begin{aligned}
&t_{\alpha_1\alpha_3}t_{\alpha_2\alpha_4} - t_{\alpha_1\alpha_4}t_{\alpha_2\alpha_3}\\
&= (1-2d_{\alpha_1\alpha_3})(1-2d_{\alpha_2\alpha_4}) - (1-2d_{\alpha_1\alpha_4})(1-2d_{\alpha_2\alpha_3})\\
&= 4(d_{\alpha_1\alpha_3}d_{\alpha_2\alpha_4} - d_{\alpha_1\alpha_4}d_{\alpha_2\alpha_3})\\
&\quad - 2(d_{\alpha_1\alpha_3} + d_{\alpha_2\alpha_4} - d_{\alpha_1\alpha_4} - d_{\alpha_2\alpha_3}).
\end{aligned}$$

With (3.3.7) the Hamming distances have the following form:

$$\begin{aligned}
d_{\alpha_1\alpha_3} &= 2(m_{0010} + m_{0011} + m_{0110} + m_{0111}),\\
d_{\alpha_1\alpha_4} &= 2(m_{0001} + m_{0011} + m_{0101} + m_{0111}),\\
d_{\alpha_2\alpha_3} &= 2(m_{0010} + m_{0011} + m_{0100} + m_{0101}),\\
d_{\alpha_2\alpha_4} &= 2(m_{0001} + m_{0011} + m_{0100} + m_{0110}).
\end{aligned}$$

With this observation, equation (3.4.57) reduces to:

$$t_{\alpha_1\alpha_3}t_{\alpha_2\alpha_4} - t_{\alpha_1\alpha_4}t_{\alpha_2\alpha_3} = 4(d_{\alpha_1\alpha_3}d_{\alpha_2\alpha_4} - d_{\alpha_1\alpha_4}d_{\alpha_2\alpha_3}),$$

and thus, (3.3.8) is derived.

For the inner edge recall the matrix product $P^{\varrho_1\alpha_3} = P^{\varrho_1\varrho_2}P^{\varrho_2\alpha_3}$. This equation results in:

$$p_{\varrho_1\alpha_3} = p_{\varrho_1\varrho_2}(1 - p_{\varrho_2\alpha_3}) + p_{\varrho_2\alpha_3}(1 - p_{\varrho_1\varrho_2}),$$

and therefore:

$$\begin{aligned}
p_{\varrho_1\varrho_2} &= \frac{p_{\varrho_1\alpha_3} - p_{\varrho_2\alpha_3}}{1 - 2p_{\varrho_2\alpha_3}} = \frac{\frac{\Delta_3}{2(1-2d_{\alpha_1\alpha_4})} - \frac{\Delta_1}{2(1-2d_{\alpha_1\alpha_2})}}{\frac{\Delta_3}{1-2d_{\alpha_1\alpha_4}}}\\
&= \frac{1}{2}\left(1 - \frac{\Delta_1(1-2d_{\alpha_1\alpha_4})}{\Delta_3(1-2d_{\alpha_1\alpha_2})}\right) = \frac{1}{2}\left(1 - \sqrt{\frac{(1-2d_{\alpha_1\alpha_4})(1-2d_{\alpha_2\alpha_3})}{(1-2d_{\alpha_1\alpha_2})(1-2d_{\alpha_3\alpha_4})}}\right).
\end{aligned}$$

With (3.3.8) the matrix product $P^{\varrho_1\alpha_4} = P^{\varrho_1\varrho_2}P^{\varrho_2\varrho_4}$ provides the same result. Therefore, no new invariant is obtained from this computations, and the proof is completed. □

Proof of Lemma 3.3.3. From the construction of the applied function of **Mathematica** follows, that the presented polynomials are indeed a basis for the needed elimination ideal. Equivalence of the statements follows from Theorem 2.2.2. □

Proof of Theorem 3.3.4. Having the conditions of Theorem 3.1.5 it remains to compute:

$$0 \le p_{xy}^{\varrho_1\varrho_2} \le 1.$$

Inserting the acquired notion of Theorem 3.3.1 yields for $p_{00}^{\varrho_1\varrho_2}$

$$\begin{aligned}
0 \le \frac{1}{2} - \frac{r_\gamma^{(i)}t_{\alpha\delta} - r_\gamma^{(iii)}t_{\alpha\beta}}{2t_{\alpha\beta}\sqrt{\chi^{(iii)}}} + \frac{t_{\alpha\delta}\sqrt{\chi^{(i)}}}{2t_{\alpha\beta}\sqrt{\chi^{(iii)}}} \le 1\\
-\frac{1}{2} \le -\frac{t_{\alpha\delta}(r_\gamma^{(i)} - \sqrt{\chi^{(i)}})}{2t_{\alpha\beta}\sqrt{\chi^{(iii)}}} + \frac{r_\gamma^{(iii)}}{2\sqrt{\chi^{(iii)}}} \le \frac{1}{2}\\
-\frac{r_\gamma^{(iii)} + \sqrt{\chi^{(iii)}}}{2t_{\alpha\delta}} \le -\frac{r_\gamma^{(i)} - \sqrt{\chi^{(i)}}}{2t_{\alpha\beta}} \le -\frac{r_\gamma^{(iii)} - \sqrt{\chi^{(iii)}}}{2t_{\alpha\delta}}.
\end{aligned}$$

The calculated terms are the notions for the transition probabilities from Theorem 3.1.4. Analogous computations for $p_{10}^{\varrho_1\varrho_2}$ yield the same bounds, i.e. condition (3.3.9) is verified.

The proof of Theorem 3.3.1 shows, that this condition translates to all leaves, i.e. also

$$\begin{aligned}
\min\{p_{00}^{\alpha,(i)}, p_{10}^{\alpha,(i)}\} \leq p_{00}^{\alpha,(iii)}, p_{10}^{\alpha,(iii)} \leq \max\{p_{00}^{\alpha,(i)}, p_{10}^{\alpha,(i)}\},\\
\min\{p_{00}^{\beta,(i)}, p_{10}^{\beta,(i)}\} \leq p_{00}^{\beta,(iv)}, p_{10}^{\beta,(iv)} \leq \max\{p_{00}^{\beta,(i)}, p_{10}^{\beta,(i)}\},\\
\min\{p_{00}^{\delta,(iii)}, p_{10}^{\delta,(iii)}\} \leq p_{00}^{\delta,(ii)}, p_{10}^{\delta,(ii)} \leq \max\{p_{00}^{\delta,(iii)}, p_{10}^{\delta,(iii)}\}
\end{aligned}$$

must be fulfilled. □

Proof of Lemma 3.3.5. According to Proposition 1.2.5 the restrictions of a Markov process on a tree $\mathcal{T}$ to the triple trees derived from $\mathcal{T}$ are again Markov processes. Hence, these restrictions need to satisfy Theorem 3.1.5 as well. Now, consider a tree $\mathcal{T}$ with $n > 3$ leaves. Whenever one assembles a triple tree the process must satisfy the sign conditions. In particular, denote by δ_1, δ_2 and δ_3 the selected leaves. Then either all leaves are positively correlated or two are positively correlated to each other and negatively correlated to the third. Now consider a fourth leaf δ_4 and the four triples generated by the four leaves. The following scenarios are compatible with Theorem 3.1.5:

1. $t_{\delta_1\delta_2}, t_{\delta_1\delta_3}, t_{\delta_2\delta_3} > 0$ and either
$$t_{\delta_i\delta_4} > 0 \quad \text{or} \quad t_{\delta_i\delta_4} < 0 \quad \text{for } i \in \{1,2,3\}.$$
2. $t_{\delta_1\delta_2} > 0$ and $t_{\delta_i\delta_3} < 0$, $i = 1,2$ and either
$$\begin{aligned}
t_{\delta_i\delta_4} > 0 \quad \text{and} \quad t_{\delta_3\delta_4} < 0,\ i = 1,2 \quad \text{or}\\
t_{\delta_i\delta_4} < 0 \quad \text{and} \quad t_{\delta_3\delta_4} > 0,\ i = 1,2.
\end{aligned}$$

Any selection of three leaves under these conditions will satisfy the sign condition for the covariance terms. Moreover, there is no other scenario that is admissible. Looking closely at the cases reveals that they always agree with the statements of Lemma 3.3.5, one always observes two sets - one of them possibly empty - which satisfy (3.3.10). This completes the proof. □

Chapter 4

Stochastic Models of Molecular Evolution in k States

The previous chapter analyzed the general two state model on a triple tree. However, looking at models with a larger number of possible states provides a better relationship to molecular evolution. To give a glimpse at the difficulties of this task , two simple symmetrical models are examined on triple trees, the Neyman N_k and the Kimura 2ST model. The obtained solutions depend on the pairwise leaf distributions only, but still need three leaves to be formed. Moreover, the number of needed phylogenetic invariants rises considerably.

Each section starts with the presentation of the basic model properties, followed by the derivation of conditions on leaf distributions for a Markov extension, the computation of characterizations for extensions and conditions for their model relevance. The sections close out with the transfer of the results to the time-continuous model specifications of rates and molecular clock. The proofs to all presented results are given at the end of the chapter.

If not defined otherwise, the basic structure on which the models are discussed, is the triple tree $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ with

$$\mathcal{V} := \{\alpha, \beta, \gamma, \varrho\}, \quad \mathcal{E} := \{(\varrho, \alpha), (\varrho, \beta), (\varrho, \gamma)\} \tag{4.0.1}$$

with leaf set $\mathcal{L} := \{\alpha, \beta, \gamma\}$.

4.1 The Neyman N_k Model

The Neyman N_k model is introduced in Example 1.3.2. Here, each edge of $\mathcal{T}$ is assigned one parameter, and the root distribution is stationary.

The following section starts with the presentation phylogenetic invariants for the existence of an algebraic solution of (LF) w.r.t. a triple leaf distribution under the model. Next, an explicit description of all possible Markov-like extensions to a given leaf distribution, and conditions for stochastic admissibility are computed. Finally, a transfer of the results to the rate model and its specification, the molecular clock, is undertaken. Throughout this section, let $k \geq 2$ and the state set is given as $\mathcal{S} := \{0, 1, \ldots, k-1\}$.

4.1.1 Basic Model Properties

Let $\underline{\mu} := (\mu_{uxyz})_{u,x,y,z \in \mathcal{S}}$ denote a Markov distribution on $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with (4.0.1) w.r.t. the Neyman N_k model, i.e. $\underline{\mu}$ is subject to (LF) and has the following properties:

(4.1.1) $\quad \mu_u := \mathbb{P}(X_\varrho = u) = 1/k \quad$ for all $u \in \mathcal{S}$,

(4.1.2) $\quad p_\delta := \mu_{x|u} = \mathbb{P}(X_\delta = x | X_\varrho = u) \quad$ for all $x, u \in \mathcal{S}$ and $\delta \in \mathcal{L}$.

Property (4.1.1) describes the stationarity of any marginal distribution, i.e. each state has the same probability to occur at a site of the sequence for a vertex. The transition is considered in (4.1.2), where the kind of transition is not distinguished, and only its occurrence is observed.

The properties show that a Markov distribution on $\mathcal{T}$ subject to the Neyman model is fully determined by a triple $(p_\alpha, p_\beta, p_\gamma)$. Observe that the transition probability p_δ cannot exceed $1/(k-1)$ in order to be an element of a transition matrix. The model properties are now used to define a Neyman extension.

Definition 4.1.1. *Let $\underline{m} := (m_{xyz})_{x,y,z \in \mathcal{S}}$ denote a leaf distribution on $\mathcal{T}$. A Markov distribution $\mu := (\mu_{uxyz})_{u,x,y,z \in \mathcal{S}}$ on $\mathcal{T}$ subject to the Neyman model with*

$$m_{xyz} = \sum_{u \in \mathcal{S}} \mu_{uxyz} \quad \text{for } x, y, z \in \mathcal{S},$$

is called Neyman extension *to $\underline{m}$ on $\mathcal{T}$.*

Applying properties (4.1.1) and (4.1.2) to the system (LF) shows that a triple leaf distribution with a Neyman extension has to obey certain relations. These relations are presented in the following lemma:

Lemma 4.1.1. *Let $k \geq 3$, and let $\underline{m}$ denote a leaf distribution on $\mathcal{L}$. If $\underline{m}$ has a Neyman extension μ on $\mathcal{T}$, it satisfies the following conditions for $x \neq y \neq z \in \mathcal{S}$:*

(4.1.3) $\quad m_{xxx} = m_{000},\ m_{xxy} = m_{001},\ m_{xyx} = m_{010},\ m_{yxx} = m_{100},\ m_{xyz} = m_{012}.$

Thus, a triple leaf distribution subject to a Neyman extension is characterized by five values. The summation condition for probabilities yields the following relationship:

$$k\,m_{000} + k(k-1)(m_{001} + m_{010} + m_{100} + (k-2)m_{012}) = 1. \tag{4.1.4}$$

These observations lead to the following insight:

Lemma 4.1.2. *Let $\underline{m}$ denote a leaf distribution with Markov extension μ. Then, finding a characterization of μ under the Neyman model by solving (LF) is equivalent to solving*

$$\begin{aligned}
k\,m_{000} &= (1-(k-1)p_\alpha)(1-(k-1)p_\beta)(1-(k-1)p_\gamma) + (k-1)p_\alpha p_\beta p_\gamma,\\
k\,m_{001} &= (1-(k-1)p_\alpha)(1-(k-1)p_\beta)p_\gamma + p_\alpha p_\beta(1-p_\gamma),\\
k\,m_{010} &= (1-(k-1)p_\alpha)p_\beta(1-(k-1)p_\gamma) + p_\alpha(1-p_\beta)p_\gamma,\\
k\,m_{100} &= p_\alpha(1-(k-1)p_\beta)(1-(k-1)p_\gamma) + (1-p_\alpha)p_\beta p_\gamma,\\
k\,m_{012} &= p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma - 2kp_\alpha p_\beta p_\gamma.
\end{aligned} \tag{4.1.5}$$

System (4.1.5) consists of five equations in three variables. For $k = 2$ one will find that

$$2(m_{000} + m_{001} + m_{010} + m_{100}) = 1,$$

i.e. m_{012} is zero under the N_2 model. This should be kept in mind for the following considerations.

According to Chapter 2, a polynomial basis in $m_{000}, m_{001}, m_{010}, m_{100}$ and m_{012} is needed to obtain a characterization of triple leaf distributions with a solution to (4.1.5). Proposition 2.6.4 states that the dimension of the algebraic variety of such distributions is equal to the dimension of the tangent space in a simple point. Using this approach yields a lower bound for the number of polynomials in the basis:

Lemma 4.1.3. *The dimension of the variety of triple leaf distributions with a Neyman-like extension is two.*

Thus at least two polynomials are needed to describe this variety. The software **Singular** (see Greuel et al. [2001]) provides the package *elim* which generates an elimination ideal for a given set of polynomials. Applying this package to (4.1.5) yields the following result, where $x_1 = m_{000}$, $x_2 = m_{001}$, $x_3 = m_{010}$, $x_4 = m_{100}$, $x_5 = m_{012}$:

$$\begin{aligned}
N_1^k(x_1,\dots,x_5) &= kx_1 + k(k-1)(x_2 + x_3 + x_4 + (k-2)x_5) - 1,\\
N_2^k(x_2,\dots,x_5) &= k^3(x_2^2x_3 + x_2^2x_4 + x_2x_3^2 + x_2x_4^2 + x_3^2x_4 + x_3x_4^2) + 2k^3x_2x_3x_4\\
&\quad + (k-2)k^3x_5(x_2^2 + x_3^2 + x_4^2) + (k-2)^3k^3x_5^3 - k(3-6k+2k^2)x_5^2\\
&\quad + x_5 + k(3(k-2)k^2x_5 - 1)(x_2x_3 + x_2x_4 + x_3x_4)\\
&\quad + 2kx_5((k-2)^2k^2x_5 - k + 1)(x_2 + x_3 + x_4).
\end{aligned}$$

It is possible to find alternative sets of polynomials, but they will not change the algebraic variety for $(x_1, x_2, x_3, x_4, x_5)$. Obviously, (4.1.4) and N_1^k are equivalent, i.e. a leaf distribution $\underline{m}$ that obeys (4.1.3) satisfies $N_1^k(\underline{m}) = 0$. For the existence of a solution of (LF) w.r.t. $\underline{m}$, also $N_2^k(\underline{m}) = 0$ must necessarily hold.

Proposition 4.1.4. *Let $\underline{m}$ denote a leaf distribution satisfying (4.1.3). If $\underline{m}$ has a Neyman-like extension, it satisfies*

$$N_1^k(m_{000}, m_{001}, m_{010}, m_{100}, m_{012}) = 0, \tag{4.1.6}$$

$$N_2^k(m_{001}, m_{010}, m_{100}, m_{012}) = 0. \tag{4.1.7}$$

The polynomials N_1^k and N_2^k are phylogenetic invariants for the Neyman N_k model. As indicated by various texts (eg. Allman and Rhodes [2003]) uniqueness of the invariants is not given although the algebraic variety spanned by the polynomials will stay the same. Since all triple leaf distributions in this section need to satisfy (4.1.3), they are also in the algebraic variety of N_1^k. Hence, only (4.1.7) will be cited in later considerations.

4.1.2 An Algebraic Extension

The first step to find a solution for (4.1.5) w.r.t. a triple leaf distribution $\underline{m}$ is to reduce the system to a number of equations equal to the number of variables. Here, the pairwise leaf distributions come into consideration. In Baake [1998] is stated that for symmetrical models of molecular evolution, pairwise leaf distributions are sufficient for a reconstruction of the initial tree. The needed pairwise leaf distributions are derived from $\underline{m}$ through the following summations:

$$\begin{aligned} m_{01\Sigma} &= m_{010} + m_{100} + (k-2)m_{012}, \\ m_{0\Sigma 1} &= m_{001} + m_{100} + (k-2)m_{012}, \\ m_{\Sigma 01} &= m_{001} + m_{010} + (k-2)m_{012}. \end{aligned} \tag{4.1.8}$$

Extending these computations using the equalities from (4.1.5) yields the following system:

$$\begin{aligned} k\, m_{01\Sigma} &= (1-(k-1)p_\alpha)p_\beta + p_\alpha(1-(k-1)p_\beta) + (k-2)p_\alpha p_\beta, \\ k\, m_{0\Sigma 1} &= (1-(k-1)p_\alpha)p_\gamma + p_\alpha(1-(k-1)p_\gamma) + (k-2)p_\alpha p_\gamma, \\ k\, m_{\Sigma 01} &= (1-(k-1)p_\beta)p_\gamma + p_\beta(1-(k-1)p_\gamma) + (k-2)p_\beta p_\gamma, \end{aligned} \tag{4.1.9}$$

For writing purpose denote the set of pairwise leaf distributions to a given triple leaf distribution $\underline{m}$ by $\underline{m}^P$, i.e.:

$$\underline{m}^P := \{(m_{xy\Sigma})_{x,y \in \mathcal{S}}, (m_{x\Sigma z})_{x,z \in \mathcal{S}}, (m_{\Sigma yz})_{y,z \in \mathcal{S}}\}. \tag{4.1.10}$$

Concerning the relationship of solutions for (4.1.5) and (4.1.9) resp. the following statements are valid.

Lemma 4.1.5. *Let $\underline{m}$ denote a triple leaf distribution and $\underline{m}^P$ its associated set of pairwise leaf distributions. Any solution of system (4.1.5) w.r.t. $\underline{m}$ is a solution of (4.1.9) w.r.t $\underline{m}^P$. Conversely, any solution of (4.1.9) w.r.t. $\underline{m}^P$ is a solution of (4.1.5) w.r.t. $\underline{m}$, if $\underline{m}$ obeys (4.1.6) and (4.1.7).*

Due to this observation the establishment of a solution to (4.1.9) provides a solution for (4.1.5) under the conditions of (4.1.6) and (4.1.7). The next result gives an idea about the number of possible solutions to (4.1.9).

Proposition 4.1.6. *Let $\underline{m}$ denote a triple leaf distribution and $\underline{m}^P$ its associated set of pairwise leaf distributions. If the vector $(p_\alpha, p_\beta, p_\gamma)$ is a solution of (4.1.9) w.r.t. $\underline{m}^P$, then $(\hat{p}_\alpha, \hat{p}_\beta, \hat{p}_\gamma)$ with $\hat{p}_\delta = 2/k - p_\delta$, $\delta \in \mathcal{L}$ is also a solution of (4.1.9).*

This result shows that two different solutions of (4.1.9) w.r.t. a set of pairwise leaf distributions $\underline{m}^P$ can be identified by one vector $(p_\alpha, p_\beta, p_\gamma)$. Hence, an algebraic solution $(p_\alpha, p_\beta, p_\gamma)$ to (4.1.9) can only be *unique up to duplicity.* When applying the insights to the initial system (4.1.5), one finds that the observed symmetry is only subject to (4.1.9). For the triple system (4.1.5) an initial distribution $\underline{m}$ will be recovered at most once. The verification of this observation will be found later in this section.

The generation of an explicit solution to (4.1.9) unearthed a well-known quantity:

Definition 4.1.2. *The* Hamming distance *between α and β is defined by*

$$d_{\alpha\beta} := \sum_{x \neq y} m_{xy\Sigma} = k(k-1)m_{01\Sigma}. \tag{4.1.11}$$

The Hamming distances $d_{\alpha\gamma}$ and $d_{\beta\gamma}$ are defined analogously.

In this section, the Hamming distance mainly occurs in the following term:

$$\tilde{d}_{\delta_1\delta_2} = 1 - \frac{k}{k-1} d_{\delta_1\delta_2}, \quad \delta_1, \delta_2 \in \mathcal{L}. \tag{4.1.12}$$

A closer examination of (4.1.12) for α, β yields in combination with (4.1.11) that $\tilde{d}_{\alpha\beta} = 1 - k^2 m_{01\Sigma}$. In addition, if the summation property

$$1 = \sum_{x,y \in \mathcal{S}} m_{xy\Sigma} = k\, m_{00\Sigma} + k(k-1)m_{01\Sigma}$$

is considered, (4.1.12) can be written as:

$$\tilde{d}_{\alpha\beta} = 1 - k(k-1)m_{01\Sigma} - k\, m_{01\Sigma} = k(m_{00\Sigma} - m_{01\Sigma}).$$

This observation suggests a treatment of $\tilde{d}_{\delta_1\delta_2}$ as the *similarity-dissimilarity-difference* of leaves δ_1 and δ_2. The above proportions and definitions are used for the following result:

Theorem 4.1.7. *Let $\underline{m}$ denote a leaf distribution on $\mathcal{L}$ which satisfies (4.1.3) and let $\underline{m}^P$ denote the associated set of pairwise leaf distributions. Then, system (4.1.9) has a unique solution up to duplicity w.r.t $\underline{m}^P$, if for each pair δ_1, δ_2 of leaves the associated Hamming distance satisfies*

$$d_{\delta_1\delta_2} \neq \frac{k-1}{k}. \tag{4.1.13}$$

The extension is determined by

$$p_\alpha = \frac{1}{k}\left(1 \pm \frac{\Delta}{\tilde{d}_{\beta\gamma}}\right), \quad p_\beta = \frac{1}{k}\left(1 \pm \frac{\Delta}{\tilde{d}_{\alpha\gamma}}\right), \quad p_\gamma = \frac{1}{k}\left(1 \pm \frac{\Delta}{\tilde{d}_{\alpha\beta}}\right), \tag{4.1.14}$$

where

$$\Delta := \sqrt{\tilde{d}_{\alpha\beta}\tilde{d}_{\alpha\gamma}\tilde{d}_{\beta\gamma}}. \tag{4.1.15}$$

If $\underline{m}$ satisfies (4.1.7), exactly one of these solutions is a solution of system (4.1.5) w.r.t. $\underline{m}$.

Hence a characterization for a Neyman-like extension is established. Real data usually provide Hamming distances smaller than $1 - 1/k$. The solutions have a similar structure compared to Theorem 3.1.4 and looking at Corollary 3.2.1 indicates that the result also holds in the N_2-case. However, in the two-state-case only one invariant, namely (4.1.6) is needed, and $\underline{m}$ has two symmetrical extensions. Next, using the terminology from the theorem, the transfer to triple leaf distributions is quantified:

Corollary 4.1.8. *Denote by $\widehat{\mathbf{m}}$ and $\widetilde{\mathbf{m}}$ the triple leaf distributions on $\mathcal{T}$ obtained by inserting the symmetrical solutions from (4.1.14) into (4.1.5). Their difference is given by:*

$$\widetilde{m}_{000} - \widehat{m}_{000} = \frac{2}{k^2}(k-1)(k-2)\Delta, \quad \widetilde{m}_{012} - \widehat{m}_{012} = \frac{4}{k^2}\Delta,$$
$$\widetilde{m}_{001} - \widehat{m}_{001} = \widetilde{m}_{010} - \widehat{m}_{010} = \widetilde{m}_{100} - \widehat{m}_{100} = -\frac{2}{k^2}(k-2)\Delta,$$

where Δ is given by (4.1.15).

Since (4.1.13) implies $\Delta \neq 0$, each set of pairwise distributions that yields a solution according to Theorem 4.1.7 returns two different triple leaf distributions each with a Neyman extension. Thus, if the initial triple distribution $\underline{m}$ admits (4.1.7), one has to check which solution returns $\underline{m}$. The following example provides an insight into all the statements from this section.

Example 4.1.1. Consider the following vector, satisfying (4.1.4):

$$\underline{m} = (100, 15, 15, 10, 5)/1000.$$

Due to its generation, the vector satisfies $N_1^4(\underline{m}) = 0$. Unfortunately, for the second invariant one computes $N_2^4(\underline{m}) = 17/125000 \neq 0$, i.e. $\underline{m}$ has no Neyman-like extension. Still the associated pairwise distributions provide two solution vectors p^1 and p^2 with

$$p^1 = \left(\frac{1}{15}, \frac{1}{10}, \frac{1}{10}\right), \quad p^2 = \left(\frac{13}{30}, \frac{2}{5}, \frac{2}{5}\right).$$

The probability vectors generated by inserting p^1 and p^2 respectively into (4.1.5) have the form

$$\begin{aligned}\underline{m}_1 &= (197, 31, 31, 21, 9)/2000,\\ \underline{m}_2 &= (49, 32, 32, 27, -12)/1000.\end{aligned}$$

Both vectors satisfy (4.1.6) and (4.1.7). p^1 is stochastically admissible, whereas $\underline{m}_2$ is not a probability distribution.

4.1.3 A Neyman Extension

After finding conditions for a Neyman-like extension, the conditions for a true Neyman extension need to be established. This is done by bounding the parameters provided in Theorem 4.1.7 between zero and $1/(k-1)$.

Theorem 4.1.9. *Let $\underline{m}$ denote a triple leaf distribution on $\mathcal{T}$ satisfying (4.1.3) and $\underline{m}^P$ its associated set of pairwise leaf distributions. If for $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$ the similarity-dissimilarity-differences satisfy*

$$0 < \frac{\tilde{d}_{\delta_1\delta_2}\tilde{d}_{\delta_1\delta_3}}{\tilde{d}_{\delta_2\delta_3}} \leq 1, \tag{4.1.16}$$

then system (4.1.9) has a stochastically admissible solution w.r.t. the associated pairwise leaf distributions. Further, if

$$0 < \frac{\tilde{d}_{\delta_1\delta_2}\tilde{d}_{\delta_1\delta_3}}{\tilde{d}_{\delta_2\delta_3}} \leq \frac{1}{(k-1)^2} \quad , \tag{4.1.17}$$

both solutions of system (4.1.9) are stochastically admissible w.r.t. the associated pairwise leaf distributions. If $\underline{m}$ admits (4.1.7), one of those solutions characterizes a Neyman extension to $\underline{m}$.

Condition (4.1.16) provides the positivity condition for the product $\tilde{d}_{\alpha\beta}\tilde{d}_{\alpha\gamma}\tilde{d}_{\beta\gamma}$ which is necessary for non-complex transition parameters. Further, (4.1.16) can be considered as a three-point-condition, since it relates the associated similarity-dissimilarity-differences. It states that the leaves have to be close enough together. Moreover, two distances can be negative. Recalling that $\tilde{d}_{\alpha\beta} = k(m_{00\Sigma} - m_{01\Sigma})$, negativity means that the sequences have more dissimilarities than similarities. Thus, either all leaves have sufficiently similar sequences or one leaf is significantly different from the other two. Comparing that insight to Theorem 3.1.5 one can observe that the statements are similar. This leads to the conjecture that an even number of pairwise negative relations is a necessary condition for all models.

Example 4.1.2. Recall the distribution from Example 4.1.1. The three similarity-dissimilarity-differences have the following values:

$$\tilde{d}_{\alpha\beta} = \tilde{d}_{\alpha\gamma} = \frac{11}{25}, \quad \tilde{d}_{\beta\gamma} = \frac{9}{25}.$$

With this values the condition (4.1.16) can be observed by

$$\frac{\tilde{d}_{\alpha\beta}\tilde{d}_{\alpha\gamma}}{\tilde{d}_{\beta\gamma}} = \frac{11^2}{15^2}, \quad \frac{\tilde{d}_{\alpha\beta}\tilde{d}_{\beta\gamma}}{\tilde{d}_{\alpha\gamma}} = \frac{\tilde{d}_{\alpha\gamma}\tilde{d}_{\beta\gamma}}{\tilde{d}_{\alpha\beta}} = \tilde{d}_{\beta\gamma} = \frac{9}{25}.$$

Thus, at least one solution characterizes a Neyman process. However, (4.1.17) is not satisfied, since all quotients exceed 1/9. As already seen, p^1 is the stochastically admissible solution and p^2 is not admissible.

