Probabilistic Analysis of Evolutionary Models with Applications to Phylogenetic Inference

by

Max Bacharach

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)

at the
University of Wisconsin-Madison
2023

Date of Final Oral Exam: 11/22/2023
The dissertation is approved by the following members of the Final Oral Committee:

Sébastien Roch, Professor, Mathematics
Ané, Cécile, Professor, Botany and Statistics
Jose Israel Rodriguez, Assistant Professor, Mathematics
Hanbaek Lyu, Assistant Professor, Mathematics

Statistical Problems in Phylogenetic Inference

Max Hill

This thesis considers a number of statistical problems in mathematical phylogenetics relating to the estimation of evolutionary trees from DNA sequence data. We consider the robustness of a variety of inference procedures under certain biological assumptions, including assumptions about intralocus recombination, gene duplication and loss, and mutation rate variability between genes. For these questions, robustness is understood in terms of identifiability, statistical consistency, and sample complexity (i.e., the number of samples required to have high probability of correct inference). In addition, we consider in detail maximum likelihood estimation of species trees from DNA sequence data on trees with three or four leaves with the aim of developing tools which, in the future, may be useful for better understanding of some of the ways that maximum likelihood can fail.

Dedication

This thesis is dedicated to Ananda Weerasinghe,
for showing me that probability is a rich and beautiful subject,
and in memory of my grandmother, Hilde Guth Bacharach,
an exemplar of strength, courage, and love.

Acknowledgements

Over the time that I have spent studying mathematics, I have developed a deep well of gratitude to the countless people who have shared their knowledge and time with me, and who have supported me over this time.

I want to thank first my advisor, Sebastien Roch, for his guidance, encouragement, patience, and support. I am profoundly grateful to have had the opportunity to work with and learn from him. In addition to the mathematical perspectives he has shared with me, and all the detailed feedback on my work, he has been both kind and patient; the year of the Covid-19 pandemic was, to put it lightly, a difficult time for me, and I would not have made it through without his understanding and support.

I would also like to thank my committee members, Cécile Ané, Jose Israel Rodriguez, and Hanbaek Lyu. I am very grateful for the many kind words, support, and insights shared with me by Professor Ané; over the last five years at Madison, she has been a role model for the kind of researcher I aspire to be. My deepest thanks also to Professor Rodriguez, who for the last year and a half has been an incredibly supportive mentor and advocate for me, and has provided many opportunities to learn and engage with the beautiful subject of algebraic statistics. The final chapter of this thesis would not have taken form without his mentorship, encouragement, and collaboration.

More than anything, it has meant a lot to me to have had people who believed in me even when, at times, I doubted myself. In this respect, I am extremely grateful to my advisor and committee members, and to Ananda Weerasinghe, Tim Sorenson, Claudia Solís-Lemus, Dave Anderson, Mike Olinick, John D. Emerson, Emily Proctor, Bill Peter-

son, Pete Schumer, Pablo Raúl Stinga, Jonas Hartwig, Justin Peters, Louis Fan, and many others. I would like to add special thanks to Mihaela Ifrim, whose support, both vocal and unwavering, has meant more to me than I have words to describe.

I would also not have managed without the love and support of my parents, as well as my friends and fellow grad students: Ben, Beth, Bowen, Brandon, Dakota, Hyounji, Jacob, Mary, Leo, Shuqi, Stefanie, Yu, Xiao, and so many others, thank you. And to all my math professors at Middlebury College, thank you for introducing me to this beautiful subject.

Parts of this thesis, namely Chapters 2 and 3 and 3, appear in:

- Hill, M., & Roch, S. (2022). Inconsistency of Triplet-Based and Quartet-Based Species Tree Estimation under Intralocus Recombination. *Journal of Computational Biology*, 29(11), 1173-1197.
- Hill, M., & Roch, S. (2022, April). On the Effect of Intralocus Recombination on Triplet-Based Species Tree Estimation. In International Conference on Research in Computational Molecular Biology (pp. 143-158). Cham: Springer International Publishing.
- Hill, M., Legried, B., & Roch, S. (2022). Species tree estimation under joint modeling of coalescence and duplication: sample complexity of quartet methods. *The Annals of Applied Probability*, 32(6), 4681-4705.

In addition, the Chapter 5 grew out of an undergraduate research headed created by my advisor, for which I was the graduate mentor. This project, "Shedding light on unexplained behavior of phylogeny estimation methods" took place under the organization of the Madison Experimental Mathematics (MXM) Lab in the Fall and Spring semesters of 2022, and I would like to acknowledge the hard work and contributions of the students involved: Achutha Balaji, Yuheng Cai, Binhao Chen, Laura Huang, Nitit Jongsawatsataporn, Megan Lolling, Mengwei Sun, and Jingde Wan.

Finally, I am thankful for the financial support I have received from my advisor and Professor Rodriguez, through NSF grants DMS-1902892 and DMS-2023239 (TRIPODS Phase II), a Vilas Associates Award (to SR). Part of this research was performed while the author was visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant No. DMS-1929348).

Contents

Chapter 1

Introduction

1.1 Overview

At the heart of mathematical phylogenetics is the goal of understanding the history of life on Earth. The great diversity of lifeforms and of biological processes at play in evolution give rise to many methodological and theoretical questions—all with the ultimate goal of better understanding and resolving the tree of life. These overarching questions include:

(1) How can genomic data, typically sampled from extant species, best be used to reconstruct relationships between their (usually long-extinct) ancestors?

(2) What are the theoretical limits of what can, in principle, be discovered about the evolutionary past, provided the sorts of data available to us? And for those biological events (speciations, hybridizations, etc) which are not in some sense 'lost to time,' how much data must be gathered to achieve adequate levels of confidence in our estimates of what happened, and when?

(3) How accurate is a given inference method under certain assumptions about the evolutionary process? And how accurate might it be when those assumptions not met?

As the term *tree of life* implies, evolutionary relationships between species are often thought to be represented by a tree graph (e.g. Fig. 1.2 or the black tree in Fig. 1.5). In this scheme, the edges represent some level of taxonomy—usually species or populations—with the vertices representing most recent common ancestors and the leaves of the tree

representing extant taxa from which genetic data can be sampled. Tree parameters such as the edge lengths and the tree topology represent hypotheses about what happened in the evolutionary past to produce the organisms observed today.

Complicating this picture, there are many examples in phylogenetics of seemingly reasonable inference methods which turn out to perform poorly when previously-unrecognized biological phenomena are taken into account. The major examples include the phenomenon of long branch attraction, a form of estimation bias affecting parsimony and maximum-likelihood methods [1, 2, 3], inconsistency of concatenation-based methods [4, 5], the 'anomaly zone' in which individual genes are most likely to exhibit patterns of ancestry which differ from that of the species [6, 7, 8], and inconsistency arising from violations of tree-likeness, due to sources like hybridization, introgression, and horizontal gene flow [9, 10, 11, 12].

There are a number of mathematical questions considered in this thesis which are relevant to the evaluation of phylogenetic inference methods. These include the following:

- *Statistical identifiability:* can the model parameters be recovered from the theoretical distribution of the data?
- *Statistical consistency:* does the method converge in probability to the true parameter value as the number of samples tends to infinity?
- *Sample complexity:* how much data (e.g. genes) must be sampled to have a high level of probability of correct inference?

This thesis considers a number of problems in mathematical phylogenetics, relating both to the ways in which various biological phenomena can complicate inference of the evolutionary past, as well as the the robustness of inference methods to these and other complications. Of special interest is the question of determining which biological phenomena need to be accounted for in models of evolution, and which—having little effect on estimation—can be safely ignored. In this thesis, the main biological phenomena considered in this way are

- (1) homologous recombination, in Chapter 2
- (2) gene duplication and loss, in Chapter 3, and
- (3) variability of mutation rates between genes, in Chapter 4.

In Chapter 5, we take a different approach, instead focusing our study on a particular kind of inference method, maximum likelihood estimation, with the goal of better understanding one of the ways which it can fail (i.e., long-branch attraction).

The structure of the remainder of this introduction is as follows. In Section 1.2, we provide a concrete example of the sort of phylogenetic analyses that we are interested in studying. In Section 1.3, we describe the underlying probabilistic models of evolution which are typically assumed in phylogenetic analyses, with particular attention paid to drawing connections between biological assumptions and mathematical ones. In Section 1.4, we present another way of looking at models of evolution—as parameterization maps—in order to precisely define the notion of identifiability and introduce an perspective on models of evolution which is more algebraic than probabilistic. Finally, in Section 1.5, we provide a brief overview of the major results in the chapters of this thesis.

1.2 Phylogenetic Reconstruction: A Motivating Example

Genomic data provides a wellspring of information about the otherwise hidden evolutionary past of living organisms. As a motivating example as to the types of problems we are concerned with, suppose we wish to estimate the evolutionary history of the following seven virus strains:

MN3, MN5, MN12, C14_CSU_034_10640, C33, C46, and
BoviShield_Gold_FP5_MLV_vaccine.

These strains are members of BHV-1, also known as bovine herpes virus type 1. The last virus is a commercial BHV-1 vaccine strain.

Associated with each of these strains is a DNA sequence, consisting of approximately 144,500 base pairs. These sequences are arranged as rows in a matrix, called a *multiple*

species alignment (MSA). Fig. 1.1 shows (a subset of) the MSA for the seven viral strains listed above. (The full dataset was obtained from NCBI taxonomy database [13]).

An MSA is arranged in such a way that the nucleotides in each column, or *site*, are assumed to have descended from a common ancestor. While the MSA mostly consists of the nucleotide letters A, T, C and G, gaps in the alignment (depicted with dashes) are common, possibly arising due to missing data or gene insertion/deletion events which occurred in the evolutionary past [14, 15].

The patterns of nucleotide polymorphism in an MSA are usually thought to convey information about the evolutionary past. In the case of the viral strains, for example, the sequences of the seven strains are all extremely similar, with 99.4% of the sites in the genome exhibit no variation of nucleotides between the strains. This indicates that the viral strains are all very closely related. Nonetheless, approximately 600 polymorphic sites exist in the genome, including two in the subset of the MSA shown in Fig. 1.1. We have identified these by marking the two sites with the symbol $*$. Based on the site pattern observed in both of these columns, and without knowing any further information, it would be reasonable to guess that the seven strains can be separated into two distinct clades consisting of the sets {C14_CSU_034_10640, C33, C46, BoviShield_Gold_FP5_MLV_vaccine} and {MN3, MN5, MN1}.

```
                                                                  *           *
>MN3  GCCGCGCGTGGAGGTGCTCTCCTCCTCCTCCTCCTCCTCCTCCGCCTCCTCGCCCGCGGCGTCGCCC
>MN5  GCCGCGCGTGGAGGTGCTCTCCTCCTCCTCCTCCTCCTCCTCCGCCTCCTCGCCCGCGGCGTCGCCC
>MN12 GCCGCGCGTGGAGGTGCTCTCCTCCTCCTCCTCCTCCTCCTCCGCCTCCTCGCCCGCGGCGTCGCCC
>C14  GCCGCGCGTGGAGGTGCTC---------TCCTCCTCCTCCTCCGCCTCCTCGTCCGCGGCGTCGTCC
>C33  GCCGCGCGTGGAGGTGCT------------CTCCTCCTCCTCCGCCTCCTCGTCCGCGGCGTCGTCC
>C46  GCCGCGCGTGGAGGTGCTC---------TCCTCCTCCTCCTCCGCCTCCTCGTCCGCGGCGTCGTCC
>Bovi GCCGCGCGTGGAGGTGCTC---------TCCTCCTCCTCCTCCGCCTCCTCGTCCGCGGCGTCGTCC
```

Figure 1.1: A subset of the MSA containing DNA sequences from the seven related viral strains.

There are myriad ways to estimate the course of evolution from data like that shown in Fig. 1.1. Using one such method (the likelihood-based software IQ-TREE [16]), and using the full genome as input, we estimate an unrooted phylogenetic tree, shown in

Fig. 1.2. This tree is a biologically interpretable hypothesis about the evolutionary history of the seven viral strains, with vertices representing common ancestors and branch lengths representing some measure of evolutionary distance. Consistent with our earlier guess, the tree contains an internal branch (marked with a the $*$ symbol) which splits the taxa into the two clades mentioned earlier.

Having produced an estimate, it is reasonable to ask how much confidence should we have in the tree in Fig. 1.2. Since the inference method used—maximum likelihood—assumes a particular model of evolution, what is the risk that the assumptions made by that model are a poor approximation of the underlying evolutionary process experienced by these strains? For example, we have assumed here that the evolution of these viral strains was *treelike*, i.e., that there was little to no transfer of genes between separate viral strains. Due to viral recombination, which is known to occur in BHV-1 viral strains [17], such an assumption may be unreasonable.

Even assuming the model of evolution is reasonably adequate, does 144,500 base pairs constitute a sufficient amount of data to have confidence in every feature of this tree? Of special interest are the very short internal edges, such as the edge separating the taxon pair $\{\text{BoviShield\_Gold}, \text{C14}\}$ from the rest of the tree. While not shown in Fig. 1.2, this length of this edge is only 0.00001446, when measured in expected number of substitutions per site. Under standard model assumptions, this means we expect to see only two substitutions occurring along that edge *in the entire genome*. Given that sampling, sequencing, and alignment methods are necessarily imperfect, how much confidence should we really have that the closest relative of BoviShield_Gold is actually C14 as the tree suggests, and not, say, C46? These are the sorts of questions which we aim to treat in a rigorous mathematical way in this thesis.

1.3 Mathematical Models of Evolution

As the questions raised in Section 1.2 suggest, we are interested in rigorously studying methods of reconstructing, as a tree or network, the evolutionary history of multiple taxa.

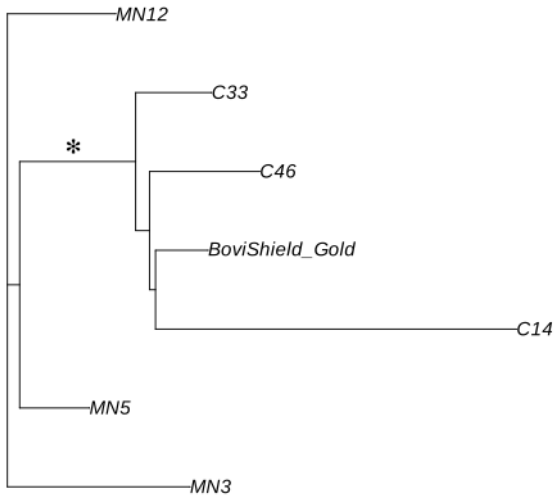


Figure 1.2: An estimated unrooted phylogenetic tree inferred from the full viral genome Fig. 1.1 using the maximum-likelihood software IQ-TREE with the default settings [16] and plotted using PhyloNetworks [18].

As a starting point, we begin with a discussion of some of the standard probabilistic models of evolution which will be utilized in this thesis, with particular attention to the connections between the sorts of biological assumptions which are made and how these assumptions justify (or not) the mathematical assumptions made by the models.

1.3.1 Trees and Networks

At a high level, the evolutionary history of related taxa or individuals can be represented by a *graphical model* (i.e., a tree or network) which depicts their evolutionary relationships, such as lineages, clades, most recent common ancestors, outgroups, hybridization events and so forth. The most important type of graphical model is the *phylogenetic tree*, a central object of study and one which provides an essential framework for thinking about the history of life from an evolutionary perspective [19]. The most commonly considered types of phylogentic trees are (1) *species trees*, which represent the evolutionary history of two or more populations, and (2) *gene trees*, which represent the genealogical history of a single homologous gene family, such as a gene which codes for a particular protein found in one or more species; by *homologous* we mean that the genes descended from a common ancestral

gene [19]. Phylogenetic trees have labeled leaves and also typically have edge weights $(d_e)_{e \in E(T)}$, called *branch lengths*, representing some biologically meaningful measure of evolutionary distance (e.g. number of generations, expected number of mutations per site).

The term *species tree*, although standard, is somewhat misleading as "species" is loaded and not applicable to all organisms. For example, we regard the tree in Fig. 1.2 as a species tree, even though it is not obvious how to define the term species with respect to virus taxonomy [20]. Similar difficulties arise when attempting to provide universal definitions to a whole host of common terms in biology, such as "population," "individual," and "gene," and owing to the fundamental complexity of life, imperfect simplifications of such concepts are unavoidable. As mentioned earlier, even the notion of a "tree of life" does not provide a perfect framework since evolutionary relationships are often not treelike, and indeed there has been a great deal of recent research motivated by the goal of relaxing that assumption.

1.3.2 Wright-Fisher

The models of evolution that we will consider are, at some level, continuous approximations of a discrete-time model known a the *Wright-Fisher model*, which itself provides a natural staring point for thinking about mathematical models of evolution.

This Wright-Fisher model considers an idealized population with discrete, non-overlapping generations each consisting of $2N$ haploid individuals (or equivalently N diploid individuals). It further assumes that the organisms are hermaphrodites which undergo random mating in each generation with no selection [21]. In this setting, the term "individual" refers not to an individual organism, but to an individual haploid gene.

Under these assumptions, the parent of any individual in generation g is chosen uniformly at random from the *previous* generation $g + 1$. Thus, starting with n individuals sampled from generation 1, one can proceed backwards in time by randomly choosing the parents from generation 1 for each sampled individual, and then choosing parents for *them*

from generation 2, and so forth. Each time, one can trace back the lines of descent, or *lineages*, as shown in Fig. 1.3. When two individuals share the same parent, their lineages are said to *coalesce*. As one goes back further in time, the occurrence of coalescence events can only reduce the number of lineages; when the most recent common ancestor of the sampled individuals has been reached, only one lineage will remain.

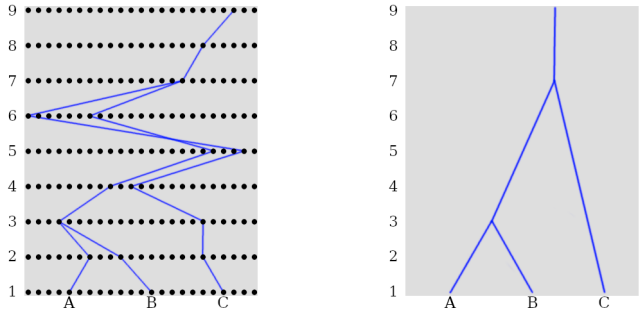


Figure 1.3: Wright-Fisher in a single population consisting of 9 generations with N=23 individuals per generation (left). Three individuals A, B, C are sampled in generation 1, and lines of descent are constructed by repeatedly choosing parents at random from the previous generation. Coalescent events among lineages from samples A, B and C occur in generations 3 and 7. The resulting genealogy for A, B, C is obtained by removing all individuals not ancestral to the sample, and untangling the lineages, and can be represented by the gene tree on the right.

Given two or more sampled individuals, how long ago was the most recent coalescence of lineages? That is, how many generations must we go back before reaching a common ancestor of some pair of individuals from our sample? To answer this question, let p_n be the probability that n individuals have exactly n distinct ancestors in the previous generation. In other words, p_n is the probability that no lineages coalesce in the previous generation. Under the assumption of random mating, the probability that two individuals share the same parent is $\frac{1}{2N}$, and hence it follows that

$$p_2 = 1 - \frac{1}{2N}.$$

By similar reasoning, the probability that *three* individuals "choose" three distinct parents

is

$$p_3 = \left(1 - \frac{1}{2N}\right)\left(1 - \frac{2}{2N}\right).$$

More generally, for all $n \geq 2$, and as $N \to \infty$, it holds that

$$\begin{aligned} p_n &= \prod_{k=1}^{n-1}\left(1 - \frac{k}{2N}\right) \\ &= 1 - \frac{1 + 2 + \ldots + (n-1)}{2N} + O\left(\frac{1}{N^2}\right) \\ &= 1 - \frac{\binom{n}{2}}{2N} + O\left(\frac{1}{N^2}\right). \end{aligned}$$

In other words, when N is large, and n is small relative to N,

$$p_n \approx 1 + \frac{\binom{n}{2}}{2N}.$$

This probability does not change from generation to generation, in the sense that depends only on the number of individuals sampled from a generation, but not on the generation itself. That is, no matter if the n individuals are sampled from the current generation, or the parent generation, or the great- great-grandparent generation, the probability of those n individuals having exactly n parents is the same. We can conclude from this that the time until the most recent coalescence is approximately geometrically distributed with success probability

$$1 - p_n \approx \frac{\binom{n}{2}}{2N}.$$

Therefore, given a sample of n individuals, if T_n is the number of generations with n

distinct ancestors of the sample, then

$$\begin{aligned}\mathbb{P}\left[T_n > 2Nt\right] &= (p_n)^{\lfloor 2Nt \rfloor} \\ &= \left(1 - \frac{\binom{n}{2}}{2N} + O\left(\frac{1}{N^2}\right)\right)^{\lfloor 2Nt \rfloor} \\ &= \left(1 - \frac{\binom{n}{2}}{2N} + O\left(\frac{1}{N^2}\right)\right)^{2N \cdot \frac{\lfloor 2Nt \rfloor}{2N}} \\ &= e^{-\binom{n}{2}t} + o(1).\end{aligned}$$

for any $t > 0$ as $N \to \infty$. It follows that if N is large, and t is rescaled so as to be measured in units of $2N$ generations, then

$$\mathbb{P}\left[T_n > t\right] \approx e^{-\binom{n}{2}t}. \tag{1.1}$$

This shows that if N is large, and n small relative to N, then T_n is approximately exponentially distributed with rate $\binom{n}{2}$. The convention of measuring time in units of $2N$ generations is common, and when time is scaled in this manner we say that it is measured *coalescent units* (e.g., one coalescent unit equals $2N$ generations).