4.1.4 Rates and Molecular Clock

This section will look into the implications of the computed properties to the very popular rate model and its even more popular special case of molecular clock. The models introduce edge lengths and have one rate matrix containing the infinitesimal rates of change across an edge.

Rates

The relationship between transition probabilities and rates is given through

$$P_\delta = \mathsf{e}^{Qt_\delta}, \quad \delta \in \mathcal{L},$$

where Q denotes the rate matrix of the associated model and t_δ the edge length of edge (ϱ, δ), $\delta \in \mathcal{L}$ on tree $\mathcal{T}$. The biggest advantage of the model is that the transition matrix to a path is the product of the transition matrices of the edges along the path, whereas the rate matrix is computed as the sum of the rates. The following rate matrix is subject to the Neyman N_k model:

$$Q = \left.\begin{pmatrix} -(k-1)q & q & \dots & q \\ q & -(k-1)q & \dots & q \\ \vdots & \vdots & \ddots & \vdots \\ q & q & \dots & -(k-1)q \end{pmatrix}\right\} k \text{ rows},$$

i.e. the Neyman model with rates on $\mathcal{T}$ consists of three edge lengths and a rate parameter q. Thus, the information obtained by solving the system for probabilities is not sufficient to obtain edge lengths as well as a rate. However, it is possible to provide a closed form for the product rate $q_\delta = q\, t_\delta,\ \delta \in \mathcal{L}$, as the following result shows.

Proposition 4.1.10. *Let $\underline{m}$ denote a triple leaf distribution on $\mathcal{T}$ satisfying (4.1.3) and $\underline{m}^P$ its associated set of pairwise leaf distributions. If the pairwise leaf distributions satisfy (4.1.17), exactly one set of transition probabilities transfers to the rate set*

$$\begin{aligned} q_\alpha &= -\frac{1}{2k}\big(\ln|\tilde{d}_{\alpha\beta}| + \ln|\tilde{d}_{\alpha\gamma}| - \ln|\tilde{d}_{\beta\gamma}|\big), \\ q_\beta &= -\frac{1}{2k}\big(\ln|\tilde{d}_{\alpha\beta}| + \ln|\tilde{d}_{\beta\gamma}| - \ln|\tilde{d}_{\alpha\gamma}|\big), \\ q_\gamma &= -\frac{1}{2k}\big(\ln|\tilde{d}_{\beta\gamma}| + \ln|\tilde{d}_{\alpha\gamma}| - \ln|\tilde{d}_{\alpha\beta}|\big). \end{aligned} \tag{4.1.18}$$

If only (4.1.16) is satisfied, the existence of admissible rates is not guaranteed.

The parameters q_δ, $\delta \in \mathcal{L}$ will be called rates only if they are subject to a set of transition probabilities. The relationship between rates and probabilities under the Neyman model is given by

$$q_\delta = -\frac{1}{k}\ln(1 - kp_\delta) = -\frac{1}{k}\ln\big(\pm\frac{\Delta}{\tilde{d}_{\delta_1\delta_2}}\big), \quad \delta \neq \delta_1 \neq \delta_2 \in \mathcal{L}.$$

Thus, only one set of transition parameters will provide the logarithm to a positive number. Therefore, if only (4.1.16) is satisfied, one has to ask whether the computed rates are subject to the transition probabilities. Hence, only (4.1.17) guarantees rates.

Molecular Clock

As introduced in Example 1.3.4, the molecular clock is a model which assumes that different species had the same time to evolve from their common ancestor.

This property provides the possibility of finding a root while further restricting the model. Let $\hat{\mathcal{T}} := (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ denote the rooted binary tree with

$$\hat{\mathcal{V}} := \{\delta_1, \delta_2, \delta_3, \varrho_1, \varrho_2\}, \quad \hat{\mathcal{E}} := \{(\varrho_2, \delta_1), (\varrho_2, \delta_2), (\varrho_1, \varrho_2), (\varrho_1, \delta_3)\}, \tag{4.1.19}$$

where $(\delta_1, \delta_2, \delta_3)$ denotes a permutation of the leaves in $\mathcal{L}$. In the case of molecular clock, the edge lengths of (ϱ_2, δ_1) and (ϱ_2, δ_2) must be equal and the edge length of (ϱ_1, δ_3) must be equal to the sum of edge lengths of (ϱ_1, ϱ_2) and (ϱ_2, δ_1) (see Figure 1.8). Incorporating these conditions, one gets the following result:

Proposition 4.1.11. *Let $\underline{m}$ denote a triple leaf distribution on $\mathcal{T}$ satisfying (4.1.3) and let $\underline{m}^P$ denote its associated set of pairwise leaf distributions. If the pairwise leaf distributions satisfy (4.1.17), and if:*

$$\tilde{d}_{\delta_1\delta_3} = \tilde{d}_{\delta_2\delta_3} \leq \tilde{d}_{\delta_1\delta_2}, \quad \delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}, \tag{4.1.20}$$

the model parameters have an extension with molecular clock on tree $\hat{\mathcal{T}} := (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ with (4.1.19). This extension is provided by the following rates:

$$q^c_{\delta_1} = q^c_{\delta_2} = -\frac{1}{2k} \ln|\tilde{d}_{\delta_1\delta_2}|, \quad q^c_{\delta_3} = -\frac{1}{2k} \ln|\tilde{d}_{\delta_1\delta_3}|, \quad q^c_{\varrho_2} = -\frac{1}{2k}\big(\ln|\tilde{d}_{\delta_1\delta_3}| - \ln|\tilde{d}_{\delta_1\delta_2}|\big).$$

The necessity for condition (4.1.20) becomes apparent when considering that the rates need to be non-negative in order to relate to transition probabilities. Thus, when looking at $q^c_{\varrho_2}$ one automatically finds that condition (4.1.20) needs to be satisfied. Figure 4.1 illustrates a working molecular clock, whereas Figure 4.2 shows a non-working case.

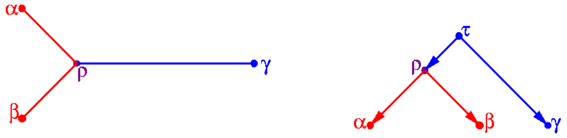

Figure 4.1: Molecular clock: Extension possible. The edge (ρ, γ) is longer than edges (ρ, α) and (ρ, β). Thus, a vertex τ can be introduced that obeys the addition rule for molecular clock.

Molecular clock has certain implications to a leaf distribution, for which (4.1.5) has a solution. Condition (4.1.20) implies with (4.1.14), that $p_{\delta_1} = p_{\delta_2} < p_{\delta_3}$ must hold for $\delta_1 \neq \delta_2 \neq \delta_3$. Applying this insight to (4.1.5) yields that one of the following three cases holds:

1. If $p_\alpha = p_\beta < p_\gamma$ then $m_{010} = m_{100}$.

2. If $p_\alpha = p_\gamma < p_\beta$ then $m_{001} = m_{100}$.
3. If $p_\beta = p_\gamma < p_\alpha$ then $m_{001} = m_{010}$.

If all three transition parameters were equal, $m_{001} = m_{010} = m_{100}$ would follow. Moreover, with $\tilde{d}_{\delta_1\delta_2} = \tilde{d}_{\delta_1\delta_3}$ the inner vertices ϱ_1 and ϱ_2 merges.

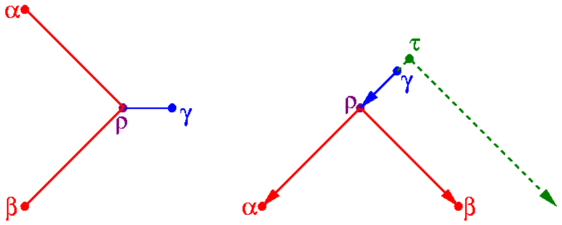

Figure 4.2: Molecular clock: Extension not possible. The edge (ρ, γ) is shorter than edges (ρ, α) and (ρ, β). The edge (τ, γ) must have negative length in order to satisfy the addition rule for molecular clock.

These observations conclude that a Neyman model with molecular clock is even more restrictive to the input and should be treated with even more care than the already quite restrictive rate model.

Example 4.1.3. Generating a leaf distribution that is subject to molecular clock is rather easy. Take two values and take them as the similarity-dissimilarity difference. To accommodate Theorem 4.1.9 and (4.1.20) choose $k = 4$ and

$$\tilde{d}_{\alpha\beta} = \tilde{d}_{\alpha\gamma} = 1/9,\ \tilde{d}_{\beta\gamma} = 1/5.$$

The resulting rates are given by

$$q^c_\beta = q^c_\gamma = 0.20118, \quad q^c_\alpha = 0.274653, \quad q^c_{\varrho_2} = 0.0734733.$$

Applying Theorem 4.1.7yields the following two sets of transition parameters:

$$p^1_\beta = p^1_\gamma = \frac{1}{4}\left(1 - \frac{1}{\sqrt{5}}\right), \quad p^1_\alpha = \frac{1}{4}\left(1 - \frac{\sqrt{5}}{9}\right),$$
$$p^2_\beta = p^2_\gamma = \frac{1}{4}\left(1 + \frac{1}{\sqrt{5}}\right), \quad p^2_\alpha = \frac{1}{4}\left(1 + \frac{\sqrt{5}}{9}\right).$$

Clearly, the second set is inadmissible. The full implication of this observation becomes visible when looking at the associated five-valued vector generated by inserting the transition parameters into (4.1.5):

$$m_1 = (0.0400751, 0.0144194, 0.0144194, 0.019975, 0.0105806),$$
$$m_2 = (0.0307582, 0.017525, 0.017525, 0.0230806, 0.00747495).$$

The structure of m_1 and m_2 happily agrees with the above statements concerning leaf distributions.

4.2 The Kimura 2ST Model

The Neyman model is the simplest model for molecular evolution. It treated any possible kind of change uniformly. The next step is to distinguish kinds of change. The simplest such distinction is the defining property of the *Kimura 2ST model.* This model is introduced in Example 1.3.3.

The following section will look at the properties of this model. The section is composed as the preceding sections. First basic properties are introduced. Then, conditions on leaf distributions for algebraic and stochastically admissible extension are established, and a closed form for the extension presented. Finally, the transfers of the results to the rate model and molecular clock is considered. Recall that $\mathcal{T} := (\mathcal{V}, \mathcal{E})$ denotes the triple tree with (4.0.1) and leaf set $\mathcal{L} := \{\alpha, \beta, \gamma\}$. Let $\mathcal{S} := \{0, 1, 2, 3\}$ and denote the class of purines by $\{0, 1\}$ and the class of pyrimidines by $\{2, 3\}$. A change within a class is called a TRANSITION, while a change from purine to pyrimidine or back is called a TRANSVERSION. To avoid confusion with the phrase transition in terms of the general transition of the process along an edge, above notation will be applied whenever the kind of state change is mentioned.

4.2.1 Basic Model Properties

Let $\underline{\mu} := (\mu_{uxyz})_{u,x,y,z \in \mathcal{S}}$ denote a Markov distribution on $\mathcal{T}$ w.r.t. Kimura 2ST model, i.e. $\underline{\mu}$ is subject to (LF), and has the following properties for $\delta \in \mathcal{L}$ (e.g. Ewens and Grant [2001, sect. 13.2]):

$$\mu_u := \mathbb{P}(X_\varrho = u) = 1/4 \quad \text{for all } u \in \mathcal{S}, \tag{4.2.1}$$

$$p_\delta := \mu_{x|u} = \mathbb{P}(X_\delta = x | X_\varrho = u) \quad \text{for } u \neq x \in \{0, 1\} \text{ or } u \neq x \in \{2, 3\}, \tag{4.2.2}$$

$$q_\delta := \mu_{y|u} = \mathbb{P}(X_\delta = y | X_\varrho = u) \quad \text{for } u \in \{0, 1\},\ y \in \{2, 3\} \text{ or vice versa.} \tag{4.2.3}$$

Equally to (4.1.1), stationarity of states in the root is proposed by (4.2.1). Properties (4.2.2) and (4.2.3) introduce the TRANSITION and TRANSVERSION parameter for the model respectively.

The properties show that such a Markov distribution is fully determined by the family $(p_\delta, q_\delta)_{\delta \in \mathcal{L}}$. Note that the sum $p_\delta + 2q_\delta$ cannot exceed 1/3 if $\underline{\mu}$ is a probability distribution. The model properties are used to define a *Kimura extension.*

Definition 4.2.1. *Let* $\underline{m} := (m_{xyz})_{x,y,z\in\mathcal{S}}$ *denote a leaf distribution on* $\mathcal{L}$*. A Markov distribution* $\underline{\mu}$ *on* $\mathcal{T}$ *subject to the Kimura 2ST model with*

$$m_{xyz} = \sum_{u\in\mathcal{S}} \mu_{uxyz}, \quad x, y, z \in \mathcal{S},$$

is called a Kimura extension *to* $\underline{m}$ *on* $\mathcal{T}$*.*

Similarly to the Neyman model, those properties have implications to the factorization property (LF). Since the model does not exactly distinguish the kind state change, the probability of some joint states is equal. In particular, there are ten groups of states and within each group, all states have the same probability. This property is presented in the next lemma.

Lemma 4.2.1. *Let* $\underline{m}$ *denote a leaf distribution on* $\mathcal{L}$*. If* $\underline{m}$ *has a Kimura extension* $\underline{\mu}$ *on* $\mathcal{T}$*, it satisfies the following conditions*

$$\begin{aligned}
&m_{000} = m_{111} = m_{222} = m_{333}, \quad m_{001} = m_{110} = m_{223} = m_{332},\\
&m_{010} = m_{101} = m_{232} = m_{323}, \quad m_{011} = m_{100} = m_{233} = m_{322},\\
&m_{002} = m_{003} = m_{112} = m_{113} = m_{220} = m_{221} = m_{330} = m_{331},\\
&m_{020} = m_{030} = m_{121} = m_{131} = m_{202} = m_{212} = m_{303} = m_{313},\\
&m_{022} = m_{033} = m_{122} = m_{133} = m_{200} = m_{211} = m_{300} = m_{311},\\
&m_{012} = m_{013} = m_{102} = m_{103} = m_{230} = m_{231} = m_{320} = m_{321},\\
&m_{021} = m_{031} = m_{120} = m_{130} = m_{203} = m_{213} = m_{302} = m_{312},\\
&m_{023} = m_{032} = m_{123} = m_{132} = m_{201} = m_{210} = m_{301} = m_{310}.
\end{aligned} \tag{4.2.4}$$

Thus, a triple leaf distribution $\underline{m}$ subject to a Kimura extension is characterized by ten probability values. The summation property for distributions provides the following relationship for those values

$$4(m_{000}+m_{001}+m_{010}+m_{100})+8(m_{002}+m_{020}+m_{200}+m_{012}+m_{102}+m_{201}) = 1. \tag{4.2.5}$$

Applying the model properties (4.2.1)-(4.2.3) to the basic factorization equation (LF) yields the following system that describes the relationship of the transition parameters to the ten probability values:

$$\begin{aligned}
4m_{000} &= (1-p_\alpha-2q_\alpha)(1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma)+p_\alpha p_\beta p_\gamma+2q_\alpha q_\beta q_\gamma,\\
4m_{001} &= (1-p_\alpha-2q_\alpha)(1-p_\beta-2q_\beta)p_\gamma+p_\alpha p_\beta(1-p_\gamma-2q_\gamma)+2q_\alpha q_\beta q_\gamma,\\
4m_{010} &= (1-p_\alpha-2q_\alpha)p_\beta(1-p_\gamma-2q_\gamma)+p_\alpha(1-p_\beta-2q_\beta)p_\gamma+2q_\alpha q_\beta q_\gamma,\\
4m_{100} &= (1-p_\alpha-2q_\alpha)p_\beta p_\gamma+p_\alpha(1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma)+2q_\alpha q_\beta q_\gamma,\\
4m_{002} &= (1-p_\alpha-2q_\alpha)(1-p_\beta-2q_\beta)q_\gamma+p_\alpha p_\beta q_\gamma+q_\alpha q_\beta(1-2q_\gamma),\\
4m_{012} &= (1-p_\alpha-2q_\alpha)p_\beta q_\gamma+p_\alpha(1-p_\beta-2q_\beta)q_\gamma+q_\alpha q_\beta(1-2q_\gamma),\\
4m_{020} &= (1-p_\alpha-2q_\alpha)q_\beta(1-p_\gamma-2q_\gamma)+p_\alpha q_\beta p_\gamma+q_\alpha(1-2q_\beta)q_\gamma,\\
4m_{021} &= (1-p_\alpha-2q_\alpha)q_\beta p_\gamma+p_\alpha q_\beta(1-p_\gamma-2q_\gamma)+q_\alpha(1-2q_\beta)q_\gamma,\\
4m_{200} &= q_\alpha(1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma)+q_\alpha p_\beta p_\gamma+(1-2q_\alpha)q_\beta q_\gamma,\\
4m_{201} &= q_\alpha(1-p_\beta-2q_\beta)p_\gamma+q_\alpha p_\beta(1-p_\gamma-2q_\gamma)+(1-2q_\alpha)q_\beta q_\gamma.
\end{aligned} \tag{4.2.6}$$

This system consists of ten equations in six variables. Again, one has to distinguish between algebraic solutions and stochastically admissible solution of the system. An algebraic solution of (4.2.6) w.r.t. $\underline{m}$ will characterize a *Kimura-type extension* to $\underline{m}$ on $\mathcal{T}$, whereas a stochastically admissible solution will characterize a *Kimura extension* to $\underline{m}$ on $\mathcal{T}$, i.e. the latter will characterize a Markov process on $\mathcal{T}$ that obeys the Kimura 2ST model.

According to Proposition 2.6.4, the dimension of the tangent space provides a lower bound for the number of polynomials needed to span the space of leaf distributions which have a solution.

Lemma 4.2.2. *The dimension of the smallest variety containing all triple leaf distributions $\{m_{xyz}\}_{x,y,z\in\mathcal{S}}$ with Kimura-type extensions is four.*

Thus at least four polynomials are needed to characterize the variety of triple leaf distributions with Kimura-type extensions. Computations with the software **Singular** returned a system of 18 polynomials in the ten variables provided by the left hand sides of (4.2.6). Writing them out would take about 24 pages. As a (very) small insight, equation (4.2.5) is included in the set of polynomials. For readability reasons those polynomials will only be presented on the accompanying CD. In accordance with Proposition 4.1.4 the following statement is given:

Proposition 4.2.3. *There is a set of 18 polynomials K_i such that if a leaf distribution $\underline{m}$ has a Kimura-type extension, it obeys (4.2.4) and*

$$K_i(\underline{m}) = 0, \quad i = 1, \ldots, 8. \tag{4.2.7}$$

Clearly, all polynomials K_i denote phylogenetic invariants. Chapter 2 showed that a given set of polynomials might not be the best basis for the sought algebraic variety. It is possible, that a smaller system could be obtained via better algorithms. But it can always get worse. For instance, the algorithm employed in **Mathematica** returned for the system (4.2.6) a system of 26 polynomials that filled more than 200 pages together.

4.2.2 An Algebraic Extension

After presenting conditions under which an algebraic extension exists, it is time to present the derivation of a characterization of such an extension. Again, the pairwise distributions are the key to the closed forms of a solution to (4.2.6). The pairwise probabilities $m_{01\Sigma}$ and $m_{02\Sigma}$ are computed via

$$m_{01\Sigma} = m_{010} + m_{100} + 2m_{012}, \quad m_{02\Sigma} = m_{020} + m_{021} + m_{200} + m_{201}.$$

The pairwise probabilities $m_{0\Sigma 1}$, $m_{0\Sigma 2}$, $m_{\Sigma 01}$ and $m_{\Sigma 02}$ are computed similarly. As in the Neyman section the set of pairwise distributions to a given triple distribution $\underline{m}$ is denoted by $\underline{m}^P$. System (4.2.6) yields the following relations between pairwise distributions and transition parameters

$$\begin{aligned}4m_{01\Sigma} &= (1-p_\alpha-2q_\alpha)p_\beta + p_\alpha(1-p_\beta-2q_\beta) + 2q_\alpha q_\beta,\\ 4m_{0\Sigma 1} &= (1-p_\alpha-2q_\alpha)p_\gamma + p_\alpha(1-p_\gamma-2q_\gamma) + 2q_\alpha q_\gamma,\\ 4m_{\Sigma 01} &= (1-p_\beta-2q_\beta)p_\gamma + p_\beta(1-p_\gamma-2q_\gamma) + 2q_\beta q_\gamma,\\ 4m_{02\Sigma} &= q_\alpha + q_\beta - 4q_\alpha q_\beta,\\ 4m_{0\Sigma 2} &= q_\alpha + q_\gamma - 4q_\alpha q_\gamma,\\ 4m_{\Sigma 02} &= q_\beta + q_\gamma - 4q_\beta q_\gamma.\end{aligned} \tag{4.2.8}$$

Apparently, the latter three equations only depend on the TRANSVERSION parameters and are of the same type as system (4.1.9). Thus, recalling Theorem 4.1.7, the system (4.2.8) has at least two solutions. The following proposition will provide a first characterization of the solutions.

Proposition 4.2.4. *If $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$ is a solution of (4.2.8) w.r.t. to a set of pairwise distributions, then the following families are also a solution to (4.2.8) w.r.t. the same set of pairwise distributions.*

$$(\hat{p}_\delta, q_\delta)_{\delta\in\mathcal{L}}, \quad \text{where} \quad \hat{p}_\delta = 1-2q_\delta-p_\delta, \tag{4.2.9}$$

$$(\tilde{p}_\delta, \tilde{q}_\delta)_{\delta\in\mathcal{L}}, \quad \text{where} \quad 2\tilde{q}_\delta = 1-2q_\delta \quad \text{and} \quad 2\tilde{p}_\delta = 1-2p_\delta. \tag{4.2.10}$$

The proposition shows that four different solutions of (4.2.8) w.r.t. a set $\underline{m}^P$ of pairwise leaf distributions can be identified by one family $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$. Hence, an algebraic solution to (4.2.8) can at most be *unique up to symmetry.* For an illustrative analysis of the alternative solutions, a look at the associated transition matrices is helpful.

Property (4.2.9) relates to the following matrix operation

$$\begin{pmatrix} \clubsuit & \spadesuit & \diamondsuit & \diamondsuit \\ \spadesuit & \clubsuit & \diamondsuit & \diamondsuit \\ \diamondsuit & \diamondsuit & \clubsuit & \spadesuit \\ \diamondsuit & \diamondsuit & \spadesuit & \clubsuit \end{pmatrix} \longrightarrow \begin{pmatrix} \spadesuit & \clubsuit & \diamondsuit & \diamondsuit \\ \clubsuit & \spadesuit & \diamondsuit & \diamondsuit \\ \diamondsuit & \diamondsuit & \spadesuit & \clubsuit \\ \diamondsuit & \diamondsuit & \clubsuit & \spadesuit \end{pmatrix}, \tag{4.2.11}$$

or literally, the probability of staying within a certain state class remains the same only the kinds of state change within swap probabilities. When inserting (4.2.9) into the initial system (4.2.6) one finds that it retains the triple leaf distribution. For property (4.2.10) the following change can be observed:

$$\begin{pmatrix} \clubsuit & \spadesuit & \diamondsuit & \diamondsuit \\ \spadesuit & \clubsuit & \diamondsuit & \diamondsuit \\ \diamondsuit & \diamondsuit & \clubsuit & \spadesuit \\ \diamondsuit & \diamondsuit & \spadesuit & \clubsuit \end{pmatrix} \longrightarrow \begin{pmatrix} 2\diamondsuit-\heartsuit & \heartsuit & \frac{\clubsuit+\spadesuit}{2} & \frac{\clubsuit+\spadesuit}{2} \\ \heartsuit & 2\diamondsuit-\heartsuit & \frac{\clubsuit+\spadesuit}{2} & \frac{\clubsuit+\spadesuit}{2} \\ \frac{\clubsuit+\spadesuit}{2} & \frac{\clubsuit+\spadesuit}{2} & 2\diamondsuit-\heartsuit & \heartsuit \\ \frac{\clubsuit+\spadesuit}{2} & \frac{\clubsuit+\spadesuit}{2} & \heartsuit & 2\diamondsuit-\heartsuit \end{pmatrix}, \tag{4.2.12}$$

or more precisely, the probability mass of staying within a class and changing the class are swapped and accordingly redistributed. Contrary to (4.2.9), this property does not keep (4.2.6) invariant to the change and leads to the conjecture that the solutions found for a set of pairwise distributions will provide two different triple distributions when inserted into (4.2.6). This conjecture is verified later on.

After having proposed the general style of a solution it is time to introduce its explicit form. Its computation unearthed certain notions whose relevance for the Kimura 2ST model resembles the relevance of the similarity-dissimilarity differences for the Neyman model.

Definition 4.2.2. *Denote by*

$$d^s_{\alpha\beta} := 1 - 8m_{01\Sigma} - 8m_{02\Sigma}, \quad d^v_{\alpha\beta} := 1 - 16m_{02\Sigma},$$

the TRANSITION- *and resp. the* TRANSVERSION difference *between α and β. The differences $d^v_{\alpha\gamma}$, $d^s_{\alpha\gamma}$, $d^v_{\beta\gamma}$, $d^s_{\beta\gamma}$ are defined accordingly.*

Under the Kimura 2ST model, the joint probabilities for a pair of leaves are related in the following way:

$$4m_{00\Sigma} + 4m_{01\Sigma} + 8m_{02\Sigma} = 1$$

With this knowledge, the TRANSITION- and the TRANSVERSION difference can respectively be written as:

$$d^v_{\alpha\beta} = 4(m_{00\Sigma} + m_{01\Sigma}) - 8m_{02\Sigma}, \tag{4.2.13}$$

$$d^s_{\alpha\beta} = 4(m_{00\Sigma} - m_{01\Sigma}), \tag{4.2.14}$$

yielding the following interpretations of the differences:

The TRANSITION difference is the difference of the probability of both leaves having states in the same class and the probability of both having states in different classes, whereas the TRANSVERSION difference is the difference of the probability of both leaves having the same state and the probability of both having different states in the same class.

With these notions, the explicit form of the characterization of a Kimura-type extension for a given set of pairwise leaf distributions is introduced:

Theorem 4.2.5. *Let $\underline{m}$ be a leaf distribution on $\mathcal{L}$ satisfying (4.2.4) and $\underline{m}^P$ its associated set of pairwise distributions. If the differences $d^v_{\delta_1\delta_2}$ and $d^s_{\delta_1\delta_2}$ satisfy for $\delta_1 \neq \delta_2 \in \mathcal{L}$*

$$d^v_{\delta_1\delta_2} \neq 0, \quad d^s_{\delta_1\delta_2} \neq 0, \tag{4.2.15}$$

then system (4.2.8) has a unique solution up to symmetry w.r.t. $\underline{m}^P$.

The solution is characterized by

$$\begin{aligned} q_\alpha &= \frac{1}{4}\left(1 \pm \frac{\Delta_v}{d^v_{\beta\gamma}}\right), & p_\alpha &= \frac{1}{2}\left(1 - 2q_\alpha \pm \frac{\Delta_s}{d^s_{\beta\gamma}}\right), \\ q_\beta &= \frac{1}{4}\left(1 \pm \frac{\Delta_v}{d^v_{\alpha\gamma}}\right), & p_\beta &= \frac{1}{2}\left(1 - 2q_\beta \pm \frac{\Delta_s}{d^s_{\alpha\gamma}}\right), \\ q_\gamma &= \frac{1}{4}\left(1 \pm \frac{\Delta_v}{d^v_{\alpha\beta}}\right), & p_\gamma &= \frac{1}{2}\left(1 - 2q_\gamma \pm \frac{\Delta_s}{d^s_{\alpha\beta}}\right), \end{aligned} \tag{4.2.16}$$

with

$$\Delta_v := \sqrt{d^v_{\alpha\beta} d^v_{\alpha\gamma} d^v_{\beta\gamma}}, \quad \Delta_s := \sqrt{d^s_{\alpha\beta} d^s_{\alpha\gamma} d^s_{\beta\gamma}}.$$

In addition, if $\underline{m}$ satisfies (4.2.7), it has a Kimura-type extension characterized by the values given above.

Together with the observations from Proposition 4.2.4, it becomes apparent that a set of pairwise leaf distributions yields two different triple leaf distribution each of which is subject to a pair of solution vectors, namely the vectors $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$ and $(1 - 2q_\delta - p_\delta, q_\delta)$. Those two triple leaf distributions only include $\underline{m}$ if it satisfies (4.2.7). The next result will show that the inferred triple leaf distributions are indeed distinct. For this purpose attach an index to a parameter family according to their sign in (4.2.16). For instance, the family $(p^{++}_\delta, q^{+}_\delta)_{\delta\in\mathcal{L}}$ denotes the parameters with

$$q^{+}_{\delta_1} = \frac{1}{4}\left(1 + \frac{\Delta_v}{d^v_{\delta_2\delta_3}}\right), \quad p^{++}_{\delta_1} = \frac{1}{2}\left(1 - 2q^{+}_{\delta_1} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right)$$

for $\delta_1 \neq \delta_2 \neq \delta_3$. Similarly, $(p_\delta^{+-}, q_\delta^+)_{\delta\in\mathcal{L}}$, $(p_\delta^{-+}, q_\delta^-)_{\delta\in\mathcal{L}}$ and $(p_\delta^{--}, q_\delta^-)_{\delta\in\mathcal{L}}$ are introduced.