The Wright-Fisher model is used extensively in population genetics to understand changes in allele frequencies due to genetic drift. While its assumptions (discrete generations, random mating, constant population size, etc) appear to be restrictive and unrealistic, a substantial relaxation of these assumptions is possible (e.g., to allow for such phenomena as overlapping generations, separate sexes, variable population size) by means of alternative rescalings of time [22].

1.3.3 The Coalescent

The *coalescent model*, due to [23], is an important continuous-time approximation of the Wright-Fischer model. In the coalescent model, one "forgets" about generations as discrete entities containing discrete individuals, and insteads models only the evolution of *lineages*

using a continuous time Markov process, with time measured in coalescent units. Without formally defining the state space of this Markov process (which can be found in [23]), the coalescent can be understood simply in the following manner as a bottom-up, backwards-in-time procedure for growing an n-leaf tree starting from the leaves, as we now describe.

First, for each $k = 2, \ldots, n$, let T_k be an independent rate $\binom{k}{2}$ exponential random variable. Then, starting at time $t = 0$ with n labelled nodes (which will become the leaves of the tree), trace lineages upwards as time proceeds. After T_n time has elapsed, choose two lineages uniformly at random and join them; this is the first a coalescence event and now only $n - 1$ lineages remain. After a further T_{n-1} time has elapsed, randomly choose two of the remaining $n-1$ lineages to coalesce. Continue in this manner so that for each k, T_k is the duration of the epoch in which there are k lineages. Then time $T_2 + T_3 + \ldots + T_n$ is the time of the most recent common ancestor of all n leaves, and after this time only a single lineage remains, so the procedure is stopped. To see that this procedure terminates in finite time, observe that

$$\begin{aligned}
\mathbb{E}\left[T_2 + \ldots + T_n\right] &= \sum_{k=2}^{n} \mathbb{E}\left[T_k\right] \\
&= \sum_{k=2}^{n} \frac{1}{\binom{k}{2}} \\
&= \sum_{k=2}^{n} \frac{2}{k(k-1)} \\
&= 2 \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) \\
&= 2 \left(1 - \frac{1}{n} \right).
\end{aligned}$$

Since this is finite, it implies that $T_2 + \ldots + T_n < \infty$ with probability one, and hence with probability one the output of the coalescent is indeed a tree.

Labelled trees generated in this manner are called *coalescent trees*, and their distribution is well-studied and understood, for example see [21, 22, 24, 25]. The justification for choosing $T_k \sim \exp\left(\binom{n}{2}\right)$ is due to Eq. (1.1), and the justification for choosing lineages

uniformly at random to coalesce is due to the assumption of random mating. From these facts alone, we can see that coalescent trees have a very particular kind of distribution; for example, since the rate of coalescences is very fast when there are many lineages and slows down as the number of lineages decreases, it follows that the lineages closer to the leaves tend to be much shorter than the lineages close to the root.

The coalescent provides a framework for thinking about genealogies, understood as the latent (i.e., unobserved) evolutionary relationships between sequences of DNA observed in different organisms [21]. In the study of population genetics, for example, coalescent trees are used to model evolution of alleles within a *single* population, and the distribution of coalescent trees, when combined with models of mutation, forms the basis for a myriad of estimators of genetic diversity [21]. For our purposes, however, we are interested in the evolutionary relationships *between* species, and so it is necessary to take yet another step, in order to generalize the coalescent to the setting where there are multiple related populations.

1.3.4 The Multispecies Coalescent

The *multi-species coalescent (MSC)*, due to [25], generalizes the coalescent to multiple related species, and is likely the most commonly-used model of gene evolution. The MSC takes as input a species tree parameter and outputs a gene tree.

The MSC model can be described informally, and we refer the reader to Fig. 1.4 for reference. In this model, a fixed *species tree* $\mathcal{S}$ is treated as a parameter representing the true phylogeny of the taxa; we regard the edges of $\mathcal{S}$ as "populations" and these are drawn as fat edges (e.g., the four-leaf tree in Fig. 1.4). Given this input, a *gene tree*, which represents the ancestry of a particular segment of the genome, is then generated according to the MSC in the following way. Starting at the leaves of $\mathcal{S}$, we draw lines in a bottom-up manner; as in the coalescent, these lines represent the genealogical lines of descent, or *lineages*, ancestral to the samples. If two lineages are in the same population, then they *coalesce* (i.e. merge into a single lineage) at rate 1, and pairs of lineages coalesce

independently. When only a single lineage remains, the most recent common ancestor of the sample has been reached, and the process terminates. The result is a gene tree whose edges are the ancestral lineages, which is the output of the MSC. Essentially, therefore, the MSC takes the idea of running the single population coalescent within each edge of the species tree in a bottom-up manner and "gluing" the results together.

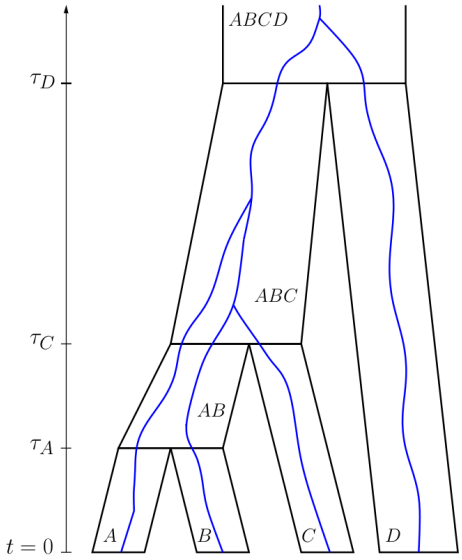


Figure 1.4: An example realization of the multi-species coalescent. The lineages of a gene tree are (shown as wavy blue lines) are regarded as evolving *within* the edges of a 4-taxa species tree S (with fat edges in black), which are regarded as populations. The species tree S, along with the divergence times τ_A, τ_C, τ_D, are treated as fixed parameters; by contrast, the coalescent times of the gene tree are random, with the main restriction being that two lineages can coalesce only if they are in the same population. The gene tree shown exhibits incomplete lineage sorting, as the lineages from A and B do not coalesce in population AB, and as a result its topology differs from that of S.

One complication inherent to the MSC is that two lineages might exit a population without having coalesced, a phenomenon called *deep coalescence*, or more commonly *incomplete lineage sorting (ILS)*. An example of this is shown occurring in population AB in Fig. 1.4. In biological terms, ILS is caused by the existence of gene polymorphism (i.e., the co-existence of two or more alleles for a gene in a population) is maintained in a population undergoing successive speciation events [26]. Since genetic drift has the effect of killing off polymorphisms via fixation, ILS is less probable when speciation events are separated by

long periods of time, and more likely when the time between speciation events is short.

ILS has recieved a great deal of attention as a potential source of error in phylogenetic estimation because it can result in gene trees not sharing the same topology as the species tree, a phenomenon known as *gene tree incongruence*. Gene tree incongruence due to ILS is thought to be widespread in the primate genome; for example, although chimps are the closest living relative to humans, nonetheless genes making up as much as 15% of our genome is more closely related to gorillas, and another 15% of the primate genome suggests that chimps are most closely related to gorillas rather humans [27, 28]. One especially striking result is that ILS can even give rise to situations, termed the "anomaly zone," in which the *most likely* gene tree topology is different from that of the species tree [29, 7, 6, 8, 24]. Concerns about the impact the ILS on inference has provided impetus for the development of a variety of 'supertree' methods which focus on recovering subtrees (or other information) from 3 or 4 leaves at a time, and piecing those results together to come up with an estimate of $\mathcal{S}$ [7, 30, 31].

At this point, it is important to note that a standard assumption underpinning most phylogenetic analyses is that gene trees are independent in the probabilistic sense. The biological assumption which underpins this assumption is that homologous recombination (i.e., crossing over) occurs between genes. The reason for this is that, although two genes located close to each other on a chromosome will usually be co-inherited (i.e., passed on together from parent to offspring), sometimes a recombination event occurs between them during meiosis, resulting in only one of them being passed on to the offspring. As a consequence, recombination *between* genes reduces linkage between their ancestries, so that over the course of many generations, the correlation between their genealogies are thought to decay to negligible levels [21].

However, this assumption of independence is in tension with another assumption implicitly made by all three models presented thus far: that individual genes—understood here as the portions of chromosomal material whose ancestry is modeled by the process—are indivisible. That is, the MSC and the models it is based on assume that individual

genes are always passed down from parent to offspring *in whole*, and not broken up by crossover events [32, 29, 24]. This assumption may be violated due to *intralocus recombination*, whereby recombination breakpoints occur *within* the genetic locus of the gene, with the result that a single observed gene may consist of several different components with distinct evolutionary histories (i.e., so that the gene itself is comprised of two DNA segments tracing back to different grandparents). Tracing a gene lineage backwards in time, an intralocus recombination event would correspond to a splitting of the lineage into two lineages. Intralocus recombination is common in real data, occuring in $\sim 80\%$ of protein-coding genes in Eukaryotes [29]. It is currently an open question how much intralocus recombination can impact phylogenetic inference when not accounted for, and there has been significant debate on this topic [33, 34, 35, 32]. It is therefore of interest to better understand the extent to which intralocus recombination may affect inference under realistic biological settings [29, 36]. This question is considered in detail in Chapter 2

1.3.5 Mutation

Up to this point, what we have described are models of a *genealogical* process, models which describe only lineages and shared ancestry; the real power of such models to allow for inference manifests when they are paired together with models of DNA *mutation* and *substitution*. After all, since data takes of form of observable DNA sequences, and genealogies are almost never directly observable, it is necessary to consider how DNA sequences of a gene might be related to its latent genealogy. This is usually formulated as an additional step, after first generating a gene tree, in which some model of nucleotide substitution involving a Markov chain on a tree [37, 38], is then run *using the gene tree as input* to obtain a DNA sequences. We begin by examining some of the assumptions of standard mutational models, which model DNA mutation by a continuous-time Markov jump process.

Each generation, the nucleotide at a specified site i in the genome has a small chance of undergoing spontaneous neutral mutation. By *neutral* we mean selectively neutral,

i.e., that the mutation has no effect on the fitness of the offspring. It is assumed that the chance of a neutral mutation occurring from generation to generation is independent (though in practice this assumption is not always true [21]). Under these conditions, it is natural to model the "arrivals" of mutations at i over many generations by a fixed-rate Poisson process.

To justify this, it is useful to recount the Poisson approximation of the binomal. Let $\tilde{\mu} \in (0,1)$, and let $p_1, p_2, \ldots \in (0,1)$ be a sequence such that

$$p_n = \frac{\tilde{\mu}}{n} + o\left(\frac{1}{n}\right). \tag{1.2}$$

For each $n \geq 1$, consider the sequence of random variables

$$X_1^{(n)}, \ldots, X_n^{(n)} \overset{iid}{\sim} \text{Bernoulli}(p_n)$$

so that

$$X_1^{(n)} + \ldots + X_n^{(n)} \sim \text{Binomial}(n, p_n).$$

Since $np_n \to \tilde{\mu} \in (0, \infty)$, the *law of rare events* applies (see, e.g., [Breiman, Chp9]) and it holds that

$$X_1^{(n)} + \ldots + X_n^{(n)} \xrightarrow{d} \text{Poisson}(\tilde{\mu}).$$

which implies that for each nonnegative integer k,

$$\mathbb{P}\left[X_1^{(n)} + \ldots + X_n^{(n)} = k\right] = \frac{\tilde{\mu}^k}{k!} e^{-\tilde{\mu}} + o(1)$$

as $n \to \infty$. Loosely speaking, when n is large, then $np_n \approx \tilde{\mu}$ and the following approximation holds

$$\text{Binomial}(n, p_n) \approx \text{Poisson}(\tilde{\mu}).$$

Returning to the question of mutations in the genome, let us assume that there exists some $\mu \in (0,1)$ such that in each generation, the probability of a neutral mutation occur-

ring at site i is μ. Therefore, going back n generations, we can represent the occurrence of mutations in each generation by n independent Bernoulli random variables:

$$Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Bernoulli}(\mu).$$

Let Y denote the number of mutations occurring at site s over the past n generations. Then

$$Y = Y_1 + \ldots + Y_n \sim \text{Binomial}(n, \mu).$$

Taking $\tilde{\mu} = n\mu$, this is precisely the regime of Eq. (1.2), and it follows that Y should be well-approximated by a Poisson distribution with rate parameter $n\mu$, provided that n is large and μ is small. Both of these conditions are realistic, as evolution occurs over deep timescales, and typical estimates for μ hover around $\mu = 10^{-8}$ or $\mu = 10^{-9}$ [21]

While we have described the evolution of only a single site, it is common to implicitly assume that all sites in the genome evolved independently in the same manner, i.e., that the mutation rate μ is constant across sites. This assumption is less defensible, resulting in a site distribution which is substantially different from real data; for example *invariant sites*, which experience no variation across many taxa, are much more common in real data than would be the case under these assumptions [21, 38]. The question of across-site rate variation is considered in detail in Chapter 4.

1.3.6 Substitution

Over the course of many generations, the nucleotide located at a particular site in the genome may have experienced more than one mutation. For example, the nucleotide might have mutated from A to C, and then back to A again. Or it might have mutated many times over many generations, e.g., from A to C to A to T. In such cases, some or all of the mutation events are not observable. What are observed, instead, are *substitutions*, understood as DNA differences observed between species [21].

Site substitution models provide a way to model the evolution of DNA sequences from

a tree (e.g., a gene tree), provided certain assumptions about the underlying mutational process. A site substitution model can be understood in a probabilistic sense, as a Markov chain on a tree, which more informally can be thought of as a procedure which takes as input a tree and outputs nucleotides at the leaves of the tree. That is, we begin with a state space S of "nucleotide sites" (e.g., $\{A, T, C, G\}$ or $\{0, 1\}$). The input is an n-leaf tree $\mathcal{T}$ with leaves labelled $1, \ldots, n$ such that each edge of $\mathcal{T}$ has an associated a transition matrix $P^{(e)} = (p_{ij}^{(e)})$, where $p_{ij}^{(e)}$ is the probability of a transition from state i to state j along edge e. Starting at the root of the tree, one assigns a nucleotide state according too some prescribed probability distribution π on S. Then independently along each edge $e \in E(\mathcal{T})$, and propagating outwards from the root, state transitions occur in a random manner according to the transition matrix $P^{(e)}$. The output of this procedure, after each edge has been considered, is the state of the process at each labelled leaf of $\mathcal{T}$, i.e., one nucleotide at each leaf of the tree.

Since fewer mutations are expected to occur in the history of closely-related organisms than distantly-related ones, it is reasonable that for each edge e, the transition matrix $P^{(e)}$ should depend in some way on the branch length of e. Although it is possible to simply assign an arbitrary probability transition matrix to each edge of e, it is much more common to construct transition matrices from a set of explicit assumptions about the mutational process.

In order to explain this latter approach, let us first consider the case of a single edge of a phylogenetic tree, having branch length $d_e > 0$, measured in coalescent units. Assume a constant per-site per generation neutral mutation rate of $\mu > 0$.

We say that a square matrix $Q = (q_{ij})$ is a *rate matrix* if the rows of Q sum to one, and $q_{ij} \geq 0$ whenever $i \neq j$. When $i \neq j$, the term q_{ij} represents the instantaneous rate at which state i might "mutate" to state j. In this thesis, we will consider only the simplest models, the Cavender-Farris-Neyman (CFN) [39, 40, 41] and Jukes-Cantor (JC) [42] models. In the Jukes-Cantor model, there are four nucleotide states A, C, T, and G, while in the Cavendar-Farris-Neyman model, there are only two (e.g., 0 and 1, representing

purine and pyrimidine). The rate matrices for these two models are

$$Q_{\text{CFN}} = \begin{bmatrix} -\mu & \mu \\ \mu & -\mu \end{bmatrix} \quad \text{and} \quad Q_{\text{JC}} = \begin{bmatrix} -\mu, \frac{\mu}{3}, \frac{\mu}{3}, \frac{\mu}{3} \\ \frac{\mu}{3}, -\mu, \frac{\mu}{3}, \frac{\mu}{3} \\ \frac{\mu}{3}, \frac{\mu}{3}, -\mu, \frac{\mu}{3} \\ \frac{\mu}{3}, \frac{\mu}{3}, \frac{\mu}{3}, -\mu \end{bmatrix}.$$

The structure of a rate matrix represents some set of assumptions about the mutational process, such as the number of nucleotide states as well as the relative likelihood of the various types of mutations. In the two examples above, since $q_{ij} = \mu$ whenever $i \neq j$, both Q_{CFN} and Q_{JC} "contain" the assumption that at a given moment in time, all types of mutations are equally likely to occur; in addition, since neither Q_{CFN} nor Q_{JC} depends on t, they also contain the assumption that the rates of mutations along each individual lineage do not vary over time. The CFN and JC models have the virtue of simplicity, but are unrealistic. For example, it is known that due to the molecular structure of DNA, transitions (mutations between A and G, or between C and T) occur more often than transversions (i.e. all other mutations) [21]. A plethora of more sophisticated substitutions models have been developed, and the reader is referred to [43] for a good overview of research trends in this area.

Using a rate matrix Q, we can construct probability transition matrix in the following way. Let $(X_t)_{t\in[0,d_e]}$ be a continuous-time stochastic process with state space S (say, $S = \{A, T, C, G\}$), so that X_t represents the nucleotide state at site i on edge e at time t. Letting

$$p_{ij}(\epsilon) := \mathbb{P}\left[X_{t+\epsilon} = j \mid X_t = i\right]$$

and $\delta_{ij} := 1_{[i=j]}$, we make the assumption that for all $i, j \in S$,

$$p_{ij}(\epsilon) = \delta_{ij} + q_{ij}\epsilon + o(\epsilon) \quad \text{as } \epsilon \to 0^+, \tag{1.3}$$

Eq. (1.3) expresses the idea that the process (X_t) is a Markov process, i.e., it represents

the assumption the future behavior of the process X_t depends not on its past history, but only on the current state at time t; this is because as $\epsilon \to 0^+$, the right-hand side of Eq. (1.3) depends asymptotically only on the current state i, as well as the state j, but not on any previous history of the process. Using the Chapman-Kolmogorov equation (see, e.g., [44, chapter 15]), it holds that for any $t, \epsilon > 0$,

$$\begin{aligned} p_{ij}(t+\epsilon) &= \sum_{k \in S} p_{ik}(t) p_{kj}(\epsilon) \\ &= \sum_{k \in S} p_{ik}(t)(\delta_{kj} + q_{kj}\epsilon) + o(\epsilon) && \text{by Eq. (1.3)} \\ &= p_{ij}(t) + \sum_{k \in S} p_{ik}(t) q_{kj} \epsilon + o(\epsilon). \end{aligned}$$

Rearranging terms,

$$\frac{p_{ij}(t+\epsilon) - p_{ij}(t)}{\epsilon} = \sum_{k \in S} p_{ik}(t) q_{kj} + o(1),$$

and sending $\epsilon \to 0^+$, we obtain a system of ODEs known as the *Forwards Equations*:

$$P'(t) = P(t)Q. \tag{1.4}$$

Utilizing the boundary condition $P(0) = I$, it follows that this system of equations has solution

$$P(t) = \exp(Qt), \tag{1.5}$$

where $\exp(\cdot)$ is the matrix exponential. Eq. (1.5) shows the general form of a probability transition matrix $P(t)$ arising from time-homogeneous continuous time Markov process, and one can use Eq. (1.4) or Eq. (1.5) to compute the entries of these matrices from a rate matrix Q. In general, transition matrices are matrices of conditional probabilities; in the

case of the Jukes-Cantor process for example,

$$P_{\mathrm{JC}}(t) = \begin{pmatrix} p_{AA}(t) & p_{AC}(t) & p_{AG}(t) & p_{AT}(t) \\ p_{CA}(t) & p_{CC}(t) & p_{CG}(t) & p_{CT}(t) \\ p_{GA}(t) & p_{GC}(t) & p_{GG}(t) & p_{GT}(t) \\ p_{TA}(t) & p_{TC}(t) & p_{TG}(t) & p_{TT}(t) \end{pmatrix}$$

where

$$p_{ij}(t) := \begin{cases} \frac{1}{4}(1 + 3e^{-\frac{4}{3}\mu t}) & : \quad i = j \\ \frac{1}{4}(1 - e^{-\frac{4}{3}\mu t}) & : \quad i \neq j \end{cases}.$$

A transition matrix can then be assigned to each edge $e \in E(\mathcal{T})$ by setting $P^{(e)} := P(d_e)$. Most commonly, the same rate matrix Q is used for all branches of the species tree $\mathcal{T}$, an assumption known as rate homogeneity [38].

1.4 Models as Maps

1.4.1 Identifiability

The practical question of reconstructing an evolutionary history from genomic data is at heart a question of recovering parameters from a parametric statistical model; in this context, the fundamental limits of what can and cannot be inferred are formulated in terms of *statistical identifiability*. To define this notion precisely, let Θ be a space of parameters and let $\Delta_{d-1} = \left\{p \in \mathbb{R}^d : p_1, \ldots, p_d \geq 1, p_1 + \ldots + p_d = 1\right\}$ be a probability simplex of dimension $d-1$, for some $d \geq 2$. A *phylogenetic model* is the image of a map

$$\phi : \Theta \to \Delta_{n-1}$$

given by

$$\theta \mapsto p_\theta$$

which sends a vector of parameters $\theta = (\theta_1, \ldots, \theta_m)$ to a probability measure representing the probability distribution of some random but observable biological feature (e.g., the site frequency spectrum).

Fundamental and practical questions about the theoretical limits of what sorts of biological events can be inferred boil down to properties of this map ϕ. If the parameterization ϕ is injective, then the parameter vector θ is said to be *statistically identifiable*, or *identifiable* for short. If θ is identifiable, then in principle it can be recovered it from p_θ by inverting the map ϕ_θ. At first glance, for someone interested in inference, identifiability of model parameters would seem to be the *very least* one might ask of a model. After all, if θ is not identifiable, then it cannot be recovered, even in the theoretical limit with infinite data.

Nonetheless, identifiablity is a very strong property, often too strong for many applications in phylogenetics, and therefore various relaxations of identifiability are used to characterize what information about the parameters can be recovered when identifiability is not satisfied. For example, it is often the case that θ is not identifiable, but that there is some *function* of the parameters

$$s : \Theta \to \mathbb{R}$$

which can be recovered. In this case, the function s, which is also called a "parameter," is said to be *identifiable* if there exists a function f such that $s(\theta) = f \circ \phi(\theta)$ for all $\theta \in \Theta$. A typical example of this is when s is a projection such as $s(\theta) = \theta_i$ for some fixed $i \in [m]$, in which case we drop the formality and simply say that the parameter θ_i is identifiable. Another common relaxation identifiability is the notion of *generic identifiability*, which holds when ϕ is injective when restricted to some dense open subset of its domain; in particular this implies that for (Lebesgue) almost every choice of model parameters, the parameters *can* be recovered from the distribution. For a good overview of these and related notions, see [45, 38].

There are essentially two types of idenfiability questions considered in this thesis: the first is *topological identifiability*, which asks whether the structure of underlying graphical

model (i.e., the tree or network) can be recovered, and the second is *parameters identifiability*, which asks whether the other parameters (e.g., branch lengths, hybridization probability, root location) can be recovered. Next we will give an example of the latter notion, which will also help elucidate an alternative (algebraic) perspective on models of site substitution as polynomial maps, which will be utilized in Chapter 5.

1.4.2 Substitution Models as Polynomial Maps

Suppose $\mathcal{T}$ is a binary tree with n leaves having topology τ and branch lengths $d_e \in (0, \infty)$ for all $e \in E(\mathcal{T})$. Let

$$X = (X_1, \ldots, X_n) \in \{0, 1\}^n$$

be a random vector such that for each $i \in [n]$, X_i represents the nucleotide observed at leaf i of $\mathcal{T}$, and let p_θ be the distribution of X under the CFN site substition model. Thus we can define the CFN model on a tree with topology τ as the image of the map

$$\phi_\tau : \Theta \to \Delta_{2^n-1} \subseteq \mathbb{R}^{2^n}$$

given by

$$(\theta_1, \ldots, \theta_{2n-3}) \mapsto p_\theta := (p_{ijkl}(\theta))_{i,j,k,l \in \{0,1\}},$$

where

$$p_{ijkl}(\theta) := \mathbb{P}_\theta \left[X_1 = i, X_2 = j, X_3 = k, X_4 = l \right].$$

It turns out that under a suitable reparameterization, detailed in Chapter 5, the function ϕ is a *polynomial map*, meaning that for all choices of $i, j, k, l \in \{0, 1\}$ the probability $p_{ijkl}(\theta)$ can be written as a polynomial function $f_{ijkl}(\theta)$ in the variables $\theta_1, \ldots, \theta_{2n-3}$. Therefore we can define a map between polynomial rings

$$\psi_\tau : \mathbb{C}[p_{0000}, p_{0001}, \ldots, p_{1111}] \to \mathbb{C}[\theta_1, \ldots, \theta_{2n-3}]$$

by

$$p_{ijkl} \mapsto f_{ijkl}(\theta).$$

Let $I_\tau = \ker(\psi_\tau)$. Here, I_τ is a polynomial ideal consisting of all polynomials in the variables $p_{0000}, p_{0001}, \ldots, p_{1111}$ which vanish for any choice of branch lengths of $\mathcal{T}$. By the Hilbert Basis theorem there exists a finite set of polynomials $g_1, \ldots, g_k \in I_\tau$, called *phylogenetic invariants*, which generate I_τ. Moreover, since

$$\mathrm{im}(\phi_\tau) \subset \left\{ x \in \mathbb{C}^{2^n} : g_1(x) = \ldots = g_k(x) = 0 \right\}.$$

the phylogenetic invariants $g_1, \ldots, g_k$ can be thought to implicitly define the model $\mathrm{im}(\phi_\tau)$.

In this example we have assumed the site substitution model to be the CFN model; in fact the same story plays out under a wide class of substitution models, known as group-based models [45]. The study of phylogenetic invariants has given rise to a powerful toolset for proving identifiability results for tree-based models [45, 46, 47, 48, 49, 50, 51], and increasingly network-based models as well [52, 53, 54]. In Chapter 5 we will utilize phylogenetic invariants as *constraints* in a constrained optimization problem to compute maximum likelihood estimates for 4-leaf trees.

1.5 Chapters Overview

In this section we present an overview of main areas of inquiry, which are divided into four chapters.

1.5.1 Recombination

As discussed in Section 1.3.4, although the MSC assumes recombination occurring *between* genes, it assumes that genes themselves are indivisible, i.e., that recombination breakpoints do not occur *within* observed genes. The inherent tension between these two assumptions gives rise to the question of how (and how much) unrecognized intralocus recombination affects phylogenetic inference.

In Chapter 2 we investigate the impact of intralocus recombination on species tree estimation. In particular, we introduce a previously unknown way in which homologous recombination can bias species tree estimation. To do this we adapt a model of evolution incorporating intralocus recombination to a multispecies setting. The model we consider, the ancestral recombination graph (ARG), due to [55], is similar to the coalescent but allows for gene lineages to periodically split in two (these are recombination events). A realization of the generalized model we consider is shown in Fig. 1.5.

The inference procedure that we consider is based on a method known as R^* (majority-vote rooted triple), which is known to be a statistically consistent estimator of the species tree topology given gene tree data generated according to the MSC (i.e., when there is no intralocus recombination) [8, 56]. For each sampled gene, and for each triplet of leaves, one infers a rooted topology for that triplet from sequence data. An estimate of the species tree topology is then assembled using a majority-vote procedure [57].

Our contribution is to provide the first rigorous analytic proof that intralocus recombination can result in inconsistent inference. To do this, we introduce a modification of the MSC which allows for intralocus recombination, and prove that under the right conditions, recombination rate heterotachy can result in R^* failing to converge to the correct topology. The proof of this theorem, which hinges on a subtle interplay between ILS and uneven rates of recombination, provided insights into a new, previously unconsidered, way that recombination might affect inference. In addition, we provide evidence from simulations to characterize this phenomenon and show that it can occur under biologically plausible mutation and recombination rate parameters.

In addition, we also implemented a simulator to help characterize the sorts of trees which produced this effect. Our simulation study suggests that, at least for small trees with 3 or 4 taxa, majority-vote methods may be reasonably robust to recombination-induced error when rates of recombination do not differ too much between different species. The impact of recombination on more commonly-used maximum-likelihood methods is not yet known, but our simulations suggest a similar effect holds.

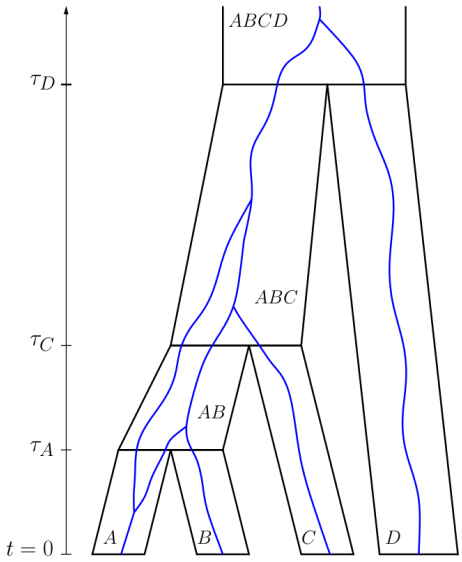


Figure 1.5: An example realization of a multi-species ARG, with lineages of a multispecies ARG shown in blue. A recombination event is shown occurring in population A where the lineage splits into two lineages. Under the MSC, such events are assumed not to happen (compare: Fig. 1.4).

1.5.2 Gene Duplication and Loss

In Chapter 3, we consider the impact of *gene duplication and loss (GDL)* on species tree inference. The precise meaning and implications of GDL are not obvious and require some explanation. Gene duplication can arise as a result of several biological mechanisms [58] and is a common event in all three domains of life [59]. For our purposes the key feature of gene duplication is that it results in the creation of individual whose genotype contains an additional copy of a particular gene; the gene copies in this individual are termed the *mother duplicate* and *daughter duplicate*. Gene duplication thus involves the creation of a new locus in a genome, i.e., the location of the daughter duplicate.

As with all other genes, the two gene copies arising from a duplication event are assumed to evolve independently due to recombination occurring between them, however there is one important difference between the mother and the daughter duplicates: since the locus of the daughter duplicate did not exist prior to the duplication event, the daughter duplicate's descendants cannot have a MRCA preceding the gene duplication event.

The same is not true of the mother duplicate. Thus, if we wish to modify the MSC to account for gene duplication, we should require that all pairs of lineages descendant to the daughter duplicate coalesce prior to reaching the time of the gene duplication.

Gene loss occurs when an individual is born containing fewer copies of a particular gene than their ancestors. This is termed a *gene deletion* event [60] and, like gene duplication, has the effect of introducing into the population an individual whose genome is structurally different from that of their parent(s). As with gene duplication, there are several biological mechanisms which can cause this to occur [61]. While the term "gene deletion" is sometimes used more broadly to include to loss of gene function due to mutation [61], we do not adopt that usage here. In the case of both gene duplication and gene loss events—both of which introduce into the population an individual with a different number of copies of a gene from their parent(s)—there are three possible outcomes which may occur due to genetic drift: the variant introduce will either (a) sweep to fixation, (b) go extinct, or (c) neither, a situation called *hemiplasy* [60]. Hemiplasy of GDL events can lead to a range of complex scenarios and, following [60], we assume for simplicity that it does not occur.

In light of the above discussion of gene duplication and loss, it is worth revisiting the concept of homology introduced in Section 1.3.1. Homologous genes—i.e., genes which are related by vertical descent from a common ancestor—can be categorized as *paralogous* or *orthologous*, depending on whether the are related via a gene duplication, or not [62]. That is, if the last common ancestor of two gene lineages generated those lineages via a gene duplication event, then the genes are said to be *paralogous*; on the other hand, genes are said to be *orthologous* if their ancestry can be traced back to a common ancestor without crossing a duplication event, and instead their ancestry separated due to speciation [19].

The type of gene homology assumed by the MSC, and indeed most current phylogenetic methods [63], is that genes are orthologous. But as we have noted, lineage coalescence should behave differently depending on whether the lineages are descendants of a mother or daughter duplicate. In order to more accurately model gene evolution in the presence

of paralogy, Rasmussen and Kellis [60] introduced a model of gene evolution incorporating gene duplication and loss into the multispecies coalescent, called the *DLCoal model*; this model has been regarded as an important methodological advance as it is the the first unified model of both GDL and ILS [64, 59].

The DLCoal model takes as input a species tree and outputs a multilocus gene tree. In the first step, a top-down birth-death process is run within the branches of the species tree to simulate gene duplication and loss events. This process produces a "locus tree" situated within the species tree, and a slightly modified version of the MSC is then run, as a bottom-up process within the locus tree to produce a gene tree. The main modification to the MSC is to condition on all pairs of lineages descending from a daughter duplicate to coalesce before reaching the time of duplication.

In Chapter 3, we analyzed the most commonly used quartet-based inference method, ASTRAL [31, 65, 66], and its performance given data coming from the DLCoal model. In particular, we considered the problem of recovering a species tree topology from many independent gene trees generated according to the DLCoal model. Through a direct analysis of the DLCoal process on 4-leaf trees, we proved four main results: (1) that the species tree topology is statistically identifiable; (2) that ASTRAL-one and (3) ASTRAL-multi both provide for statistically consistent estimation of the species tree topology, generalizing the result of [67] which considered GDL alone without ILS; and (4) we obtained an estimate regarding the sample complexity of ASTRAL-one in terms of the minimum branch length f, the duplication and loss rates λ, μ, total tree depth Δ, and number of taxa n—in particular, showing that as $f \to \infty$, the number of genes needed to accurately estimate a species tree topology is on the order of f^{-2}.

1.5.3 Variable Mutation Rates

One assumption which is usually made phylogenetic analyses is that of *homogeneity* of the mutational process [38]. This includes the assumption that the mutation rate μ and rate matrix Q is constant in time, as was assumed in Section 1.3.5, meaning that mutation

rates do not change from generation to generation. The term homogeneity is also used to refer to the assumption that mutation rates do not very between different branches on the species tree, which in practice means that the same rate matrix Q is used for all edges of the species tree [38].

Yet another way that mutation rates might vary is *across* the genome, a phenomenon known as *among-site rate variation* (ASRV). That is, mutation rates may be higher for some sites or genes than for others. In addition to this, regions of the genome which are conserved (e.g., by purifying selection) may exhibit few if any substitutions between taxa [38]. The use of the gamma distribution for modeling mutation rates was suggested by [68], and the most commonly-considered distributions for this purpose are either either a discrete gamma distribution or a gamma with an atom at zero (i.e., mixture of a gamma distribution with a Dirac point mass at zero) [69, 70, 71, 38]. The consideration of a point mass at zero is important to model genes which are invariant, for example genes which are sufficiently conserved that they do not change at all over long periods of time [69].

The evolutionary distance used for gene tree branch lengths is typically measured in expected number of mutations per site. This is a function of both the number of generations elapsed as well as the per-generation per-site mutation rate μ, neither of which is independently identifiable from DNA sequence data alone [21, 24]. Assumptions about the mutational process, and in particular about the rates of mutation, may therefore have a significant impact on estimation of gene tree branch lengths. In fact there is also evidence that ignoring ASRV decreases accuracy of topological estimation as well [72].

In Chapter 4, we investigate the effect of ASRV on the sample complexity of species tree estimation by considering a relaxation of the assumption that all genes undergo mutation at the same rate. Specifically, we consider gene trees generated according to the MSC in which each gene tree is assumed to have a mutation rate drawn randomly from some fixed distribution. This work utilizes techniques developed in [73] which established sample complexity bounds relating the number and lengths of genes needed for accurate estimation. We show that in order to distinguish two branches differing by length $f > 0$

(here f is small), the number of genes required for accurate estimation grows on the order of at least $\frac{1}{f^2}$ as $f \to 0$, assuming mutation rates are drawn according to a gamma distribution with an atom at zero. We also prove a lower bound under much more general assumptions about the distribution of mutation rates, almost certainly encompassing any biologically plausible distribution, however in this latter case the the proof is more involved and the story is somewhat different: the lower bound on the number of samples in this case is either on the order of $-\frac{1}{f^2 \log f}$ or $\frac{1}{f^2}$ depending on whether or not the mutation rates are bounded away from zero.

1.5.4 Long-Branch Attraction

In Chapter 5, we take a somewhat different approach. Rather than considering a specific biological assumption, we focus on inference method known to fail for reasons not fully understood. Specifically, we consider maximum likelihood inference for trees with three or four leaves with data generated according to the CFN model of site substitution. Given some data, the *maximum likelihood estimate (MLE)* consists of the parameter values which maximize the probability of the data under a specified model. Maximum likelihood is among the most widely-used methods for inferring phylogenetic trees from sequence data, however it is thought to perform poorly when the underlying tree parameter contains two or more distantly-related or fast-evolving taxa. In particular, it is thought that maximum likelihood tends to incorrectly infer very distantly related taxa to be more closely related than they are; a phenomenon which is thought to be widely observed in real data, and which has the potential to result in major errors in species tree estimation [74]. While it has been widely-studied, the exact nature of *long-branch attraction* (LBA) is not fully understood [2, 3, 75, 29]. With the aim of better understanding LBA, we explore the problem of computing closed form and exact solutions to the maximum likelihood problem for 3- and 4-leaf trees when data is generated according to the CFN model of site substitution.

In particular, Chapter 5 contains two main results. First, using Hadamard conjugation,

which combines a nonlinear change of coordinates and the discrete Fourier transformation, we compute a closed-form solution to the maximum likelihood problem for unrooted 3-leaf trees, given generic data. That is, we provide a formula for the numerical branch-length parameter, as a function of the data, which maximizes the log-likelihood function. For the 4-leaf case, by contrast, it is unlikely that a closed-form solution exists. Instead, we utilize the results for the 3-leaf case, along with techniques from numerical algebraic geometry and phylogenetic invariants, to implement a fast algorithm to compute an *exact solution* to the 4-leaf MLE problem. This algorithm computes the maximum likelihood tree to arbitrary precision without the need to execute a heuristic (e.g. hill-climbing) search. Our second main result is a proof of the correctness of this algorithm. In both cases, particular attention is paid to account for submodels in which one or more branch of the underlying tree parameter has infinite length, as it is believed that such cases may provide special insight into the problem of LBA.

Chapter 2

Inconsistency of Triplet-Based and Quartet-Based Species Tree Estimation under Intralocus Recombination

Abstract: We consider species tree estimation from multiple loci subject to intralocus recombination. We focus on R^*, a summary coalescent-based method using rooted triplets, as well as a related quartet-based inference method. We demonstrate analytically that in both cases intralocus recombination gives rise to an inconsistency zone, in which correct inference is not assured even in the limit of infinite amount of data. In addition, we validate and characterize this inconsistency zone through a simulation study that suggests that differential rates of recombination between closely related taxa can amplify the effect of incomplete lineage sorting and contribute to inconsistency.

2.1 Introduction

Species tree estimation from genomic data is complicated by various biological phenomena which generate phylogenetic conflict, among them hybridization, horizontal gene transfer, gene duplication and loss, and incomplete lineage sorting (ILS) ([76]). In particular, ILS may cause phylogenetic conflict in which a gene tree exhibits a different topology from that of the species tree, and is of greatest concern for species trees with short internal branches [76]. Of some interest is the existence of an anomaly zone for species trees, in which the most probable topology in the gene tree distribution differs from the topology of the species tree ([6, 8, 7]) [see also 24, 29, for a more recent discussion of these and othe relevant issues].

The existence of an anomaly zone has served as an impetus for the development of summary coalescent-based methods, such as R^*, MP-EST, BUCKy, ASTRAL, and others [57, 77, 78, 31]. Some of these methods are based on the fact that rooted triples and unrooted quartets are special cases in which no anomaly zone exists [7, 30] and also provide sufficient information to reconstruct the full phylogeny [79, 80]. Provided that the gene trees are estimated without error, such methods can provide statistically consistent methods of estimating species tree topology [56].

A common assumption of coalescent-based models based on the multispecies coalescent (MSC) [25, 24] is that recombination occurs between genes (or loci)—so that gene trees may be assumed unlinked or statistically independent—but that *intralocus recombination* (i.e., recombination occurring *within* gene sequences), does not occur [32, 29]. The significance of the latter assumption—that is, the impact of intralocus recombination on phylogenetic inference—is a matter of present interest [36, 29] and much debate about its significance when unaccounted for [33, 34, 32]. One justification for assuming no intralocus recombination is that within-gene recombination may break gene function [76].

An influential simulation study argued that even high levels of intralocus recombination do not present a significant challenge for species tree estimation relative to other biological phenomena [33]. On the other hand, the authors of [35] suggest the absence of intralocus

recombination may be an unreasonable assumption in real data, such as protein-coding genes in eukaryotes [81, 29], and particularly in the case of species phylogenies with many taxa [34]. In particular, the potential for intralocus recombination to distort gene tree frequencies has been recognized as a challenge to summary coalescent-based methods, and the study of [33] has been critiqued for its focus on shallow divergences and limitation to a low number of loci and taxa [34].

In this chapter we take an analytical approach to investigate the effect of intralocus recombination. We prove that intralocus recombination has the potential to confound R^*, a summary coalescent-based methods based on inferring rooted triples. That is, we show that correct inference of rooted triplets cannot be guaranteed in the presence of intralocus recombination, assuming a distance-based approach is used for gene tree reconstruction. We then present a simulation study which characterizes the "inconsistency zone", i.e. the regime of parameters for S in which rooted triple inference does not converge to S as $m \to \infty$. We also demonstrate similar results for inference of unrooted quartets. In both cases, we find that the effect arises only when recombination occurs at different rates along different edges of the species phylogeny, and in particular only when differential rates of recombination are exhibited between closely-related taxa.

2.1.1 Key Definitions

A *species phylogeny* $S = (V_S, E_S; r, \bar{\rho}, \bar{\tau}, \bar{\theta})$ is a directed binary tree with vertex set V_S, edge set E_S, root $r \in V_S$, and n labeled leaves $L_S = [n]$, such that each edge $e \in E_S$ is associated with a length $\tau_e \in (0, \infty)$, expressed in coalescent units, a recombination rate $\rho_e \in [0, \infty)$, and a mutation rate $\theta_e \in [0, \infty)$. It is assumed that there exists an ancestral population common to all leaves of S, i.e., a population above the root, with respective mutation and recombination parameters. Mutation rates are assumed to be per site per coalescent unit (a coalescent unit being $2N_e$ generations for diploid organisms, where N_e is the effective population size); recombination rates are per locus per coalescent unit.

The general question considered here is how to reconstruct the topology of the species

phylogeny from gene sequence data sampled from its leaves. This sequence data takes the form of multiple sequence alignments; a *multiple sequence alignment* (MSA) is an $n \times k$ matrix M whose entries are letters in the nucleotide alphabet $\{A, T, C, G\}$ such that entries in the same column are assumed to share a common ancestor. The phylogenetic reconstruction problem in this chapter is to recover the topology of S from m independent samples of M.