Corollary 4.2.6. *Let $(p_\delta^{++}, q_\delta^+)_{\delta\in\mathcal{L}}$, $(p_\delta^{+-}, q_\delta^+)_{\delta\in\mathcal{L}}$, $(p_\delta^{-+}, q_\delta^-)_{\delta\in\mathcal{L}}$ and $(p_\delta^{--}, q_\delta^-)_{\delta\in\mathcal{L}}$ denote the solutions to system (4.2.8) w.r.t. a set of pairwise leaf distributions satisfying condition (4.2.15). Further, let $\underline{m}^1, \underline{m}^2, \underline{m}^3$ and $\underline{m}^4$ denote the associated triple leaf distributions obtained by inserting the solutions into system (4.2.6). Then, $\underline{m}^1 = \underline{m}^2$ and $\underline{m}^3 = \underline{m}^4$ and the difference of the two distributions is given by*

$$m_{000}^3 - m_{000}^1 = \frac{\Delta_v}{16}\left(\frac{d_{\alpha\beta}^s}{d_{\alpha\beta}^v} + \frac{d_{\alpha\gamma}^s}{d_{\alpha\gamma}^v} + \frac{d_{\beta\gamma}^s}{d_{\beta\gamma}^v}\right), \quad m_{001}^3 - m_{001}^1 = \frac{\Delta_v}{16}\left(\frac{d_{\alpha\beta}^s}{d_{\alpha\beta}^v} - \frac{d_{\alpha\gamma}^s}{d_{\alpha\gamma}^v} - \frac{d_{\beta\gamma}^s}{d_{\beta\gamma}^v}\right),$$

$$m_{010}^3 - m_{010}^1 = \frac{\Delta_v}{16}\left(\frac{d_{\alpha\gamma}^s}{d_{\alpha\gamma}^v} - \frac{d_{\alpha\beta}^s}{d_{\alpha\beta}^v} - \frac{d_{\beta\gamma}^s}{d_{\beta\gamma}^v}\right), \quad m_{100}^3 - m_{100}^1 = \frac{\Delta_v}{16}\left(\frac{d_{\beta\gamma}^s}{d_{\beta\gamma}^v} - \frac{d_{\alpha\gamma}^s}{d_{\alpha\gamma}^v} - \frac{d_{\alpha\beta}^s}{d_{\alpha\beta}^v}\right),$$

$$m_{002}^3 - m_{002}^1 = m_{012}^1 - m_{012}^3 = \frac{\Delta_v}{16}\frac{d_{\alpha\beta}^s}{d_{\alpha\beta}^v}, \quad m_{020}^3 - m_{020}^1 = m_{021}^1 - m_{021}^3 = \frac{\Delta_v}{16}\frac{d_{\alpha\gamma}^s}{d_{\alpha\gamma}^v},$$

$$m_{200}^3 - m_{200}^1 = m_{201}^1 - m_{201}^3 = \frac{\Delta_v}{16}\frac{d_{\beta\gamma}^s}{d_{\beta\gamma}^v}.$$

The triple distributions $\underline{m}^1$ and $\underline{m}^3$ are distinct since (4.2.15) demands that the differences are not zero. Hence, all solutions have to be considered in order to distinguish the solution for a given triple distribution $\underline{m}$ that satisfies (4.2.7). Additionally it has to be noted that a triple leaf distribution $\underline{m}$ that does not satisfy (4.2.7) still provides two distinct triple distributions with Kimura-type extensions.

Example 4.2.1. Consider the following vector:

$$\underline{m} = (17, 7, 2, 7, 7, 0, 2, 0, 2, 0)/200.$$

This vector satisfies (4.2.5) but is no root of the established phylogenetic invariants. However, as already mentioned, its associated pairwise leaf distributions will provide sets of transition parameters. These are presented here:

$$p_\alpha^{++} = p_\beta^{++} = p_\gamma^{++} = \frac{1}{2}\left(1 - \frac{\sqrt{17} - \sqrt{28}}{5}\right), \quad p_\alpha^{+-} = p_\beta^{+-} = p_\gamma^{+-} = \frac{1}{2}\left(1 - \frac{\sqrt{17} + \sqrt{28}}{5}\right),$$

for $q_\alpha^+ = q_\beta^+ = q_\gamma^+ = \frac{1}{4}\left(1 + \sqrt{17}/5\right)$, and

$$p_\alpha^{-+} = p_\beta^{-+} = p_\gamma^{-+} = \frac{1}{2}\left(1 + \frac{\sqrt{17} + \sqrt{28}}{5}\right), \quad p_\alpha^{--} = p_\beta^{--} = p_\gamma^{--} = \frac{1}{2}\left(1 + \frac{\sqrt{17} - \sqrt{28}}{5}\right)$$

for $q_\alpha^- = q_\beta^- = q_\gamma^- = \frac{1}{4}\left(1 - \sqrt{17}/5\right)$. Reinserting these parameters into (4.2.6) yields the following vectors:

$$\begin{aligned}200\,\underline{m}^- &= (19.0793, 6.30691, 1.30691, 6.30691, 6.30691,\\ &\quad 0.693087, 1.30691, 0.693087, 1.30691, 0.693087),\\ 200\,\underline{m}^+ &= (10.4207, 9.19309, 4.19309, 9.19309, 9.19309,\\ &\quad -2.19309, 4.19309, -2.19309, 4.19309, -2.19309).\end{aligned}$$

Both vectors are roots of the phylogenetic invariants mentioned previously, but apparently $\underline{m}^+$ is no distribution vector.

4.2.3 A Kimura Extension

Next, as in the preceding sections, conditions for the stochastic admissibility of the solutions are sought. Similarly to the previous cases, such conditions are obtained by bounding the terms derived through algebraic computations between zero and the upper bound admissible for the model. The next theorem proposes necessary conditions:

Theorem 4.2.7. *Let $\underline{m}$ denote a triple leaf distribution on $\mathcal{L}$ and $\underline{m}^P$ its associated set of pairwise leaf distributions. A necessary condition for stochastic admissibility is*

$$0 < d^v_{\delta_1\delta_2} d^v_{\delta_1\delta_3} \le d^v_{\delta_2\delta_3},\ 0 < d^s_{\delta_1\delta_2} d^s_{\delta_1\delta_3} \le d^s_{\delta_2\delta_3}, \quad \delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}. \tag{4.2.17}$$

If for $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$

$$-\frac{1}{2} \le \left(\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right) \le \frac{3}{2} \quad \textit{and} \quad -\frac{1}{2} \le \left(\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right) \le \frac{3}{2} \tag{4.2.18}$$

holds, the parameter families $(q^-_\delta, p^{--}_\delta)_{\delta\in\mathcal{L}}$ and $(q^-_\delta, p^{-+}_\delta)_{\delta\in\mathcal{L}}$ are stochastically admissible. If

$$-\frac{3}{2} \le \left(\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right) \le \frac{1}{2} \quad \textit{and} \quad -\frac{3}{2} \le \left(\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right) \le \frac{1}{2} \tag{4.2.19}$$

holds, the parameter families $(q^+_\delta, p^{+-}_\delta)_{\delta\in\mathcal{L}}$ and $(q^+_\delta, p^{++}_\delta)_{\delta\in\mathcal{L}}$ are stochastically admissible. All four families are stochastically admissible if

$$0 \le \left|\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right| \le \frac{1}{2} \quad \textit{and} \quad 0 \le \left|\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right| \le \frac{1}{2}, \tag{4.2.20}$$

Proposition 4.2.4 shows that if family $(q^+_\delta, p^{++}_\delta)_{\delta\in\mathcal{L}}$ is stochastically admissible, also $(q^+_\delta, p^{+-}_\delta)_{\delta\in\mathcal{L}}$ necessarily must be stochastically admissible. This observation is verified by (4.2.18) and (4.2.19). (4.2.20) is a simple conclusion from (4.2.18) and (4.2.19).

Example 4.2.2. Recall Example 4.2.1. Relating the distribution $\underline{m}$ to (4.2.17) yields

$$\begin{aligned}\frac{d^v_{\alpha\beta}d^v_{\alpha\gamma}}{d_{\beta\gamma^v}} &= \frac{d^v_{\alpha\beta}d^v_{\beta\gamma}}{d_{\alpha\gamma^v}} = \frac{d^v_{\alpha\gamma}d^v_{\beta\gamma}}{d_{\alpha\beta^v}} = \frac{17}{25},\\ \frac{d^s_{\alpha\beta}d^s_{\alpha\gamma}}{d_{\beta\gamma^s}} &= \frac{d^s_{\alpha\beta}d^s_{\beta\gamma}}{d_{\alpha\gamma^s}} = \frac{d^s_{\alpha\gamma}d^s_{\beta\gamma}}{d_{\alpha\beta^s}} = \frac{7}{25}.\end{aligned}$$

Hence, the necessary condition is satisfied. Consider the following differences:

$$\begin{aligned}\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &= \frac{31}{50} > \frac{1}{2},\\ \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &= \frac{3}{50}.\end{aligned}$$

According to this result only (4.2.18) holds, i.e. the parameter families to q_δ^- are stochastically admissible. This agrees with the observations in Example 4.2.1, where m^+ is no leaf distribution and the associated parameter families are not admissible.

4.2.4 Rates and Molecular Clock

As previously done for to the Neyman model, this section will present a transfer of the results to the popular rate model. The particularities of this specific model were given in Example 1.3.4.

Rates

Similarly to the observations in Section 4.1.4, the parameters provided in Theorem 4.2.5 are not sufficient to explicitly return edge lengths for the model but need to be incorporated into the rates. Thus, each edge has its own rate matrix. For the Kimura 2ST model such a rate matrix has the following form:

$$Q_\delta = \begin{pmatrix} -r_{p_\delta} - 2r_{q_\delta} & r_{p_\delta} & r_{q_\delta} & r_{q_\delta}\\ r_{p_\delta} & -r_{p_\delta} - 2r_{q_\delta} & r_{q_\delta} & r_{q_\delta}\\ r_{q_\delta} & r_{q_\delta} & -r_{p_\delta} - 2r_{q_\delta} & r_{p_\delta}\\ r_{q_\delta} & r_{q_\delta} & r_{p_\delta} & -r_{p_\delta} - 2r_{q_\delta}\end{pmatrix}, \quad \delta \in \mathcal{L}$$

The relationship of rates to probabilities is given by $P_\delta = \mathsf{e}^{Q_\delta}$, $\delta \in \mathcal{L}$. Using this, the rates are computed:

Proposition 4.2.8. *Let $\underline{m}$ be a triple leaf distribution on $\mathcal{L}$ and $\underline{m}^P$ its associated set of pairwise leaf distributions. If $\underline{m}^P$ satisfies (4.2.17) and (4.2.18), the transition probabilities correspond to the following rates for $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$*

$$r_{q_{\delta_1}} = -\frac{1}{8}(\ln|d^v_{\delta_1\delta_2}| + \ln|d^v_{\delta_1\delta_3}| - \ln|d^v_{\delta_2\delta_3}|), \tag{4.2.21}$$

$$r_{p_{\delta_1}} = -\frac{1}{4}\left(\ln\frac{|d^s_{\delta_1\delta_2}|}{|d^v_{\delta_1\delta_2}|} + \ln\frac{|d^s_{\delta_1\delta_3}|}{|d^v_{\delta_1\delta_3}|} - \ln\frac{|d^s_{\delta_2\delta_3}|}{|d^v_{\delta_2\delta_3}|}\right). \tag{4.2.22}$$

Observe that contrary to the transition probabilities, the rates provide only one family of admissible parameters. This is due to the following fact

Lemma 4.2.9. *Only the triple leaf distribution $\underline{m}$ generated by the parameter set $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$ with*

$$q_{\delta_1} = \frac{1}{4}\left(1 - \sqrt{\frac{d^v_{\delta_1\delta_2}d^v_{\delta_1\delta_3}}{d^v_{\delta_2\delta_3}}}\right), \quad p_{\delta_1} = \frac{1}{2}\left(1 - 2q_{\delta_1} - \sqrt{\frac{d^s_{\delta_1\delta_2}d^s_{\delta_1\delta_3}}{d^s_{\delta_2\delta_3}}}\right), \tag{4.2.23}$$

for every permutation $(\delta_1, \delta_2, \delta_3)$ of the leaves in $\mathcal{L}$, has an extension to the rate model.

Thus, the rate model loses three alternative solutions and one associated leaf distribution cannot be extended to the rate specification of the Kimura 2ST model. In effect, an input leaf distribution $\underline{m}$ has to satisfy an enormous amount of conditions to be subject to the rate model.

Molecular Clock

Now the rate model is transferred to the molecular clock framework. The molecular clock extends a given leaf distribution to the tree $\hat{\mathcal{T}} := (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ with (4.1.19). Molecular clock demands that the two shorter terminal edges in $\hat{\mathcal{T}}$ have the same lengths and that the rate for the long terminal edge is equal to the sum of the length of the short terminal edges and the length of the inner edge. The illustration of this property is given in Figure 4.1. Using the length properties, the rates are obtained.

Proposition 4.2.10. *Let $\underline{m}$ be a triple leaf distribution on $\mathcal{L}$ and $\underline{m}^P$ its associated set of pairwise leaf distributions. If for $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$ the pairwise distributions satisfy (4.2.17), (4.2.18) or (4.2.19) and*

$$\ln|d^s_{\delta_1\delta_3}| = \ln|d^s_{\delta_2\delta_3}| < \ln|d^s_{\delta_1\delta_2}|, \quad \ln|d^v_{\delta_1\delta_3}| = \ln|d^v_{\delta_2\delta_3}| < \ln|d^v_{\delta_1\delta_2}|, \tag{4.2.24}$$

then the model parameters have an extension with molecular clock on tree $\hat{\mathcal{T}}$ with

(4.1.19). This extension is characterized by the following rates:

$$\begin{aligned}
r_{q_{\delta_1}} &= r_{q_{\delta_2}} = -\frac{1}{8}\ln|d^v_{\delta_1\delta_2}|, & r_{p_{\delta_1}} &= r_{p_{\delta_2}} = -\frac{1}{8}\ln\frac{|d^s_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|},\\
r_{q_{\delta_3}} &= -\frac{1}{8}\ln|d^v_{\delta_1\delta_3}|, & r_{q_{\varrho_2}} &= -\frac{1}{8}(\ln|d^v_{\delta_1\delta_3}| - \ln|d^v_{\delta_1\delta_2}|),\\
r_{p_{\delta_3}} &= -\frac{1}{8}\ln\frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|}, & r_{p_{\varrho_2}} &= -\frac{1}{8}\left(\ln\frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|} - \ln\frac{|d_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|}\right).
\end{aligned}$$

Similarly to (4.1.20) condition (4.2.24) is needed to prevent tree structures, as illustrated by Figure 4.2. The next example presents a Kimura leaf distribution $\underline{m}$ that is subject to the Kimura 2ST model with molecular clock.

Example 4.2.3. Generate a set of differences according to (4.2.24):

$$\begin{aligned}
d^s_{\alpha\beta} &= d^s_{\alpha\gamma} = 3/10, & d^s_{\beta\gamma} &= 2/5,\\
d^v_{\alpha\beta} &= d^v_{\alpha\gamma} = 1/2, & d^v_{\beta\gamma} &= 7/10.
\end{aligned}$$

The associated rates are given by:

$$\begin{aligned}
r_{q_\beta} &= r_{q_\gamma} = 0.0445844, & r_{q_\alpha} &= 0.0866434, & r_{q_{\varrho_2}} &= 0.042059,\\
r_{p_\beta} &= r_{p_\gamma} = 0.184488, & r_{p_\alpha} &= 0.21435, & r_{p_{\varrho_2}} &= 0.0298615,
\end{aligned}$$

and the associated transition parameters from (4.2.16) have the following values:

$$\begin{aligned}
q_\beta &= q_\gamma = 0.040835, & q_\alpha &= 0.100596,\\
p_\beta &= p_\gamma = 0.142937, & p_\alpha &= 0.162233.
\end{aligned}$$

Inserting these parameters into (4.2.6) yields the triple distribution:

$$\begin{aligned}
\underline{m} = (&0.0965951, 0.0222173, 0.00621881, 0.0222173, 0.0277203,\\
&0.00315619, 0.00621881, 0.00315619, 0.0159673, 0.00590768),
\end{aligned}$$

which provides a nice insight into the structure of a leaf distributions subject to the Kimura 2ST model with molecular clock.

4.3 Proofs

This section cumulates all proofs of the results of this chapter. Before starting with proofs consider the following equation system:

$$\begin{aligned}
y_{12} &= a_2x_1 + a_1x_2 - cx_1x_2,\\
y_{13} &= a_3x_1 + a_1x_3 - cx_1x_3,\\
y_{23} &= a_3x_2 + a_2x_3 - cx_2x_3.
\end{aligned}\tag{4.3.1}$$

Deriving the solution of this system helps to derive solutions for the systems regarded in this chapter.

Lemma 4.3.1. *Assume that the parameters* $a_1, a_2, a_3, y_{12}, y_{13}, y_{23} \in \mathbb{C}$ *satisfy*

$$c \neq 0 \quad \text{and} \quad a_i a_j \neq c y_{ij}, \quad i \neq j.$$

Then, the system (4.3.1) has two solutions. The solutions have the following explicit form:

$$x_i^+ = \frac{1}{c}\left(a_i + \sqrt{\frac{(a_i a_j - c y_{ij})(a_i a_k - c y_{ik})}{a_j a_k - c y_{jk}}}\right) \tag{4.3.2}$$

and $x_i^- = 2a_i/c - x_i^+$ *for* $(i, j, k) \in \pi(1, 2, 3)$.

Here, π denotes the permutation mapping. The signs in (x_1, x_2, x_3) must be the same, i.e. only the vectors (x_1^+, x_2^+, x_3^+) and (x_1^-, x_2^-, x_3^-) are admissible solutions. This becomes apparent when inserting the notions from (4.3.2) into (4.3.1), as it will be done in the proof below.

Proof. Change over the first two equations of (4.3.1) after x_1 to get

$$x_2(a_1 - cx_1) = y_{12} - a_2 x_1, \quad x_3(a_1 - cx_1) = y_{13} - a_3 x_1.$$

Inserting these terms into the third equation yields the equation

$$\begin{aligned} y_{23}(a_1 - cx_1)^2 &= a_3(y_{12} - a_2 x_1)(a_1 - cx_1) + a_2(y_{13} - a_3 x_1)(a_1 - cx_1) \\ &\quad - c(y_{12} - a_2 x_1)(y_{13} - a_3 x_1) \\ &= a_3(a_1 y_{12} - x_1(c y_{12} + a_1 a_2) + a_2 c x_1^2) \\ &\quad + a_2(a_1 y_{13} - x_1(c y_{13} + a_1 a_3) + a_3 c x_1^2) \\ &\quad - c(y_{12} y_{13} - x_1(a_2 y_{12} + a_3 y_{13}) + a_2 a_3 x_1^2). \end{aligned}$$

Changing over and summarizing the terms returns in the following quadratic equation:

$$0 = (c x_1^2 - 2a_1 x_1)(c y_{23} - a_2 a_3) + a_1^2 y_{23} - a_1 a_3 y_{12} - a_1 a_2 y_{13} + c y_{12} y_{13}.$$

Therefore, one arrives at the form:

$$x_1^\pm = \frac{a_1}{c} \pm \sqrt{\frac{a_1^2}{c^2} - \frac{a_1^2 y_{23} - a_1 a_3 y_{12} - a_1 a_2 y_{13} + c y_{12} y_{13}}{c^2 y_{23} - c a_2 a_3}}.$$

Considering the root term more closely provides the following computations:

$$\begin{aligned}
&\frac{a_1^2(cy_{23}-a_2a_3)-ca_1^2y_{23}+ca_1a_3y_{12}+ca_1a_2y_{13}-c^2y_{12}y_{13}}{c^2(cy_{23}-a_2a_3)}\\
=\;&\frac{-a_1^2a_2a_3+ca_1a_3y_{12}+ca_1a_2y_{13}-c^2y_{12}y_{13}}{c^2(cy_{23}-a_2a_3)}\\
=\;&\frac{(a_1a_2-cy_{12})(a_1a_3-cy_{13})}{c^2(a_2a_3-cy_{23})},
\end{aligned}$$

thus yielding the root term proposed in (4.3.2). Since the computations for x_2 and x_3 are analogously, the form (4.3.2) is observed.

For the completion of the proof insert the proposed terms into system (4.3.1). Again, for symmetry reasons inserting into one equation is sufficient for validity on the whole system. Therefore,

$$y_{12}=a_2x_1+a_1x_2-cx_1x_2$$

and

$$\begin{aligned}
cy_{12}=\;&a_2\left(a_1\pm\sqrt{\frac{(a_1a_2-cy_{12})(a_1a_3-cy_{13})}{a_2a_3-cy_{23}}}\right)+a_1\left(a_2\pm\sqrt{\frac{(a_1a_2-cy_{12})(a_2a_3-cy_{23})}{a_1a_3-cy_{13}}}\right)\\
&-\left(a_1\pm\sqrt{\frac{(a_1a_2-cy_{12})(a_1a_3-cy_{13})}{a_2a_3-cy_{23}}}\right)\left(a_2\pm\sqrt{\frac{(a_1a_2-cy_{12})(a_2a_3-cy_{23})}{a_1a_3-cy_{13}}}\right)\\
=\;&a_1a_2\pm(a_1a_2-cy_{12}).
\end{aligned}$$

The equality is only observed if both terms have the same sign. Hence, the demand for same signs for the parameters x_1, x_2 and x_3 is verified. This completes the proof. □

4.3.1 Proofs for Section 4.1

Here, the Neyman N_k model on triple trees was analyzed.

Proof of Lemma 4.1.1. With (4.1.1) the notion of (LF) for m_{xyz}, $x,y,z\in\mathcal{S}$ changes to:

$$m_{xyz}=\sum_{u\in\mathcal{S}}\mu_{uxyz}=\sum_{u\in\mathcal{S}}q_u^{\varrho}p_{ux}^{\varrho\alpha}p_{uy}^{\varrho\beta}p_{uz}^{\varrho\gamma}=\frac{1}{k}\sum_{u\in\mathcal{S}}p_{ux}^{\varrho\alpha}p_{uy}^{\varrho\beta}p_{uz}^{\varrho\gamma}. \tag{4.3.3}$$

Property (4.1.2) gives the following specification for the transition parameters

$$p_{ux}^{\varrho\delta}=\begin{cases}p_\delta, & u\neq x,\\ 1-(k-1)p_\delta, & u=x\end{cases},\quad u,x\in\mathcal{S},$$

i.e. the transition parameters are independent of the states. Inserting this observation into (4.3.3) thus yields

$$\begin{aligned}
km_{xxx} &= (1-(k-1)p_\alpha)(1-(k-1)p_\beta)(1-(k-1)p_\gamma) + (k-1)p_\alpha p_\beta p_\gamma,\\
km_{xxy} &= (1-(k-1)p_\alpha)(1-(k-1)p_\beta)p_\gamma + p_\alpha p_\beta(1-p_\gamma),\\
km_{xyx} &= (1-(k-1)p_\alpha)p_\beta(1-(k-1)p_\gamma) + p_\alpha(1-p_\beta)p_\gamma,\\
km_{xyy} &= (1-p_\alpha)p_\beta p_\gamma + p_\alpha(1-(k-1)p_\beta)(1-(k-1)p_\gamma),\\
km_{xyz} &= p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma - 2kp_\alpha p_\beta p_\gamma,
\end{aligned}$$

for all $x \neq y \neq z \in \mathcal{S}$. Therefore, for all $x \neq y \neq z \in \mathcal{S}$ set

$$m_{xxx} = m_{000}, \quad m_{xxy} = m_{001}, \quad m_{xyx} = m_{010}, \quad m_{xyy} = m_{100}, \quad m_{xyz} = m_{012}.$$

Thus, (4.1.3) is derived an the lemma thus proven. □

Proof of Lemma 4.1.2. The previous proof already derived (4.2.6) and the naming of the left hand sides is given by (4.1.3). □

Proof of Lemma 4.1.3. The dimension is obtained by deriving the functional matrix to (4.1.5) and computing its rank at a rational point. Each column of the matrix contains the following entries:

$$\begin{aligned}
f_{1,i_1} &= (k-1)p_{i_2}p_{i_3} - (k-1)(1-(k-1)p_{i_2})(1-(k-1)p_{i_3}),\\
f_{2,i_1} &= p_{i_2}(1-p_{i_3}) - (k-1)p_{i_3}(1-(k-1)p_{i_2}),\\
f_{3,i_1} &= p_{i_3}(1-p_{i_2}) - (k-1)p_{i_2}(1-(k-1)p_{i_3}),\\
f_{4,i_1} &= (1-(k-1)p_{i_2})(1-(k-1)p_{i_3}) - p_{i_2}p_{i_3},\\
f_{5,i_1} &= p_{i_2} + p_{i_3} - 2kp_{i_2}p_{i_3},
\end{aligned}$$

with $i_1 \neq i_2 \neq i_3 \in \mathcal{L}$. Since the functional matrix is a 5×3 matrix, it cannot have a rank larger than tree. Select three rows from the matrix and compute the determinant to this 3×3 submatrix. W.l.o.g. consider, rows two, three and five:

$$\det(f_2, f_3, f_5) = -(1-kp_\alpha)(1-kp_\beta)(1-kp_\gamma)(p_\beta + p_\gamma).$$

Apparently, the roots in k of this polynomial are $1/p_\alpha$, $1/p_\beta$ and $1/p_\gamma$. Hence for $(p_\alpha, p_\beta, p_\gamma) = (2/7, 2/7, 2/7)$ the determinant has no integer valued root and the rank of the functional matrix in this point is three for any $k \in \mathbb{Z}$, i.e. the tangent space has dimension two. With Proposition 2.6.4 this is also the dimension of the variety and the proposition is thus proven. □

Proof of Proposition 4.1.4. Though **Singular** returns the polynomials for fixed k only, these polynomials are used to derive the polynomials presented in (4.1.6) and (4.1.7). Computations for a sufficient set of k return the proposed polynomials for arbitrary k. The remaining statements follow from Proposition 2.2.1 and this completes the proof. □

Proof of Lemma 4.1.5. Let $f_1, \dots, f_5$ and g_1, g_2, g_3 denote the polynomials generated by (4.1.5) and (4.1.9) respectively by subtracting the right hand sides from their respective left hand sides. (4.1.8) implies $\langle g_1, g_2, g_3 \rangle \subset \langle f_1, \dots, f_5 \rangle$ and with Proposition 2.6.1.1 $\mathbf{V}(g_1, g_2, g_3) \supset \mathbf{V}(f_1, \dots, f_5)$. This proves the first statement. The second statement follows from the insight that $\langle g_1, g_2, g_3, N_1^k, N_2^k \rangle = \langle f_1, \dots, f_5 \rangle$ and from Lemma 2.6.2. □

Proof of Proposition 4.1.6. The system (4.1.9) is equivalent to the following system:

$$km_{01\Sigma} = p_\alpha + p_\beta - kp_\alpha p_\beta, \tag{4.3.4}$$

$$km_{0\Sigma 1} = p_\alpha + p_\gamma - kp_\alpha p_\gamma, \tag{4.3.5}$$

$$km_{\Sigma 01} = p_\beta + p_\gamma - kp_\beta p_\gamma. \tag{4.3.6}$$

To show that $(\hat{p}_\alpha, \hat{p}_\beta, \hat{p}_\gamma)$ is a solution to the same left hand sides as $(p_\alpha, p_\beta, p_\gamma)$, it is sufficient to insert it into the right hand side of (4.3.4):

$$\begin{aligned}
&\hat{p}_\alpha + \hat{p}_\beta - k\hat{p}_\alpha \hat{p}_\beta = (\frac{2}{k} - p_\alpha) + (\frac{2}{k} - p_\beta) - k(\frac{2}{k} - p_\alpha)(\frac{2}{k} - p_\beta) \\
&= \frac{4}{k} - (p_\alpha + p_\beta) - (\frac{4}{k} - 2(p_\alpha + p_\beta) + kp_\alpha p_\beta) \\
&= p_\alpha + p_\beta - kp_\alpha p_\beta = km_{01\Sigma}.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 4.1.7. System (4.1.9) resembles system (4.3.1) with

$$a_1 = a_2 = a_3 = 1, \quad c = k, \quad y_{12} = km_{01\Sigma},\ y_{13} = km_{0\Sigma 1},\ y_{23} = km_{\Sigma 01}.$$

Thus, according to Lemma 4.3.1 the derived parameters have the form presented in (4.1.14) and the established solution is unique up to duplicity. □