We define a *rooted triple* to be a rooted binary phylogenetic tree with label set of size three; we use the notation $XY|Z$ (or equivalently $YX|Z$) to denote a rooted triple with leaves X, Y, Z having the property that the path from X to Y does not intersect the path from Z to the root [79]. The term *species triplet* refers to a restriction of S to three of its leaves. A rooted triple $XY|Z$ is said to be *uniquely favored* if it appears in more gene samples than either of the other two rooted triples $XZ|Y$ or $YZ|X$.

2.1.2 Inference Methods

This chapter considers *Majority-Rule Rooted Triple*, or R^*, a consensus-based pipeline for species tree estimation. R^* utilizes the fact that the full topology of S is uniquely determined by, and hence can be recovered from, its rooted triples [80]. The R^* pipeline has three steps: first, for each gene, infer a rooted triple for each triplet of leaves $X, Y, Z \in L_S$. Second, make a list of uniquely favored triples from the m sampled genes. Finally, construct the most-resolved topology containing only uniquely favored triples. When gene trees are drawn independently according to the MSC, it holds that for every set of three taxa, the most probable rooted triple in the gene tree distribution matches the rooted triple obtained by restricting the species tree S to that set of three taxa; for this reason, the topology of the R^* consensus tree converges to that of S [57].

Since we are interested in the inference of the species-tree topology from *sequence data*, we consider a distance-based approach in which a species triplet with leaves X, Y, Z is inferred to have topology $XY|Z$ if

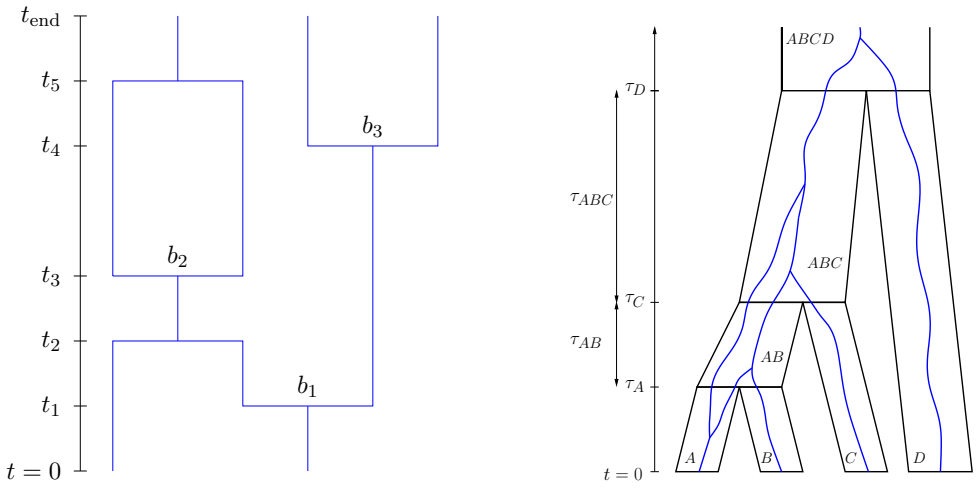
$$\delta_{XY} < \delta_{XZ} \wedge \delta_{YZ}. \tag{2.1}$$

where $\delta_{XY} = \delta_{XY}(M_k)$ is the number of mismatching nucleotides between sequences $\mathbf{s}_X$ and $\mathbf{s}_Y$ $(X, Y \in L_S)$. We refer to this inference procedure as $\boldsymbol{R}^*$ **with sequence distances**.

2.1.3 Multispecies Coalescent with Recombination

The model considered here, which we term the *Multispecies Coalescent with Intralocus Recombination*, or MSCR, uses the ancestral recombination graph (ARG) model from [55] [see also 82] within the framework of the multi-species coalescent (MSC) [25, 24, 8]. In the single-population ARG specified by [55], ancestors are represented by edges in the graph (see Figure 2.1a), and the number N of ancestors, or *gene lineages*, at time t is a bottom-up birth-death process in which births (recombination events) occur at rate ρN and deaths (coalescent events) occur at rate $N(N-1)/2$. When a coalescent event happens, two edges are chosen at random and merged into one. When recombination occurs, a randomly chosen lineage splits into two parent lineages. Each recombination vertex is labeled by a number b, chosen uniformly on [0,1]; this number is the *breakpoint* of the recombination.

The single-population ARG can be extended to multiple species in a manner similar to the MSC: at time $t = 0$, each leaf of S begins with a single lineage, and these lineages evolve in a bottom-up manner according to the ARG process along each edge of a fixed species tree (see Figure 2.1b). If $\mathcal{G}$ is a rooted directed graph with edge lengths and leaf and breakpoint labels obtained in this manner, then we say that $\mathcal{G}$ is **generated according to the MSCR process on** $\boldsymbol{S}$. In this scheme, the locus is modeled by the unit interval, and for each site $x \in [0, 1]$, a *marginal gene tree* $\mathcal{T}(x)$ can be obtained by tracing upward along the edges of $\mathcal{G}$ starting from the leaves; if a recombination vertex is reached with breakpoint b, take the left path if $x \leq b$ and the right path if $x > b$. This yields a collection of rooted edge-weighted binary trees; a simple example is shown in Figure 2.2. The set of marginal gene trees $\mathcal{M} := \{\mathcal{T}(x) : 0 \leq x \leq 1\}$ is almost surely finite [55]. For each $T_g \in \mathcal{M}$, define $I(T_g) = \{x \in [0, 1] : \mathcal{T}(x) = T_g\}$, and define $w_g = |I(T_g)|$,



(a) A single-population ARG

(b) A multi-species ARG

Figure 2.1: Two depictions of an ARG, in a single population (left) and in the multispecies case (right). In Fig. 2.1a, two lineages enter the population at time 0 and three exit at time t_{end}. Coalescent events occurred at times t_2 and t_5. Recombinations with breakpoints b_1, b_2, b_2 occurred at times t_1, t_3, and t_4. In Fig. 2.1b, the lineages of a multispecies ARG are shown in blue within a 4-taxa species tree S (the thick tree) with fixed edge lengths $\tau_A, \tau_B, \ldots, \tau_{ABC}$.

where $|\cdot|$ denotes Lebesgue measure. In words, $I(T_g)$ is the identical-by-descent segment of the locus having genealogy T_g, and w_g is the proportion of sites with genealogy T_g.

Measuring time in coalescent units, this chapter assumes that the per-site mutation rate is given by a fixed number $\theta > 0$ which does not vary on S. For each $x \in [0, 1]$, site x evolves independently according to the Jukes-Cantor process [42, 80] on the tree $\mathcal{T}(x)$. A somewhat more general description of this algorithm can be found in [83].

Thus, to model the evolution of a genetic locus consisting of k sites in which recombination breakpoints are distributed uniformly between them, a two-step process is followed. First, a multispecies ARG $\mathcal{G}$ is generated according to the MSCR process on S, from which a marginal gene tree $\mathcal{T}(x)$ is obtained for each $x \in [0, 1]$. Second, for each $x \in [0, 1]$ the Jukes-Cantor process is run with input tree $\mathcal{T}(x)$ in order to generate a nucleotide $\mathcal{N}(i, x) \in \{A, T, C, G\}$ for each $i \in L_S$. The MSA M_k is then defined as the $n \times k$ random matrix with rows $\mathbf{s}_1, \ldots, \mathbf{s}_n$ where for each $X \in [n]$, $\mathbf{s}_X = (s_X(1), \ldots, s_X(k))$ where $s_X(j) = \mathcal{N}(X, \frac{j}{k-1})$, $j = 0, 1, \ldots, k-1$. In this case, we say that M_k is **generated**

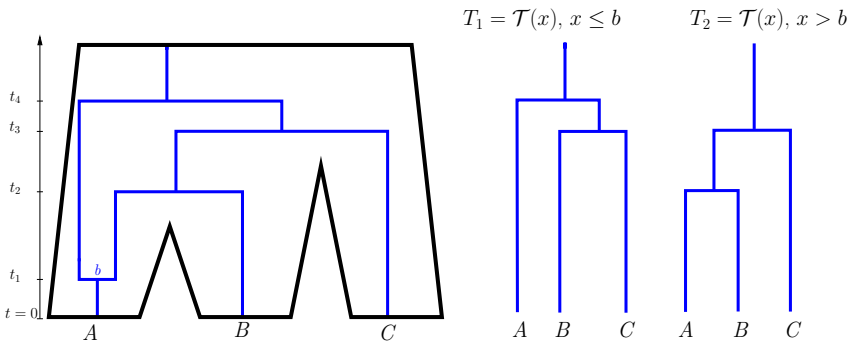


Figure 2.2: On the left, an ancestral recombination graph (in blue) is shown within a 3-taxa tree S (in black). The times of coalescence and recombination events are labeled $t_1, \ldots t_4$ on the time axis, and the breakpoint associated with the recombination event is labeled $b \in [0, 1]$. On the right, the corresponding marginal gene trees T_1 and T_2 are shown. This particular example also illustrates how intralocus recombination may contribute to phylogenetic conflict by allowing for 'partial' ILS, whereby one or more of the marginal gene trees (in this case T_1) exhibits a topology different from that of S.

according to the MSCR-JC(k) process on S.

In words, the MSCR-JC(k) process models the evolution of n homologous genes situated at a common genetic locus consisting of k sites, and which may have experienced intralocus recombination; these homologous genes are assumed to have been drawn from n distinct species whose true species phylogeny is represented by S. The resulting homologous aligned DNA sequences are the rows of the $n \times k$ matrix M_k. The phylogenetic reconstruction problem considered here pertains to whether the topology of S can be recovered from sequence data generated in this manner, or more precisely:

> **Problem:** Let S be a species phylogeny with leaf labels $L_S = [n]$. Fix $k \geq 2$. Given m independent samples $M_k^{(1)}, \ldots, M_k^{(m)}$, each generated according to the MSCR-JC(k) process on S, recover the topology of S.

2.1.4 Estimating Sequence Distances

Let $\mathcal{G}$ be generated according to the MSCR process on S, and $\mathcal{M}$ the corresponding set of marginal gene trees. Given a marginal gene tree $T_g \in \mathcal{M}$, let $d_{XY}^{T_g}$ be the *evolutionary distance* between leaves X and Y on T_g, defined as the expected number of mutations per site along the unique path between X and Y. It follows from the assumptions about the

mutation process that $d_{XY}^{T_g} = 2\theta t$, where t is the time of the most recent common ancestor of X and Y on T_g. For example in Figure 2.2, $d_{AB}^{T_1} = 2\theta t_4$ and $d_{AB}^{T_2} = 2\theta t_2$. Define the *breakpoint-weighted uncorrected distance* by

$$\Delta_{XY} := \frac{3}{4} \sum_{T_g \in \mathcal{M}} w_g \left(1 - e^{-\frac{4}{3} d_{XY}^{T_g}}\right). \tag{2.2}$$

This formula, due to [84], generalizes the uncorrected Jukes-Cantor distance to the setting of intralocus recombination; if no intralocus recombination occurs, then the right-hand side has only a single summand and reduces to the inverse of the Jukes-Cantor distance correction formula for a single non-recombining locus.

Our first lemma shows that δ_{XY} can be approximated by $k\Delta_{XY}$ when k is large.

Lemma 1. *If M_k is generated according to the MSCR-JC(k) process on S then for all $X, Y \in L_S$, conditioned on $\mathcal{G}$, $\delta_{XY}(M_k) = k\Delta_{XY} + o(k)$ almost surely as $k \to \infty$.*

In order to prove Lemma 1, we will make use of the following two auxiliary lemmas.

Lemma 2. *Suppose $\mathcal{G}$ is generated according to the MSCR process on S and $\mathcal{M}$ is its associated collection of marginal gene trees. With probability one, for all $T \in \mathcal{M}$, $I(T)$ is a finite union of nondegenerate subintervals of $[0,1]$ such that the endpoints of each of these intervals are elements of $B(\mathcal{G}) \cup \{0,1\}$ where $B(\mathcal{G})$ is the set of recombination breakpoints of $\mathcal{G}$.*

Proof. Let $\mathcal{H}$ denote the collection consisting of the empty set together with all nondegenerate subintervals of $[0,1]$ whose endpoints are elements of $B(\mathcal{G})$, and define

$$\Pi = \{S \subset [0,1] : S = I_1 \cup \cdots \cup I_j \text{ with } I_1, \ldots, I_j \in \mathcal{H} \text{ for some } j = 0, 1, \ldots\}.$$

For each $e \in E(\mathcal{G})$ define $\mathcal{P}_e = \{x \in [0,1] : e \in E(\mathcal{T}(x))\}$. As noted in [55], with probability one $\mathcal{P}_e$ is either empty or is a finite union of intervals such that the endpoints of these intervals are a subset of $B(\mathcal{G})$. Therefore $\mathcal{P}_e \in \Pi$ for all e. Let $T \in \mathcal{M}$ and denote the edge set of T by $E(T)$. Let $x \in [0,1]$. Then $\mathcal{T}(x) = T$ if and only if $e \in E(\mathcal{T}(x))$ for

all $e \in E(T)$. Therefore

$$\begin{aligned} I(T) &= \{x \in [0,1] : \mathcal{T}(x) = T\} \\ &= \{x \in [0,1] : E(T) \subset E(\mathcal{T}(x))\} \\ &= \bigcap_{e \in E(T)} \mathcal{P}_e \end{aligned}$$

Since $E(T)$ is finite, it will suffice to show that Π is closed under finite intersections, as the statement of the lemma will then follow from the inclusion $I(T) \in \Pi$. Suppose $I = I_1 \cup \cdots \cup I_p$ and $J = J_1 \cup \cdots \cup J_q$ with $I_i, J_j \in \mathcal{H}$ for all i, j. Then by the distributive property of unions and intersections,

$$I \cap J = \bigcup_{i=1}^{p} \bigcup_{j=1}^{q} I_i \cap J_j$$

and we note that each set $I_i \cap J_j$ is either empty or is an interval with endpoints in $B(\mathcal{G})$. Therefore $I_i \cap J_j \in \mathcal{H}$ and hence $I \cap J \in \Pi$. The conclusion that Π is closed under finite intersections follows by induction. □

Lemma 3. *With probability one, if $\mathcal{G}$ is generated according the MSCR process on S, then*

$$\frac{1}{k} \# \{i : i \in S_k \cap I(T_g)\} \to |I(T_g)|$$

as $k \to \infty$ for all $T_g \in \mathcal{M}$.

Proof. We first prove the following claim: If I is a subinterval of [0,1] with endpoints $a, b \in B(\mathcal{G})$, where $a \leq b$, then

$$\lim_{k \to \infty} \frac{1}{k} \# \{i : i \in S_k \cap I\} = b - a. \tag{2.3}$$

To prove this claim, first observe that

$$\frac{1}{k} \# \{i : i \in S_k \cap I\} = \frac{p_k - q_k}{k}$$

where $p_k = \sup\left\{\ell \geq 0 : \frac{\ell}{k} < b\right\}$ and $q_k = \sup\left\{\ell \geq 0 : \frac{\ell}{k} < a\right\}$. Moreover, by definition of supremum

$$\frac{p_k}{k} < b \leq \frac{p_k + 1}{k} \quad \text{and} \quad \frac{q_k}{k} < a \leq \frac{q_k + 1}{k}.$$

These inequalities imply that $\lim_{k\to\infty} \frac{p_k - q_k}{k} = b - a$. This proves the claim.

We now prove the statement of the lemma. Using the result of Lemma 2, with probability one there exists a pairwise disjoint sequence of intervals $I_1, \ldots, I_h \in \mathcal{H}$ such that $I(T_g) = \cup_{i=1}^{h} I_j$. Therefore

$$\{i : i \in S_k \cap I(T_g)\} = \bigcup_{j=1}^{h} \{i : i \in S_k \cap I_j\}$$

and hence

$$\frac{1}{k}\#\{i : i \in S_k \cap I(T_g)\} = \sum_{j=1}^{h} \frac{1}{k}\#\{i : i \in S_k \cap I_j\}.$$

By Equation (2.3), $\frac{1}{k}\#\{i : i \in S_k \cap I_j\} \to |I_j|$ for all j, and hence

$$\lim_{k\to\infty} \frac{1}{k}\#\{i : i \in S_k \cap I(T_g)\} = \sum_{i=1}^{h} |I_j| = |I(T_g)|.$$

□

Finally, we now use the previous two lemmas to prove Lemma 1.

Proof of Lemma 1. Let $k \geq 2$, $S_k = \{\frac{j}{k-1} : j = 0, 1, \ldots, k-1\}$, and $\xi_i = \mathbf{1}_{[\mathcal{N}(X,i) \neq \mathcal{N}(Y,i)]}$, $i \in S_k$. Then $\xi_i, i \in S_k$ are independent and $\delta_{XY}(M_k) = \sum_{i \in S_k} \xi_i$. For all $T_g \in \mathcal{M}$,

$$\mathbb{E}[\xi_i] = \mathbb{P}[\mathcal{N}(X,i) \neq \mathcal{N}(Y,i) | \mathcal{T}(i) = T_g] = \frac{3}{4}\left(1 - e^{-\frac{4}{3} d_{XY}^{T_g}}\right) \tag{2.4}$$

a.s. for all $i \in I(T_g)$ by definition of the Jukes-Cantor process [see, e.g., 85]. Since $\mathcal{M}$ is

a.s. finite,

$$\frac{1}{k}\delta_{XY}(M_k) = \frac{1}{k}\sum_{i\in S_k}\xi_i = \frac{1}{k}\sum_{T_g\in\mathcal{M}}\sum_{i\in S_k\cap I(T_g)}\xi_i$$
$$= \sum_{T_g\in\mathcal{M}:N_k(T_g)>0}\frac{N_k(T_g)}{k}\left(\frac{1}{N_k(T_g)}\sum_{i\in S_k\cap I(T_g)}\xi_i\right) \tag{2.5}$$

where $N_k(T_g) = \#\{i : i \in S_k \cap I(T_g)\}$. Since $I(T_g)$ is a.s. finite union of nondegenerate intervals with endpoints in S_k for all $T_g \in \mathcal{M}$, it follows that $\lim_{k\to\infty} N_k(T_g) = \infty$ and $\lim_{k\to\infty} N_k(T_g)/k = w_g$. (By Lemmas 2 and 3.) From these limits, together with (2.4) and the strong law of large numbers, the right-hand side of (2.5) converges to Δ_{XY} almost surely. □

2.2 Inconsistency of R^*

In this section, we state and prove our main result about R^* using sequence distances. Related results for quartets are in Section 2.3.

2.2.1 Statement and Overview

Our main claim is the following:

Theorem 1. *For k sufficiently large, R^* using sequence distances is not statistically consistent under the MSCR-JC(k) model. That is, there exists a species phylogeny S and a $k^0 \geq 1$ such that, for all $k \geq k^0$, the topology of the output of R^* using sequence distances does not converge in probability to the topology of the species tree.*

To prove Theorem 1, it suffices to consider a species tree S with $L_S = \{A, B, C\}$ and topology $AB|C$ [see, e.g., 79]. Denote edges of S, or *populations*, by the letters A, B, C, AB, and ABC as depicted in Figure 2.3 where A, B, C correspond to the leaf populations, AB is the parent edge of A and B, and ABC is edge extending above the root. The key idea is to allow recombination only in population A. In order to keep the analysis tractable,

the recombination rate and length of edge A are chosen so that with high probability the number of recombinations is 0 or 1, so that the number of lineages on the ARG exiting population A (backwards-in-time) is either one or two. By choosing the internal branch length τ_{AB} sufficiently small, ILS occurs along that edge with high probability, so that all coalescent events on the ancestral recombination graph occur in the root population ABC. In that case, as long as the mutation rate is not too large, we show that, on the event R_1C_0 (see Figure 2.3), taxa B and C are more likely to be inferred as more closely related than taxa A and B, so that R^* converges to the wrong topology $BC|A$ as the number m of samples grows.

The mutation rate θ is assumed to be the same in all populations. The vector of recombination rates $\bar{\rho}$ is defined by setting $\rho_A = \rho > 0$ and $\rho_X = 0$ for all $X \neq A$. Assume S to be ultrametric. The populations A and B have length $\tau_A = \tau_B > 0$, the internal population AB has length τ_{AB}, the age of the root t_{root} is given by $t_{\text{root}} = \tau_A + \tau_{AB} = \tau_C$. For now assume that $\tau_{AB} > 0$ and $\tau_A > 0$; their precise values will be determined later in the proof.

Let M_k be generated according to the MSCR-JC(k) process on S, and let $E_{XY|Z}$ be the event that the rooted triple inferred from M_k using (2.1) is $XY|Z$. The following lemma implies that to prove Theorem 1, it will suffice to prove

$$\mathbb{P}\left[E_{YZ|X}\right] > \mathbb{P}\left[E_{XY|Z}\right]. \tag{2.6}$$

The *consistency zone* for R^* with sequence distances under the MSCR-JC(k) model is the set of species phylogenies S such that the topology of the R^* consensus tree converges in probability to the topology of S as $m \to \infty$. We will use the following notation: $\mathcal{R}(S) = \{S|J : J \subseteq L_S, |J| = 3, \text{ and } S|J \text{ is binary}\}$ is the set of restricted rooted triples of S [see, e.g., 79].