Proof of Corollary 4.1.8. The proof starts with inserting the insights from the proof of Proposition 4.1.6 into (4.1.5):

$$
\begin{aligned}
&\widetilde{m}_{000} - \widehat{m}_{000} = \left(\frac{k-2}{k} - (k-1)p_\alpha\right)\left(\frac{k-2}{k} - (k-1)p_\beta\right)\left(\frac{k-2}{k} - (k-1)p_\gamma\right) \\
&\quad + (1-(k-1)p_\alpha)(1-(k-1)p_\beta)(1-(k-1)p_\gamma) + (k-1)p_\alpha p_\beta p_\gamma \\
&\quad - (k-1)(\frac{2}{k} - p_\alpha)(\frac{2}{k} - p_\beta)(\frac{2}{k} - p_\gamma) \\
&= \frac{(k-2)^3}{k^3} - \frac{(k-2)^2(k-1)}{k^2}(p_\alpha + p_\beta + p_\gamma) + \frac{(k-2)(k-1)^2}{k}(p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) \\
&\quad - (k-1)^3 p_\alpha p_\beta p_\gamma + 1 - (k-1)(p_\alpha + p_\beta + p_\gamma) + (k-1)^2(p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) \\
&\quad - (k-1)^3 p_\alpha p_\beta p_\gamma + (k-1)p_\alpha p_\beta p_\gamma - \frac{8(k-1)}{k^3} + \frac{4(k-1)}{k^2}(p_\alpha + p_\beta + p_\gamma) \\
&\quad - \frac{2(k-1)}{k}(p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) + (k-1)p_\alpha p_\beta p_\gamma \\
&= 2(k-1)(k-2)\left(\frac{1}{k^2} - \frac{1}{k}(p_\alpha + p_\beta + p_\gamma) + (p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) - k p_\alpha p_\beta p_\gamma\right).
\end{aligned}
$$

Next, the notions from (4.1.14) are inserted into the brackets. Note that the factor $1/k^2$ comes with every summand.

$$
\begin{aligned}
&1 - 3 + \frac{\Delta}{\tilde{d}_{\beta\gamma}} + \frac{\Delta}{\tilde{d}_{\alpha\gamma}} + \frac{\Delta}{\tilde{d}_{\beta\gamma}} + \big(1 - \frac{\Delta}{\tilde{d}_{\beta\gamma}}\big)\big(1 - \frac{\Delta}{\tilde{d}_{\alpha\gamma}}\big) + \big(1 - \frac{\Delta}{\tilde{d}_{\beta\gamma}}\big)\big(1 - \frac{\Delta}{\tilde{d}_{\alpha\beta}}\big) \\
&\quad + \big(1 - \frac{\Delta}{\tilde{d}_{\alpha\gamma}}\big)\big(1 - \frac{\Delta}{\tilde{d}_{\alpha\beta}}\big) - \big(1 - \frac{\Delta}{\tilde{d}_{\beta\gamma}}\big)\big(1 - \frac{\Delta}{\tilde{d}_{\alpha\gamma}}\big)\big(1 - \frac{\Delta}{\tilde{d}_{\alpha\beta}}\big) \\
&= \Delta.
\end{aligned}
\tag{4.3.7}
$$

Thus the proposed result

$$\widetilde{m}_{000} - \widehat{m}_{000} = \frac{2}{k^2}(k-1)(k-2)\Delta$$

is returned. Now for the next equality:

$$
\begin{aligned}
&\widetilde{m}_{012} - \widehat{m}_{012} = p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma - 2k p_\alpha p_\beta p_\gamma - \big(\frac{2}{k} - p_\alpha\big)\big(\frac{2}{k} - p_\beta\big) \\
&\quad - \big(\frac{2}{k} - p_\alpha\big)\big(\frac{2}{k} - p_\gamma\big) - \big(\frac{2}{k} - p_\beta\big)\big(\frac{2}{k} - p_\gamma\big) + 2k\big(\frac{2}{k} - p_\alpha\big)\big(\frac{2}{k} - p_\beta\big)\big(\frac{2}{k} - p_\gamma\big) \\
&= 4\left(\frac{1}{k^2} - \frac{1}{k}(p_\alpha + p_\beta + p_\gamma) + (p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) - k p_\alpha p_\beta p_\gamma\right)
\end{aligned}
$$

The term in the bracket was already computed in (4.3.7) and hence,

$$\widetilde{m}_{012} - \widehat{m}_{012} = \frac{4}{k^2}\Delta$$

i.e. the proposed result. For the remaining equalities:

$$\begin{aligned}\widetilde{m}_{001} - \widehat{m}_{001} &= p_\gamma + p_\alpha p_\beta - (k-1)p_\gamma(p_\alpha + p_\beta) + k(k-2)p_\alpha p_\beta p_\gamma - \frac{2}{k} + p_\gamma \\ &\quad - \big(\frac{2}{k} - p_\alpha\big)\big(\frac{2}{k} - p_\beta\big) + (k-1)\big(\frac{2}{k} - p_\gamma\big)\big(\frac{4}{k} - (p_\alpha + p_\beta)\big) \\ &\quad - k(k-2)\big(\frac{2}{k} - p_\alpha\big)\big(\frac{2}{k} - p_\beta\big)\big(\frac{2}{k} - p_\gamma\big) \\ &= -2(k-2)\bigg(\frac{1}{k^2} - \frac{1}{k}(p_\alpha + p_\beta + p_\gamma) + (p_\alpha p_\beta + p_\alpha p_\gamma + p_\beta p_\gamma) - kp_\alpha p_\beta p_\gamma\bigg).\end{aligned}$$

Again, the bracket was computed in (4.3.7) and from this the proposed equality

$$\widetilde{m}_{001} - \widehat{m}_{001} = -\frac{2}{k^2}(k-2)\Delta$$

is established. The computations for the remaining differences are similar to the final computation and thus the proof is finished. □

Proof of Theorem 4.1.9. The parameter p_δ, $\delta \in \mathcal{L}$ yields a transition probability if $0 \le p_\delta \le 1 - 1/k$. Inserting (4.1.14) returns

$$\begin{aligned} 0 \le \frac{1}{k}\bigg(1 \pm \frac{\Delta}{\tilde{d}_{\beta\gamma}}\bigg) &\le \frac{1}{k-1} \\ 0 \le 1 \pm \frac{\Delta}{\tilde{d}_{\beta\gamma}} &\le \frac{k}{k-1} \\ -1 \le \pm \frac{\Delta}{\tilde{d}_{\beta\gamma}} &\le \frac{1}{k-1}. \end{aligned}$$

Applying (4.1.15) gives

$$-1 \le -\sqrt{\frac{\tilde{d}_{\alpha\beta}\tilde{d}_{\alpha\gamma}}{\tilde{d}_{\beta\gamma}}} \le 0, \quad \text{and} \quad 0 \le \sqrt{\frac{\tilde{d}_{\alpha\beta}\tilde{d}_{\alpha\gamma}}{\tilde{d}_{\beta\gamma}}} \le \frac{1}{k-1}.$$

The first inequality assures that p_α^- is a probability; the second applies to p_α^+. Finally, squaring the inequalities yields (4.1.16) and (4.1.17). Positivity follows from (4.1.13). This concludes the proof. □

Proof of Proposition 4.1.10. From $P_\delta = \exp(Q_\delta)$, $\delta \in \mathcal{L}$ derive

$$p_\delta = \frac{1}{k}\big(1 - \mathsf{e}^{-kq_\delta}\big), \quad q_\delta = -\frac{1}{k}\ln(1 - kp_\delta).$$

Let $\delta_1, \delta_2 \in \mathcal{L}$ denote the remaining two leaves. Then inserting (4.1.14) yields

$$q_\delta = -\frac{1}{k}\ln\bigg(1 - \bigg(1 \pm \frac{\Delta}{\tilde{d}_{\delta_1\delta_2}}\bigg)\bigg) = -\frac{1}{k}\ln\bigg(\pm\frac{\Delta}{\tilde{d}_{\delta_1\delta_2}}\bigg).$$

Only the positive part is usable. W.l.o.g., let $\tilde{d}_{\delta_1\delta_2} > 0$. Thus, one computes

$$q_\delta = -\frac{1}{k}\ln\sqrt{\frac{\tilde{d}_{\delta\delta_1}\tilde{d}_{\delta\delta_2}}{\tilde{d}_{\delta_1\delta_2}}} = -\frac{1}{2k}\ln\frac{\tilde{d}_{\delta\delta_1}\tilde{d}_{\delta\delta_2}}{\tilde{d}_{\delta_1\delta_2}} = -\frac{1}{2k}\big(\ln|\tilde{d}_{\delta\delta_1}| + \ln|\tilde{d}_{\delta\delta_2}| - \ln|\tilde{d}_{\delta_1\delta_2}|\big),$$

i.e. the proposed notions. □

Proof of Proposition 4.1.11. The rates for (ϱ_2, δ_1) and (ϱ_2, δ_2) must be the same. Thus, from (4.1.14)

$$\frac{\Delta}{\tilde{d}_{\delta_1\delta_3}} = \frac{\Delta}{\tilde{d}_{\delta_2\delta_3}},$$

and therefore, $d_{\delta_1\delta_3} = d_{\delta_2\delta_3}$, and further, from (4.1.18)

$$q^c_{\delta_1} = -\frac{1}{2k}\big(\ln|\tilde{d}_{\delta_1\delta_2}| + \ln|\tilde{d}_{\delta_1\delta_3}| - \ln|\tilde{d}_{\delta_2\delta_3}|\big) = -\frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}|.$$

The rate for edge (ϱ_2, δ_3) from $\mathcal{T}$ may be denoted by q_{δ_3}. Then according to (4.1.18) it has the form

$$q_{\delta_3} = -\frac{1}{2k}\big(\ln|\tilde{d}_{\delta_1\delta_3}| + \ln|\tilde{d}_{\delta_2\delta_3}| - \ln|\tilde{d}_{\delta_1\delta_2}|\big) = -\frac{1}{k}\ln|\tilde{d}_{\delta_1\delta_3}| + \frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}|.$$

To compute $q^c_{\delta_3}$ for edge (ϱ_1, δ_3) and $q^c_{\varrho_2}$ for edge (ϱ_1, ϱ_2) the molecular clock offers the following equations:

$$q_{\delta_3} = q^c_{\delta_3} + q^c_{\varrho_2}, \quad q_{\delta_1} = q^c_{\delta_1} = q^c_{\delta_3} - q^c_{\varrho_2}.$$

Thus, compute $q^c_{\delta_3}$ from

$$2q^c_{\delta_3} = q_{\delta_1} + q_{\delta_3} = -\frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}| - \frac{1}{k}\ln|\tilde{d}_{\delta_1\delta_3}| + \frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}| = -\frac{1}{k}\ln|\tilde{d}_{\delta_1\delta_3}|$$

and $q^c_{\varrho_2}$ from

$$2q^c_{\varrho_2} = q_{\delta_3} - q_{\delta_1} = -\frac{1}{k}\ln|\tilde{d}_{\delta_1\delta_3}| + \frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}| + \frac{1}{2k}\ln|\tilde{d}_{\delta_1\delta_2}| = -\frac{1}{k}\big(\ln|\tilde{d}_{\delta_1\delta_3}| - \ln|\tilde{d}_{\delta_1\delta_2}|\big).$$

These are the proposed terms and the proof is completed. □

4.3.2 Proofs for Section 4.2

Here, the Kimura 2ST model on triple trees was examined.

Proof of Lemma 4.2.1. Insert the model properties into (LF) to obtain:

$$4m_{xxx} = (1 - p_\alpha - 2q_\alpha)(1 - p_\beta - 2q_\beta)(1 - p_\gamma - 2q_\gamma) + p_\alpha p_\beta p_\gamma + 2q_\alpha q_\beta q_\gamma, \quad x \in \mathcal{S}.$$

But the right hand side is independent of the choice of state $x \in \mathcal{S}$, thus $m_{xxx} = m_{000}$. The argument applies to all statements given in (4.2.4) and finally, (LF) yields (4.2.6) under the Kimura 2ST model. □

Proof of Lemma 4.2.2. To determine the dimension of the variety compute the functional matrix to system (4.2.6) in $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$ and look at the rank of this matrix. The rows for p_α and q_α have the following form:

$$\begin{pmatrix} p_\beta p_\gamma - (1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma) & 2q_\beta q_\gamma - 2(1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma) \\ p_\beta(1-2q_\beta) - p_\gamma(1-2q_\beta) & 2(q_\beta q_\gamma - p_\gamma(1-p_\beta-2q_\beta)) \\ p_\gamma(1-2q_\beta) - p_\beta(1-2q_\gamma) & 2(q_\beta q_\gamma - p_\beta(1-p_\gamma-2q_\gamma)) \\ (1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma) - p_\beta p_\gamma & 2(q_\beta q_\gamma - p_\beta p_\gamma) \\ -q_\gamma(1-2p_\beta-2q_\beta) & q_\beta(1+2q_\gamma) - 2q_\gamma(1-p_\beta) \\ q_\gamma(1-2p_\beta-2q_\beta) & q_\beta(1-2q_\gamma) - 2p_\beta q_\gamma \\ -q_\beta(1-2p_\gamma-2q_\gamma) & q_\gamma(1+2q_\beta) - 2q_\beta(1-p_\gamma) \\ q_\beta(1-2p_\gamma-2q_\gamma) & q_\gamma(1-2q_\beta) - 2q_\beta p_\gamma \\ 0 & p_\beta p_\gamma - 2q_\beta q_\gamma + (1-p_\beta-2q_\beta)(1-p_\gamma-2q_\gamma) \\ 0 & p_\beta + p_\gamma - 2p_\beta p_\gamma - 2p_\beta q_\gamma - 2q_\beta p_\gamma - 2q_\beta q_\gamma \end{pmatrix},$$

with similar rows for the other variables. For the vector

$$(p_\alpha, q_\alpha, p_\beta, q_\beta, p_\gamma, q_\gamma) = (1/20, 3/100, 2/50, 1/50, 1/10, 1/500)$$

the rank of the functional matrix

$$\nabla F = \begin{pmatrix} -0.82032 & -1.64856 & -0.79244 & -1.59476 & -0.8168 & -1.6364 \\ -0.05616 & -0.18392 & -0.0442 & -0.17788 & 0.8168 & -0.0028 \\ -0.00176 & 0.01624 & -0.00168 & 0.02632 & 0 & 0.8196 \\ 0.05616 & -0.0716 & 0.79244 & -0.00988 & 0.0104 & -0.07 \\ 0.82032 & -0.00792 & 0.0442 & -0.08948 & -0.0104 & -0.0908 \\ 0.00176 & 0.01976 & 0.00168 & 0.02968 & 0 & 0.0804 \\ -0.01788 & -0.03392 & 0 & 0.79742 & -0.0178 & -0.0058 \\ 0.01592 & -0.00208 & 0 & 0.13368 & 0.0168 & 0.0268 \\ 0 & 0.82824 & -0.02388 & -0.05188 & -0.0264 & -0.0364 \\ 0 & 0.12776 & 0.02388 & -0.00412 & 0.0264 & 0.0164 \end{pmatrix},$$

is six. Thus the dimension of the tangent space is four and with Proposition 2.2.1 this is also the dimension of the variety. This completes the proof. □

Proof of Proposition 4.2.3. The polynomials $(K_i)_{i=1}^{18}$ are computed with **Singular** . They are sufficient for the description of the sought variety. The remaining statements follow from Proposition 2.2.1. □

Proof of Proposition 4.2.4. The statement is, that not only $(p_\delta, q_\delta)_{\delta\in\mathcal{L}}$ but also $(\hat{p}_\delta, q_\delta)_{\delta\in\mathcal{L}}$ and $(\tilde{p}_\delta, \tilde{q}_\delta)_{\delta\in\mathcal{L}}$ with

$$\hat{p}_\delta = 1 - 2q_\delta - p_\delta, \quad 2\tilde{q}_\delta = 1 - 2q_\delta \quad \text{and} \quad 2\tilde{p}_\delta = 1 - 2p_\delta$$

are solutions of (4.2.8). Similar to the previously considered models the proof of the statement is sufficiently completed if only the following equations obey the propositions:

$$4m_{01\Sigma} = p_\alpha(1-2q_\beta) + p_\beta(1-2q_\alpha) + 2q_\alpha q_\beta - 2p_\alpha p_\beta, \tag{4.3.8}$$

$$4m_{02\Sigma} = q_\alpha + q_\beta - 4q_\alpha q_\beta. \tag{4.3.9}$$

Start with (4.3.9) and insert $\tilde{q}_\alpha$ and $\tilde{q}_\beta$:

$$\begin{aligned}2\tilde{q}_\alpha + 2\tilde{q}_\beta - 8\tilde{q}_\alpha\tilde{q}_\beta &= 1-2q_\alpha+1-2q_\beta-2(1-2q_\alpha)(1-2q_\beta)\\ &= 2-2(q_\alpha+q_\beta)-2+4(q_\alpha+q_\beta)-8q_\alpha q_\beta = 2q_\alpha+2q_\beta-8q_\alpha q_\beta,\end{aligned}$$

i.e. equivalence is attained. Now for equation (4.3.8) and vector $(\hat{p}_\delta, q_\delta)_{\delta\in\mathcal{L}}$:

$$\begin{aligned}&\hat{p}_\alpha(1-2q_\beta)+\hat{p}_\beta(1-2q_\alpha)+2q_\alpha q_\beta-2p_\alpha p_\beta\\ &=(1-2q_\alpha-p_\alpha)(1-2q_\beta)+(1-2q_\beta-p_\beta)(1-2q_\alpha)+2q_\alpha q_\beta\\ &\quad-2(1-2q_\alpha-p_\alpha)(1-2q_\beta-p_\beta)\\ &=2(1-2q_\alpha)(1-2q_\beta)-p_\alpha(1-2q_\beta)-p_\beta(1-2q_\alpha)+2q_\alpha q_\beta\\ &\quad-2(1-2q_\alpha)(1-2q_\beta)+2p_\alpha(1-2q_\beta)+2p_beta(1-2q_\alpha)-2p_\alpha p_\beta\\ &=p_\alpha(1-2q_\beta)+p_\beta(1-2q_\alpha)+2q_\alpha q_\beta-2p_\alpha p_\beta,\end{aligned}$$

and thus this configuration is verified. Finally apply $(\tilde{p}_\delta, \tilde{q}_\delta)_{\delta\in\mathcal{L}}$ to (4.3.8):

$$\begin{aligned}&2\tilde{p}_\alpha(1-2\tilde{q}_\beta)+2\tilde{p}_\beta(1-2\tilde{q}_\alpha)+4\tilde{q}_\alpha\tilde{q}_\beta-4\tilde{p}_\alpha\tilde{p}_\beta\\ &=2(1-2p_\alpha)q_\beta+2(1-2p_\beta)q_\alpha+(1-2q_\alpha)(1-2q_\beta)-(1-2p_\alpha)(1-2p_\beta)\\ &=2q_\beta-4p_\alpha q_\beta+2q_\alpha-4p_\beta q_\alpha+1-2q_\alpha-2q_\beta+4q_\alpha q_\beta-1+2p_\alpha+2p_\beta-4p_\alpha p_\beta\\ &=2p_\alpha(1-2q_\beta)+2p_\beta(1-2q_\alpha)+4q_\alpha q_\beta-4p_\alpha p_\beta.\end{aligned}$$

Hence, also this configuration solves (4.2.8) and the proof is thus completed. □

Proof of Theorem 4.2.5. The notions for q_δ, $\delta\in\mathcal{L}$ immediately follow from Theorem 4.1.7, since their defining equations are similar to (4.1.9). The defining equations for parameters p_δ, $\delta\in\mathcal{L}$ also resemble system (4.3.1) with

$$\begin{aligned}&a_1=1-2q_\alpha,\ a_2=1-2q_\beta,\ a_3=1-2q_\gamma,\ c=2,\\ &y_{12}=4m_{01\Sigma}-2q_\alpha q_\beta,\ y_{13}=4m_{0\Sigma1}-2q_\alpha q_\gamma,\ y_{23}=4m_{\Sigma01}-2q_\beta q_\gamma.\end{aligned}$$

Hence, Lemma 4.3.1 provides the general structure of the proposed parameters. To complete the proof, insert above conventions into the (4.3.2). The terms $a_ia_j - cy_{ij}$, $i\neq j$ are of particular interest. Take a closer look for $i=1, j=2$:

$$\begin{aligned}&(1-2q_\alpha)(1-2q_\beta)-8m_{01\Sigma}+2q_\alpha 2q_\beta\\ &=\frac{1}{4}\left(1+\frac{\Delta_v}{d^v_{\beta\gamma}}\right)\left(1+\frac{\Delta_v}{d^v_{\alpha\gamma}}\right)-8m_{01\Sigma}+\frac{1}{4}\left(1-\frac{\Delta_v}{d^v_{\beta\gamma}}\right)\left(1-\frac{\Delta_v}{d^v_{\alpha\gamma}}\right)\\ &=\frac{1}{2}+\frac{d^v_{\alpha\beta}}{2}-8m_{01\Sigma}=1-8m_{01\Sigma}-8m_{02\Sigma}=d^s_{\alpha\beta}.\end{aligned}$$

Analogue computations yield $a_1a_3 - cy_{13} = d^s_{\alpha\gamma}$ and $a_2a_3 - cy_{23} = d^s_{\beta\gamma}$. Hence, also these terms are observed and the proposed notions for p_δ, $\delta \in \mathcal{L}$ verified. Therefore, with (4.3.1) the theorem is proven. □

Proof of Corollary 4.2.6. For ease of reading set $q^1_\delta := q^+_\delta$, $q^2_\delta := q^-_\delta$, $p^1_\delta := p^{++}_\delta$, $p^2_\delta := p^{+-}_\delta$, $p^3_\delta := p^{-+}_\delta$, $p^4_\delta := p^{--}_\delta$, $\delta \in \mathcal{L}$. First, the equalities $\underline{m}^1 = \underline{m}^2$ and $\underline{m}^3 = \underline{m}^4$ are verified. Start with the first equality by inserting the terms from Theorem 4.2.5 into (4.2.6):

$$1 - p^1_\alpha - 2q^1_\alpha = 1 - \frac{1}{2} + q^1_\alpha - \frac{\Delta_s}{2d^s_{\beta\gamma}} - 2q^1_\alpha = \frac{1}{2}(1 - 2q^1_\alpha - \frac{\Delta_s}{d^s_{\beta\gamma}}) = p^2_\alpha. \tag{4.3.10}$$

This property will be used to verify that solutions (p^1_δ, q^1_δ) and (p^2_δ, q^1_δ) yield the same triple leaf distribution.

$$\begin{aligned}
&(1 - p^1_\alpha - 2q^1_\alpha)(1 - p^1_\beta - 2q^1_\beta)(1 - p^1_\gamma - 2q^1_\gamma) + p^1_\alpha p^1_\beta p^1_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma\\
&= p^2_\alpha p^2_\beta p^2_\gamma + p^1_\alpha p^1_\beta p^1_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma\\
&= (1 - p^2_\alpha - 2q^1_\alpha)(1 - p^2_\beta - 2q^1_\beta)(1 - p^2_\gamma - 2q^1_\gamma) + p^2_\alpha p^2_\beta p^2_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma.
\end{aligned}$$

Next, consider the equalities $m^1_{001} = m^2_{001}$, $m^1_{002} = m^2_{002}$ and $m^1_{012} = m^2_{012}$:

$$\begin{aligned}
&(1 - p^1_\alpha - 2q^1_\alpha)(1 - p^1_\beta - 2q^1_\beta)p^1_\gamma + p^1_\alpha p^1_\beta(1 - p^1_\gamma - 2q^1_\gamma) + 2q^1_\alpha q^1_\beta q^1_\gamma\\
&= p^2_\alpha p^2_\beta p^1_\gamma + p^1_\alpha p^1_\beta p^2_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma,\\
&(1 - p^1_\alpha - 2q^1_\alpha)(1 - p^1_\beta - 2q^1_\beta)q^1_\gamma + p^1_\alpha p^1_\beta q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma)\\
&= p^2_\alpha p^2_\beta q^1_\gamma + p^2_\alpha p^2_\beta q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma),\\
&(1 - p^1_\alpha - 2q^1_\alpha)p^1_\beta q^1_\gamma + p^1_\alpha(1 - p^1_\beta - 2q^1_\beta)q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma)\\
&= p^2_\alpha p^1_\beta q^1_\gamma + p^1_\alpha p^2_\beta q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma).
\end{aligned}$$

Similar to the above computations these results already provide equality. The remaining equalities are computed similarly and thus the equality $\underline{m}^1 = \underline{m}^2$ is observed. Through similar computations $\underline{m}^3 = \underline{m}^4$ is obtained.
Next, the difference $\underline{m}^1 - \underline{m}^3$ will be computed. Using above notions examine the differences. For symmetry reasons, it is sufficient to consider the following differences:

$$\begin{aligned}
4(m^1_{000} - m^3_{000}) &= p^2_\alpha p^2_\beta p^2_\gamma + p^1_\alpha p^1_\beta p^1_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma - p^4_\alpha p^4_\beta p^4_\gamma - p^3_\alpha p^3_\beta p^3_\gamma - 2q^2_\alpha q^2_\beta q^2_\gamma,\\
4(m^1_{001} - m^3_{001}) &= p^2_\alpha p^2_\beta p^1_\gamma + p^1_\alpha p^1_\beta p^2_\gamma + 2q^1_\alpha q^1_\beta q^1_\gamma - p^4_\alpha p^4_\beta p^3_\gamma - p^3_\alpha p^3_\beta p^4_\gamma - 2q^2_\alpha q^2_\beta q^2_\gamma,\\
4(m^1_{002} - m^3_{002}) &= p^2_\alpha p^2_\beta q^1_\gamma + p^1_\alpha p^1_\beta q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma) - p^4_\alpha p^4_\beta q^2_\gamma - p^3_\alpha p^3_\beta q^2_\gamma - q^2_\alpha q^2_\beta(1 - 2q^2_\gamma),\\
4(m^1_{012} - m^3_{012}) &= p^2_\alpha p^1_\beta q^1_\gamma + p^1_\alpha p^2_\beta q^1_\gamma + q^1_\alpha q^1_\beta(1 - 2q^1_\gamma) - p^4_\alpha p^3_\beta q^2_\gamma - p^3_\alpha p^4_\beta q^2_\gamma - q^2_\alpha q^2_\beta(1 - 2q^2_\gamma).
\end{aligned}$$

Consider the triple products separately:

$$\begin{aligned}
8p^1_\alpha p^1_\beta p^1_\gamma &= (1-2q^1_\alpha+\frac{\Delta_s}{d^s_{\beta\gamma}})(1-2q^1_\beta+\frac{\Delta_s}{d^s_{\alpha\gamma}})(1-2q^1_\gamma+\frac{\Delta_s}{d^s_{\alpha\beta}})\\
&= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)-8q^1_\alpha q^1_\beta q^1_\gamma\\
&\quad+\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d_{\beta\gamma}}+4q^1_\alpha q^1_\beta\frac{\Delta_s}{d^s_{\alpha\beta}}+4q^1_\alpha q^1_\gamma\frac{\Delta_s}{d^s_{\alpha\gamma}}+4q^1_\beta q^1_\gamma\frac{\Delta_s}{d^s_{\beta\gamma}}+\Delta_s\\
&\quad-2q^1_\alpha(\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\alpha\gamma}})-2q^1_\beta(\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\beta\gamma}})-2q^1_\gamma(\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d^s_{\beta\gamma}})\\
&\quad+(1-2q^1_\alpha)d^s_{\beta\gamma}+(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta},\\
8p^2_\alpha p^2_\beta p^2_\gamma &= (1-2q^1_\alpha-\frac{\Delta_s}{d^s_{\beta\gamma}})(1-2q^1_\beta-\frac{\Delta_s}{d^s_{\alpha\gamma}})(1-2q^1_\gamma-\frac{\Delta_s}{d^s_{\alpha\beta}})\\
&= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)-8q^1_\alpha q^1_\beta q^1_\gamma\\
&\quad-\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\alpha\gamma}}-\frac{\Delta_s}{d_{\beta\gamma}}-4q^1_\alpha q^1_\beta\frac{\Delta_s}{d^s_{\alpha\beta}}-4q^1_\alpha q^1_\gamma\frac{\Delta_s}{d^s_{\alpha\gamma}}-4q^1_\beta q^1_\gamma\frac{\Delta_s}{d^s_{\beta\gamma}}-\Delta_s\\
&\quad+2q^1_\alpha(\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\alpha\gamma}})+2q^1_\beta(\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\beta\gamma}})+2q^1_\gamma(\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d^s_{\beta\gamma}})\\
&\quad+(1-2q^1_\alpha)d^s_{\beta\gamma}+(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta}.
\end{aligned}$$

Thus, summing both left hand sides yields:

$$\begin{aligned}
4(p^1_\alpha p^1_\beta p^1_\gamma+p^2_\alpha p^2_\beta p^2_\gamma) &= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)\\
&\quad-8q^1_\alpha q^1_\beta q^1_\gamma+(1-2q^1_\alpha)d^s_{\beta\gamma}+(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta}.
\end{aligned}$$

Inserting those notions and their analogous versions for the other terms into the initial difference one gets

$$\begin{aligned}
8(m^1_{000}-m^3_{000}) &= q^2_\alpha-q^1_\alpha+q^2_\beta-q^1_\beta+q^2_\gamma-q^1_\gamma\\
&\quad+2(q^1_\alpha q^1_\beta-q^2_\alpha q^2_\beta+q^1_\alpha q^1_\gamma-q^2_\alpha q^2_\gamma+q^1_\beta q^1_\gamma-q^2_\beta q^2_\gamma)\\
&\quad+d^s_{\alpha\beta}(q^2_\gamma-q^1_\gamma)+d^s_{\alpha\gamma}(q^2_\beta-q^1_\beta)+d^s_{\beta\gamma}(q^2_\alpha-q^1_\alpha)
\end{aligned}\tag{4.3.11}$$

Now a closer observation of the differences is appropriate:

$$4(q^2_\alpha-q^1_\alpha)=1-\frac{\Delta_v}{d^v_{\beta\gamma}}-1-\frac{\Delta_v}{d^v_{\beta\gamma}}=-2\frac{\Delta_v}{d^v_{\beta\gamma}}\tag{4.3.12}$$

$$16(q^1_\alpha q^1_\beta-q^2_\alpha q^2_\beta)=(1+\frac{\Delta_v}{d^v_{\beta\gamma}})(1+\frac{\Delta_v}{d^v_{\alpha\gamma}})-(1-\frac{\Delta_v}{d^v_{\beta\gamma}})(1-\frac{\Delta_v}{d^v_{\alpha\gamma}})=2(\frac{\Delta_v}{d^v_{\alpha\gamma}}+\frac{\Delta_v}{d^v_{\beta\gamma}}).$$

Reconstructing this yields:

$$m^1_{000}-m^3_{000}=-\left(\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}+\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}+\frac{d^s_{\beta\gamma}}{d^v_{\alpha\gamma}}\right)\frac{\Delta_v}{16}.$$

For the next differences consider the triple products:

$$\begin{aligned}
8p^1_\alpha p^1_\beta p^2_\gamma &= (1-2q^1_\alpha+\frac{\Delta_s}{d^s_{\beta\gamma}})(1-2q^1_\beta+\frac{\Delta_s}{d^s_{\alpha\gamma}})(1-2q^1_\gamma-\frac{\Delta_s}{d^s_{\alpha\beta}})\\
&= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)-8q^1_\alpha q^1_\beta q^1_\gamma\\
&\quad -\frac{\Delta_s}{d^s_{\alpha\beta}}+\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d_{\beta\gamma}}-4q^1_\alpha q^1_\beta\frac{\Delta_s}{d^s_{\alpha\beta}}+4q^1_\alpha q^1_\gamma\frac{\Delta_s}{d^s_{\alpha\gamma}}+4q^1_\beta q^1_\gamma\frac{\Delta_s}{d^s_{\beta\gamma}}-\Delta_s\\
&\quad +2q^1_\alpha(\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\alpha\gamma}})+2q^1_\beta(\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\beta\gamma}})-2q^1_\gamma(\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d^s_{\beta\gamma}})\\
&\quad -(1-2q^1_\alpha)d^s_{\beta\gamma}-(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta},\\
8p^2_\alpha p^2_\beta p^1_\gamma &= (1-2q^1_\alpha-\frac{\Delta_s}{d^s_{\beta\gamma}})(1-2q^1_\beta-\frac{\Delta_s}{d^s_{\alpha\gamma}})(1-2q^1_\gamma+\frac{\Delta_s}{d^s_{\alpha\beta}})\\
&= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)-8q^1_\alpha q^1_\beta q^1_\gamma\\
&\quad +\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\alpha\gamma}}-\frac{\Delta_s}{d_{\beta\gamma}}+4q^1_\alpha q^1_\beta\frac{\Delta_s}{d^s_{\alpha\beta}}-4q^1_\alpha q^1_\gamma\frac{\Delta_s}{d^s_{\alpha\gamma}}-4q^1_\beta q^1_\gamma\frac{\Delta_s}{d^s_{\beta\gamma}}+\Delta_s\\
&\quad -2q^1_\alpha(\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\alpha\gamma}})-2q^1_\beta(\frac{\Delta_s}{d^s_{\alpha\beta}}-\frac{\Delta_s}{d^s_{\beta\gamma}})+2q^1_\gamma(\frac{\Delta_s}{d^s_{\alpha\gamma}}+\frac{\Delta_s}{d^s_{\beta\gamma}})\\
&\quad -(1-2q^1_\alpha)d^s_{\beta\gamma}-(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta}.
\end{aligned}$$

Summing both left hand sides yields:

$$\begin{aligned}
4(p^1_\alpha p^1_\beta p^2_\gamma + p^2_\alpha p^2_\beta p^1_\gamma) &= 1-2(q^1_\alpha+q^1_\beta+q^1_\gamma)+4(q^1_\alpha q^1_\beta+q^1_\alpha q^1_\gamma+q^1_\beta q^1_\gamma)\\
&\quad -8q^1_\alpha q^1_\beta q^1_\gamma-(1-2q^1_\alpha)d^s_{\beta\gamma}-(1-2q^1_\beta)d^s_{\alpha\gamma}+(1-2q^1_\gamma)d^s_{\alpha\beta}.
\end{aligned}$$

Now consider the difference $m^1_{001}-m^3_{001}$. Derive the description for the other terms accordingly to get:

$$\begin{aligned}
8(m^1_{001}-m^3_{001}) &= q^2_\alpha-q^1_\alpha+q^2_\beta-q^1_\beta+q^2_\gamma-q^1_\gamma\\
&\quad +2(q^1_\alpha q^1_\beta-q^2_\alpha q^2_\beta+q^1_\alpha q^1_\gamma-q^2_\alpha q^2_\gamma+q^1_\beta q^1_\gamma-q^2_\beta q^2_\gamma)\\
&\quad +d^s_{\alpha\beta}(q^2_\gamma-q^1_\gamma)-d^s_{\alpha\gamma}(q^2_\beta-q^1_\beta)-d^s_{\beta\gamma}(q^2_\alpha-q^1_\alpha).
\end{aligned}$$

As in (4.3.11) the first two lines cancel each other and the remaining terms yield:

$$m^1_{001}-m^3_{001} = -\left(\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}-\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}-\frac{d^s_{\beta\gamma}}{d^v_{\beta\gamma}}\right)\frac{\Delta_v}{16}$$

and similarly

$$\begin{aligned}
m^1_{010}-m^2_{010} &= -\left(\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}-\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}-\frac{d^s_{\beta\gamma}}{d^v_{\beta\gamma}}\right)\frac{\Delta_v}{16},\\
m^1_{100}-m^2_{100} &= -\left(\frac{d^s_{\beta\gamma}}{d^v_{\beta\gamma}}-\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}-\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}\right)\frac{\Delta_v}{16}.
\end{aligned}$$

For the remaining two cases the pairwise products are of interest:

$$\begin{aligned}
4p^1_\alpha p^1_\beta &= (1 - 2q^1_\alpha + \frac{\Delta_s}{d^s_{\beta\gamma}})(1 - 2q^1_\beta + \frac{\Delta_s}{d^s_{\alpha\gamma}}) \\
&= 1 - 2(q^1_\alpha + q^1_\beta) + 4q^1_\alpha q^1_\beta + \frac{\Delta_s}{d^s_{\beta\gamma}}(1 - 2q^1_\beta) + \frac{\Delta_s}{d^s_{\alpha\gamma}}(1 - 2q^1_\alpha) + d^s_{\alpha\beta}, \\
4p^2_\alpha p^2_\beta &= (1 - 2q^1_\alpha - \frac{\Delta_s}{d^s_{\beta\gamma}})(1 - 2q^1_\beta - \frac{\Delta_s}{d^s_{\alpha\gamma}}) \\
&= 1 - 2(q^1_\alpha + q^1_\beta) + 4q^1_\alpha q^1_\beta - \frac{\Delta_s}{d^s_{\beta\gamma}}(1 - 2q^1_\beta) - \frac{\Delta_s}{d^s_{\alpha\gamma}}(1 - 2q^1_\alpha) + d^s_{\alpha\beta}, \\
4p^1_\alpha p^2_\beta &= (1 - 2q^1_\alpha + \frac{\Delta_s}{d^s_{\beta\gamma}})(1 - 2q^1_\beta - \frac{\Delta_s}{d^s_{\alpha\gamma}}) \\
&= 1 - 2(q^1_\alpha + q^1_\beta) + 4q^1_\alpha q^1_\beta + \frac{\Delta_s}{d^s_{\beta\gamma}}(1 - 2q^1_\beta) - \frac{\Delta_s}{d^s_{\alpha\gamma}}(1 - 2q^1_\alpha) - d^s_{\alpha\beta}, \\
4p^2_\alpha p^1_\beta &= (1 - 2q^1_\alpha - \frac{\Delta_s}{d^s_{\beta\gamma}})(1 - 2q^1_\beta + \frac{\Delta_s}{d^s_{\alpha\gamma}}) \\
&= 1 - 2(q^1_\alpha + q^1_\beta) + 4q^1_\alpha q^1_\beta - \frac{\Delta_s}{d^s_{\beta\gamma}}(1 - 2q^1_\beta) + \frac{\Delta_s}{d^s_{\alpha\gamma}}(1 - 2q^1_\alpha) - d^s_{\alpha\beta}.
\end{aligned}$$

Inserting these computations into the difference $m^1_{002} - m^3_{002}$ returns:

$$\begin{aligned}
8(m^1_{002} - m^3_{002}) &= 2(q^1_\alpha q^1_\beta - q^2_\alpha q^2_\beta - q^1_\alpha q^1_\gamma + q^2_\alpha q^2_\gamma - q^1_\beta q^1_\gamma + q^2_\beta q^2_\gamma) \\
&\quad + q^1_\gamma - q^2_\gamma + d^s_{\alpha\beta}(q^1_\gamma - q^2_\gamma).
\end{aligned} \tag{4.3.13}$$

Applying the notions from (4.3.12) to (4.3.13) yields:

$$m^1_{002} - m^3_{002} = \frac{\Delta_v}{16}\,\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}$$

and accordingly

$$m^1_{020} - m^3_{020} = \frac{\Delta_v}{16}\,\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}, \quad m^1_{200} - m^3_{200} = \frac{\Delta_v}{16}\,\frac{d^s_{\beta\gamma}}{d^v_{\beta\gamma}}.$$

Finally, for the difference $m^1_{012} - m^3_{012}$ compute

$$\begin{aligned}
8(m^1_{012} - m^3_{012}) &= 2(q^1_\alpha q^1_\beta - q^2_\alpha q^2_\beta - q^1_\alpha q^1_\gamma + q^2_\alpha q^2_\gamma - q^1_\beta q^1_\gamma + q^2_\beta q^2_\gamma) \\
&\quad + q^1_\gamma - q^2_\gamma + d_{\alpha\beta^s}(q^2_\gamma - q^1_\gamma).
\end{aligned}$$

Consistently with (4.3.13) this yields for the final differences:

$$m^1_{012} - m^3_{012} = -\frac{\Delta_v}{16}\,\frac{d^s_{\alpha\beta}}{d^v_{\alpha\beta}}, \quad m^1_{021} - m^3_{021} = -\frac{\Delta_v}{16}\,\frac{d^s_{\alpha\gamma}}{d^v_{\alpha\gamma}}, \quad m^1_{201} - m^3_{201} = -\frac{\Delta_v}{16}\,\frac{d^s_{\beta\gamma}}{d^v_{\beta\gamma}}.$$

This completes the proof. □

Proof of Theorem 4.2.7. Set $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$. The conditions $0 < d^v_{\delta_1\delta_2} d^v_{\delta_1\delta_3} d^v_{\delta_2\delta_3}$ and $0 < d^s_{\delta_1\delta_2} d^s_{\delta_1\delta_3} d^s_{\delta_2\delta_3}$ follow immediately from the necessity to avoid complex solutions.

To get a stochastic solution the following conditions must be satisfied:

$$0 \leq 2q_{\delta_1} \leq 1, \quad 0 \leq p_{\delta_1} \leq 1, \quad 0 \leq p_{\delta_1} + 2q_{\delta_1} \leq 1.$$

The first two conditions are necessary to guarantee that the transition parameters are actually probabilities and the third condition is necessary to guarantee that $P^{\varrho\delta_i} = (p^{\varrho\delta_i}_{ux})_{u,x\in\mathcal{S}}$ is a transition matrix. First, consider the inequality for q_δ:

$$0 \leq \frac{1}{2} \pm \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} \leq 1, \quad -\frac{1}{2} \leq \pm\frac{\Delta_v}{2d^v_{\delta_2\delta_3}} \leq \frac{1}{2},$$

and thus $0 \leq d^v_{\delta_1\delta_2} d^v_{\delta_1\delta_3} \leq d^v_{\delta_2\delta_3}$.

For the remaining conditions consider each parameter family separately. Start with $(q^+_\delta, p^{++}_\delta)_{\delta\in\mathcal{L}}$:

$$\begin{aligned} 0 \leq 1 - 2q^+_{\delta_1} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq 1 + 2q^+_{\delta_1} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ 0 \leq \frac{1}{2} - \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq \frac{3}{2} + \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ -\frac{3}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{1}{2}, & -\frac{3}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{1}{2}. \end{aligned}$$

Consider $(q^+_\delta, p^{+-}_\delta)_{\delta\in\mathcal{L}}$:

$$\begin{aligned} 0 \leq 1 - 2q^+_{\delta_1} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq 1 + 2q^+_{\delta_1} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ 0 \leq \frac{1}{2} - \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq \frac{3}{2} + \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ -\frac{3}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{1}{2}, & -\frac{3}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{1}{2}. \end{aligned}$$

Continue with $(q^-_\delta, p^{--}_\delta)_{\delta\in\mathcal{L}}$:

$$\begin{aligned} 0 \leq 1 - 2q^-_{\delta_1} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq 1 + 2q^-_{\delta_1} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ 0 \leq \frac{1}{2} + \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, & 0 \leq \frac{3}{2} - \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq 2, \\ -\frac{1}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{3}{2}, & -\frac{1}{2} \leq \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\leq \frac{3}{2}. \end{aligned}$$

Finish the task with $(q_\delta^-, p_\delta^{-+})_{\delta\in\mathcal{L}}$:

$$\begin{aligned}
0 \le 1 - 2q_{\delta_1}^- + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le 2, & 0 \le 1 + 2q_{\delta_1}^- + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le 2,\\
0 \le \frac{1}{2} + \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le 2, & 0 \le \frac{3}{2} - \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le 2,\\
-\frac{1}{2} \le \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} + \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le \frac{3}{2}, & -\frac{1}{2} \le \frac{\Delta_v}{2d^v_{\delta_2\delta_3}} - \frac{\Delta_s}{d^s_{\delta_2\delta_3}} &\le \frac{3}{2}.
\end{aligned}$$

Therefore, in accordance with Proposition 4.2.4 either both families to q_ϱ^+ resp. q_ϱ^- are stochastically admissible or not and thus there are either zero, two or four stochastically admissible solutions. This completes the proof.

□

Proof of Proposition 4.2.8. The relationship between rates and probabilities is given through $P_\delta = \mathsf{e}^{Qt_\delta}$, $\delta \in \mathcal{L}$. In our case $Q_\delta := Qt_\delta$. This relationship yields the following equations:

$$q_\delta = \frac{1}{4}(1 - \mathsf{e}^{-4r_{q_\delta}}), \quad p_\delta = \frac{1}{4}(1 + \mathsf{e}^{-4r_{q_\delta}} - 2\mathsf{e}^{2(r_{q_\delta}+r_{p_\delta})}), \quad \delta \in \mathcal{L}.$$

Restructuring the equations yields:

$$\text{(4.3.14)} \qquad r_{q_\delta} = -\frac{1}{4}\ln(1 - 4q_\delta), \quad r_{p_\delta} = -r_{q_\delta} - \frac{1}{2}\ln(1 - 2q_\delta - 2p_\delta).$$

An examination of the probabilities shows that for $\delta_1 \neq \delta_2 \neq \delta_3 \in \mathcal{L}$ the following can be observed:

$$\text{(4.3.15)} \quad 1 - 4q_{\delta_1} = \pm\frac{\Delta_v}{d^v_{\delta_2\delta_3}}, \quad 1 - 2q_{\delta_1} - 2p_{\delta_1} = 1 - 2q_{\delta_1} - 1 + 2q_{\delta_1} \pm \frac{\Delta_s}{d^s_{\delta_2\delta_3}} = \pm\frac{\Delta_s}{d^s_{\delta_2\delta_3}}.$$

Looking at (4.3.14) indicates that only the constellation $(q_\delta, p_\delta)_{\delta\in\mathcal{L}}$ with

$$q_{\delta_1} = \frac{1}{4}\left(1 - \sqrt{\frac{d^v_{\delta_1\delta_2}d^v_{\delta_1\delta_3}}{d^v_{\delta_2\delta_3}}}\right), \quad p_{\delta_1} = \frac{1}{2}\left(1 - 2q_{\delta_1} - \sqrt{\frac{d^s_{\delta_1\delta_2}d^s_{\delta_1\delta_3}}{d^s_{\delta_2\delta_3}}}\right)$$

can provide a rate, because all other constellations return the logarithm of a negative number. Inserting the insights from (4.3.15) into (4.3.14) returns

$$\begin{aligned}
r_{q_{\delta_1}} &= -\frac{1}{4}\ln\left|\frac{\Delta_v}{d^v_{\delta_2\delta_3}}\right| = -\frac{1}{8}(\ln|d^v_{\delta_1\delta_2}| + \ln|d^v_{\delta_1\delta_3}| - \ln|d^v_{\delta_2\delta_3}|),\\
r_{p_{\delta_1}} &= -r_{q_{\delta_1}} - \frac{1}{2}\ln\left|\frac{\Delta_s}{d^s_{\delta_2\delta_3}}\right| = -\frac{1}{8}\left(\ln\frac{|d^s_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|} + \ln\frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|} - \ln\frac{|d^s_{\delta_2\delta_3}|^2}{|d^v_{\delta_2\delta_3}|}\right).
\end{aligned}$$

Thus the proposed rates are established and the proof is completed. □

Proof of Lemma 4.2.9. The lemma proposed that from the probability parameter families obtained in Theorem 4.2.5 only one family allows a transfer to the rate model. This follows immediately from the previous proof when looking at the generation of the rates, since the logarithm is defined only for the parameter family given in (4.2.23). This completes the proof. □

Proof of Proposition 4.2.10. As already stated, the molecular clock model is based on the tree $\hat{\mathcal{T}} := (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ with

$$\hat{\mathcal{V}} := \{\delta_1, \delta_2, \delta_3, \varrho_1, \varrho_2\}, \quad \hat{\mathcal{E}} := \{(\varrho_1, \varrho_2), (\varrho_1, \delta_3), (\varrho_2, \delta_1), (\varrho_2, \delta_2)\},$$

where $(\delta_1, \delta_2, \delta_3)$ is a permutation of $\mathcal{L}$ and $\varrho_2 = \varrho \in \mathcal{V}$, the root of the initially considered tree $\mathcal{T}$. Molecular clock demands that the rates for edges (ϱ_2, δ_1) and (ϱ_2, δ_2) are equal. Hence,

$$0 = r_{q_{\delta_1}} - r_{q_{\delta_2}} = -\frac{1}{8}\left(\ln \left| \frac{\Delta_v}{d^v_{\delta_2\delta_3}} \right| - \ln \left| \frac{\Delta_v}{d^v_{\delta_1\delta_3}} \right| \right) = \frac{1}{4}(\ln |d^v_{\delta_1\delta_3}| - \ln |d^v_{\delta_2\delta_3}|),$$

i.e. $\ln |d^v_{\delta_1\delta_3}| = \ln |d^v_{\delta_3\delta_3}|$. Thus under the molecular clock the transversion rate for δ_1, δ_2 is given by:

$$r_{q_{\delta_1}} = r_{q_{\delta_2}} = -\frac{1}{8} \ln |d^v_{\delta_1\delta_2}|. \tag{4.3.16}$$

Consider the transition rates. Together with (4.3.16) they yield:

$$0 = r_{p_{\delta_1}} - r_{p_{\delta_2}} = -r_{q_{\delta_1}} - \frac{1}{2} \ln \left| \frac{\Delta_s}{d^s_{\delta_2\delta_3}} \right| + r_{q_{\delta_2}} + \frac{1}{2} \ln \left| \frac{\Delta_s}{d^s_{\delta_1\delta_3}} \right| = \frac{1}{2}(\ln |d^s_{\delta_1\delta_3}| - \ln |d^s_{\delta_2\delta_3}|),$$

i.e. $\ln |d^s_{\delta_1\delta_3}| = \ln |d^s_{\delta_2\delta_3}|$ and thus, the transition rate for δ_1, δ_2 is given by

$$r_{p_{\delta_1}} = r_{p_{\delta_2}} = -\frac{1}{8} \ln \frac{|d^s_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|}. \tag{4.3.17}$$

These insights provide the rates for edge (ϱ_2, δ_3) as

$$r_{q_{\delta_3}} = -\frac{1}{4} \ln |d^v_{\delta_1\delta_3}| + \frac{1}{8} \ln |d^v_{\delta_1\delta_2}|, \quad r_{p_{\delta_3}} = -\frac{1}{4} \ln \frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|} + \frac{1}{8} \ln \frac{|d^s_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|}. \tag{4.3.18}$$

For the remaining two rates the advantage of the rate model becomes apparent. Whereas the transition matrices are related through multiplication, the rate matrices are related through addition due to the property of the exponential function. Denote the needed rates by $r^c_{q_{\delta_3}}$, $r^c_{p_{\delta_3}}$, $r^c_{q_{\varrho_2}}$ and $r^c_{p_{\varrho_2}}$. Then, they are obtained by using the following relationships:

$$\begin{aligned} r^c_{q_{\delta_3}} + r^c_{q_{\varrho_2}} &= r_{q_{\delta_3}}, & \qquad r^c_{p_{\delta_3}} + r^c_{p_{\varrho_2}} &= r_{p_{\delta_3}}, \\ r^c_{q_{\delta_3}} - r^c_{q_{\varrho_2}} &= r_{q_{\delta_1}}, & \qquad r^c_{p_{\delta_3}} - r^c_{p_{\varrho_2}} &= r_{p_{\delta_1}} \end{aligned}$$

and thus,

$$\begin{aligned} 2r^c_{q_{\delta_3}} &= r_{q_{\delta_3}} + r_{q_{\delta_1}}, & 2r^c_{p_{\delta_3}} &= r_{p_{\delta_3}} + r_{p_{\delta_1}}, \\ 2r^c_{q_{\varrho_2}} &= r_{q_{\delta_3}} - r_{q_{\delta_1}}, & 2r^c_{p_{\varrho_2}} &= r_{p_{\delta_3}} - r_{p_{\delta_1}}. \end{aligned}$$

Applying (4.3.16),(4.3.17) and (4.3.18) to this notions yields:

$$\begin{aligned} r^c_{q_{\delta_3}} &= -\frac{1}{8}\ln|d^v_{\delta_1\delta_3}|, & r^c_{q_{\varrho_2}} &= -\frac{1}{8}(\ln|d^v_{\delta_1\delta_3}| - \ln|d^v_{\delta_1\delta_2}|), \\ r^c_{p_{\delta_3}} &= -\frac{1}{8}\ln\frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|}, & r^c_{p_{\varrho_2}} &= -\frac{1}{8}\left(\ln\frac{|d^s_{\delta_1\delta_3}|^2}{|d^v_{\delta_1\delta_3}|} - \ln\frac{|d^s_{\delta_1\delta_2}|^2}{|d^v_{\delta_1\delta_2}|}\right). \end{aligned}$$

This completes the proof. □

Chapter 5

Some Statistical Tools

Chapters 3 and 4 examined the algebraic properties of the equation system (LF) w.r.t. three specifications. In particular, the examinations of Chapter 4 showed that finding a Markov extension to a given leaf distribution is difficult.

The purpose of this chapter is to develop methods for the generation of satisfying approximations of a given leaf distribution by a Markov process. To achieve this goal, some estimators are presented, and various confidence regions are discussed. In addition, an algorithm for finding a phylogenetic tree is presented. It makes use of the parameters established for simple trees as developed in Chapters 3 and 4.

The chapter starts with the introduction of the Likelihood Scoring Functions, which is used as a decision criterion for Maximum Likelihood methods. Its global maximum is presented, and problems in the generation of a maximum under the constraints of an underlying Markov process are discussed. The structure of the function is well-known (e.g. Yang [1994]) and suggests to treat estimated leaf distributions as a random vector for the parameters of a multinomial distribution.

Section 5.2 presents a couple of estimators. In Subsection 5.2.1 a consistent estimator for real-valued vectors, whose elements sum to one, is introduced. In terms of phylogenetic reconstruction its purpose is to manipulate inadmissible transition parameters obtained from solving system (LF). The particular form of the estimator for the models discussed in this work are presented in Subsection 5.2.2. Subsection 5.2.3 provides a Bayesian estimator. Its purpose is the manipulation of input distributions to provide admissible approximations. The main factor for applying this estimator are joint states of probability zero in the joint leaf distribution. As the constructions in Section 5.3 will indicate, most confidence regions provide only marginal coverage if one of the entries of the vector, on which the region is generated, is zero. In particular, such confidence regions contain only those vectors which have entries of value zero for each state that has probability zero in the initially considered distribution. In the light of the observation that joint states of probability zero will result in inadmissible solutions of (LF) such a probability mass redistribution has its own right.

Section 5.3 introduces several kinds of confidence regions and presents statements concerning their applicability. Most statements are taken from Brown et al. [2001]. In Subsection 5.3.1 the Central Limit Theorem is used to consider square Gaussian distributed random variables, which are chi square distributed. In Subsection 5.3.2 the *Clopper-Pearson* confidence interval is presented. In this approach all entries of vector $\underline{m}$ are treated as independent random variables for a parameter p of a Binomial distribution $B_{N,p}$. The confidence region is generated by computing for each vector entry m_i the boundaries $\max_k B_{N,m_i}(0,k) < \eta$ and $\min_k B_{N,m_i}(k,N)$, $i = 1,\dots,K$. Subsection 5.3.3 introduces confidence regions which also treat the vector entries as independent random variables for a parameter p of a Binomial distribution $B_{N,p}$. These types are often compared w.r.t. their average coverage probability (e.g. Brown et al. [2001], Jhun and Jeong [2000], May and Johnson [1997] or Agresti and Coull [1998]). The results of these comparisons are part of the discussion in Subsection 5.3.4.

In Section 5.4 an algorithm is introduced that uses the transition parameters obtained from the triple tree restrictions of the input data. The structure of the algorithm is similar to methods presented in Pearl and Tarsi [1986] or Chang [1996]. Moreover, the method will incorporate the estimator presented in Subsection 5.2.1. The section closes with the application of the algorithm to the well-known *Great Ape* $\{0,1\}-$*data set*.

As before, the end of the chapter is reserved for the proofs.

5.1 The Likelihood Scoring Function

The chapter starts with a motivational consideration of some aspects of Maximum Likelihood methods for phylogenetic reconstruction. Usually Maximum Likelihood methods look for the model configuration which best estimates the observed frequency vector $\underline{n} = N\underline{m}$ (cf. Felsenstein [1981] or Guindon and Gascuel [2003]). The best ML estimator is defined as the maximum of the function:

$$\text{LS}(\underline{n},\underline{p}) = \sum_{\underline{x}\in\mathcal{S}^n} n_{\underline{x}} \ln(p_{\underline{x}}), \qquad \sum_{\underline{x}\in\mathcal{S}^n} p_{\underline{x}} = 1. \tag{5.1.1}$$

This function is usually called the loglikelihood scoring function. The maximum of this function is well-known:

Lemma 5.1.1. *Let* $\underline{n} = N\underline{m}$ *denote a frequency vector.* $\text{LS}(\underline{n},\underline{p})$ *is maximal in* $\underline{p} = \underline{m}$.

Generally, $\underline{p} = \underline{p}(\mathbf{X},\mathcal{T})$ denotes the leaf distribution subject to the model configuration of a Markov process $\mathbf{X}$ on a tree $\mathcal{T}$. Therefore, Maximum Likelihood methods aim for

$$\mathrm{argmax}_{\mathbf{X},\mathcal{T}}\mathrm{LS}(\underline{n},\underline{p}(\mathbf{X},\mathcal{T})).$$

If $\underline{m}$ is subject to a Markov process, any Maximum Likelihood method should return $\underline{m}$ as the best approximation of itself.

The following example provides a visualization of the proposals for the results of Corollary 4.1.8 by using (5.1.1) as a tool of comparison of possible estimations of an initial relative frequency vector $\underline{m}$:

Example 5.1.1. Recall from Example 4.1.1 that on a triple tree the initial relative frequency vector

$$\underline{m} = (100, 15, 15, 10, 5)/1000. \tag{5.1.2}$$

does not have a Neyman extension, but by Theorem 4.1.7 its pairwise distributions yield the following parameters of a possible leaf distribution

$$\underline{m}_1 = (197, 31, 31, 21, 9)/2000, \tag{5.1.3}$$

$$\underline{m}_2 = (49, 32, 32, 27, -12)/1000. \tag{5.1.4}$$

To compare these distribution vectors, their loglikelihood score is computed. For the Neyman N_k model the score function is specified by:

$$\mathrm{LS}_{\underline{a}}(\underline{b}) = ka_1\ln(b_1) + k(k-1)(a_2\ln(b_2) + a_3\ln(b_3) + a_4\ln(b_4) + (k-2)a_5\ln(b_5)),$$

where $\underline{a}$ and $\underline{b}$ satisfy (4.1.4). Example 4.1.1 was considered for $k = 4$ states. The loglikelihood scores for the three distributions are

$$\mathrm{LS}_{\underline{m}}(\underline{m}) = -3621.35, \quad \mathrm{LS}_{\underline{m}}(\underline{m}_1) = -3622.38, \quad \mathrm{LS}_{\underline{m}}(\underline{m}_2) = -3409.67 + 376.991\mathrm{i}.$$

Therefore, in terms of the Likelihood score, the Markov process with triple leaf distribution $\underline{m}_1$ provides a very good approximation of $\underline{m}$. Since $\underline{m}_2$ contains a negative value, the value $\mathrm{LS}_{\underline{m}}(\underline{m}_2)$ cannot be compared with the other families. This problem is treated in Section 5.2.