Lemma 4. *A necessary and sufficient condition for S to lie in the consistency zone for*

R^* with sequence distances under the MSCR-JC(k) model is that for all $XY|Z \in \mathcal{R}(S)$,

$$\mathbb{P}\left[E_{XY|Z}\right] > \mathbb{P}\left[E_{XZ|Y}\right] \vee \mathbb{P}\left[E_{YZ|X}\right]. \tag{2.7}$$

Proof of Lemma 4. Suppose $M_k^{(1)}, \ldots, M_k^{(m)}$ are generated independently according the MSCR-JC(k) process on S. For each $\ell = 1, \ldots, m$ and each triplet $X, Y, Z \in L$, let $E_{XY|Z}^{(\ell)}$ to be the event that the rooted triple inferred from $M^{(\ell)}$ is $XY|Z$. Let $U_S^{(m)}$ be the event that the topology of S is successfully reconstructed from these m independent samples using R^* with sequence distances. It suffices to show $\lim_{m\to\infty} \mathbb{P}\left[U_S^{(m)}\right] = 1$ if and only if (2.7) holds. By definition of the R^* pipeline, $U_S^{(m)} = \bigcap_{XY|Z\in\mathcal{R}(S)} U_{XY|Z}^{(m)}$ where

$$U_{XY|Z}^{(m)} = \left[\sum_{\ell=1}^{m} \mathbf{1}_{E_{XY|Z}^{(\ell)}} > \max\left\{\sum_{\ell=1}^{m} \mathbf{1}_{E_{XZ|Y}^{(\ell)}}, \sum_{\ell=1}^{m} \mathbf{1}_{E_{YX|X}^{(\ell)}}\right\}\right]. \tag{2.8}$$

Since the samples $M_k^{(1)}, \ldots, M_k^{(m)}$ are i.i.d., the law of large numbers implies

$$\lim_{m\to\infty} \frac{1}{m} \sum_{\ell=1}^{m} \mathbf{1}_{E_{XY|Z}^{(\ell)}} = \mathbb{P}\left[E_{XY|Z}^{(1)}\right] \tag{2.9}$$

for all triplets $X, Y, Z \in L$. It follows from (2.8) and (2.9) that (2.7) holds if and only if $\lim_{m\to\infty} \mathbb{P}\left[U_{XY|Z}^{(m)}\right] = 1$ for all $XY|Z \in \mathcal{R}(S)$. Since $|\mathcal{R}(S)|$ is finite, the equivalence of (2.7) and $\lim_{m\to\infty} \mathbb{P}\left[U_S^{(m)}\right] = 1$ is established. □

By Lemma 1, with probability one, an ancestral recombination graph $\mathcal{G}$ generated according to the MSCR process has the property that sequences of increasing length k generated on it by the Jukes-Cantor process satisfy the almost sure limit $\frac{1}{k}\delta_{XY}(M_k) \to \Delta_{XY}$ as $k \to \infty$. Since almost sure convergence implies convergence in distribution, it holds that under the joint process which combines both genealogical and mutational processes, $\frac{1}{k}\delta_{XY}(M_k) \Rightarrow \Delta_{XY}$ as $k \to \infty$ for all $X, Y \in L_S$. Therefore, since the distribution function of Δ_{XY} is continuous, $\mathbb{P}[E_{XY|Z}] \to \mathbb{P}[E]$ and $\mathbb{P}[E_{YZ|X}] \to \mathbb{P}[F]$ as $k \to \infty$, where

$$E := [\Delta_{AB} < \Delta_{AC} \wedge \Delta_{BC}] \quad \text{and} \quad F := [\Delta_{BC} < \Delta_{AB} \wedge \Delta_{AC}].$$

Therefore inequality (2.6) will hold for sufficiently large k provided that

$$\mathbb{P}[F] > \mathbb{P}[E]. \tag{2.10}$$

To prove this, we will make use of the lemmas detailed in the next sub-section.

2.2.2 Key Lemmas

In what follows, set intersection is denoted with product notation (i.e., so that $XY = X \cap Y$ for events X, Y) and the important events to be considered are

$$\begin{aligned} R_i &= [\text{exactly } i \text{ recombinations occur in the time interval } (0, \tau_A)] \\ C_i &= [\text{exactly } i \text{ coalescences occur during the time interval } (0, t_{\text{root}})] \\ C_{0,X} &= [\text{no coalescence occurs in population } X]. \end{aligned}$$

Since recombination occurs only in population A, the number of recombination events is governed by the recombination rate ρ and the duration τ_A of population A. The following lemma shows that τ_A can be chosen sufficiently small that with high probability, zero or one recombination occurs.

Lemma 5 (Recombination Probabilities)**.** *For all* $\rho, \tau_A \geq 0$, $\mathbb{P}[R_0] = e^{-\rho\tau_A}$ *and* $\mathbb{P}[R_1] \geq \mathbb{P}[R_1 C_{0,A}] \geq \rho\tau_A e^{-(1+2\rho)\tau_A}$. *As* $\tau_A \to 0^+$, $\mathbb{P}[\cup_{k\geq 2} R_k] = O(\rho^2\tau_A^2)$.

Proof of Lemma 5. Since recombination occurs only in population A of S, the event R_i occurs if and only if exactly i recombinations occur in population A. Since the process starts with a single lineage in A at time t, the first event (if an event occurs in A) must be a recombination, which occurs with rate ρ. Therefore there is an exponentially distributed waiting time e_ρ with rate ρ such that a the lineage in A recombines at time e_ρ if $e_\rho < \tau_A$ and not otherwise. Therefore $\mathbb{P}[R_0] = \mathbb{P}[e_\rho > \tau_A] = e^{-\rho\tau_A}$.

Given $e_\rho < \tau_A$, population A has two lineages at time e_ρ, so the next event (coalescence or recombination) occurs at rate $1 + 2\rho$. Therefore there is an exponentially distributed

waiting time $e_{1+2\rho}$ with rate $1+2\rho$ and independent of e_ρ such that if $e_{1+2\rho} < \tau_A - e_\rho$ then an event occurs at time $e_\rho + e_{1+2\rho}$ and not otherwise. Therefore $R_1 C_{0,A} = [e_\rho < \tau_A, e_{1+2\rho} > \tau_A - e_\rho]$, and hence by the numerical inequality $e^x - 1 \geq x$,

$$\begin{aligned}
\mathbb{P}\left[R_1 C_{0,A}\right] &= \int_0^{\tau_A} \mathbb{P}\left[e_{1+2\rho} > \tau_A - s | e_\rho = s\right] \rho e^{-\rho s} ds \\
&= \int_0^{\tau_A} e^{(1+2\rho)(s-\tau_A)} \rho e^{-\rho s} ds \\
&= \frac{\rho}{1+\rho} e^{-(1+2\rho)\tau_A} \left(e^{\tau_A(1+\rho)} - 1\right) \\
&\geq \rho \tau_A e^{-(1+2\rho)\tau_A}.
\end{aligned}$$

To bound $\mathbb{P}\left[\cup_{i=2}^{\infty} R_i\right]$, we may restrict our analysis to the single-population case since the recombination rate is positive only in population A. Define

$$\begin{aligned}
\tau_1 &= \inf\left\{t \geq 0 : \text{ recombination occurs at time } t\right\} \quad \text{and} \\
\tau_2 &= \inf\left\{t > \tau_1 : \text{ recombination occurs at time } t\right\},
\end{aligned}$$

so that $\cup_{i=2}^{\infty} R_i \subset \left[\tau_2 < \tau_A\right]$. When $t = 0$, there is only one lineage in population A, and hence if an event occurs it must be a recombination event, which occurs at rate ρ. Therefore

$$\mathbb{P}\left[\tau_2 < \tau_A\right] = \int_0^{\tau_A} \mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z\right] \rho e^{-\rho z} dz.$$

Conditional on $\tau_1 < \tau_A$, let s_1 denote the time until the next event. There are two lineages at time τ_1, so that events happen at rate $1+2\rho$, and hence $s_1 \sim \exp(1+2\rho)$, with the event at $\tau_1 + s_1$ being either a recombination with probability $\frac{2\rho}{1+2\rho}$ or a coalescence with probability $\frac{1}{1+2\rho}$. Therefore

$$\mathbb{P}\left[\tau_2 < \tau_A\right] = \int_0^{\tau_A} \int_0^{\tau_A - z} \mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z, s_1 = y\right] (1+2\rho) e^{-(1+2\rho)y} dy \rho e^{-\rho z} dz. \quad (2.11)$$

Conditional on $s_1 + \tau_1 < \tau_A$, let H_{coal} be the event that a coalescence occurs at time $\tau_1 + s_1$ rather than a recombination. Assume that $z \in (0, \tau_A)$ and $y \in (0, \tau_A - z)$. Then by the

Law of Total Probability,

$$\mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z, s_1 = y\right] = \frac{1}{1+2\rho}\mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z, s_1 = y, H_{\text{coal}}\right] + \frac{2\rho}{1+2\rho} \cdot 1. \quad (2.12)$$

Moreover, conditional on $[\tau_1 = z, s_1 = y] \cap H_{\text{coal}}$, there is exactly one lineage at time $s_1 + \tau_1$ in population A, so the time s_2 until the next event is an independent exponential with rate ρ and the next event must be a recombination. Therefore

$$\begin{aligned}\mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z, s_1 = y, H_{\text{coal}}\right] &= \mathbb{P}\left[s_2 < \tau_A - y - z\right] \\ &\leq \mathbb{P}\left[s_2 < \tau_A\right] = 1 - e^{-\rho\tau_A} \\ &\leq \rho\tau_A\end{aligned}$$

By this inequality and (2.12), $\mathbb{P}\left[\tau_2 < \tau_A | \tau_1 = z, s_1 = y\right] \leq \frac{\rho}{1+2\rho}(\tau_A + 2)$. Therefore by (2.11),

$$\mathbb{P}\left[\tau_2 < \tau_A\right] \leq \int_0^{\tau_A}\int_0^{\tau_A} \rho^2(\tau_A + 2)dydz = \rho^2\tau_A^2(\tau_A + 2) = O(\rho^2\tau_A^2),$$

which establishes the final statement in the lemma. □

For the case where no recombination occurs, the probabilities of E and F are estimated in the following lemma using elementary MSC calculations.

Lemma 6 (No Recombination Case)**.** $\mathbb{P}[E|R_0] - \mathbb{P}[F|R_0] \leq \tau_{AB}$.

Proof of Lemma 6. Since the recombination rate is zero except in population A, it follows that conditional on R_0, the ARG is a tree, so that the analysis is exactly the same as in the case of the multi-species coalescent. The lineages from A and B coalesce in population AB with probability $1 - e^{-\tau_{AB}}$. If that happens, the lineages from A and B coalesce at some time $t_c < t_{\text{root}}$ so that

$$\Delta_{AB} = \frac{3}{4}\left(1 - e^{-\frac{8}{3}\theta t_c}\right) < \frac{3}{4}\left(1 - e^{-\frac{8}{3}\theta t_{\text{root}}}\right) < \Delta_{AC} \wedge \Delta_{BC}$$

hence E occurs with probability one on that event. On the other hand, if the lineages A and

B do *not* coalesce in population AB, then by a symmetry argument (each pair of lineages being equally likely to coalesce in population ABC), the random variables Δ_{AB}, Δ_{AC}, and Δ_{BC} are identically distributed, so that $\mathbb{P}[E|R_0C_0] = \mathbb{P}[F|R_0C_0] = 1/3$. It follows by the law of total probability that $\mathbb{P}[E|R_0] - \mathbb{P}[F|R_0] = \mathbb{P}[C_1|R_0] = 1 - e^{-\tau_{AB}} \leq \tau_{AB}$. □

For the case in which *exactly one* recombination occurs, the following lemma characterizes the behavior of coalescent events occurring below the root of S. Intuitively, it says that coalescence in population AB is rare when τ_{AB} is small.

Lemma 7 (Effect of Small Internal Edge)**.** *As* $\tau_{AB} \to 0^+$, $\mathbb{P}[C_0|R_1] = K + O(\tau_{AB})$, $\mathbb{P}[C_{0,A}|R_1C_1] = O(\tau_{AB})$, *and* $\mathbb{P}[C_2|R_1] = O(\tau_{AB})$, *where* $K = \mathbb{P}[C_{0,A}|R_1] \in (0,1)$ *depends only on* τ_A *and* ρ, *and satisfies* $\lim_{\tau_A \to 0} K = 1$ *for any fixed* $\rho > 0$.

Proof of Lemma 7. Since $K \in (0,1)$ depends only on the portion of the ARG in population A, it is clear that K does not depend on τ_{AB}. By Lemma 5, $\mathbb{P}[R_1C_{0,A}] \geq \rho\tau_A e^{-\tau_A(1+2\rho)}$ and $\mathbb{P}[R_1] \leq 1 - \mathbb{P}[R_0] = 1 - e^{-\rho\tau_A} \leq \rho\tau_A$. Therefore,

$$K = \mathbb{P}[C_{0,A}|R_1] = \frac{\mathbb{P}[R_1C_{0,A}]}{\mathbb{P}[R_1]} \geq e^{-\tau_A(1+2\rho)}.$$

Therefore $\lim_{\tau_A \to 0} K = 1$ for all $\rho > 0$ uniformly in τ_{AB}.

Let $e_2 \sim \exp(1)$ and $e_3 \sim \exp(3)$ be independent exponential clocks for the arrival of the next coalescence event given 2 and 3 lineages respectively. Then

$$\begin{aligned}
\mathbb{P}[C_0|R_1] &= \mathbb{P}[C_{0,A}C_{0,AB}|R_1] \\
&= K\mathbb{P}[C_{0,AB}|R_1C_{0,A}] \\
&= K\mathbb{P}[e_3 > \tau_{AB}] \\
&= Ke^{-3\tau_{AB}},
\end{aligned}$$

which implies

$$\mathbb{P}[C_0|R_1] = K + O(\tau_{AB}). \tag{2.13}$$

Next we claim that

$$\mathbb{P}\left[C_1|R_1C_{0,A}\right] = O(\tau_{AB}). \tag{2.14}$$

To prove this claim, observe that conditional on $R_1C_{0,A}$ three lineages enter population AB (i.e., two lineages from A and one from B) and one leaves (backwards in time). Thus, the event C_1 occurs if and only if exactly one of the three pairs of lineages entering population AB coalesces and no further coalescence events occur in population AB (i.e., during the time interval $(\tau_A, t_{\text{root}})$). Letting $\tilde{e}_3$ be the time until the first coalescence given three lineages, and $\tilde{e}_2$ be the (independent) clock for two lineages, it follows that C_1 occurs if and only if $\tilde{e}_3 < \tau_{AB}$ and $\tilde{e}_3 + \tilde{e}_2 > \tau_{AB}$. Therefore since $\tilde{e}_3 \sim \exp(3)$ and $\tilde{e}_2 \sim \exp(1)$,

$$\begin{aligned}
\mathbb{P}\left[C_1|R_1C_{0,A}\right] &= \mathbb{P}\left[\tilde{e}_3 < \tau_{AB}, \tilde{e}_2 + \tilde{e}_3 > \tau_{AB}\right] \\
&= 3\int_0^{\tau_{AB}} e^{-3x}\mathbb{P}\left[\tilde{e}_2 > \tau_{AB} - x\right] dx \\
&= 3\int_0^{\tau_{AB}} e^{-2x-\tau_{AB}} dx \\
&= \frac{3}{2}e^{-\tau_{AB}}\left(1 - e^{-2\tau_{AB}}\right).
\end{aligned}$$

This proves (2.14). On the other hand, conditional on $R_1C_{0,A}^c$, the event C_1 occurs if and only if the two lineages (A and B) entering population AB fail to coalesce during the time interval $(\tau_A, t_{\text{root}})$. Therefore

$$\mathbb{P}\left[C_1|R_1C_{0,A}^c\right] = \mathbb{P}\left[e_2 > \tau_{AB}\right] = e^{-\tau_{AB}}. \tag{2.15}$$

By the Law of Total Probability, $\mathbb{P}[C_1|R_1] = \mathbb{P}[C_1|R_1C_{0,A}]K + \mathbb{P}[C_1|R_1C_{0,A}^c](1-K)$, and hence by (2.14) and (2.15),

$$\mathbb{P}\left[C_1|R_1\right] = 1 - K + O(\tau_{AB}). \tag{2.16}$$

Conditional on R_1, exactly one of the events C_0, C_1, and C_2 occurs. Therefore $\mathbb{P}\left[C_2|R_1\right] + \mathbb{P}\left[C_0|R_1\right] + \mathbb{P}\left[C_1|R_1\right] = 1$. Therefore by (2.13) and (2.16), $P\left[C_2|R_1\right] = O(\tau_{AB})$. Finally,

applying Bayes' rule along with the estimates from (2.14), (2.16), one obtains

$$\mathbb{P}[C_{0,A}|R_1C_1] = \frac{\mathbb{P}\left[C_{0,A}|R_1\right]\mathbb{P}\left[C_1|R_1C_{0,A}\right]}{\mathbb{P}\left[C_1|R_1\right]} = O(\tau_{AB}),$$

uniformly in ρ, τ_A as $\tau_{AB} \to 0^+$. □

Next we apply Lemma 7 to show that $\mathbb{P}[E|R_1C_1] - \mathbb{P}[F|R_1C_1]$ is small, tending to zero as $\tau_{AB} \to 0^+$.

Lemma 8. $\mathbb{P}\left[E|R_1C_1\right] - \mathbb{P}\left[F|R_1C_1\right] = O(\tau_{AB})$ *as* $\tau_{AB} \to 0^+$,

Proof of Lemma 8. By a symmetry argument similar to that given in the proof of Lemma 6, it follows that conditional on $R_1C_1C_{0,A}^c$, the events E and F are equally likely. Therefore by the Law of Total Probability,

$$\begin{aligned}\mathbb{P}[E|R_1C_1] - \mathbb{P}[F|R_1C_1] &= \left(\mathbb{P}[E|R_1C_1C_{0,A}] - \mathbb{P}[F|R_1C_1C_{0,A}]\right)\mathbb{P}[C_{0,A}|R_1C_1] \\ &\leq \mathbb{P}[C_{0,A}|R_1C_1],\end{aligned}$$

and the right-hand side is $O(\tau_{AB})$ by Lemma 7. □

We now come to a key part of the calculation: the event R_1C_0, depicted in Figure 2.3. The next lemma demonstrates that as long as θ is not too large, conditional on R_1C_0, the event F is more likely than E.

Lemma 9. *The quantity* $\bar{\alpha} := \mathbb{P}\left[F|R_1C_0\right] - \mathbb{P}\left[E|R_1C_0\right]$ *depends only on* θ *and is positive if* $\theta \in (0, 3/4)$.

Proof of Lemma 9. Conditional on R_1C_0, four distinct lineages enter population ABC at time t_{root}. Denote these lineages by A_1, A_2, B, and C, with B and C with the letter corresponding the originating population, as shown in Figure 2.3. Since no recombination occurs in population ABC, the order in which the lineages coalesce determines a *labeled history* (defined as an ultrametric rooted binary tree with labeled tips and internal nodes rank-ordered according to age [24]), whose tips are taken to be the lineages A_1, A_2, B and

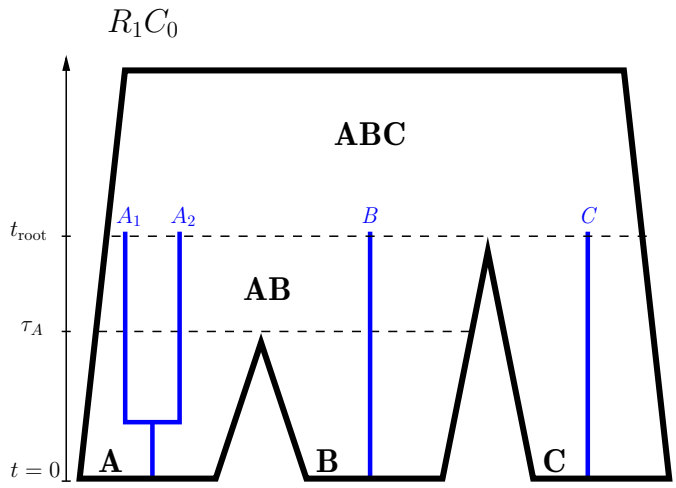


Figure 2.3: A depiction of the event R_1C_0. The portion of the ancestral recombination graph more ancient than t_{root} is not shown. Intuitively, since there are *two* lineages from A and only one from each of B and C, at least one of the A lineages is more likely to be included in the final coalescing pair, favoring greater pairwise distances between A and the other two taxa than those between B and C.

C at time t_{root}. Denote by $((WX)Y)Z$ the event that the labeled history exhibited in population ABC has a caterpillar topology, W and X are the first pair of lineages to coalesce, and the most recent common ancestor of W and Z coalesces with the lineage ancestral to Y before coalescing with that of Z. Denote by $(WX)YZ$ the event that the labeled history has balanced topology with lineages W, X coalescing first.

Let t_k be the duration of the epoch in which population ABC has k distinct lineages. Then t_k is an independent exponential random variable with parameter $k(k-1)/2$. Conditional on R_1C_0, the set $\mathcal{M}$ has two distinct elements, T_1 and T_2, where T_i is the triplet with leaves A_i, B, and C for $i = 1, 2$. Hence $w_i := |I(T_i)|$ is uniformly distributed on $[0, 1]$ for $i = 1, 2$ and $w_1 + w_2 = 1$. Since recombination is assumed only to occur in population A, it follows that $d_{BC}^{T_1} = d_{BC}^{T_2}$, and we shall denote this quantity by d_{BC}.