Concerning the number of maxima of LS for a given vector $\underline{n} = N\underline{m}$ observe the following fact:

Lemma 5.1.2. *Let $\underline{m}$ denote a relative frequency vector over $\mathcal{S}^n$ on a leaf set $\mathcal{L}$ of an unknown tree $\mathcal{T}$. If $\underline{p}(\mathbf{X},\mathcal{T})$ is a maximal leaf distribution with Markov extension for (5.1.1), then also $\underline{p}(\tilde{\mathbf{X}},\mathcal{T})$ with*

$$p_x(\mathbf{X},\mathcal{T}) = p_x(\tilde{\mathbf{X}},\mathcal{T}) \quad \textit{for } x \in \{y \in \mathcal{S}^n : m_y \neq 0\}. \tag{5.1.5}$$

is a maximal leaf distribution with Markov extension.

This observation is due to the fact that for the set $M = \{y \in \mathcal{S}^n : m_y = 0\}$ the equation

$$\sum_{y \in M} n_y \ln(p_y(\mathbf{X}, \mathcal{T})) = 0 \tag{5.1.6}$$

holds for all parameter sets $(\mathbf{X}, \mathcal{T})$. That such cases easily occur becomes clear, when five species and four states are considered. In that case, $4^5 = 1024$ possible joint states exist. Hence, to have a strictly positive relative frequency vector, one needs a set of aligned sequences of length of at least 1024. However, multiple aligned sequences of such lengths are usually not available (cf. Waterman [1995]).

The structure of (5.1.1) also suggests to regard $\underline{n}$ as a random vector for the parameters of a multinomial distribution. This property will be used in the next two sections.

5.2 Consistency and Bayesian Estimation

This section proposes two kinds of estimators. The first estimator tackles the problem of inadmissible transition parameters by providing a consistent quadratic approximation mapping. The second estimator tackles the problem of possible joint states that do not occur in the observed data. Joint states of zero frequency pose problems for inference methods, and also when considering certain kinds of confidence regions.

5.2.1 Estimation with Least Squares

Chapters 3 and 4 showed that under the considered models almost any input vector $\underline{m}$ generates an algebraic solution of (LF). However, only few input vectors generate a stochastically admissible solution. As a consequence, most frequencies derived from the data won't be subject to an Markov process. Therefore, one has to adjust the parameters obtained from the frequencies to fulfill the admissibility conditions.

Given an (inadmissible) set of parameters $(p_1, \ldots, p_k)$, the estimator is constructed as the solution of a least squares problem. In particular, one wants to solve:

$$\min_{\underline{q}=(q_1,\ldots,q_k)} F(\underline{q}), \quad F(\underline{q}) := \sum_{i=1}^{k} |q_i - p_i|^2 \tag{5.2.1}$$

under the constraints

$$q_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{k} q_i = 1. \tag{5.2.2}$$

Since

$$|q_i - p_i|^2 = (q_i - \mathrm{Re}(p_i)^2 + (\mathrm{Im}(p_i))^2$$

it suffices to solve the problem for real $p_1, \ldots, p_k$.

Proposition 5.2.1. *Let $(p_i)_{i=1}^k$ denote a family of real numbers with $p_1 + \cdots + p_k = 1$ and consider the minimization problem (5.2.1) under the constraints (5.2.2). The problem has a unique minimum $\underline{q} = \underline{q}(\underline{p})$. The mapping $\underline{p} \mapsto \underline{q}(\underline{p})$ is continuous.*

The numerical computation of the minimum is subject of the next results:

Proposition 5.2.2. *Let $(p_i)_{i=1}^k$ denote a family of real numbers with $p_1 + \cdots + p_k = 1$. Then there is an index set $I \subset \{1, \ldots, k\}$ such that $\underline{q}$ with $q_i = p_i + c$ for $i \in I$ and $q_i = 0$ for $i \in I^c$, where*

$$c := \frac{1}{\sharp(I)} \sum_{i \in I^c} p_i \tag{5.2.3}$$

is the minimum of (5.2.1).

Hence the estimator $\underline{q}$ can be retrieved by determining I. For this purpose the following properties can be observed:

Corollary 5.2.3. *Let $(p_i)_{i=1}^k$ denote a family of real numbers with $p_1 + \cdots + p_k = 1$ and let I be the index set generated by (5.2.3). Then one observes:*

$$p_i > -c \quad \textit{for all } i \in I \textit{ and } p_j \leq -c \textit{ for all } i \in I^c, \tag{5.2.4}$$

and the following order relation is found:

1. *If $p_i \geq p_j$ and $j \in I$, then also $i \in I$.*
2. *If $p_i \leq p_j$ and $j \in I^c$, then also $i \in I^c$.*
3. *If $p_i \leq 0$, then $i \in I^c$.*

Corollary 5.2.3 permits an ordering of the family $(p_i)_{i=1}^k$ with $p_i < p_{i+1}$ and l is the index for which $p_{l-1} \in I^c$ and $p_l \in I$. Moreover, 3 shows that all negative values are projected into zero. These observations permit the following introduction of an algorithm for the generation of the estimator family:

Algorithm 5.1. Let $(\hat{p}_i)_{i=1}^k$ denote a real-valued parameter family with $\hat{p}_1 + \cdots + \hat{p}_k = 1$.

1. Sort the family $(\hat{p}_i)_{i=1}^k$ such that $p_i < p_{i+1}$ for $i = 1, \ldots, k-1$. Then an index

$1 \leq l \leq k$ exists with $p_{l-1} \leq 0$ and $p_l > 0$. Set $I^c := \{1, \ldots, l-1\}$ and $j := l - 1$.

2. Compute c as given in (5.2.3), set $q_j = 0$ for all $j \in I^c$ and set $j := j + 1$.

3. If $\hat{p}_j \leq -c$ set $I^c := I^c \cup \{j\}$ and go to step 3, else set $q_i := \hat{p}_i + c$ for all $i \geq j$ and STOP.

The algorithm has the nice property that it returns the minimum proposed in Proposition 5.2.2 as a side effect.

Proposition 5.2.4. *The family $(q_i)_{i=1}^k$ obtained from $(p_i)_{i=1}^k$ through Algorithm 5.1 is the minimum of (5.2.1) under the side conditions (5.2.2).*

Coming back to the initial problem of estimating the true process of evolution from a given relative frequency vector $\underline{\widehat{m}}$ at the leaves regard a set of inadmissible transition parameters $(\underline{\hat{p}}^\alpha)_{\alpha \in \mathcal{L}}$ and $\underline{\hat{q}}^\varrho$. An application of Algorithm 5.1 yields a consistent estimator:

Theorem 5.2.5. *Let $\underline{\widehat{m}}$ denote an estimated leaf distribution to a tree $\mathcal{T}$, $(\underline{\hat{p}}^\alpha)_{\alpha \in \mathcal{V} \setminus \{\varrho\}}$ and $\underline{\hat{q}}^\varrho$ the parameters retrieved by solving system (LF) w.r.t. $\underline{\widehat{m}}$. Then, the projections $(\underline{p}^\alpha)_{\alpha \in \mathcal{V} \setminus \{\varrho\}}$ and $\underline{q}^\varrho$ to $(\underline{\hat{p}}^\alpha)_{\alpha \in \mathcal{V} \setminus \{\varrho\}}$ and $\underline{\hat{q}}^\varrho$, respectively, obtained by applying Proposition 5.2.1, are consistent estimators for the true process.*

Note, that the proposed consistency result only states that the boundary points of the estimator sequence are a solution of the true equations. Moreover, this solution is either the true parameter or one of the permutations (see Lemma 2.4.1 for an explanation concerning uniqueness up to permutation). Generally, consistency is a difficult property when considering phylogenetic reconstruction (cf. Chang [1996]).

5.2.2 Some Examples

Chapters 3 and 4 discussed three different models. This subsection addresses the least squares estimator from Subsection 5.2.1 for each model. The two state model and the Neyman N_k model do not really provide a challenge when establishing the quadratic estimator but the Kimura 2ST model is more demanding. Sometimes generated transition parameters are complex. Since the estimators are only constructed for real parameters, only the real part of a complex number is considered.

The General Two-State-Case

As shown in Theorem 3.1.4, the two symmetric solutions for a given triple leaf distribution are related through row permutation in the transition matrices. This implies that inadmissibility of one solutions results in inadmissibility of the other.

Thus, either no optimization is needed, or both solutions need to be optimized simultaneously again yielding the same leaf distribution. For arbitrary trees the following conclusion from Proposition 5.2.1 can be drawn:

Corollary 5.2.6. *Let $(\hat{\underline{p}}^{\alpha})_{\alpha\in\mathcal{V}\setminus\{\varrho\}}$ and $\hat{\underline{q}}^{\varrho}$ denote the transition parameters derived from a joint leaf distribution on a rooted tree $\mathcal{T}_{\varrho}$ under the two state model. If $\hat{p}^{\delta}_{xy} < 0$ for some $\delta \in \mathcal{V} \setminus \{\varrho\}$ and $x, y \in \{0,1\}$ then $\hat{p}^{\delta}_{x(1-y)} > 1$ and the estimator is $p^{\delta}_{xy} = 0$ and $p^{\delta}_{x(1-y)} = 1$. An analogue observation holds for the estimator $\underline{q}^{\varrho}$ for $\hat{\underline{q}}^{\varrho}$.*

The estimator projects inadmissible values into the boundaries of admissibility. To illustrate the stated properties consider the following example:

Example 5.2.1. The software package Tree-Puzzle(cf. Schmidt et al. [2002]) presents a tool to derive phylogenies from sequence data using Maximum Likelihood methods on quartet trees. It also presented the author of this text with his first set of aligned sequence data, namely the Great Ape $\{0,1\}$-data set of five species and of length 895. For the purpose of illustration consider the triple Human-Chimp-Orangutan. The frequency vector of the sequence has the following form:

$$\underline{m}_{HCO} = (506, 18, 0, 2, 1, 2, 14, 352)/895.$$

The inferred transition parameters using Theorem 3.1.4 have the following form:

$$P^H = \begin{pmatrix} 0.998182 & 0.00181834 \\ 0.00565767 & 0.994342 \end{pmatrix}, \qquad P^C = \begin{pmatrix} 1.00016 & -0.000157454 \\ 0.00555766 & 0.994442 \end{pmatrix},$$
$$P^O = \begin{pmatrix} 0.965669 & 0.0343305 \\ 0.0382517 & 0.961748 \end{pmatrix}, \qquad \underline{q}^{\varrho} = \begin{pmatrix} 0.586436 \\ 0.413564 \end{pmatrix}.$$

The second solution is attained through row permutation for every matrix and an element switch for the root vector. Clearly, P^C is not stochastically admissible. According to Corollary 5.2.6 the estimation with the identity matrix is consistent. The remaining matrices are stochastically admissible. The resulting estimated leaf distribution has the following form:

$$\widehat{\underline{m}}_{HCO} = (505.92, 17.9972, 0.0796589, 2.00283, 0.999855, 1.99999, 14.0001, 352)/895.$$

The mentioned row permutation for the estimated matrix yields a parameter set for the same distribution vector. Simple comparison shows that $\underline{m}_{HCO}$ and $\widehat{\underline{m}}_{HCO}$ differ only slightly. Qualitative analysis of such results is subject of Section 5.3.

The Neyman N_k Model

From the point of transition parameters this model is similar to the two state model. In particular, the estimator from Proposition 5.2.1 has the following property under the model:

Corollary 5.2.7. *Let* $(\hat{p}_\alpha)_{\alpha\in\mathcal{V}\setminus\{\varrho\}}$ *denote the transition parameters derived from a joint leaf distribution on a rooted tree* $\mathcal{T}_\varrho$ *under the Neyman* N_k *model.*

1. *If* $\hat{p}_\delta < 0$ *for a* $\delta \in \mathcal{V}\setminus\{\varrho\}$*, then the estimator is* $p_\delta = 0$*.*
2. *If* $\hat{p}_\delta > 1/(k-1)$ *for a* $\delta \in \mathcal{V}\setminus\{\varrho\}$*, then the estimator is* $p_\delta = 1/(k-1)$*.*

Again, inadmissible transition parameters are projected into the boundaries of admissibility. However, as shown in Corollary 4.1.8 and Example 4.1.1 the two established parameter families yield different distribution vectors, $\underline{\widehat{m}}$ and $\underline{\widetilde{m}}$, where neither is necessarily equal to the initial vector $\underline{m}$.

Example 5.2.2. Recall the relative frequency vector $\underline{m}_2$ from Example 5.1.1. The inadmissibility of $\underline{m}_2$ provides an opportunity to apply Corollary 5.2.7. Example 4.1.1 showed that the second parameter set p^2 had three inadmissible values. Therefore by Corollary 5.2.7 all are transferred to 1/3. Such a parameter set yields the triple leaf distribution

$$\underline{\widehat{m}}_2 = (3,2,2,2,1)/108 \tag{5.2.5}$$

with a loglikelihood score of $ls_{\underline{m}}(\underline{\widehat{m}}_2) = -3909.98$, i.e. smaller than the score for $\underline{m}_1$.

The Kimura 2ST Model

The previous two models were simple projections into the boundaries of admissibility. In contrast to this, parameters under the Kimura 2ST model provide the opportunity to observe the least squares estimator in a nontrivial fashion.

Corollary 5.2.8. *Let* $(\hat{p}_\alpha, \hat{q}_\alpha)_{\alpha\in\mathcal{V}\setminus\{\varrho\}}$ *denote the transition parameters derived from a joint leaf distribution on a rooted tree* $\mathcal{T}_\varrho$ *under the Kimura 2ST model. Further, denote by* $\hat{r}_\delta = 1 - \hat{p}_\delta - 2\hat{q}_\delta$ *for* $\delta \in \mathcal{V}\setminus\{\varrho\}$ *the diagonal parameter of the associated transition matrix. If for* $\delta \in \mathcal{V}\setminus\{\varrho\}$ *the transition parameters are inadmissible, then one of the following scenarios is possible:*

1. *If* $\hat{r}_\delta < 0$ *then* $r_\delta = 0$ *and the remaining estimated parameters have either the form* $p_\delta = \hat{p}_\delta + \hat{r}_\delta/3$, $q_\delta = \hat{q}_\delta + \hat{r}_\delta/3$ *or* $p_\delta = 1$, $q_\delta = 0$ *or* $p_\delta = 0$, $q_\delta = 1/2$*.*
2. *If* $\hat{p}_\delta < 0$ *then* $p_\delta = 0$ *and the remaining estimated parameters have either the form* $r_\delta = \hat{r}_\delta + \hat{p}_\delta/3$, $q_\delta = \hat{q}_\delta + \hat{p}_\delta/3$ *or* $r_\delta = 1$, $q_\delta = 0$ *or* $r_\delta = 0$, $q_\delta = 1/2$*.*

3. *If $\hat{q}_\delta < 0$ then $q_\delta = 0$ and the remaining estimated parameters have either the form $r_\delta = \hat{r}_\delta + \hat{q}_\delta$, $p_\delta = \hat{p}_\delta + \hat{q}_\delta$ or $r_\delta = 1$, $p_\delta = 0$ or $r_\delta = 0$, $p_\delta = 1$.*

Hence if one parameter is less than zero, its value is distributed on the remaining two parameters or if one parameter value falls under the condition of step 3 in the algorithm the remaining parameter has the whole probability mass. With Theorem 4.2.5 and Corollary 4.2.6 one observes that for a given triple leaf distribution a possible application of the estimation algorithm would affect two parameter families simultaneously. Again, a qualitative consideration of the derived leaf distributions is provided in Section 5.3.

5.2.3 A Bayesian Estimator

The following estimator manipulates the observed frequency distributions. The advantage of the presented estimator is, that it is the best Bayesian estimator for any possible leaf distribution w.r.t. the maximal quadratic distance. Consider the state space $\{1, 2, \ldots, K\}$ and a frequency vector $\underline{n} = (n_i)_{i=1}^K$ with $N = n_1 + \cdots + n_K$. The goal of this section is the selection of a vector $\underline{p} = (p_i)_{i=1}^K$ from the space of multinomial probability vectors of which the relative frequency vector $\underline{m} = \underline{n}/N$ is a good approximation.

The goal is to derive a good approximation for the normalized vector $\underline{m} = \underline{n}/N$ from the space of multinomial probability vectors $\underline{p} = (p_i)_{i=1}^K$.

Lemma 5.2.9. *Let $\tilde{\underline{p}}$ denote a multinomial vector and $\underline{n} = N\tilde{\underline{p}}$. Define the vector $\hat{\underline{p}}$ by*

$$\hat{p}_i = \hat{p}_i(\underline{n}) = \frac{\frac{\sqrt{N}}{K} + n_i}{\sqrt{N} + N}. \tag{5.2.6}$$

Then one observes

$$\mathbb{E}_{\underline{p}} \|\hat{\underline{p}} - \underline{p}\|^2 = c = const$$

for all vectors $\underline{p} = (p_i)_{i=1}^K$.

Therefore, the estimator $\hat{\underline{p}}$ has constant risk. In the terminology of Ferguson [1967], it is an *equalizer rule*. Moreover, $\hat{\underline{p}}$ is a Bayesian estimator with the following interesting property:

Lemma 5.2.10. *The estimator $\hat{\underline{p}}$ defined by (5.2.6) is a Bayesian estimator with*

$$\mathbb{E}_{\underline{p}} \|\hat{\underline{p}} - \underline{p}\|^2 = \inf_{\tilde{\underline{p}}} \sup_{\underline{p}} \mathbb{E}_{\underline{p}} \|\tilde{\underline{p}} - \underline{p}\|^2.$$

The vector $\hat{\underline{p}}$ is the Bayesian estimator with the smallest maximal distance to all multinomial vectors. For decreasing N the value given to a state of zero frequency increases. For example, if one wants to approximate three aligned sequences of length 500 under the two state model with $\hat{\underline{p}}$, a zero frequency state will be assigned with probability 0.0054 which is quite a huge jump in probability.

Section 5.3 will introduce several simultaneous confidence regions as tools of evaluating estimated leaf distributions for their approximations of a given relative frequency vector $(p_1, \ldots, p_K)$. One such confidence regions is given by:

$$\left\{\pi_1, \ldots, \pi_K : \sum_{i=1}^{K} \pi_i = 1,\ n \sum_{i=1}^{K} \frac{(p_i - \pi_i)^2}{p_i} \leq \chi^2_{K-1}(1-\eta)\right\}, \tag{5.2.7}$$

where $\chi^2_{K-1}(1-\eta)$ is the upper $(1-\eta)$ quantile of the chi-squared distribution with $K-1$ degrees of freedom. This region was suggested in Jhun and Jeong [2000]. Clearly, if $p_i = 0$ for a $i \in \{1, \ldots, K\}$, then this confidence region cannot be computed. Hence, a manipulation of the initial distribution can be advantageous for later considerations. Moreover, it was observed that zero frequency states generally do not allow admissible solutions. For an illustration consider the following example:

Example 5.2.3. Let $\mathcal{T}$ denote a triple tree and let $\mathcal{S} = \{0,1\}$. The observed joint frequency distribution at the leaves $\mathcal{L} := \{1,2,3\}$ is given by:

$$\underline{m} = \{550, 10, 0, 15, 0, 2, 5, 418\}/1000.$$

With (5.2.6) the following estimated distribution is generated:

$$\widehat{\underline{m}} = (536.972, 13.5251, 3.83168, 18.3719, 3.83168, 5.77037, 8.67841, 409.019)/1000.$$

Computing the transition parameters for both vectors yields

$$p^1_{\underline{m}} = \begin{pmatrix} 1.00004 & -0.00004 \\ 0.0346493 & 0.965351 \end{pmatrix}, \qquad p^2_{\underline{m}} = \begin{pmatrix} 1.00033 & -0.00033 \\ 0.00476293 & 0.995237 \end{pmatrix},$$
$$p^3_{\underline{m}} = \begin{pmatrix} 0.982269 & 0.0177312 \\ 0.0118203 & 0.98818 \end{pmatrix}, \qquad q^{\varrho}_{\underline{m}} = \begin{pmatrix} 0.55972 \\ 0.44028 \end{pmatrix},$$

for $\underline{m}$, where $p^1_{\underline{m}}$ and $p^2_{\underline{m}}$ are obviously no transition matrices. For $\widehat{\underline{m}}$ one computes:

$$p^1_{\widehat{\underline{m}}} = \begin{pmatrix} 0.993135 & 0.00686456 \\ 0.0427956 & 0.957204 \end{pmatrix}, \qquad p^2_{\widehat{\underline{m}}} = \begin{pmatrix} 0.993626 & 0.00637429 \\ 0.0136935 & 0.986307 \end{pmatrix},$$
$$p^3_{\widehat{\underline{m}}} = \begin{pmatrix} 0.975881 & 0.0241191 \\ 0.020721 & 0.979279 \end{pmatrix}, \qquad q^{\varrho}_{\widehat{\underline{m}}} = \begin{pmatrix} 0.557596 \\ 0.442404 \end{pmatrix},$$

i.e. a stochastically admissible solution. Finally, a look at the confidence region (5.2.7) shows that:

$$13.737 \approx 1000 \sum_{i=1}^{8} \frac{(\widehat{m}_i - m_i)^2}{\widehat{m}_i} \leq \chi^2_7(0.05) \approx 14.1.$$

Therefore, the initial leaf distribution $\underline{m}$ is in the 0.95-quantile of the Bayesian estimate $\widehat{\underline{m}}$.

5.3 Simultaneous Confidence Regions

After providing ways to manipulate generated inadmissible transition parameters, the manipulated parameters will be evaluated. Probably the most popular evaluation approach is the declaration of simultaneous confidence regions. Several papers (e.g. Goodman [1964] or Jhun and Jeong [2000]) deal with the comparison of different choices of confidence regions for the parameters of multinomial distributions. This section will introduce some of the suggested intervals and apply them to estimated distributions for the models discussed in Chapters 3 and 4.

As usual, $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ denotes a tree, $n = \sharp(\mathcal{L})$ the number of leaves and $k = \sharp(\mathcal{S})$ the number of states. The estimated leaf distribution taken from the input data is denoted by $\widehat{\underline{m}}$.

5.3.1 A Chi Square Approach

Simultaneous confidence regions permit the comparison of a selected process with the input data by checking if the derived leaf distribution is in a $(1-\eta)$-confidence region of the leaf frequency distribution, and the acceptance or rejection of the selected process accordingly.

The first approach uses the Central Limit Theorem to generate an η-confidence region.

Consider a family $(Y^i)_{i=1}^N$ of i.i.d. random variables with values in $\{e_1, \ldots, e_{k^n}\}$, the unit vectors of $\mathbb{R}^{k^n}$, distributed according to the estimated leaf distribution $\underline{m}$, i.e. $\mathbb{P}(Y^i = e_l) = m_l$, $l \in \{1, \ldots, k^n\}$. This family has the mean vector $\mathbb{E}(Y^i) = \underline{m}$ and the covariance matrix

$$\text{Cov}(Y^i) = \mathcal{C} \quad \text{with} \quad \mathcal{C}_{st} = m_s(\delta_{st} - m_t),\ s, t \in \{1, \ldots, k^n\}. \tag{5.3.1}$$

According to the Central Limit Theorem (e.g. Witting and Müller-Funk [1995, Satz 5.105]) the following distribution assumption is reasonable:

Lemma 5.3.1. *Let $(Y^i)_{i=1}^{\infty}$ denote a family of i.i.d. random variables with mean vector $\underline{m}$ and covariance matrix $\mathcal{C}$ as given in (5.3.1). Then, the empirical mean*

$$\bar{Y}_N = \frac{1}{N}\sum_{j=1}^{N} Y^j.$$

of $(Y^i)_{i=1}^{N}$ is asymptotical Gaussian with mean vector 0 *and covariance matrix $\mathcal{C}$, i.e.*

$$\mathfrak{D}(\sqrt{N}(\bar{Y}_N - \underline{m})) \overset{N\to\infty}{\Longrightarrow} \mathcal{N}(0, \mathcal{C}), \tag{5.3.2}$$

where the symbol $\mathfrak{D}(X)$ denotes the distribution of the random variable X. □

The vector $\bar{Y}$ is subject to the general model. Certain model specifications will focus on linear transformations. For such occurrences the following observation can be made:

Corollary 5.3.2. *Let $(Y^i)_{i=1}^{N}$ denote a family of i.i.d. random variables with mean vector $\underline{m} \in [0,1]^K$ and covariance matrix $\mathcal{C} \in \mathbb{R}^{K\times K}$. Further, let $A : \mathbb{R}^K \to \mathbb{R}^M$, $M > 0$ denote a linear mapping. Then $\bar{Z}_N = A\bar{Y}_N$ is asymptotically Gaussian distributed with mean vector $\underline{0}$ and covariance matrix $A\mathcal{C}A^{\mathrm{T}}$, i.e.*

$$\mathfrak{D}(\sqrt{N}(\bar{Z}_N - A\underline{m})) \overset{N\to\infty}{\Longrightarrow} \mathcal{N}(0, A\mathcal{C}A^{\mathrm{T}}).$$

The matrix A for the Neyman N_k model and the Kimura 2ST model on a triple tree are characterized by (4.1.3) and (4.2.4), respectively. The explicit form of the covariance matrices is introduced later. The asymptotic behavior of the distribution vectors permits the following statement for confidence regions:

Theorem 5.3.3. *Let $(Y^i)_{i=1}^{N}$ denote a family of i.i.d. random variables with mean vector $\underline{m} \in [0,1]^K$ and covariance matrix $\mathcal{C} \in \mathbb{R}^{K\times K}$ of rank l. The asymptotic η-confidence region for $\underline{m}$ is given by:*

$$\mathrm{CI}_{\chi^2}(\underline{m}, \eta) := \left\{\underline{p} \in [0,1]^K : N\|\underline{p} - \underline{m}\|_{\mathcal{C}}^2 < \chi_l^2(1-\eta)\right\},$$

where $\chi_l^2(1-\eta)$ denotes the $(1-\eta)$-quantile of the chi square distribution with l degrees of freedom.

To verify whether a leaf distribution obtained through a phylogenetic method is in such a confidence region one needs to compute the Pseudo-Inverse of the covariance matrix $\mathcal{C}^{-1}$. For problems of high dimensions this might pose difficulties. The confidence region presented in (5.2.7) provides an alternative chi square approach where the covariance matrix is replaced by the diagonal matrix $\mathcal{C} = \operatorname{diag}(m_1, \ldots, m_{k^n})$. In this alternative approach the degree of freedom is fixed at $k^n - 1$ contrary to (5.5.4) where the degree of freedom depends on the rank of $\hat{\mathcal{C}}$. On the other hand, if a leaf distribution $\underline{m}$ assigns probability zero to certain states, (5.2.7) cannot be applied.

Example 5.3.1. Here the Neyman distributions will be considered. As promised, the covariance matrix for the Neyman distribution on a triple tree is the first thing to be introduced. For lack of space, the matrix is divided into off-diagonal- and diagonal-elements $\big((a_1, a_2, a_3, a_4, a_5) := (m_{000}, m_{001}, m_{010}, m_{100}, m_{012})\big)$:

$$\begin{aligned} &A_{11}^{\text{Ney}} = \frac{a_1(1-ka_1)}{k}, \quad A_{ii}^{\text{Ney}} = \frac{a_i(1-k(k-1)a_i)}{k(k-1)}, \quad i = 2,3,4, \\ &A_{55}^{\text{Ney}} = \frac{a_5(1-k(k-1)(k-2)a_5)}{k(k-1)(k-2)}, \quad A_{ij}^{\text{Ney}} = -a_i a_j, \quad i \neq j. \end{aligned} \tag{5.3.3}$$

With this insight apply Example 4.1.1 to consider the presented confidence region. Here the number of states is $k = 4$. The initial leaf distribution $\underline{m}$ is given by (5.1.2), and the leaf distributions $\underline{m}_1$ and $\underline{m}_2$ inferred using Theorem 4.2.5 are given by (5.1.3) and (5.1.4), respectively. The covariance matrix $\mathcal{C}$ for $\underline{m}$ can easily be computed with (5.3.3). With this one computes (recall $N = 1000$):

$$N\|\underline{m}_1 - \underline{m}\|^2_{\mathcal{C}^{\underline{m}}} = \frac{199}{100}, \quad \|\underline{m}_2 - \underline{m}\|^2_{\mathcal{C}^{\underline{m}}} = \frac{57511}{25}, \quad \|\widehat{\underline{m}}_2 - \underline{m}\|^2_{\mathcal{C}^{\underline{m}}} = \frac{293500}{729}.$$

For $\eta = 0.05$ the upper $(1-\eta)$-quantile for a chi square distribution with four degrees of freedom is 9.48773. Consequently, $\underline{m}_1$ lies well inside the confidence region whereas $\underline{m}_2$ and $\widehat{\underline{m}}_2$ miss it by quite a large margin.

5.3.2 Clopper-Pearson Confidence Regions

This subsection presents a class of confidence regions where each entry of the vector $\underline{m} = (m_i)_{i=1}^K$ is independently considered as the probability parameter p of a Binomial distribution $B_{N,p}$. This approach was first introduced in Clopper and Pearson [1934]. Consider the following region:

$$\mathrm{CI}(p, \eta) = \{\nu \in \mathbb{R} : \, k_l^\eta(p) \leq \nu \leq k_u^\eta(p)\}.$$

Let f_i and g_i, $i = 1, \ldots, N$ be functions with $q \in [0,1]$ and

$$f_i(q) := B_{N,q}(\{0, \ldots, i\}) = \sum_{j=0}^{i} \binom{N}{j} q^j (1-q)^{N-j},$$

$$g_i(q) := B_{N,q}(\{i+1, \ldots, N\}) = \sum_{j=i}^{N} \binom{N}{j} q^j (1-q)^{N-j}.$$

Clearly, $f_i(0) = 1$ and $f_i(1) = \delta_{Ni}$ for all $i \in \{0, \dots, N\}$. Thus, f is decreasing in q and increasing in i. Further, $g_i(0) = \delta_{0i}$ and $g_i(1) = 1$ for all i, i.e. g is increasing in q and decreasing in i. For a given confidence level η compute the following integers:

$$k_l^\eta(p) := \text{argmax}_i \big\{ i \in \{0, \dots, N\} : f_i(p) < \eta/2 \big\},$$
$$k_u^\eta(p) := \text{argmax}_i \big\{ i \in \{0, \dots, N\} : g_i(p) < \eta/2 \big\}.$$

Using these boundary terms the confidence interval $\text{CI}(p, \eta)$ is rewritten as:

$$\text{CI}(p, \eta) := \{\nu : k_l^\eta(p) \leq N\nu \leq k_u^\eta(p)\}. \tag{5.3.4}$$

In honor of the initial contributors the interval $\text{CI}(p, \eta)$ is also known as the *Clopper-Pearson* interval. Accordingly, the *Binomial* or *Clopper-Pearson confidence region* $\text{CI}_{\text{CP}}(\underline{m}, \eta)$ for a vector $\underline{m} \in \mathbb{R}_+^K$ and a confidence level η is given by:

$$\text{CI}_{\text{CP}}(\underline{m}, \eta) = \text{CI}(m_1, \eta/K) \times \cdots \times \text{CI}(m_K, \eta/K). \tag{5.3.5}$$

Brown et al. [2001] propose that the presented bounds k_l^η and k_u^η are nothing more than the $\eta/2$ quantile of a beta distribution $\beta(x, N - x + 1)$, and the $1 - \eta/2$ quantile of a beta distribution $\beta(x + 1, N - x)$, respectively. The considerations of this subsection are concluded by the following example:

Example 5.3.2. The saga of the vectors $\underline{m}, \underline{m}_1, \underline{m}_2$ and $\widehat{\underline{m}}_2$ given by (5.1.2) to (5.2.5) continues. First, the values $k_l^\eta(p)$ and $k_u^\eta(p)$ need to be computed for some $p \in [0, 1]$. Since each vector entry m_i, $(i = 1, \dots, K)$ is considered independently, no weights need to be considered. Due to the negative value in $\underline{m}_2$ this vector cannot be found in any Clopper-Pearson region, and is thus immediately rejected.
For $\eta = 0.05$ one computes the following boundaries for $\text{CI}_{\text{CP}}(\underline{m}, \eta)$:

$$b_l^{\text{CP}}(\underline{m}, \eta) = (76, 6, 6, 3, 0),$$
$$b_u^{\text{CP}}(\underline{m}, \eta) = (126, 27, 27, 20, 13).$$

As stated above, the exact boundaries are given by considering the $\eta/(2K)$ quantile for a beta distribution $\beta(Nm_i, N(1 - m_i) + 1)$ and the $1 - \eta/(2K)$ quantile for a beta distribution $\beta(Nm_i + 1, N(1 - m_i))$ for the lower and upper boundary, respectively. Using this approach, the exact Clopper-Pearson confidence region for $N\underline{m}$ for the confidence level $\eta = 0.05$ is given by boundary vectors

$$b_l^{\text{CP}\beta}(\underline{m}, \eta) = (77.0217, 6.91801, 6.91801, 3.72678, 1.07951), \quad \text{and}$$
$$b_u^{\text{CP}\beta}(\underline{m}, \eta) = (126.88, 27.9788, 27.9788, 21.2761, 14.0851).$$

The vectors $N\underline{m}_1$ and $N\widehat{\underline{m}}_2$ have the following form:

$$N\underline{m}_1 = (98.5, 15.5, 15.5, 10.5, 4.5),$$
$$N\widehat{\underline{m}}_2 = (27.7778, 18.5185, 18.5185, 18.5185, 9.25926).$$

Thus, relating $\underline{m}_1$ and $\widehat{\underline{m}}_2$ to these confidence regions again shows that for $\underline{m}$, $\underline{m}_2$ is a better estimate than $\widehat{\underline{m}}_2$. Hence $\underline{m}_1$ is accepted for all considered confidence regions whereas $\widehat{\underline{m}}_2$ is rejected.

5.3.3 Simultaneous Confidence Regions for Binomial Proportions

This subsection presents another type of confidence region where again each entry of the vector $\underline{m} = (m_i)_{i=1}^{K}$ is independently considered as the probability parameter p of a Binomial distribution $B_{N,p}$. In particular, the following type of confidence interval is considered for parameter p and a confidence level $\eta > 0$:

$$\mathrm{CI}(p,\eta) := \left\{ \nu \in p \pm Q\big(\frac{\eta}{2}\big)\sqrt{\frac{p(1-p)}{N}} \right\}, \tag{5.3.6}$$

where $Q(\eta/2)$ is the $1-\eta/2$-quantile to a chosen symmetric distribution, and $\sqrt{p(1-p)/N}$ is the estimated standard deviation (see (5.3.1)). Such interval types are the center of numerous discussions, most notably in Brown et al. [2001]. If the standard gaussian distribution is chosen as the quantile distribution, the interval is called the *Wald confidence interval* for p (cf. Agresti and Coull [1998]). Another choice is the Student t distribution in $K-1$ degrees of freedom. The simultaneous confidence region for a vector $\underline{m}$ is constructed similarly to (5.3.5). Such types are called *Bonferroni simultaneous confidence regions* to confidence level $1-\eta$ (cf. Jhun and Jeong [2000]). For a comparison consider the following example:

Example 5.3.3. Recall $\underline{m}, \underline{m}_1$ and $\widehat{\underline{m}}_2$ given by (5.1.2) to (5.2.5). The confidence region $\mathrm{CI}(\underline{m},\eta)$ given by (5.3.5) and (5.3.6) is considered for a couple of popular choices of quantile distribution.
Considering Wald confidence intervals to the confidence level $\eta = 0.05$, i.e. using the quantile of the standard gaussian distribution, yields the following boundaries for the confidence region $\mathrm{CI}_{\mathrm{Wald}}(\underline{m},\eta)$:

$$\begin{aligned} b_l^{\mathrm{Wald}}(\underline{m},\eta) &= (75.5635, 5.09896, 5.09896, 1.89534, -0.745312),\\ b_u^{\mathrm{Wald}}(\underline{m},\eta) &= (124.436, 24.901, 24.901, 18.1047, 10.7453).\end{aligned}$$

If $Q(\eta/10)$ denotes the $(1-\eta/10)$-quantile for the Student t distribution with four degrees of freedom the following boundaries for the confidence region $\mathrm{CI}_{\mathrm{T}}(\underline{m},\eta)$ are observed:

$$\begin{aligned} b_l^{T}(\underline{m},\eta) &= (56.3217, -2.69734, -2.69734, -4.48645, -5.2693),\\ b_u^{T}(\underline{m},\eta) &= (143.678, 32.6973, 32.6973, 24.4864, 15.2693).\end{aligned}$$

Hence, $\underline{m}_1$ is in the 0.95-confidence regions whereas $\widehat{\underline{m}}_2$ misses the regions by quite a margin.

A comparison the boundaries of the presented intervals with the boundaries from Example 5.3.2 reveals that these confidence intervals $\mathrm{CI}(p)$ are not centered around

p but contain a marginal shift. This shift was already observed in Wilson [1927]. In Brown et al. [2001] the *Wilson confidence interval* is presented as:

$$\mathrm{CI}_{\mathrm{W}}(p,\kappa) = \left\{ \nu \in \frac{Np + \kappa^2/2}{N + \kappa^2} \pm \frac{\kappa\sqrt{N}}{N+\kappa^2}\sqrt{p(1-p) + \kappa^2/(4N)} \right\}$$

where κ is the $(1-\eta)$-quantile of the standard normal distribution. κ can also be interpreted as the number of successes and failures added to the data set. The approach was recalled in Agresti and Coull [1998]. Brown et al. [2001] present the *Agresti-Coull confidence interval* as:

$$\mathrm{CI}_{\mathrm{AC}}(p,\kappa) = \left\{ \nu \in \tilde{p} \pm \kappa\sqrt{\frac{\tilde{p}(1-\tilde{p})}{N+\kappa^2}} \right\},$$

where $\tilde{p} = (Np + \kappa^2/2)/(N + \kappa^2)$. Both intervals are centered around $\tilde{p}$, and are recommended by Brown et al. [2001]. Other works, like Jhun and Jeong [2000], use a constant *continuity correction* to move the interval into a more appropriate center. The following example tests the performance of CI_{W} and $\mathrm{CI}_{\mathrm{AC}}$ on the already often used vector $\underline{m}$:

Example 5.3.4. Recall $\underline{m}, \underline{m}_1$ and $\widehat{\underline{m}}_2$ given by (5.1.2) to (5.2.5). Also, set $\eta = 0.05$ and $\kappa = 1.96$. With these settings the modified vector $\widehat{\underline{m}}$ has the form:

$$\widehat{\underline{m}} = (101.531, 16.856, 16.856, 11.8752, 6.89431)/1000.$$

The associated Wilson confidence region $\mathrm{CI}_{\mathrm{W}}(\underline{m},\kappa)$ has the boundaries:

$$\begin{aligned} b_l^{\mathrm{W}}(\underline{m},\kappa) &= (82.9092, 9.1109, 9.1109, 5.4407, 2.1375), \\ b_u^{\mathrm{W}}(\underline{m},\kappa) &= (120.152, 24.6012, 24.6012, 18.3097, 11.6511). \end{aligned}$$

Similarly, the Agresti-Coull confidence region $C_{\mathrm{AC}}(\underline{m},\kappa)$ has the boundaries:

$$\begin{aligned} b_l^{\mathrm{AC}}(\underline{m},\kappa) &= (82.8107, 8.87716, 8.87716, 5.16117, 1.76571), \\ b_u^{\mathrm{AC}}(\underline{m},\kappa) &= (120.251, 24.8349, 24.8349, 18.5892, 12.0229). \end{aligned}$$

As in Example 5.3.3 the vector $\underline{m}_1$ is in both confidence regions whereas $\widehat{\underline{m}}_2$ misses them by quite a margin in the first variable.

5.3.4 Discussion

This section presented various criteria of quality management. Subsections 5.3.1, 5.3.2 and 5.3.3 introduced several types of confidence regions.

The chi square confidence region provided in Subsection 5.3.1 provides good testing criteria, but computing the needed pseudo-inverse of the covariance matrix poses big problems for long vectors $\underline{m}$. In addition, every entry of $\underline{m}$ that is zero lessens the rank of the covariance matrix and therefore the degrees of freedom for the chi square distribution. Thus, less states will make the confidence region smaller. The alternatively suggested interval (5.2.7) is not applicable as soon as $\underline{m}$ has zero entries.

In Subsection 5.3.2 the Clopper-Pearson interval for Binomial proportions is introduced. This approach proposes confidence intervals for parameters p to a Binomial distribution $B_{N,p}$. The leaf distributions $\underline{m}$ are related to this intervals by treating every entry of $\underline{m} = (m_1, \ldots, m_K)$ as a parameter for a Binomial distribution B_{N,m_i}, $i = 1, \ldots, K$. In Brown et al. [2001] the following statements are made concerning such intervals:

> page 113: ...Some authors refer to this as the "exact" procedure because of its derivation from the binomial distribution...
>
> ...The Clopper-Pearson interval is wastefully conservative and is not a good choice for practical use, unless strict adherence to the prescription $C(\underline{m}) \geq 1 - \eta$ is demanded...

In other words, Clopper-Pearson intervals are often too large to be helpful. Brown et al. [2001] suggest other intervals as a better estimation for Binomial proportions. In particular, the Wilson interval (Wilson [1927]) and the Agresti-Coull interval (Agresti and Coull [1998]) are both recommended for their change of the considered parameter $\underline{m}$ by adding or subtracting additional events.

These confidence regions are presented in Subsection 5.3.3. The subsection starts with a general definition of such intervals, called the Bonferroni confidence region. In addition to Wilson and Agresti-Coull some confidence regions without shift correction are introduced. It has to be noted that the presented Bonferroni confidence region with a Student t distribution is particularly unsatisfying when the number of possible states rises. In that case, the probability values become smaller, and the Student t distribution becomes unreliable. Moreover, every entry of $\underline{m}$ that is zero lessens the degree of freedom and thus, increases the considered confidence region. Example 5.3.3 indicates that even for the Neyman distribution the presented region is much larger than any other region.

All confidence regions of Bonferroni type share one disadvantage. If a probability m_i, $i \in \{1, \ldots, K\}$ is zero or one, the associated confidence interval has length zero since $m_i(1-m_i) = 0$ and therefore, any estimation $\widehat{\underline{m}}$ with $\widehat{m}_i > 0$ will not lie in such a Bonferroni confidence region for $\underline{m}$ with $m_i = 0$. Unfortunately, the case $m_i = 0$ occurs with certainty if the number of sequence sites of the input data is smaller than the number of possible states. For five nucleotide sequences this happens when the sequence is shorter than 1024 sites, a very common case when deriving aligned sequences.

A comparison of these regions in terms of average coverage probability is presented in various papers (e.g. May and Johnson [1997], Jhun and Jeong [2000] or Brown et al. [2001]). Their purpose w.r.t. molecular evolution is to decide whether a proposed approximation can be accepted or rejected.

Example. All proposed regions were tested on the transition vectors $\underline{m}$, $\underline{m}_1$ and $\widehat{\underline{m}}_2$ as defined in (5.1.2) to (5.2.5). When comparing the Wald boundaries for $\underline{m}$ from Example 5.3.3 with the "exact" boundaries from Example 5.3.2 one finds that b_l^{Wald} as well as b_u^{Wald} are smaller than their respective parts for the Clopper-Pearson intervals.
All confidence regions agreed on accepting $\underline{m}_1$ and rejection $\widehat{\underline{m}}_2$ as a good approximation of $\underline{m}$. In terms of coverage the following ranking can be made:

$$b_l^T(\underline{m}) < b_l^{\text{Wald}}(\underline{m}) < b_l^{CP}(\underline{m}) < b_l^{CP\beta}(\underline{m}) < b_l^{AC}(\underline{m}) < b_l^W(\underline{m}),$$
$$b_u^W(\underline{m}) < b_u^{AC}(\underline{m}) < b_u^{\text{Wald}}(\underline{m}) < b_u^{CP}(\underline{m}) < b_u^{CP\beta}(\underline{m}) < b_u^T(\underline{m}).$$

Hence, the Student t distribution provides the largest region, whereas the Wilson confidence region provides the best approximation.

Remark 5.3.5. When regarding simultaneous confidence regions for estimated leaf distributions $\underline{p}$ using Maximum Likelihood methods, it appears a good idea to use the so-called *Kullback-Leibler distance* to compare them to a given relative frequency vector $\underline{m}$:

$$d_{\text{kl}}(\underline{p}, \underline{m}) = \text{ls}(\underline{m}, \underline{m}) - \text{ls}(\underline{p}, \underline{m}) = \sum_{i=1}^{K} m_i(\ln(m_i) - \ln(p_i)).$$

Obviously, $d_{\text{kl}}(\underline{m}, \underline{m}) = 0$ and by Lemma 5.1.1 $d_{\text{kl}}(\underline{p}, \underline{m}) > 0$ for all $\underline{p} \neq \underline{m}$. However, the Kullback-Leibler distance is no real distance (see e.g. Cover and Thomas [1991, Section 2.3]), and for the use in a confidence region one has to find an acceptable distribution for this random variable.

5.4 Derivation of a Tree Structure from Triples

This section introduces an algorithm for deriving a tree structure for a given set of input data together with the characterization of a Markov process on the tree. The structure of the algorithm is similar to most reconstruction algorithms presented in the literature, eg. *neighbor joining* (cf. Saitou and Nei [1987]).

5.4.1 The Algorithm

Algorithm 5.2. Let n denote the number of considered leaves, and let $\underline{m}$ denote the joint leaf distribution of these leaves over an finite alphabet $\mathcal{S}$ of cardinality $k > 2$. For a triple tree $\mathcal{T}_i$ denote by $\underline{m}^i$ the restriction of $\underline{m}$ to the three leaves of the triple tree.

1. Compute the transition parameters and, if wanted, apply Algorithm 5.1 to receive admissible parameters.
2. For every leaf $\delta \in \mathcal{L}$ sort the triple trees containing it in descending order to $\max_{x \in \mathcal{S}} p^{\delta}_{xx}$.
3. Cluster a pair of leaves β_i, β_j for which the first $n-2$ triples contain both leaves. The new cluster point will be denoted by $\widehat{\beta}_{ij}$. If no pair is found, cluster the remaining leaves to a star tree and STOP.
4. Set $\mathcal{L} := \mathcal{L} \setminus \{\beta_i, \beta_j\}$ and $\mathcal{L} := \mathcal{L} \cup \{\widehat{\beta}_{ij}\}$. Update the triple transition parameters accordingly. Set $n := n-1$. If $n > 3$ go to 2 else STOP.

Now an analysis of the steps of Algorithm 5.2 is presented.

In this thesis, step one is applicable only for the three models presented in Chapters 3 and 4.

Step two provides the opportunity to take different order functions. However, for the models considered the diagonal elements are as good an ordering criterium as any other choice.

Steps three and four need a more thorough discussion since here the actual work is done.

5.4.2 Selection of a Pair of Leaves

Step three is only presented in an ideal case and would always apply, if input data were subject to the assumed model. However, computations for certain data sets showed that the identification of a cluster pair of leaves with this method is not always possible.

Hence, a modification of the step is due. Firstly, it is possible, that no pair of leaves can be found for which another leaf is part of all $n-2$ considered triple trees. To accommodate this, the selection looks for the pair of leaves that shares the maximal number of triple trees in the first $n-2$. Secondly, it is possible that no maximal pair of leaves can be found. Hence, the number of considered triple trees is extended until a pair is found. This cumulates in the following modified step:

3'. Set $j = -2$.

(i) Select for each leaf β the leaf $\hat{\beta}$ that occurred in most of the first $n+j$ triple trees to β.

(ii) If above step yields a pair (β_1, β_2) with $\beta_1 = \hat{\beta}_2$ and $\beta_2 = \hat{\beta}_1$, cluster them and go to 4 else if $j = \max\{n(n-5)/2, -1\}$ cluster the remaining leaves and STOP else set $j = j+1$ and go to (i).

For a fixed leaf β one sorts $(n-1)(n-2)/2$ different triple trees. If j is such that $n+j = (n-1)(n-2)/2$ one definitely won't find a maximum, since all possible triple trees are considered, and thus each leaf occurs equally often in the set of triple trees. Except for the case $n = 4$ the term $n(n-5)/2$ will be larger than -1. In Subsection 5.4.4 the number j will be added to the output to give a measure of quality. The smaller j is in each step the better the result.

5.4.3 Re-Estimation of Transition Parameters

The fourth step of Algorithm 5.2 updates old transition parameters and generates new ones for the new vertex. This section discusses mechanisms for the derivation of the values. The starting point is the description of the common scenario:
Let $\mathcal{L} := \{\beta_1, \ldots, \beta_n\}$ denote an arbitrary leaf set and $\underline{m}$ a joint leaf distribution on $\mathcal{L}$ over a sample set $\mathcal{S}$ of cardinality k. Step three selected the index pair i, j for clustering and introduced a new leaf β_{ij}. Denote by

$$\mathfrak{T}_{ij} := \{\mathcal{T}_{ijl} : \mathcal{L}(\mathcal{T}_{ijl}) = \{\beta_i, \beta_j, \beta_l\},\ l \neq i, j\}$$

the set of all triple trees that contain both leaves β_i and β_j. All these triple trees contain the edges (β_{ij}, β_i) and (β_{ij}, β_j), and thus the transition matrices $P^{\beta_{ij}\beta_i}$ and $P^{\beta_{ij}\beta_j}$ will be derived from the information content of the $n-2$ forms provided by $\mathfrak{T}_{ij}$ (see Figure 5.1).

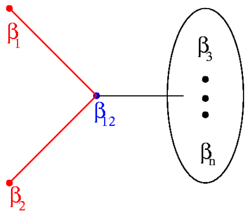

Figure 5.1: Derivation of transition parameters for clustered leaves. The transition matrices for β_1 and β_2 are subject to $n-2$ different triple trees connecting them with the rest of the leaves through β_{12}.

The method used in the example below uses the following function to derive the final parameters. Denote by $TM(k)$ the set of all transition matrices of dimension $k \times k$, and $P^{\beta_{ij}\beta_i,l}$ denotes the transition matrix for edge (β_{ij},β_i) in triple tree $\mathcal{T}_{ijl} \in \mathfrak{T}_{ij}$. The argument of

$$\min_{Q\in TM(k)} F_{ij}(Q,z) := \sum_{l\neq i,j}\sum_{x=1}^{k}(q_{zx}-p_{zx}^{\beta_{ij}\beta_i,l})^2, \quad z\in\mathcal{S}$$

is chosen as the transition matrix for β_i. Other approaches could contain restrictions of the selection to fewer triple trees with the culmination of selecting exactly one triple tree and its parameters. The established transition matrices are needed to derive the transition matrices subject to the new vertex β_{ij}.

Consider the triple tree set for β_{ij}

$$\hat{\mathfrak{T}}_{ij} := \{\mathcal{T}_{ijl_1l_2} : \mathcal{L}(\mathcal{T}_{ijl_1l_2}) = \{\beta_{ij},\beta_{l_1},\beta_{l_2},\, l_1\neq l_2\neq i,j\}.$$

The parameters of each triple tree $\mathcal{T}_{ijl_1l_2} \in \hat{\mathfrak{T}}_{ij}$ are subject to the parameters of the two triple trees $\mathcal{T}_{il_1l_2}$ and $\mathcal{T}_{jl_1l_2}$ (see Figure 5.2).

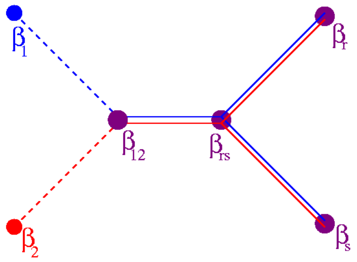

Figure 5.2: Derivation of transition parameters for updated triple tree set. The transition matrices for the new vertex β_{12} and the remaining leaves $\beta_3,\ldots,\beta_n$ need to be updated from the two triples, where two leaves β_r,β_s, $r\neq s\geq 3$ are connected to either β_1 or β_2. The two triples are merged to one, where β_{12} takes the place of the two removed leaves.

For the method used in the example below, the following selection was chosen. The transition matrices for β_{ij} derived from triple trees $\mathcal{T}_{il_1l_2}$ and $\mathcal{T}_{jl_1l_2}$ with inner vertex ϱ_{ij} are obtained from the equation

$$P^{\varrho_{ij}\beta_i} = P^{\varrho_{ij}\beta_{ij},i}P^{\beta_{ij}\beta_i}, \tag{5.4.1}$$

$$P^{\varrho_{ij}\beta_j} = P^{\varrho_{ij}\beta_{ij},j}P^{\beta_{ij}\beta_j}. \tag{5.4.2}$$

The new transition matrix for a leaf β_l, $l\neq i,j$ to a triple tree $\mathcal{T}_{ijl\hat{l}} \in \hat{\mathfrak{T}}_{ij}$, $\hat{l}\neq i,j,l$ is the argument of

$$\min_{Q\in TM(k)} \sum_{x=1}^{k}(q_{zx}-p_{zx}^{\beta_l,i})^2 + \sum_{x=1}^{k}(q_{zx}-p_{zx}^{\beta_l,j})^2, \tag{5.4.3}$$

where the $P^{\beta_{l},i}$ and $P^{\beta_{l},j}$ denote the transition matrix to the triple tree $\mathcal{T}_{il\hat{l}}$ $\mathcal{T}_{jl\hat{l}}$, respectively. Finally, the new matrix for β_{ij} is derived in similar fashion using the transition matrices introduced in (5.4.1) and (5.4.2). Again, different approaches might derive the new matrices from one triple only instead of considering both. This concludes the discussion of step four of Algorithm 5.2.

5.4.4 A Short Example

This section presents an example with five species represented by two state data.

The great ape data set used in Example 5.2.1 contains two state data for five ape species, gibbon(1), human(2), chimp(3), gorilla(4) and orangutan(5). The gibbon was selected as the outgroup. The expected structure of the tree is given in Figure 5.3.

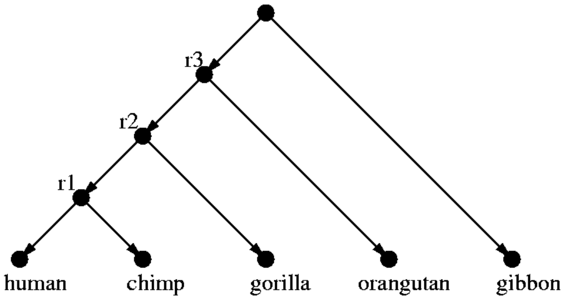

Figure 5.3: The great-apes-tree. The picture shows a root between Orangutan and Gibbon as is the generally acknowledged convention. The algorithm will only return the unrooted structure.

The whole computation for this model was rather fast. But compared to the recently developed algorithms it is quite slow, even though it provides the characterization for a process that returns a leaf distribution that is close to the initial distribution. A qualitative analysis will be provided later.

All sequences have a length of 895, and are fully aligned. The five species yield ten different triple trees. From the input data the following ten triple frequency vectors are observed:

$$\begin{aligned}
\underline{m}_{123} &= (494, 1, 2, 12, 30, 1, 1, 354), & \underline{m}_{124} &= (493, 2, 0, 14, 27, 4, 2, 353),\\
\underline{m}_{125} &= (486, 9, 3, 11, 20, 11, 12, 343), & \underline{m}_{134} &= (493, 3, 0, 13, 27, 4, 2, 353),\\
\underline{m}_{135} &= (487, 9, 2, 11, 20, 11, 12, 343), & \underline{m}_{145} &= (485, 8, 4, 12, 20, 9, 12, 345),\\
\underline{m}_{234} &= (520, 4, 0, 2, 0, 3, 2, 364), & \underline{m}_{235} &= (506, 18, 0, 2, 1, 2, 14, 352),\\
\underline{m}_{245} &= (505, 15, 1, 5, 0, 2, 15, 352), & \underline{m}_{345} &= (505, 15, 2, 5, 0, 2, 14, 352).
\end{aligned}$$

The results from Theorem 3.1.4 are applied to all triple trees generating the parameter matrix:

$$\begin{aligned}
P_{123} &= (0.942752, 0.967223, 0.996028, 0.997348, 0.998044, 0.997519, 0.588959),\\
P_{124} &= (0.948117, 0.961852, 1.00016, 0.989075, 0.996271, 0.994354, 0.583061),\\
P_{125} &= (0.961175, 0.969054, 0.994637, 0.969885, 0.982493, 0.966485, 0.578104),\\
P_{134} &= (0.948117, 0.96448, 1.00015, 0.989229, 0.994235, 0.994355, 0.584261),\\
P_{135} &= (0.961254, 0.969005, 0.996688, 0.969883, 0.982529, 0.966375, 0.578033),\\
P_{145} &= (0.960963, 0.966541, 0.992653, 0.975438, 0.984373, 0.966781, 0.577093,),\\
P_{234} &= (1.00003, 0.994535, 1.00002, 0.991825, 0.992398, 0.994536, 0.585425),\\
P_{235} &= (0.998182, 0.994342, 1.00016, 0.994442, 0.965669, 0.961748, 0.586436),\\
P_{245} &= (1.00017, 0.986059, 0.998442, 0.994343, 0.971207, 0.959128, 0.581781),\\
P_{345} &= (1.00016, 0.986142, 0.996442, 0.994344, 0.971206, 0.961748, 0.582955).
\end{aligned}$$

The values of each vector have the following meaning. The first two entries are the diagonal elements of the transition matrix for the first ape of the associated triple, entries three and four are the analogue elements for species two and entries five, and six identify the transition matrix for the third species. The seventh entry defines the root distribution. For the ordering of the triple trees w.r.t. to each leaf the sum of the diagonal elements is employed.

The algorithm returned the following array:

$$\begin{array}{cccccc}
2 & 0.99808 & 0.995409 & r1 & 0.58694 & 1\\
3 & 0.999408 & 0.994596 & r1 & 0.58694 & 1\\
4 & 0.996348 & 0.994349 & r2 & 0.583015 & 1\\
r1 & 1. & 0.992594 & r2 & 0.583015 & 1\\
1 & 0.961089 & 0.967785 & r3 & 0.577581 & 0\\
5 & 0.983442 & 0.966606 & r3 & 0.577581 & 0\\
r2 & 0.995864 & 0.98147 & r3 & 0.577581 & 0.
\end{array} \tag{5.4.4}$$

The values of each row should be interpreted in the following way. The first entry is the vertex, the next two entries are the diagonal elements for the transition matrix for the connecting edge to the new vertex, denoted by entry four with marginal distribution given by entry five. Entry six denotes the number of tries to obtain the pair of vertices to the vertex given in entry four. In the first four rows this value is equal to one, i.e. the pair was selected in the first step. In the last three rows this value is equal to zero because a triple tree is already unique.