There are 18 labeled histories $\gamma_1, \ldots, \gamma_{18}$ in which the four labeled lineages A_1, A_2, B, and C may coalesce. Since pairs of lineages are chosen to coalesce uniformly at random,

$\mathbb{P}[\gamma_j|R_1C_0] = \frac{1}{18}$ for all j, and hence

$$\mathbb{P}[F|R_1C_0] - \mathbb{P}[E|R_1C_0] = \frac{1}{18}\sum_{j=1}^{18}\left(\mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j]\right). \tag{2.17}$$

The task will therefore be to calculate $\mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j]$ for $j = 1, \ldots, 18$, which as we shall see, we allow us to conclude that the right hand side of (2.17) is positive provided that not too much signal is lost by a high mutation rate. In particular, to perform these calculations, it will be convenient to group together those labeled histories involving similar calculations into five claims.

Claim 1. *Let* $\gamma_1 = ((A_1A_2)B)C$, $\gamma_2 = ((A_1B)A_2)C$, *and* $\gamma_3 = ((A_2B)A_1)C$. *Then*

$$\mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j] = -1 \quad \textit{for } j = 1, 2, 3.$$

Proof of Claim 1. We prove the case for $j = 1$ as the other two cases are similar. The labeled history corresponding to γ_1 is depicted in Figure 2.4. Since C is the last lineage to coalesce, $d_{AB}^{T_i} < 2\theta(t_{\text{root}} + t_4 + t_3 + t_2) = d_{AC}^{T_i} = d_{BC}^{T_i}$ for $i = 1, 2$. Letting $\phi(x) := \frac{3}{4}\left(1 - e^{-\frac{4}{3}x}\right)$, it follows that

$$\begin{aligned}
\Delta_{AB} &= w\phi\left(d_{AB}^{T_1}\right) + (1-w)\phi\left(d_{AB}^{T_2}\right) \\
\Delta_{AC} &= w\phi\left(d_{AC}^{T_1}\right) + (1-w)\phi\left(d_{AC}^{T_2}\right) \\
\Delta_{BC} &= \phi\left(d_{BC}\right).
\end{aligned} \tag{2.18}$$

Therefore, since ϕ is increasing, $\Delta_{AB} < \Delta_{AC} \wedge \Delta_{BC}$ a.s., hence $\mathbb{P}[E|R_1C_0\gamma_1] = 1$. This completes the proof of Claim 1. □

Claim 2. *Let* $\gamma_4 = ((A_1A_2)C)B$, $\gamma_5 = ((A_1C)A_2)B$, *and* $\gamma_6 = ((A_2C)A_1)B$. *Then*

$$\mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j] = 0 \quad \textit{for } j = 4, 5, 6.$$

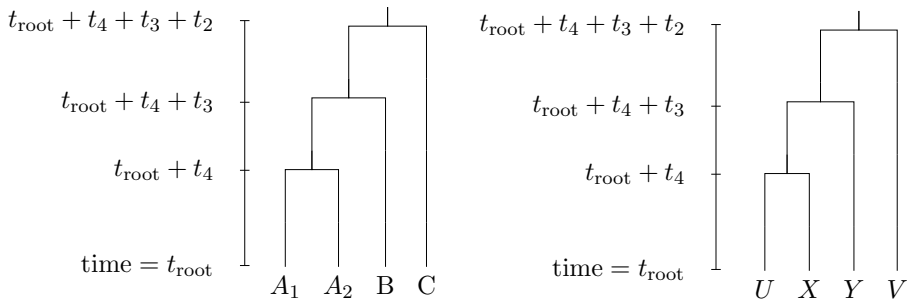


Figure 2.4: The labeled histories corresponding to Claims 1 (left) and 4 (right).

Proof of Claim 2. The proof is similar to that of Claim 1. Since B is the last lineage to coalesce, $d_{AC}^{T_i} < 2\theta(t_{\text{root}} + t_4 + t_3 + t_2) = d_{AB}^{T_i} = d_{BC}^{T_i}$ for $i = 1, 2$. Therefore $\Delta_{AC} < \Delta_{AB} \wedge \Delta_{BC}$ almost surely, so that $\mathbb{P}[E|R_1C_0\gamma_j] = \mathbb{P}[F|R_1C_0\gamma_j] = 0$. This proves Claim 2. □

Claim 3. *Let* $\gamma_7 = (A_1A_2)BC$, $\gamma_8 = (BC)A_1A_2$, $\gamma_9 = ((BC)A_1)A_2$, *and* $\gamma_{10} = ((BC)A_2)A_1$. *Then*

$$\mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j] = 1 \quad \textit{for } j = 7, 8, 9, 10.$$

Proof of Claim 3. It suffices to show $\Delta_{BC} < \Delta_{AB} \wedge \Delta_{AC}$ a.s., conditional on $R_1C_0\gamma_j$. Therefore by (2.18) it suffices to show that $d_{BC} < d_{AB}^{T_i} \wedge d_{AC}^{T_i}$ almost surely for $i = 1, 2$. Indeed, if $j = 7, 8$ then $d_{BC} \leq 2\theta(t_{\text{root}} + t_4 + t_3) < 2\theta(t_{\text{root}} + t_4 + t_3 + t_2) = d_{AB}^{T_i} \wedge d_{AC}^{T_i}$, and if $j = 9, 10$ then $d_{BC} = 2\theta(t_{\text{root}} + t_4) < 2\theta(t_{\text{root}} + t_4 + t_2) = d_{AB}^{T_i} \wedge d_{AC}^{T_i}$. This proves Claim 3. □

Claim 4. *Let* $\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{14}$ *be* $((A_1C)B)A_2, ((A_2C)B)A_1, ((A_1B)C)A_2$ *and* $((A_2B)C)A_1$ *respectively. Then*

$$\sum_{j=11}^{14} \mathbb{P}[F|R_1C_0\gamma_j] - \mathbb{P}[E|R_1C_0\gamma_j] = 6\alpha - 2$$

where

$$\alpha = 3\int_0^\infty \int_0^\infty \frac{1 - e^{-\frac{8}{3}\theta x}}{e^{\frac{8}{3}\theta y} - e^{-\frac{8}{3}\theta x}} e^{-3y-x} dx dy. \tag{2.19}$$

Proof of Claim 4. Consider the general case $((UX)Y)V$ where $\{U, V\} = \{A_1, A_2\}$ and $\{X, Y\} = \{B, C\}$. This corresponds to the labeled history depicted in Figure 2.4.

Let $u = \begin{cases} w_1 & : \quad U = A_1 \\ 1 - w_1 & : \quad U = A_2 \end{cases}$, it follows that

$$\begin{aligned} \Delta_{AX} &= u\phi(2\theta(t_{\text{root}} + t_4)) + (1-u)\phi(2\theta(t_{\text{root}} + t_4 + t_3 + t_2)) \\ \Delta_{AY} &= u\phi(2\theta(t_{\text{root}} + t_4 + t_3)) + (1-u)\phi(2\theta(t_{\text{root}} + t_4 + t_3 + t_2)) \\ \Delta_{BC} &= \Delta_{XY} = \phi(2\theta(t_{\text{root}} + t_4 + t_3)). \end{aligned} \tag{2.20}$$

Since ϕ is increasing, $\Delta_{BC} < \Delta_{AY}$ almost surely, and hence

$$\mathbb{P}\left[E_{AY|X} | R_1 C_0 \gamma_j\right] = 0. \tag{2.21}$$

Therefore there exists some real number α with $0 \leq \alpha \leq 1$ such that

$$\mathbb{P}\left[E_{BC|A} | R_1 C_0 \gamma_j\right] = \alpha \quad \text{and} \quad \mathbb{P}\left[E_{AX|Y} | R_1 C_0 \gamma_j\right] = 1 - \alpha. \tag{2.22}$$

Using (2.20) and (2.22),

$$\begin{aligned} \alpha &= \mathbb{P}\left[\Delta_{BC} < \Delta_{AX} \mid R_1 C_0 \gamma_j\right] \\ &= \mathbb{P}\Big[1 - e^{-\frac{8}{3}\theta(t_{\text{root}} + t_4 + t_3)} < u(1 - e^{-\frac{8}{3}\theta(t_{\text{root}} + t_4)}) \\ &\qquad + (1-u)(1 - e^{-\frac{8}{3}\theta(t_{\text{root}} + t_4 + t_3 + t_2)}) \mid R_1 C_0 \gamma_j\Big] \\ &= \mathbb{P}\left[1 > u e^{\frac{8}{3}\theta t_3} + (1-u) e^{-\frac{8}{3}\theta t_2} \mid R_1 C_0 \gamma_j\right]. \end{aligned}$$

Conditional on $R_1 C_0$, the random variables t_2, t_3 are exponentially distributed with rates 1 and 3 respectively and are independent of γ_j and each other. Conditional on $R_1 C_0 \gamma_j$, u

is uniformly distributed on $(0,1)$ for any j. Therefore

$$\begin{aligned}\alpha &= \int_0^1 \int_0^\infty \int_0^\infty \mathbf{1}\Big[1 > w_1 e^{\frac{8}{3}\theta y} + (1-w_1)e^{-\frac{8}{3}\theta x}\Big] e^{-x}dx\ 3e^{-3y}dy\ dw_1 \\ &= 3\int_0^\infty \int_0^\infty \int_0^1 \mathbf{1}\left[w_1 < \frac{1-e^{-\frac{8}{3}\theta x}}{e^{\frac{8}{3}\theta y} - e^{-\frac{8}{3}\theta x}}\right] dw_1\ e^{-3y-x}dx\ dy \\ &= 3\int_0^\infty \int_0^\infty \frac{1-e^{-\frac{8}{3}\theta x}}{e^{\frac{8}{3}\theta y} - e^{-\frac{8}{3}\theta x}} e^{-3y-x}\ dx\ dy.\end{aligned}$$

This shows that the value of α does not depend on j. Therefore by (2.21) and (2.22),

$$\mathbb{P}\left[E|R_1C_0\gamma_j\right] = \begin{cases} 1-\alpha & \text{if } j = 11,12 \\ 0 & \text{if } j = 13,14 \end{cases} \tag{2.23}$$

and by (2.22),

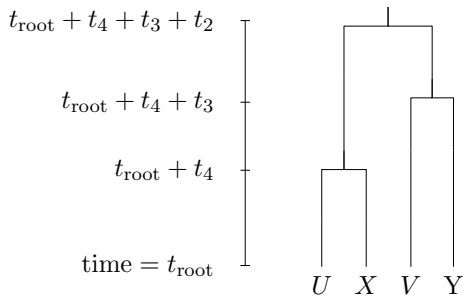
$$\mathbb{P}\left[F|R_1C_0\gamma_j\right] = \alpha \quad \text{for } j = 11,12,13,14. \tag{2.24}$$

The statement of the claim then follows from (2.23) and (2.24). This completes the proof of Claim 4. □

Claim 5. *Let* $\gamma_{15}, \gamma_{16}, \gamma_{17}$ *and* γ_{18} *be* $(A_1B)A_2C, (A_2B)A_1C, (A_1C)A_2B$, *and* $(A_2C)A_1B$ *respectively. Then*

$$\sum_{j=15}^{18} \mathbb{P}\left[F|R_1C_0\gamma_j\right] - \mathbb{P}\left[E|R_1C_0\gamma_j\right] = -2. \tag{2.25}$$

Proof of Claim 5. We consider the general case $(UX)VY$, where $\{U,V\} = \{A_1, A_2\}$ and $\{X,Y\} = \{B,C\}$. This corresponds to the following labeled history:



Letting $u = \begin{cases} w_1 & : \ U = A_1 \\ 1 - w_1 & : \ U = A_2 \end{cases}$, it follows that

$$\begin{aligned} \Delta_{AX} &= u\phi(2\theta(t_{\text{root}} + t_4)) + (1-u)\phi(2\theta(t_{\text{root}} + t_4 + t_3 + t_2)) \\ \Delta_{AY} &= u\phi(2\theta(t_{\text{root}} + t_4 + t_3 + t_2)) + (1-u)\phi(2\theta(t_{\text{root}} + t_4 + t_3)) \\ \Delta_{BC} &= \Delta_{XY} = \phi(2\theta(t_{\text{root}} + t_4 + t_3 + t_2)). \end{aligned} \tag{2.26}$$

Since $\Delta_{BC} > \Delta_{AX}$, it follows that

$$\mathbb{P}\left[F | R_1 C_0 \gamma_j\right] = \mathbb{P}\left[E_{BC|A} | R_1 C_0 \gamma_j\right] = 0 \tag{2.27}$$

whenever $15 \leq j \leq 18$. Therefore

$$\mathbb{P}\left[E_{AX|Y} | R_1 C_0 \gamma_j\right] = \beta \quad \text{and} \quad \mathbb{P}\left[E_{AY|X} | R_1 C_0 \gamma_j\right] = 1 - \beta. \tag{2.28}$$

where $\beta = \mathbb{P}\left[d_{AX}^{(\ell)} < d_{AY}^{(\ell)} | R_1 C_0 \gamma_j\right]$. By (2.26),

$$\beta = \mathbb{P}\left[u + (1-2u)e^{-\frac{8}{3}\theta(t_3+t_2)} - (1-u)e^{-\frac{8}{3}\theta t_3} > 0 | R_1 C_0 \gamma_j\right].$$

Moreover, similarly as in Claim 4, conditional on $R_1 C_0$, the random variables t_2, t_3 are exponentially distributed with rates 1 and 3 respectively and are independent of γ_j and each other, and conditional on $R_1 C_0 \gamma_j$, u is uniformly distributed on $(0,1)$ for any j. Therefore

$$\beta = \mathbb{P}\left[\hat{u} + (1-2\hat{u})e^{-\frac{8}{3}\theta(\hat{t}_3+\hat{t}_2)} - (1-u)e^{-\frac{8}{3}\theta\hat{t}_3} > 0\right]$$

where $\hat{u} \sim$unif$(0,1)$, $\hat{t}_2 \sim$exp(1) and $\hat{t}_3 \sim$exp(3), with $\hat{u}, \hat{t}_2, \hat{t}_3$ independent. In particular, this implies β does not depend on j. Moreover, equation (2.28) implies $\mathbb{P}\left[E | R_1 C_0 \gamma_{15}\right] = \mathbb{P}\left[E | R_1 C_0 \gamma_{16}\right] = \beta$ and $\mathbb{P}\left[E | R_1 C_0 \gamma_{17}\right] = \mathbb{P}\left[E | R_1 C_0 \gamma_{18}\right] = 1-\beta$. Therefore $\sum_{j=15}^{18} \mathbb{P}\left[E | R_1 C_0 \gamma_j\right] = 2$. Combining this with (2.27) implies equation (2.25). This completes the proof of Claim 5. $\square$

We now return to the proof of Lemma 9. Combining the results from Claims 1-5 with (2.17) yields

$$\mathbb{P}\left[F|R_1C_0\right] - \mathbb{P}\left[E|R_1C_0\right] = \frac{1}{18}\left(6\alpha - 3\right) = \frac{\alpha}{3} - \frac{1}{6}$$

and substituting the formula for α from (2.19) gives

$$\mathbb{P}\left[F|R_1C_0\right] - \mathbb{P}\left[E|R_1C_0\right] = \int_0^\infty \int_0^\infty \frac{1 - e^{-8\theta x/3}}{e^{8\theta y/3} - e^{-8\theta x/3}} e^{-3y-x} dxdy - \frac{1}{6}. \tag{2.29}$$

It will suffice to show that the right hand side of (2.29) is a decreasing function of θ and equals zero when $\theta = 3/4$, as this will imply that it is positive for all $0 < \theta < 3/4$. By the substitution $c = 8\theta/3$, it suffices to show that the function

$$\Psi(c) := \int_0^\infty \int_0^\infty \frac{1 - e^{-cx}}{e^{cy} - e^{-cx}} e^{-3y-x} dxdy$$

satisfies $\Psi(2) = 1/6$ and is strictly decreasing. Numerical plots of Ψ are given in Figure 2.5. Making the substitution $u = e^{-(x+y)}$, $du = -e^{-(x+y)}dx$,

$$\Psi(2) = \int_0^\infty \int_0^\infty \frac{1 - e^{-2x}}{e^{2y} - e^{-2x}} e^{-3y-x} dxdy = \int_0^\infty \int_0^{e^{-y}} \frac{e^{-4y}}{1-u^2} - \frac{u^2 e^{-2y}}{1-u^2} dudy,$$

and by the Fubini-Tonelli theorem,

$$\begin{aligned}
\Psi(2) &= \int_0^1 \frac{1}{1-u^2} \int_0^{-\log u} e^{-4y} dy - \frac{u^2}{1-u^2} \int_0^{-\log u} e^{-2y} dydu \\
&= \int_0^1 \frac{1}{1-u^2} \left[\frac{1}{4}\left(1-u^4\right)\right] - \frac{u^2}{1-u^2} \left[\frac{1}{2}\left(1-u^2\right)\right] dudy \\
&= \int_0^1 \frac{1}{4}\left(1+u^2\right) - \frac{1}{2}u^2 du = \frac{1}{6}.
\end{aligned}$$

It remains to show that $\Psi(c) > \Psi(2)$ for all $0 < c < 2$. Fix $x, y > 0$ and let $\psi(c) =$

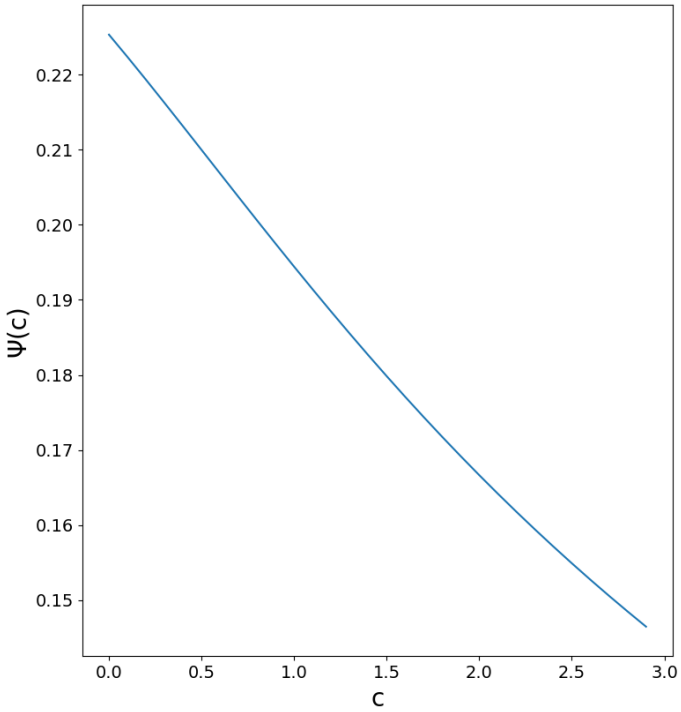


Figure 2.5: Numerical plot of the function $\Psi(c)$ with $0 < c < 3$. Note that Ψ is decreasing and $\Psi(2) = 1/6$.

$\frac{1-e^{-cx}}{e^{cy}-e^{-cx}}e^{-3y-x}$. Differentiating ψ gives

$$\begin{aligned}\psi'(c) &= \frac{(x+y)e^{c(y-x)} - ye^{cy} - xe^{-cx}}{\left(e^{cy}-e^{-cx}\right)^2} \\ &= \frac{e^{c(y-x)}}{\left(e^{cy}-e^{-cx}\right)^2}\left[y(1-e^{cx}) + x(1-e^{-cy})\right].\end{aligned}$$

Since $1 - e^a < -a$ for al $a \neq 0$, $y(1-e^{cx}) + x(1-e^{-cy}) < y(-cx) + x(cy) = 0$, it follows that $\psi'(c) < 0$ for all $c > 0$ and hence Ψ is strictly decreasing on $(0, \infty)$, as required. This completes the proof of Lemma 9. □

The next lemma applies Lemmas 7, 8, and 9 to show that $P[F|R_1] > \mathbb{P}[E|R_1]$ when the internal branch length τ_{AB} is small and the mutation rate θ is not too large.

Lemma 10. *If $\theta \in (0, 3/4)$, then $\mathbb{P}[E|R_1] - \mathbb{P}[F|R_1] = -\bar{\alpha}K + O(\tau_{AB})$ as $\tau_{AB} \to 0^+$ (where the term $-\bar{\alpha}K$ does not depend on τ_{AB}).*

Proof of Lemma 10. By Law of Total Probability, since R_1C_0, R_1C_1, and R_1C_2 partition

R_1,

$$\mathbb{P}[E|R_1] - \mathbb{P}[F|R_1] = \sum_{i=0}^{2} \left(\mathbb{P}[E|R_1C_i] - \mathbb{P}[F|R_1C_i]\right) \mathbb{P}[C_i|R_1]. \tag{2.30}$$

By Lemma 9, $\bar{\alpha} = \mathbb{P}[F|R_1C_0] - \mathbb{P}[E|R_1C_0]$ is positive and does not depend on τ_{AB}. Moreover, by Lemma 7, $K = \mathbb{P}[C_{0,A}|R_1]$ does not depend on τ_{AB} and $\mathbb{P}[C_0|R_1] = K + O(\tau_{AB})$, hence

$$\left(\mathbb{P}[E|R_1C_0] - \mathbb{P}[F|R_1C_0]\right) \mathbb{P}[C_0|R_1] = -\bar{\alpha}K + O(\tau_{AB}).$$

It remains to show that the other two summands on the right hand side of (2.30) are of magnitude at most $O(\tau_{AB})$. Indeed, the $i = 1$ term is at most $O(\tau_{AB})$ by Lemma 8. Moreover, by Lemma 7, $\mathbb{P}[C_2|R_1] = O(\tau_{AB})$, so the $i = 2$ term is at most $O(\tau_{AB})$ as well. □

We are now ready to prove Theorem 1.