The array (5.4.4) shows that human(2) and chimp(3) are closest joined by inner vertex $r1$. Next, the gorilla(4) has a common ancestor with $r1$ in $r2$. Finally, orangutan(5) and gibbon(1) are joined with $r2$ at $r3$. This structure is equivalent to the structure presented in Figure 5.3 with the exception that the overall root is placed in $r3$ instead on the edge between gibbon and $r3$. The obtained tree is visualized in Figure 5.4.

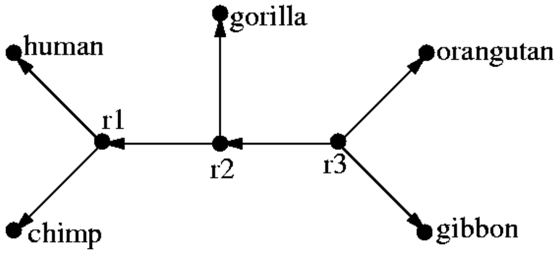

Figure 5.4: The tree obtained by Algorithm 5.2 from the Great-Ape sequences. The root is in $r3$ instead of the edge between $r3$ and the gibbon, cf. Figure 5.3.

To finish the example the presented tools of comparison are employed to observe the performance of the associated Markov process w.r.t. to the initial frequency distribution $\underline{\widehat{m}}$. The Markov process is represented by its leaf distribution $\underline{m}$.

The loglikelihood score for $\underline{\widehat{m}}$ is -952.919, and the loglikelihood score for $\underline{m}$ is -962.499. Accordingly, the Kullback-Leibler distance of the observed frequency distribution is 9.58041.

Moreover, the computed leaf distribution $\underline{m}$ lies well within the Agresti-Coull 95%-region and the Clopper-Pearson 95%-region. Overall, the result obtained by Algorithm 5.2 is acceptable, though one may assume that the computational error for a larger number of sequences will eventually return unacceptable Markov-processes.

5.5 Proofs

As usual, the chapter closes with the proofs of the results presented throughout the section.

5.5.1 Proofs of Section 5.1

Section 5.1 regarded the loglikelihood scoring function.

Proof of Lemma 5.1.1. The aim is to compute the maximum of:

$$f(x_1,\dots,x_{K-1}) = \sum_{i=1}^{K-1} y_i \ln x_i + \big(1 - \sum_{i=1}^{K-1} y_i\big) \ln \big(1 - \sum_{i=1}^{K-1} x_i\big)$$

in x_i, $i = 1, \dots, K-1$. The first derivative in x_i yields:

$$\frac{\partial}{\partial x_i} f(x_1, \dots, x_{K-1}) = \frac{y_i}{x_i} - \frac{1 - \sum_{j=1}^{K-1} y_j}{1 - \sum_{j=1}^{K-1} x_j}.$$

The root of this partial derivative in x_i is given by:

$$x_i = y_i \frac{1 - \sum_{j \neq i}^{K-1} x_j}{1 - \sum_{j \neq i}^{K-1} y_j}.$$

Since this equality must be given for all $i = 1, \dots, K-1$ it follows that $\underline{x} = \underline{y}$ is an extreme point of $f(x_1, \dots, x_{K-1})$. Looking at the second partial derivative yields the inequality

$$\frac{\partial^2}{\partial x_i^2} f(y_1, \dots, y_{K-1}) = -\frac{1}{y_i} - \frac{1}{1 - \sum_{j=1}^{K-1} y_j} < 0$$

if $y_i \geq 0$, $i = 1, \dots, K-1$ and $1 - \sum_{i=1}^{K-1} y_i \geq 0$, i.e. if $\underline{y}$ describes a joint distribution. Hence, $\widehat{\underline{m}}$ is the maximum of the loglikelihood scoring function ls. This completes the proof. □

Proof of Lemma 5.1.2. The statement follows immediately with (5.1.6), since the state set in which $\underline{m}(P, \mathcal{T})$ and $\underline{m}(P, \hat{\mathcal{T}})$ differ has a factor of one in the Likelihood score (5.1.1). □

5.5.2 Proofs for Section 5.2

This section presented an estimator. It needs to be shown, that the associated index set is unique and that the estimator is consistent.

Proof of Proposition 5.2.1. The existence of a minimum is observed if the minimized function is continuous over a compact space. Clearly, the function in (5.2.1) is continuous and the constraints describe the compact set $\{x \in \mathbb{R}_+^k : x_1 + \cdots + x_k = 1\}$. Therefore, the minimization problem has a solution. Uniqueness is attained if the considered function is strictly convex. This is true for quadratic functions and hence, the uniqueness is also observed. That $p \mapsto q(p)$ is a continuous mapping follows from Propositions 3.4 and 5.5 Deutsch [2001]. This completes the proof. □

Proof of Proposition 5.2.2. Let $F(q) := \sum_{i=1}^m (q_i - p_i)^2$ with $\sum_{i=1}^m q_i = 1$ and $q_i \geq 0$ for all $i \in \{1, \dots, m\}$. According to Kuhn-Tucker the minimum of F satisfies the following conditions:

$$\left(\frac{\partial}{\partial q_i} - \frac{\partial}{\partial q_j}\right) F(q) = 0, \quad q_i, q_j > 0,$$
$$\left(\frac{\partial}{\partial q_i} - \frac{\partial}{\partial q_j}\right) F(q) \geq 0, \quad q_i = 0, q_j > 0.$$

Conducting the necessary derivations yields the inequalities

$$(5.5.1) \qquad q_i - p_i = q_j - p_j \text{ for } q_i, q_j > 0, \quad \text{and} \quad q_j \leq p_j - p_i \text{ for } q_i = 0, q_j > 0.$$

$F(q)$ is minimal if the amount of shuffled mass is distributed equally on the $i \in \{1, \dots, m\}$ with $q_i > 0$, thus $q_i = p_i + c$ for $i \in I := \{i \in \{1, \dots, m\} : q_i > 0\}$. Now, the mass distribution is arranged by setting

$$c := \frac{1}{\sharp(I)} \sum_{i \in I^c} p_i,$$

since for all $i \in I^c$ the parameters are brought to zero and thus, their mass p_i needs to be redistributed among the $i \in I$. Since q is obtained by checking the minima conditions it is a minimum of (5.2.1). With Proposition 5.2.1 it is unique. This completes the proof. □

Proof of Corollary 5.2.3. (5.2.3) and (5.5.1) yield $p_i \leq -c$ for $i \in I^c$ and $p_i > -c$ for $i \in I$, i.e. (5.2.4). Consider the order relations: If $j \in I$ then $p_j > -c$. Since $p_i \geq p_j$ also $p_i > -c$, and thus $i \in I$. This verifies relation 1. On the other hand, if $j \in I^c$ then $p_j \leq -c$ and with $p_i \leq p_j$ also $i \in I^c$. Thus, relation 2 is observed. Assume, for $p_i \leq 0$ for some $i \in I$. Then $p_i > -c$ must hold. The order relation states, that all indices $j \in I^c$ must satisfy $p_j < p_i$. Then, $c < 0$ according to (5.2.3) and therefore $0 > p_i > -c > 0$, i.e. a contradiction. Hence relation 3 holds. This completes the proof. □

Proof of Proposition 5.2.4. The construction of $\underline{q} := (q_i)_{i=1}^k$ is done using Corollary 5.2.3. Hence, the index set obtained by Algorithm 5.1 is the index set required in (5.2.3). Therefore, with Proposition 5.2.2 the retrieved vector $\underline{q}$ is the unique minimum of (5.2.1) under the constraints (5.2.2), and the proof is complete. □

Proof of Theorem 5.2.5. Consistency follows from Proposition 5.2.1 since $p \mapsto q(p)$ is continuous. □

Proof of Corollary 5.2.6. The minimization problem (5.2.1) reduces for transition parameters $(\hat{p}_{xy}^{\delta})_{x,y \in \{0,1\}}$, $\delta \in \mathcal{V} \setminus \{\varrho\}$ to:

$$\min_{p_{x0}^{\delta}, p_{x1}^{\delta}} (p_{x0}^{\delta} - \hat{p}_{x0}^{\delta})^2 + (p_{x1}^{\delta} - \hat{p}_{x1}^{\delta})^2, \quad x \in \{0, 1\}$$

which is due to the constraints equivalent to:

$$\min_{p_{xy}^{\delta} \in [0,1]} (p_{xy}^{\delta} - \hat{p}_{xy}^{\delta})^2, \quad x, y \in \{0, 1\}.$$

Clearly, if $\hat{p}_{xy}^{\delta} > 1$ then the best approximation is $p_{xy}^{\delta} = 1$ and if $\hat{p}_{xy}^{\delta} < 0$ then $p_{xy}^{\delta} = 0$ returns the smallest squared difference. If $\hat{p}_{xy}^{\delta} \in [0, 1]$ the $p_{xy}^{\delta} = \hat{p}_{xy}^{\delta}$. This completes the proof. □

Proof of Corollary 5.2.7. For the Neyman N_k model admissible parameters are in the interval $[0, 1/(k-1)]$. Hence, if the returned parameters $\hat{p}_\delta$, $\delta \in \mathcal{V} \setminus \{\varrho\}$ are in this interval, they are admissible. However, in case of violation the parameters are subjected to

$$\min_{p_\delta} (p_\delta - \hat{p}_\delta)^2, \quad \delta \in \mathcal{V} \setminus \{\varrho\}$$

which, as above, results in a projection into the bounds and the statement is thus proven. □

Proof of Corollary 5.2.8. The presented scenarios are derived from Proposition 5.2.2 by looking at the possible cases of c in (5.2.3). Verifying one scenario is sufficient to verify all scenarios. The Kimura 2ST model has two parameters, $\hat{p}_\delta$ and $\hat{q}_\delta$ with a third parameter $\hat{r}_\delta$ to satisfy $\hat{r}_\delta + \hat{p}_\delta + 2\hat{q}_\delta = 1$. Assume $\hat{r}_\delta < 0$. Then, according to Corollary 5.2.3.3 its estimate is $r_\delta = 0$ and the probability mass of 1 needs to be distributed among the remaining parameters. If $3\hat{p}_\delta > -\hat{r}_\delta$ and $3\hat{q}_\delta > -\hat{r}_\delta$ the indices for both parameters are contained in the index set I. In this case, applying Proposition 5.2.2 yields the estimates $p_\delta = \hat{p}_\delta + \hat{r}_\delta/3$ and $q_\delta = \hat{q}_\delta + \hat{r}_\delta/3$. The sum $r_\delta + p_\delta + 2q_\delta$ again is one. Now assume $3\hat{p}_\delta < -\hat{r}_\delta$. Then its index is transferred to I^c, i.e. its estimate is $p_\delta = 0$ and $c = (\hat{p}_\delta + \hat{r}_\delta)/2$. For the remaining parameter $\hat{q}_\delta$ one computes the estimate

$$q_\delta = \hat{q}_\delta + (\hat{p}_\delta + \hat{r}_\delta)/2 = (\hat{r}_\delta + \hat{p}_\delta + 2\hat{q}_\delta)/2 = 1/2.$$

The remaining cases are treated similarly, and the corollary is thus proven. □

Proof of Lemma 5.2.9. Let $\underline{X}$ denote a random variable for multinomial parameters with $\mathbb{E}\underline{X} = N\underline{p}$ and $\mathrm{Cov}(\underline{X}) = N\big(p_i(\delta_{ij} - p_j)\big)_{i,j=1}^K$. Further, let $\tilde{\underline{p}} = \alpha\underline{X} + \beta$ with $\alpha, \beta > 0$. Setting $\alpha = 1/(\sqrt{N} + N)$ and $\beta = \sqrt{N}/(K(\sqrt{N} + N))$ yields the vector $\hat{\underline{p}}$ defined in (5.2.6). The aim is to find α and β such that $\mathbb{E}\|\tilde{\underline{p}} - \underline{p}\|^2 = c$ =const. for all $\underline{p}$.

$$\begin{aligned}
\mathbb{E}\|\tilde{\underline{p}} - \underline{p}\|^2 &= \sum_{i=1}^K \mathbb{E}(\alpha X_i + (\beta - p_i))^2 \\
&= \sum_{i=1}^K \left[\alpha^2 \mathbb{E}X_i^2 + 2\alpha(\beta - p_i)\mathbb{E}X_i + (\beta - p_i)^2\right] \\
&= \sum_{i=1}^K \left[\alpha^2 N p_i(1 - (1-N)p_i) + 2\alpha(\beta - p_i)Np_i + (\beta - p_i)^2\right] \\
&= \sum_{i=1}^K p_i^2\big(1 - 2\alpha N - \alpha^2 N(1-N)\big) + \alpha^2 N + 2\alpha\beta N + K\beta^2 - 2\beta.
\end{aligned}$$

This term is independent of p_i, $i = 1, \dots, K$ if $1 - 2\alpha N + \alpha^2 N(N-1) = 0$. This demand yields:

$$\begin{aligned}
\alpha^{\pm} &= \frac{1}{N-1} \pm \sqrt{\frac{1}{(N-1)^2} - \frac{1}{N(N-1)}} = \frac{1}{N-1} \pm \frac{1}{\sqrt{N}(N-1)} \\
&= \frac{\sqrt{N} \pm 1}{\sqrt{N}(\sqrt{N}-1)(\sqrt{N}+1)} = \frac{1}{\sqrt{N}(\sqrt{N} \pm 1)}.
\end{aligned} \tag{5.5.2}$$

The choice of β does not have an effect on $\mathbb{E}\|\hat{\underline{p}} - \underline{p}\|^2 = c$. Since $\hat{\underline{p}}$ uses α^+ the Lemma is verified. □

Proof of Lemma 5.2.10. Assume that $\underline{p}$ is Dirichlet distributed with parameters $\alpha_1, \dots, \alpha_K$ and $\alpha_0 = \alpha_1 + \dots + \alpha_K$. Then, one has:

$$\begin{aligned}
\mathbb{E}_{\underline{p}}\|\tilde{\underline{p}} - \underline{p}\|^2 &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^K \Gamma(\alpha_i)} \int_0^1 \mathrm{d}\, p_1\, p_1^{\alpha_1 - 1} \dots \int_0^{1-\sum_{i=1}^{K-2} p_i} \mathrm{d}\, p_{K-1}\, p_{K-1}^{\alpha_{K-1}-1} p_K^{\alpha_K - 1} \|\tilde{\underline{p}} - \underline{p}\|^2 \\
&= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^K \Gamma(\alpha_i)} \sum_{i=1}^K \int_{\sum_{j=1}^K p_j = 1} \mathrm{d}\, \underline{p}\, p_1^{\alpha_1 - 1} \dots p_K^{\alpha_K - 1} (\tilde{p}_i - p_i)^2.
\end{aligned}$$

This integral is minimal if $\tilde{p}_i = \mathbb{E}_{\underline{p}}(p_i)$ since:

$$\mathbb{E}(X - \mathbb{E}(X))^2 = \min_y \mathbb{E}(X - y)^2.$$

Hence, compute the mean:

$$\begin{aligned}
\mathbb{E}_{\underline{p}}(p_i) &= \frac{\Gamma(\alpha_0)}{\prod_{j=1}^K \Gamma(\alpha_j)} \int_{\sum_{j=1}^K p_j = 1} \mathrm{d}\, \underline{p} \prod_{j \neq i}^K p_j^{\alpha_j - 1} p_i^{\alpha_i} \\
&= \frac{\Gamma(\alpha_0)}{\prod_{j=1}^K \Gamma(\alpha_j)} \frac{\Gamma(\alpha_i + 1) \prod_{j \neq i}^K \Gamma(\alpha_j)}{\Gamma(\alpha_0 + 1)} = \frac{\alpha_i}{\alpha_0}.
\end{aligned}$$

Set $\alpha_i = \alpha + n_i$ with $n_i = N p_i$. Then, $\alpha_0 = K\alpha + N$ and

$$\mathbb{E}_{\underline{p}}(p_i) = \frac{\alpha + n_i}{K\alpha + N}.$$

Thus, equality with $\hat{p}_i$ from (5.2.6) is given by $\alpha = \sqrt{N}/K$. In that case, Ferguson [1967, Thm. 2.11.3] states, that $\hat{\underline{p}}$ is a Bayesian estimator. This completes the proof. □

5.5.3 Proofs of Section 5.3

Section 5.3 presented several types of confidence regions. The following proofs are associated with the presented results:

For the proofs regarding Subsection 5.3.1 further results are needed:

Lemma 5.5.1.(Hilfssatz 1.90b in Witting [1985]) *Let X denote an N-dimensional random variable with probability distribution $\mathcal{N}(\mu, \mathcal{C})$. Then, the following statements hold:*

1. *The covariance matrix is positive semi-definite and symmetrical and has the following representation:*

$$\mathcal{C} = \mathcal{C}^{\frac{1}{2}}\mathcal{C}^{\frac{1}{2}} \quad \textit{with } rank(\mathcal{C}) = rank(\mathcal{C}^{\frac{1}{2}}).$$

2. *If $rank(\mathcal{C}) = l < N$ then a $N \times l$-matrix A exists with $rank(A) = l$ and $\mathcal{C} = AA^T$.*

3. *There is a orthonormal system $U = (u_1, \dots, u_N)$ and a positive semi-definite and symmetrical matrix $\hat{\mathcal{C}} \in \mathbb{R}^{l \times l}$ with*

$$U^T \mathcal{C} U = \begin{pmatrix} \hat{\mathcal{C}} & 0 \\ 0 & 0 \end{pmatrix}.$$

Consequently, matrix A from 2 can be depicted by

$$A = U \begin{pmatrix} \hat{\mathcal{C}}^{\frac{1}{2}} \\ 0 \end{pmatrix}.$$

4. *A pseudo-inverse matrix $\mathcal{C}^{-1}$ to $\mathcal{C}$ exists such that*

$$\mathcal{C}^{-1}\mathcal{C} = \mathrm{Diag}(\underbrace{1, \dots, 1}_{\mathrm{k-times}}, 0, \dots, 0)$$

The pseudo-inverse is given by

$$U^T \mathcal{C}^{-1} U = \begin{pmatrix} \hat{\mathcal{C}}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

5. *There is a $\mathcal{N}(0, \mathbb{1}_l)$-distributed random variable W with $X = AW + \mu$. Using the descriptions from (4) yields $W = A^{-1}(X - \mu)$.*

Proof of Lemma 5.5.1. The results are common knowledge. For a full proof refer to Witting [1985]. □

With these tools the statements of Subsection 5.3.1 can be proven.

Proof of Lemma 5.3.1. Follows immediately from Lemma 5.5.1. □

Proof of Corollary 5.3.2. Also an immediate consequence from Lemma 5.5.1 if accompanied by some basic matrix computation rules. □

Proof of Theorem 5.3.3. The goal is to identify a value $\varepsilon > 0$ for the following conditions:

$$\mathbb{P}(\|\bar{Y} - \underline{m}\|^2_{\mathcal{C}} < \varepsilon) > 1 - \eta, \quad \text{where } \|x\|^2_A = \langle x, A^{-1}x \rangle, \tag{5.5.3}$$

where A^{-1} is the pseudo inverse of A and η denotes the chosen confidence level. With the properties from Lemma 5.5.1 the following computations can be applied to (5.3.2):

$$\mathbb{P}(\|\bar{Y} - \underline{m}\|^2_{\mathcal{C}} < \varepsilon) = \mathbb{P}(\|\sqrt{N}(\bar{Y} - \underline{m})\|^2_{\mathcal{C}} < N\varepsilon).$$

Consider the norm using (5.5.3) yields:

$$\begin{aligned} &\|\sqrt{N}(\bar{Y} - \underline{m})\|^2_{\mathcal{C}} = \sqrt{N}(\bar{Y} - \underline{m})^T \mathcal{C}^{-1}[\sqrt{N}(\bar{Y} - \underline{m})] \\ &= \sqrt{N}(\bar{Y} - \underline{m})(A^{-1})^T A^{-1}[\sqrt{N}(\bar{Y} - \underline{m}] = \sqrt{N}[A^{-1}(\bar{Y} - \underline{m})]^T \sqrt{N}[A^{-1}(\bar{Y} - \underline{m})] \\ &=: W^T W = \|W\|^2. \end{aligned}$$

Thus, coming back to (5.5.3) one has

$$\mathbb{P}(\|\bar{Y} - \underline{m}\|^2_{\mathcal{C}} < \varepsilon) = \mathbb{P}(\|W\|^2_{\mathbb{1}_l} < N\varepsilon).$$

Since W is $\mathcal{N}(0, \mathbb{1}_l)$-distributed, its squared norm is χ^2_l-distributed (cf. Def. 1.43a in Witting [1985]) and thus:

$$\mathbb{P}(\|\bar{Y} - \underline{m}\|^2_{\mathcal{C}} < \varepsilon) = \chi^2_l((0, N\varepsilon)) > 1 - \eta,$$

i.e,

$$\varepsilon = \frac{1}{N} Q^{\chi^2_l}(1 - \eta), \tag{5.5.4}$$

where $Q^{\chi^2_l}(1 - \eta)$ denotes the η-quantile of the χ^2_l-distribution with l degrees of freedom. The structure of the joint leaf distribution $\underline{m}$ implies $l = k^n - 1$ if $\underline{m}$ is strictly positive. This completes the proof. □

Bibliography

Alan Agresti and Brent A. Coull. Approximate is better than "exact" for interval estimation of Binomial proportions. *The American Statistician*, 52(2):119–126, May 1998.

Elizabeth S. Allman and John A. Rhodes. Phylogenetic invariants for the general Markov model of sequence mutation. *Mathematical Biosciences*, 186(2):113–144, December 2003.

Ellen Baake. What can and what cannot be inferred from pairwise sequence comparisons? *Mathematical Biosciences*, 154(1):1–21, 1998. ISSN 0025-5564.

Olaf R.P. Bininda-Emonds, John L. Gittleman, and Mike A. Steel. The (super)tree of life: Procedures, problems, and prospects. *Annu. Rev. Ecol. Syst.*, 33:265–289, 2002.

D.J. Brooks, Jacques R. Fresco, and Mona Singh. A novel method for estimating ancestral amino acid composition and its application to proteins of the last universal ancestor. *Bioinformatics*, 20(14):2251–2257, April 2004.

Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a Binomial proportion. *Statistical Science*, 16(2):101–133, May 2001. With comments from Alan Agresti and Brent A. Coull (pp. 117-120), George Casella (pp. 120-122), Chris Corcoran and Cyrus Mehta (pp. 122-124), Malay Ghosh (pp. 124-125) and Thomas J. Santner (pp. 126-128) and a rejoinder by the authors (pp. 128-133).

David Bryant and Vincent Moulton. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. In Roderic Guigó and Dan Gusfield, editors, *WABI*, volume 2452 of *Lecture Notes in Computer Science*, pages 375–391. Springer, 2002.

Joseph T. Chang. Full reconstruction of Markov models on Evolutionary Trees: Identifiability and consistency. *Mathematical Biosciences*, 137:51–73, 1996.

C. J. Clopper and E. S. Pearson. The use of confidence for fiducial limits illustrated in the case of the Binomial. *Biometrika*, 26(4):404–413, 1934.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991. ISBN 0-471-06259-6.

David Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms.* Undergraduate Texts in Mathematics. Springer Verlag, second edition, 1997. ISBN 0-387-97847.

A. Philip Dawid. Conditional independence for statistical operations. *The Annals of Statistics*, 8(3):598–617, May 1980.

Frank Deutsch. *Best Approximation in Inner Product Spaces.* Number 7 in CMS Books in Mathematics. Canadian Mathematical Society, 2001. ISBN 0-387-95156-3.

Warren J. Ewens and Gregory R. Grant. *Statistical Methods in Bioinformatics: An Introduction.* Statistics for Biology and Health. Springer Verlag Berlin, New York, Heidelberg, 2001. ISBN 0-387-95229-2.

Joe Felsenstein. Evolutionary Trees from DNA sequences: A Maximum Likelihood approach. *Journal of Molecular Evolution*, 17(6):368–76, 1981.

Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach.* Probability and Mathematical Statistics. Academic Press, 1967. ISBN 0-12-253750-5.

Leo A. Goodman. Simultaneous confidence intervals for contrasts among Multinomial populations. *The Annals of Mathematical Statistics*, 35(2):716–725, June 1964.

Gert-Martin Greuel, Gerhard Pfister, and Hans Schönemann. Singular 2.0. A Computer Algebra System for Polynomial Computations, Centre for Computer Algebra, University of Kaiserslautern, 2001. `http://www.singular.uni-kl.de`.

John Gribbin and Jeremy Cherfas. *The First Chimpanzee.* Penguin Books Ltd., 2001. ISBN 0140294813. Public Literature in Science.

Stéphane Guindon and Olivier Gascuel. A simple, fast and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Systems Biology*, 52(5):696–704, October 2003.

Thomas R. Hagedorn and Laura F. Landweber. Phylogenetic invariants and geometry. *Journal of theorethical Biology*, 205:365–376, 2000.

J. M. Hammersley and P. E. Clifford. Markov fields on finite graphs and lattices. Technical Report, 1971.

Morris W. Hirsch. *Differential Topology.* Graduate Texts in Mathematics. Springer Verlag, New York, Heidelberg, Berlin, 1976. ISBN 3-540-90148-5.

Birk Huber and Bernd Sturmfels. Bernstein's theorem in affine space. *Discrete and Computational Geometry*, 17:137–141, 1997.

John P. Huelsenbeck and Jonathan P. Bollback. Empirical and hierarchical Bayesian estimation of ancestral states. *Systematic Biology*, 50(3):351–366, 2001.

Myoungshic Jhun and Hyeong-Chul Jeong. Applications of bootstrap methods for categorical data analysis. *Computational Statistics and Data Analysis*, 35:83–91, 2000.

James A. Lake. Phylogenetic inference: How much evolutionary history is knowable? *Molecular Biology and Evolution*, 14(3):213–219, 1997.

Steffen L. Lauritzen. *Graphical Models.* Oxford Stastical Science Series. Clarendon Press, Oxford, 1996. ISBN 0-19-852219-3.

Steffen L. Lauritzen. Causal inference from graphical models. In *Complex Stochastic Systems*, pages 63–107. Chapman & Hall, London, Boca Raton, 2001.

Paul F. Lazarfeld. Latent structure analysis. In Stouffer, Guttman, Slachman, Lazarfeld, Star, and Claussen, editors, *Measurement and Prediction*, chapter 10, pages 362–412. Wiley, New York, 1966.

Frantisek Matús. On equivalence of Markov properties over undirected graphs. *Journal of Applied Probability*, 29:745–749, 1992.

Warren L. May and William D. Johnson. Properties of simultaneous confidence intervals for Multinomial proportions. *Communications in Statistics*, 26(2):495–518, 1997.

Julien Meunier and Laurent Duret. Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.*, 21(6):984–990, February 2004.

Lior Pachter and Bernd Sturmfels. Tropical geometry of statistical models. Not yet published, May 2004.

Judea Pearl and Michael Tarsi. Structuring causal trees. *Journal of Complexity*, 2: 60–77, 1986.

Naruya Saitou and Masatoshi Nei. The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4(4):406–425, 1987.

Heiko A. Schmidt, Korbinian Strimmer, Martin Vingron, and Arndt von Haeseler. TREE-PUZZLE: Maximum Likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, 18:502–504, 2002. `http://www.tree-puzzle.de/`.

Charles Semple and Mike Steel. *Phylogenetics.* Oxford Lectures Series in Mathematics and its Applications, J. Ball and D. Welsh (eds.). Oxford University Press, 2003. ISBN 0-19-850942-1.

Igor F. Shafarevich. *Basic Algebraic Geometry.* Number 213 in Die Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg New York, 1974. ISBN 3-540-06691-8.

Mike A. Steel, Mike D. Hendy, and David Penny. Reconstructing phylogenies from nucleotide pattern frequencies - a survey and some new results. *Discrete Applied Mathematics*, 88:367–396, 1998.

Korbinian Strimmer and Arndt von Haeseler. Quartet puzzling: A quartet Maximum Likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution*, 13(7):964–969, 1996.

Bryan Sykes. *The Seven Daughters of Eve: The Science That Reveals Our Genetic Ancestry.* W.W. Norton & Company, July 2001. ISBN 0393020185.

Michael S. Waterman. *Introduction to computational biology. Maps, sequences and genomes.* Interdisciplinary Statistics. Chapman & Hall, London, 1995. ISBN 0-412-99391-0.

Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, June 1927.

Hermann Witting. *Mathematische Statistik I.* B.G.Teubner Stuttgart, 1985.

Hermann Witting and Ulrich Müller-Funk. *Mathematische Statistik II.* B.G.Teubner Stuttgart, 1995.

Stephen Wolfram. *Mathematica 5.0.* Wolfram Research, Inc., Champaign, Illinois, 2003. `http://support.wolfram.com/mathematica/`.

Ziheng Yang. Statistical properties of the Maximum Likelihood method of phylogenetic estimation and comparison with distance matrix methods. *Syst. Biol.*, 43 (3):329–342, 1994.

Von Bing Yap and Terry Speed. Rooting a phylogenetic tree with nonreversible substitution models. *BMC Evolutionary Biology*, 5(2), January 2005. `http://www.biomedcentral.com/1471-2148/5/2`.