2.2.3 Proof of Theorem 1

Proof of Theorem 1. It suffices to prove (2.10) for some choice of parameters ρ, θ, τ_A, and τ_{AB}. Let $\rho > 0$ and $\theta \in (0, 3/4)$ be arbitrary; we will show that τ_A, and τ_{AB} can be chosen sufficiently small that (2.10) holds. Conditioning on the number of recombination events in population A,

$$\begin{aligned}&\mathbb{P}[F] - \mathbb{P}[E] > \\ &\quad \left(\mathbb{P}[F|R_0] - \mathbb{P}[E|R_0]\right) \mathbb{P}[R_0] + \left(\mathbb{P}[F|R_1] - \mathbb{P}[E|R_1]\right) \mathbb{P}[R_1] - \mathbb{P}\left[\cup_{k\geq 2} R_k\right].\end{aligned}$$

Therefore by Lemma 6 and the trivial inequality $\mathbb{P}[R_0] \leq 1$,

$$\mathbb{P}[F] - \mathbb{P}[E] > -\tau_{AB} + \left(\mathbb{P}[F|R_1] - \mathbb{P}[E|R_1]\right) \mathbb{P}[R_1] - \mathbb{P}\left[\cup_{k\geq 2} R_k\right].$$

By Lemma 10, there exists $\delta > 0$ such that $\mathbb{P}[F|R_1] - \mathbb{P}[E|R_1] > \bar{\alpha}K/2$ whenever $0 < \tau_{AB} < \delta$. Assume further that $\tau_{AB} \in (0, \delta)$. Then

$$\mathbb{P}[F] - \mathbb{P}[E] > -\tau_{AB} + \frac{\bar{\alpha}K}{2}\mathbb{P}[R_1] - \mathbb{P}[\cup_{k\geq 2} R_k].$$

By Lemma 5, there exists constants $C, D > 0$ not depending on τ_{AB} such that $\mathbb{P}[R_1] \geq C\rho\tau_A$ and $\mathbb{P}[\cup_{i\geq 2} R_i] \leq D\rho^2\tau_A^2$, so that

$$\mathbb{P}[F] - \mathbb{P}[E] > -\tau_{AB} + \left(\frac{1}{2}\bar{\alpha}KC - D\rho\tau_A\right)\rho\tau_A.$$

Since K does not depend on τ_{AB} and $K \to 1$ as $\tau_A \to 0$ by Lemma 7, there exists $\tau_A > 0$ sufficiently small that both $K > 1/2$ and $\epsilon := \bar{\alpha}C/4 - D\rho\tau_A > 0$. It follows that $\mathbb{P}[F] - \mathbb{P}[E] > -\tau_{AB} + \epsilon\rho\tau_A$. Since ϵ does not depend on τ_{AB}, it follows that $\mathbb{P}[F] - \mathbb{P}[E] > 0$ for τ_{AB} sufficiently small. □

2.3 Inconsistency of Unrooted Quartet Majority

The main result of Theorem 1, that there exists an inconsistency zone for majority-rule inference of rooted triples, extends to inference of unrooted quartets as well. In particular, we consider the standard *four-point method* [see, e.g., 56]. That is, a gene tree restricted to four leaves $W, X, Y, Z \in L$ is inferred to have unrooted quartet topology $WX|YZ$ if $\delta_{WX} + \delta_{YZ} < (\delta_{WY} + \delta_{XZ}) \wedge (\delta_{WZ} + \delta_{XY})$. The case we consider is that of an ultrametric species phylogeny S with leaves $L = \{A, B, C, D\}$ and rooted topology $(((AB)C)D)$. Then the unrooted quartet topology of S is $AB|CD$, but we will show that this need not be the most likely unrooted gene tree quartet topology to be inferred when the data M_k is generated according to the MSCR-JC(k) process on S for k sufficiently large.

Specifically, take S' to be a three-leaf species phylogeny as established in the proof of Theorem 1. We may append a distantly related fourth leaf D to an edge above the root as depicted in Figure 2.1b. Then the resulting four-leaf tree S has rooted topology $(((AB)C)D)$ and unrooted quartet topology $AB|CD$. However the residual bias in favor

of the $BC|A$ rooted triple rather than $AB|C$ in subtree S' causes the four-point method to infer the unrooted gene tree quartet topology $BC|AD$ more often than $AB|CD$. This idea motivates the following corollary.

Corollary 1. *For k sufficiently large, there exists a species phylogeny S with four leaves such that the most likely unrooted quartet topology to be inferred using the four-point method from an MSA generated according to the MSCR-JC(k) process on S is different from the unrooted quartet topology of S.*

Proof of Corollary 1. Let S be the four-leaf species tree described above, let M_k be generated according to the MSCR-JC(k) process on S, and let q_k be the unrooted quartet topology inferred from M_k by the four-point method. It suffices to show that for sufficiently large k,

$$\mathbb{P}\left[q_k = AB|CD\right] < \mathbb{P}\left[q_k = AD|BC\right]. \tag{2.31}$$

For each choice of leaves (distinct) leaves $W, X, Y, Z \in L$, define

$$\mathcal{E}_{WX|YZ} = \left[\Delta_{WX} + \Delta_{YZ} < (\Delta_{WY} + \Delta_{XZ}) \wedge (\Delta_{WZ} + \Delta_{XY})\right].$$

Then $\lim_{k\to\infty} \mathbb{P}\left[q_k = WX|YZ\right] = \mathbb{P}\left[\mathcal{E}_{WX|YZ}\right]$ by Lemma 1. Therefore it suffices to show that

$$\mathbb{P}\left[\mathcal{E}_{AB|CD}\right] < \mathbb{P}\left[\mathcal{E}_{AD|BC}\right],$$

since this implies 2.31 must hold for sufficiently large k. Let $\epsilon = \mathbb{P}\left[F\right] - \mathbb{P}\left[E\right]$, where

$$E = \left[\Delta_{AB} < \Delta_{AC} \wedge \Delta_{BC}\right] \quad \text{and} \quad F = \left[\Delta_{BC} < \Delta_{AB} \wedge \Delta_{AC}\right].$$

Since the restricted subtree $S|\{A, B, C\} = S'$ satisfies the assumptions of Theorem 1, it follows that for k sufficiently large there exists a choice of parameters ρ_X and $\theta = \theta_X$ for all $X \in \{A, B, C, AB, ABC\}$, along with τ_A, τ_{AB} such that $\epsilon > 0$ for all τ_{ABC} sufficiently large. See Figure 2.1b.

Further assign $\rho_{ABC} = \rho_D = 0$ and $\theta_{ABCD} = \theta$. Since S is ultrametric, $\tau_D = \tau_A +$

$\tau_{AB} + \tau_{ABC}$, which is the age of the root of S. The only parameter of S which remains to be chosen is τ_{ABC}. Let z be the age of the MRCA on the ancestral recombination graph of the lineages originating in populations A, B, and C. Since $\tau_D \to \infty$ as $\tau_{ABC} \to \infty$,

$$\lim_{\tau_{ABC}\to\infty} \mathbb{P}\left[X | z < \tau_D\right] = \lim_{\tau_{ABC}\to\infty} \frac{\mathbb{P}\left[X, z < \tau_D\right]}{\mathbb{P}\left[z < \tau_D\right]} = \mathbb{P}\left[X\right]$$

for $X \in \{E, F\}$. Therefore we may choose $\tau_{ABC} > 0$ sufficiently large that

$$\mathbb{P}\left[F | z < \tau_D\right] - \mathbb{P}\left[E | z < \tau_D\right] > \frac{\epsilon}{2} \quad \text{and} \quad \mathbb{P}\left[z < \tau_D\right] \geq \frac{2}{2+\epsilon}. \tag{2.32}$$

If $z < \tau_D$, then since S is ultrametric and $\theta_X = \theta$ for all X, it holds that for all $T_g \in \mathcal{M}$, $d_{AD}^{T_g} = d_{BD}^{T_g} = d_{CD}^{T_g} = 2h$ where h is the height of the ancestral recombination graph. Therefore $\Delta_{AD} = \Delta_{BD} = \Delta_{CD}$. Using this identity, it is easy to check that, conditional on $z < \tau_D$, the events E and $\mathcal{E}_{AB|CD}$ are equivalent, and the events F and $\mathcal{E}_{AC|BC}$ are equivalent. Therefore by the Law of Total Probability and (2.32),

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}_{AD|BC}\right] - \mathbb{P}\left[\mathcal{E}_{AB|CD}\right] &= \left(\mathbb{P}\left[\mathcal{E}_{AD|BC} | z < \tau_D\right] - \mathbb{P}\left[\mathcal{E}_{AB|CD} | z < \tau_D\right]\right) \mathbb{P}\left[z < \tau_D\right] \\
&\quad + \left(\mathbb{P}\left[\mathcal{E}_{AD|BC} | z \geq \tau_D\right] - \mathbb{P}\left[\mathcal{E}_{AB|CD} | z \geq \tau_D\right]\right) \mathbb{P}\left[z \geq \tau_D\right] \\
&\geq \left(\mathbb{P}\left[F | z < \tau_D\right] - \mathbb{P}\left[E | z < \tau_D\right]\right) \mathbb{P}\left[z < \tau_D\right] - \mathbb{P}\left[z \geq \tau_D\right] \\
&> \frac{\epsilon}{2}\left(\frac{2}{2+\epsilon}\right) - \left(1 - \frac{2}{2+\epsilon}\right) = 0.
\end{aligned}$$

□

2.4 Simulation Study

We performed a simulation study to characterize the inconsistency zones established in Theorem 1 and Corollary 1. Code, documentation, and reproducible scripts, as well as information about simulation run times and all of the simulated data can be found at https://github.com/max-hill/MSCR-simulator.git.

2.4.1 Triplet Simulations

In the simulations described in this section, sequence data was generated according to the MSCR-JC(k) process on an ultrametric species phylogeny S with three species A, B, C, and rooted topology $AB|C$. In all cases, $k = 500$, $\tau_A = 1$ and θ does not vary among populations. We use the notation $\hat{p}_{XY|Z}$ to denote the proportion of the m samples from which the rooted triple $XY|Z$ was inferred, and $\hat{t}$ to denote the R^* uniquely favored rooted triple of the m samples. By the strong law of large numbers, $\hat{p}_{XY|Z}$ serves as an estimate of $\mathbb{P}\left[E_{XY|Z}\right]$ for large m, where $E_{XY|Z}$ is defined as in Lemma 4.

The range of recombination rates considered in these simulations are comparable to those in [33], who suggest they encompass biologically plausible values. As for mutation rates, typical rates in eukaryotes are on the order of $\mu = 10^{-9}$ to 10^{-8} per site per generation [21, 86] and effective eukaryotic population sizes N_e range from 10^4 to 10^8 [87], making the values considered here of $\theta = 2N_e\mu \in \{0.01, 0.1\}$ plausible as well. Computational constraints limited the ability to consider mutation rates lower than these, as doing so would have necessitated an increase in k or m to compensate; however the analytic results here predict that the inconsistency zone will persist, and may grow, for smaller values of θ: the computed difference $\bar{\alpha} = \mathbb{P}[F|R_1C_0] - \mathbb{P}[E|R_1C_0]$ actually increases as $\theta \to 0$, suggesting that phylogenetic conflict may be greater under regimes with smaller mutation rates than those simulated here.

In the first experiment, we simulated the MSCR-JC(k) process under a variety of parameter regimes in order to characterize the anomaly zone and evaluate the robustness of triplet-based inference in the presence of intralocus recombination. In particular $m = 10^5$ replicates were generated independently under each parameter regime, with the aim of estimating how frequently the correct topology was inferred. The parameters used were $\theta = 0.01$, $\tau_{AB} \in \{0.01, 0.02, \ldots, 0.15\}$, $\rho_A \in \{0, \ldots, 20\}$, and $\rho_X =$ for all $X \neq A$, so that recombination occurred only in population A. Figure 2.6 shows the value of $\hat{t}$ for each simulated parameter regime, and Figure 2.7 plots the surface $z = \hat{p}_{AB|C} - \hat{p}_{BC|A}$ as a function of ρ_A and τ_{AB}, so that parameter regimes with negative z values indicates

inconsistent inference.

We also evaluated R^* inference with rooted triples inferred not by equation (2.1), but rather by maximum-likelihood under the (false) assumption of no intralocus recombination; in this mode, which we call $\boldsymbol{R^*}$ **with maximum likelihood**, binary sequences were simulated and the maximum likelihood rooted triple was computed analytically using the method in [88]. A plot almost identical to Figure 2.6 was obtained. For the very short internal branch length $\tau_{AB} = 0.01$, simulations were run with similar parameters and higher number of replicates ($m = 15,000$), with inference performed using both R^* with sequence distances and R^* with maximum likelihood. Figure 2.8 plots the difference $y = \hat{p}_{BC|A} - \hat{p}_{AB|C}$ as a function of ρ_A obtained from these simulations.

These results show that the combination of intralocus recombination in population A along with a very short internal branch length τ_{AB} resulted in the rooted triple $BC|A$ being more slightly likely to be inferred than the correct topology $AB|C$. Figure 2.8 shows clearly that this effect increases for larger values of ρ_A. Nonetheless, as both Figures 2.7 and 2.8 show, the magnitude of this effect is relatively small: even when $\hat{p}_{BC|A} - \hat{p}_{AB|C}$ is positive, it is never greater than 0.1. Moreover, as Figures 2.6 and 2.7 show, this effect disappears when τ_{AB} is increased (ILS being less likely to occur on longer edges of S). Notably, even for high rates of recombination, R^* under both sequence distance mode and maximum likelihood mode always correctly inferred the topology of S when $\tau_{AB} > 0.1$ coalescent units.

In our second experiment, we relaxed the assumption that recombination occurs only in population A by allowing for recombination in population B as well. For this simulation, $\tau_{AB} = 0.01$ and $\theta = 0.01$, with inference performed using R^* with sequence distances. Figure 2.9 shows the uniquely favored rooted triple for each choice of ρ_A and ρ_B, with each estimate obtained from $m = 10^5$ samples. When this experiment was repeated with $\tau_{AB} = 0.1$, all but one parameter regimes resulted in correct inference; the exception was when $\rho_A = 0$ and $\rho_B = 20$, in which case $\hat{t} = AC|B$. These results support the hypothesis that taxa exhibiting higher rates of recombination relative to other taxa are more likely

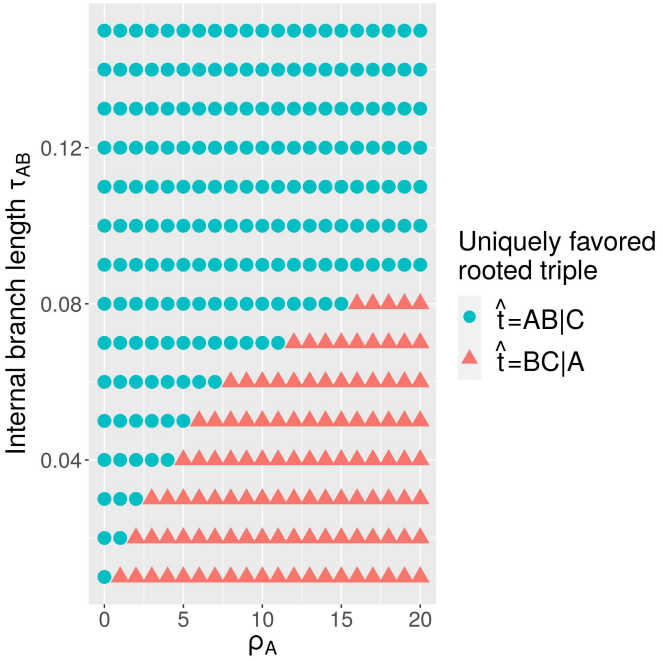


Figure 2.6: R^* inconsistency zone. Each point represents a simulation of $m = 10^5$ replicates.

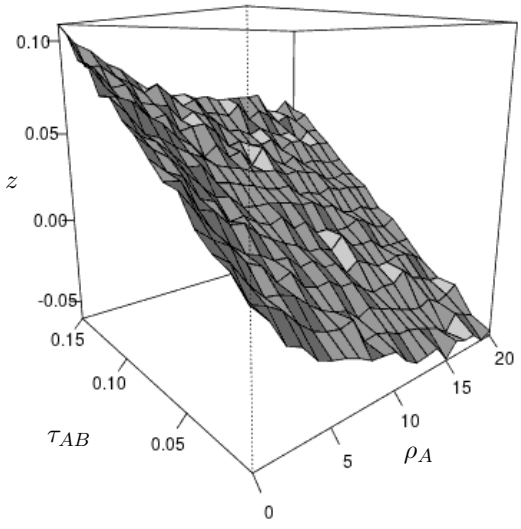


Figure 2.7: The surface $z = \hat{p}_{AB|C} - \hat{p}_{BC|A}$ as a function of τ_{AB} and ρ_A.

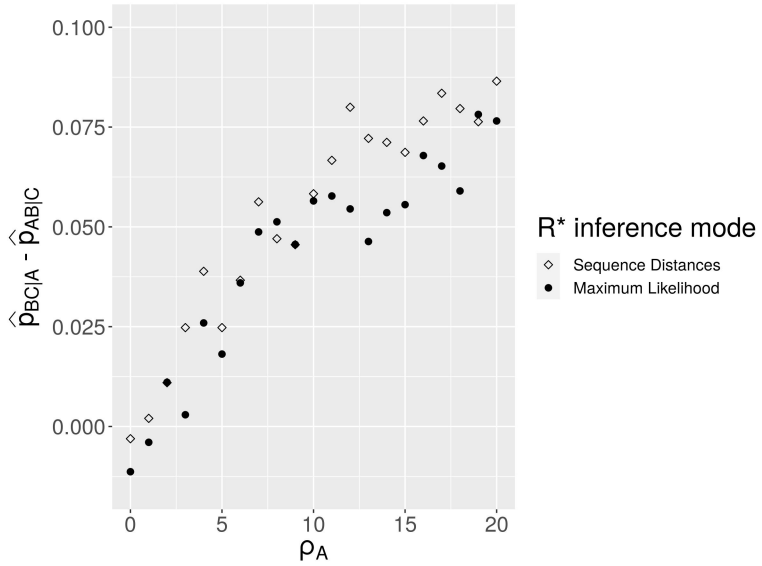


Figure 2.8: The effect of increasing ρ_A on inference using R^* with sequence distances and maximum likelihood.

to be inferred as more distantly related, but that the effect is small and manifests only in species triplets with very short internal branches.

The third experiment tested the effect when all populations in S (excluding the root population ABC) experience recombination at comparable rates. The simulation parameters were $\rho := \rho_A = \rho_B = \rho_C = \rho_{AB} \in \{0, 1, \ldots, 20\}$ and $\rho_{ABC} = 0$, along with $\theta = 0.1$, $\tau_{AB} = 0.01$, and $m = 10^6$, with inference performed using R^* with sequence distances. The results, shown in Figure 2.10, suggest that when recombination rates are similar on the edges of S, greater recombination rates does *not* lead to incorrect inference of rooted triples: in all cases, $\hat{p}_{AB|C} > \hat{p}_{AC|B} \vee \hat{p}_{BC|A}$, suggesting consistent inference despite the extremely short internal branch length, a result which agrees with the conclusions of [33] that even high recombination rates are not a significant source of error, at least when rates are comparable across species. Thus, the existence of differential rates of recombination between closely related taxa appears to be a necessary condition for a species tree S to lie in the inconsistency zone.

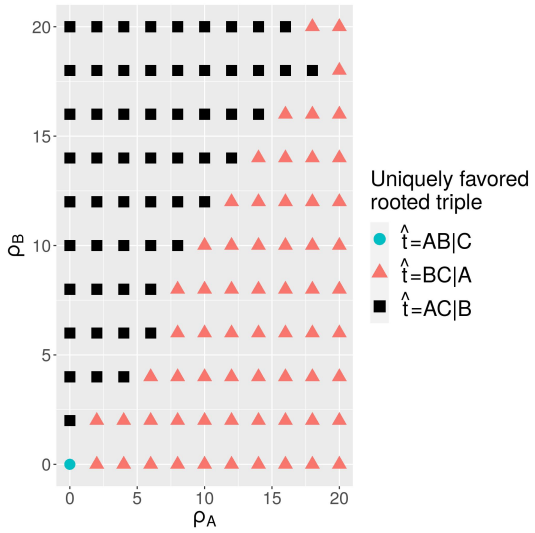


Figure 2.9: R^* inference with recombination in both populations A and B.

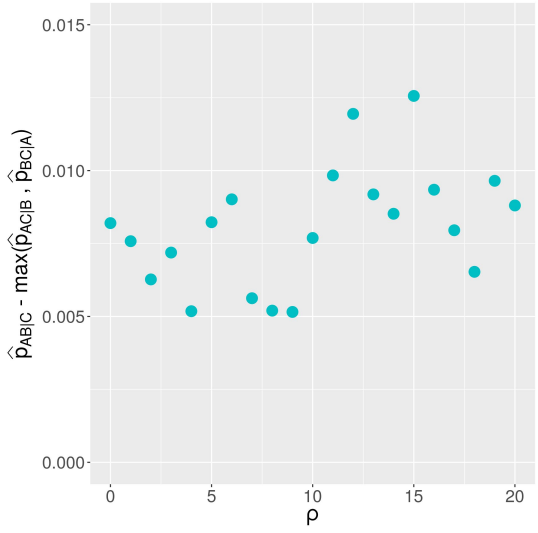


Figure 2.10: Equal recombination rates in populations A, B, C and AB of a three-leaf species phylogeny. Recombination does not negatively impact inference in this case; regardless of recombination rate, the rooted triple most likely to be inferred is the one matching the topology of the species tree.

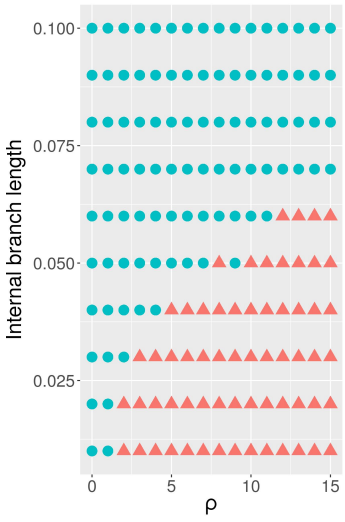
2.4.2 Quartet Simulations

Motivated by the analytic result of Corollary 1, we also ran simulations on various ultrametric species phylogenies with four leaves A, B, C, and D in order to better characterize how intralocus recombination might affect the inference of unrooted quartet topologies. In particular we considered the inference procedure detailed in Section 2.3. All simulations described in this section used parameters $k = 500$ and $\theta = 0.1$.

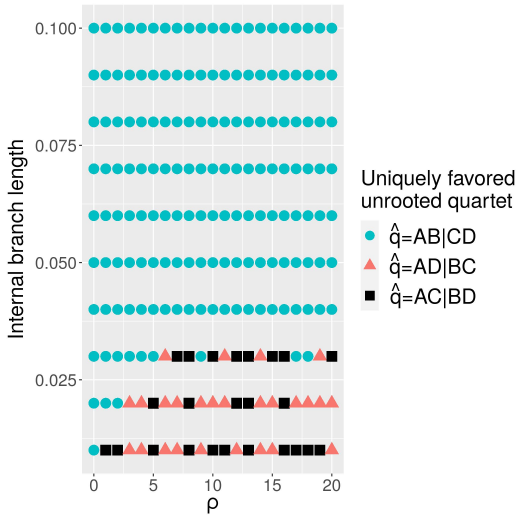
First we considered the case most similar to that of Corollary 1, in which the species phylogeny has rooted topology $(((AB)C)D)$ and in which recombination occurs only in population A (at a rate ρ ranging from 0 to 15) but with no recombination in any other populations. The internal branch corresponding to population AB was taken to be very short, ranging from 0.01 to 0.1 coalescent units, while the internal branch corresponding to population ABC was taken to be 1 coalescent unit. The leaves A and B were chosen to each have length 1, while the lengths of C and D were chosen so as to make the phylogeny ultrametric. For each parameter regime, $m = 5 \times 10^4$ samples were generated according to the MSCR-JC(500) process and the uniquely favored quartet, denoted $\hat{q}$, was recorded; the results are shown in Figure 2.11a. These results provide experimental validation of the result of Corollary 1 and are consistent with the results obtained with triplets shown in Figure 2.6.

We also considered the case in which both populations A and B exhibit high (but equal) recombination rates, and in which no recombination occurs in any other population. The same four-leaf species phylogeny was used as in the previous simulation, with the length of the internal branch corresponding to population AB ranging from 0.01 to 0.1 coalescent units. For each parameter regime, $m = 10^5$ MSAs were simulated and from these samples the uniquely favored quartet $\hat{q}$ was obtained for each regime, as shown in Figure 2.11b. The results are similar to those obtained in the previous simulation, however the addition of recombination in population B, and not just A, resulted in an overall reduction in the size of the inconsistency zone.

Next, to test the impact of intralocus recombination when rates are equal across all non-



(a) Recombination in only a single leaf A. In this plot, ρ is the recombination rate in populations A.



(b) Equal recombination rates in leaves A and B. Here ρ is the recombination rate in populations A and B.

Figure 2.11: Quartet inference with data generated on a four-leaf species phylogeny with unbalanced topology $(((AB)C)D)$, in which recombination was allowed only in specified leaves of the tree. In both plots, the vertical axis is the length, in coalescent units, of the internal edge corresponding to population AB.

root edges of the species phylogeny, we performed quartet inference with data simulated on an ultrametric species tree with four taxa A, B, C, D and rooted topology $(((AB)C)D)$. The recombination rate in the root population was set to zero in order to limit run time. This simulation was intended to be a "most challenging test case" for inference of the quartet topology because it assumed negligible internal branch lengths: both internal branches were taken to be extremely short, measuring just 0.01 coalescent units. Leaves A and B were chosen to have length 1 and the lengths of leaves C and D were chosen so that the tree was ultrametric. For each of $m = 5 \times 10^4$ samples, an MSA was generated according to the MSCR-JC(500) process and an unrooted quartet topology was inferred from the sequence distances; we denote by $\hat{p}_{AB|CD}$ the proportion of samples favoring the correct quartet $AB|BC$, and define $\hat{p}_{AC|BD}$ and $\hat{p}_{AD|BC}$ similarly for the other two unrooted quartet topologies $AC|BD$ and $AD|BC$. The results are shown in Figure 2.12a. In all 21 simulated cases, the unrooted topology of the species tree was correctly recovered. This is consistent with the results of the analogous simulation with triplets shown in Figure

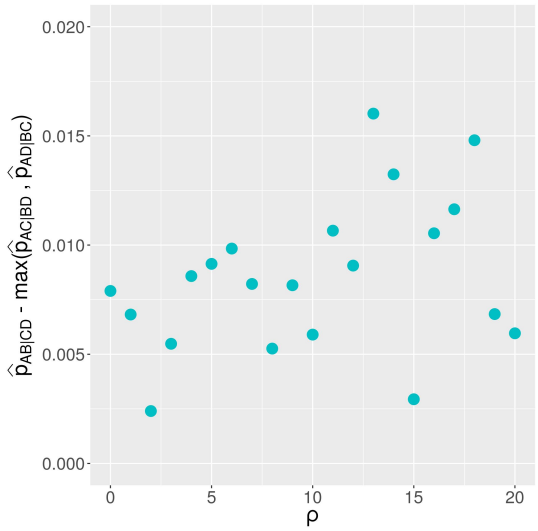
2.10.

The previous simulation was then repeated on a four-leaf tree with balanced topology $((AB)(CD))$ in which both internal branches were of length 0.01. The results are shown in Figure 2.12b. Thus, in both balanced and unbalanced cases, inference was performed on a "worst case" quartet with extremely short internal branches of only 0.01 coalescent units. Nonetheless in both cases, the unrooted quartet topology of the species tree was always correctly recovered. Under the model considered in this chapter, recombination rate heterogeneity across different lineages of the species tree appears to be a *necessary* condition for recombination to negatively impact inference.

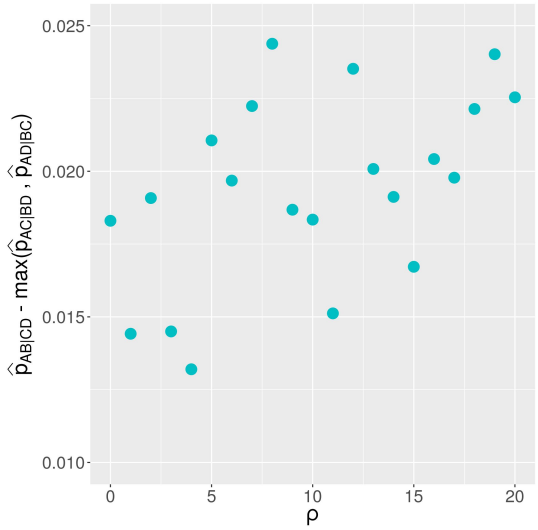
2.5 Discussion

The primary focus of this study is the effect of intralocus recombination on the inference of rooted triples. In contrast to previous simulation studies [33, 89, 90], the current work considers the effect of intralocus recombination on inference of species phylogenies *with recombination rate heterogeneity across taxa*. Our main result is a proof that within the parameter space of species phylogenies there exists a subset—the inconsistency zone—in which phylogenetic conflict between the topology of the species phylogeny and the topology of inferred gene trees is of a sufficient level to render certain majority vote methods statistically inconsistent. We further quantify and characterize this inconsistency zone through simulations, showing it includes biologically plausible recombination and mutation rates for eukaryotes, and suggesting that it arises on species phylogenies exhibiting both (1) very short internal branch lengths (less than 0.1 coalescent units) and (2) differential rates of recombination between closely related taxa. These conclusions, which hold for both rooted triples and unrooted quartets, highlight a way in which intralocus recombination can exacerbate ILS and lead to overestimation of the divergence times of those taxa exhibiting disproportionately high intralocus recombination rates relative to other taxa on the same species tree.

These findings do not necessarily contradict the conclusions of [33] that the effect of un-



(a) Equal recombination rates ρ in populations A, B, C, D, AB, and ABC on a four-taxa tree with unbalanced topology



(b) Equal recombination rates ρ in populations A, B, C, D, AB, and ABC on a four-taxa tree with balanced topology

Figure 2.12: Quartet inference is unaffected by intralocus recombination provided that recombination rates do not vary between edges of the species tree.

recognized intralocus recombination is minor, and indeed, in some important ways support such a conclusion. Our simulation experiments provide evidence that inference of rooted triples and unrooted quartets is hampered by unrecognized intralocus recombination only in cases where the internal branch length of the species tree is short, that is in cases where ILS is already high. The size of the observed effect is also relatively small; even when the uniquely favored rooted triple does not agree with the species tree, it is usually only slightly more common than the true rooted triple, a pattern which was also observed for unrooted quartets. Finally, and perhaps most importantly, if differential rates of recombination between closely-related taxa are rare, then summary coalescent-based methods which take no account of intralocus recombination may nonetheless indeed be robust even when recombination rates are high. Establishing theoretical guarantees in this direction would be of interest.

Our results raise a number of questions for future study. Our analysis focused on a simple idealized case consisting of a rooted ultrametric three-taxa species phylogeny with mutations modeled by the Jukes-Cantor process. The nature and significance of the inconsistency zone may be affected by factors such as variable population sizes as well as elements of mutation and recombination rate heterogeneity not considered here. In addition, our theoretical results only consider distance-based gene tree estimation. Extending these results to likelihood-based inference would be of interest.

Chapter 3

Species Tree Estimation under Joint Modeling of Coalescence and Duplication: Sample Complexity of Quartet Methods

Abstract: We consider species tree estimation under a standard stochastic model of gene tree evolution that incorporates incomplete lineage sorting (as modeled by a coalescent process) and gene duplication and loss (as modeled by a branching process). Through a probabilistic analysis of the model, we derive sample complexity bounds for widely used quartet-based inference methods that highlight the effect of the duplication and loss rates in both subcritical and supercritical regimes.

3.1 Introduction

Estimating phylogenies from the molecular sequences of existing species is a fundamental problem in computational biology that has been the subject of significant practical and

theoretical work [37, 91, 92, 93, 94, 95, 96]. Rigorous statistical guarantees for inference methodologies often involve the probabilistic analysis of Markov models on trees. In particular, these analyses have uncovered deep connections with phase transitions in related statistical physics models [97, 103, 104, 105, 106, 107, 108, 109, 110, 98, 99, 100, 101, 102].

In modern datasets, however, phylogeny estimation is confounded by heterogeneity across the genome from processes such as incomplete lineage sorting (ILS), gene duplication and loss (GDL), lateral gene transfer (LGT), and others [26]. Inferred trees depicting the evolution of individual loci in the genome are referred to as *gene trees*, while the tree representing the speciation history is called the *species tree*. Current sequencing technology allows phylogenetic estimates of species relationships for many genes, and a major challenge in reconstructing species trees is that gene trees often disagree for the reasons mentioned above. There is a burgeoning literature on the many ways of extracting speciation histories from collections of gene trees [111, 112, 94, 95, 113].

In this phylogenomic context, the design and analysis of species tree estimation methods require the use of a variety of stochastic processes beyond Markov models on trees, including coalescent processes [114], branching processes [115], random subtree prune-and-regraft operations [116, 117, 118], and tree mixtures [119, 120]. In fact, there is increasing realization in the phylogenetics community that ILS, GDL, LGT, etc. should not be studied in isolation [121, 11, 122, 64] and, as a result, there has been a push to consider more complex models that combine many sources of uncertainty and discordance [123, 127, 60, 128, 4, 129, 75, 130, 131, 124, 125, 126]. We study here a joint coalescent and branching process unifying ILS and GDL, as introduced in [60].

Much is known about estimating species trees in the presence of ILS alone [132], as modeled by the multispecies coalescent (MSC) [114]. The latter model posits that, on a fixed species tree, gene trees evolve backwards in time on each branch according to the Kingman coalescent [23] (see Section 3.2 for more details). Bayesian approaches are a natural choice under such complex models of evolution [133]. However they do not scale well to large datasets and more computationally efficient procedures have been developed

that combine inferred gene trees, sometimes referred to as summary methods [95].

One such method is to deduce the species tree by a plurality vote across gene trees. Unfortunately, that approach is not statistically consistent, that is, it may not converge to the true species tree as the number of gene trees grows to infinity. Indeed, for unrooted species trees with more than four species, the most frequently occurring gene tree topology need not coincide with that of the species tree [134]. However, for every 4-tuple of species—also referred to as quartet—and every locus, the most probable unrooted gene tree topology matches the species tree topology [135]. This implies in particular that the species tree topology is identifiable from the distribution of gene trees. That is, different species tree topologies necessarily produce different gene tree distributions. Further, quartet-based algorithms for combining gene trees [136, 137] are known to be statistically consistent. Tight bounds on their sample complexity, that is, how many gene trees are needed to recover the true species tree with high probability, have also been established [138].

Less is known about estimating species trees in the presence of GDL alone. The model in [115] posits that, on a fixed species tree, the number of copies in a gene family evolves forward in time on each branch according to continuous-time branching process [139] (see Section 3.2 for more details). Recently, the identifiability of the species tree in the presence of GDL alone was established in [67] by showing that, similarly to the ILS alone case, for every quartet the most frequent unrooted gene tree topology matches that of the species tree. As a result, quartet-based inference methods [66] were also shown to be statistically consistent. To date, no sample complexity results have been derived under this GDL model however.

In this chapter, we investigate the gene tree evolution model of [60], which unifies the multispecies coalescent and the branching process model of gene duplication and loss discussed above. Given that quartet-based methods have strong guarantees under these models separately, it is natural to consider their performance under the joint model as well (see Section 3.2 for more details on the methods). Numerical experiments in [65, 67] provide some evidence for the accuracy of certain quartet-based methods. In [126], the

authors give a proof of statistical consistency for one such method. In our main result, we give the first known upper bounds on the sample complexity of species tree estimation methods under the joint effect of ILS and GDL. Our proof, which highlights the somewhat counter-intuitive role played by the duplication and loss rates in the supercritical regime (see Section 3.2), is complicated by the simultaneous forward-in-time/backward-in-time nature of the process.

The rest of the chapter is organized as follows. In Section 3.2, after defining the model and inference methods, we state and discuss our results. In Section 3.3, we give a proof of identifiability including new quantitative estimates that play a role in our proof of sample complexity. The rest of the proof can be found in Section 3.4.

3.2 Background and main results

We first describe the model and then state our results formally.

3.2.1 Problem and model

Our input is a collection $\mathcal{T} = \{t_i\}_{i=1}^k$ of k multi-labeled gene trees given without estimation error. We explain the terminology. The process of gene duplication creates multiple copies of a gene within the same individual in a species. We refer to these duplicated genomic segments as *gene copies* and we refer to collections of gene copies from different unrelated genes as *gene families*. A *gene tree* is a depiction of the parental lineages of a gene or multiple gene copies from individuals across several species. Roughly speaking, it shows the joint ancestral history of these related gene copies. In contrast, a *species tree* is a depiction of the evolutionary relationships of a group of species. Roughly speaking, it shows the sequence of speciation events that have produced the current species. By *multi-labeled*, we mean that each leaf of a gene tree is associated with a label from the set S of n species of interest and that the leaf labels need not be unique. That is, multiple leaves of a gene tree may be associated with the same species; this corresponds to observing multiple paralogous copies (i.e., which have arisen from gene duplication) of a gene within the genome. In

practice, gene trees are estimated from the molecular sequences of the corresponding genomic segments using a variety of phylogenetic reconstruction methods [37, 91, 92, 93, 94, 95]. For the sake of our results, we assume that gene trees are provided *without estimation error* for a large number of gene families. Our main modeling assumption is that these gene trees have been drawn independently from a distribution for which some n-species tree T is taken as a fixed (i.e., non-random)—but unknown—parameter. Our goal is to output this unknown n-species tree T. We define the model more precisely next.

Model We assume that the gene trees in $\mathcal{T}$ are independent and identically distributed, with each gene tree generated under the DLCoal model [60], a unified model of gene duplication, loss, and coalescence. The process for generating a gene tree under DLCoal involves two steps which are described below; an example realization of these steps is provided in Figure 3.1. In particular, we introduce yet another tree, a locus tree, which is an unobserved intermediate step in the model generating each gene tree. The process below is repeated independently for each $i = 1, \ldots, k$, thereby producing an independent gene tree t_i.

1. *Locus tree: Birth-death process of gene duplication and loss with daughter edges.*

   We fix a rooted n-species tree T with edges E directed from root to leaves (i.e. from top to bottom) and edge lengths $\{\eta_e\}_{e \in E}$. In Figure 3.1, this is the four-species tree on the top left. Note that the species tree is unknown to us.

   Starting with a single ancestral copy of a gene at the root of T, a tree is generated by a top-down birth-death process [115] *within* the species tree. That is, on every edge in T, each gene copy independently duplicates at exponential rate $\lambda \geq 0$ and is lost at exponential rate $\mu \geq 0$. In addition, whenever a gene copy reaches a speciation event T, it bifurcates into two child copies, one for each of the descendant edges on T. Both duplications and speciations are indicated in the resulting locus tree by a bifurcation. The locus tree is then pruned of lost copies to give an (unobserved) rooted tree, which we refer to as *locus tree*. In this manner, we obtain a rooted

n'-individual locus tree L with edge lengths. Species labels are associated to each leaf of L from the species set S. These steps for generating L are illustrated in the top-middle and top-right trees of Figure 3.1.

Furthermore, for each vertex in L which corresponds to a gene duplication event—but *not* vertices corresponding to speciation bifurcations—we distinguish between the two child edges by choosing one of them uniformly at random to be the *daughter* edge and the other to be the *mother* edge. This distinction plays an important role in the next step.

2. *Gene tree: Coalescent process on a locus tree.*

   Gene trees are generated by a backward-in-time (i.e., bottom-up) coalescent process [23, 114] within the locus tree L. The coalescent process begins with exactly one gene copy in each leaf of L. Copies at the bottom of a directed edge in the locus tree undergo the Kingman coalescent process for a time equal to the length of the directed edge. Edge lengths here are assumed to be in so-called coalescent time units, which means that each pair of lineages independently coalesces after an exponential time with mean 1, unless the end of the edge is reached first. Further, we condition on the event that all gene copies at the bottom of any *daughter* edge of the locus tree necessarily coalesces underneath the top of the edge. Continuing upward along ancestral edges of the locus tree, this process eventually yields a gene tree.

   This process, which generates a gene tree from a locus tree, is termed the *multilocus coalescent* (MLC) in [60]. A realization of this coalescent process within a locus tree is illustrated in the bottom left of Figure 3.1, and the tree at the bottom right depicts the resulting gene tree without the overlying locus tree.

Intuitively, in Step 1 above the species tree T is a fat tree that 'contains' the birth-death process which generates the skinny tree L. In Step 2, we forget about the species tree and 'zoom in' on L so that L becomes the fat tree, with each edge of L containing

one or more edges (or parts of edges) of the gene tree.

Species tree estimation methods Next we describe two quartet-based species tree methods: ASTRAL-one and ASTRAL-multi [137, 65, 66]. Recall that a *quartet* refers to a 4-tuple of species. It is known that the unrooted topology of a species tree is entirely characterized by the collection of its quartet topologies (see, e.g., [37]), and hence a large number of quartet-based approaches to species tree estimation have been developed. Both ASTRAL-one and ASTRAL-multi are practical variants of an intuitive idea (which in the ILS/GDL context is motivated by the results of [135, 67]; see also Propositions 1 and 2 below): (1) for each quartet of species, find the most common topology across gene trees and (2) reconstruct an n-species tree that coincides with as many resulting quartet topologies as possible. The input is a collection of k multi-labeled gene trees $\mathcal{T} = \{t_i\}_{i=1}^{k}$. Let S be the set of n species and Σ be the set of m labels (or gene copies). The tree t_i is labeled by the set $\Sigma_i \subset \Sigma$. For any species tree T labeled by S, the extended tree T_{ext} labeled by Σ is built by adding to each leaf of T all gene copies corresponding to that species as a polytomy (i.e., as a vertex with degree possibly higher than 3).

Under ASTRAL-one, we pick one gene copy of each species uniformly at random and restrict the gene tree to these copies, producing a new gene tree $\tilde{t}_i$. For any collection of gene copies $\mathcal{J} = \{a, b, c, d\}$, let $T^{\mathcal{J}}$ be the restriction of any tree T to those copies. Then the quartet score of every candidate species tree $\widetilde{T}$ is

$$Q_k\left(\widetilde{T}\right) = \sum_{i=1}^{k} \sum_{\mathcal{J}=\{a,b,c,d\}\subset\Sigma_i} \mathbf{1}(\widetilde{T}_{ext}^{\mathcal{J}}, \tilde{t}_i^{\mathcal{J}}),$$

where $\mathbf{1}(T_1, T_2)$ is the indicator for the event that T_1 and T_2 have the same topology. ASTRAL-one selects the candidate tree $\widetilde{T}$ that maximizes that score.

ASTRAL-multi treats copies of a gene in a species as multiple alleles within the species. So, we do not replace t_i with any restricted gene tree $\tilde{t}_i$. The quartet score of T with respect

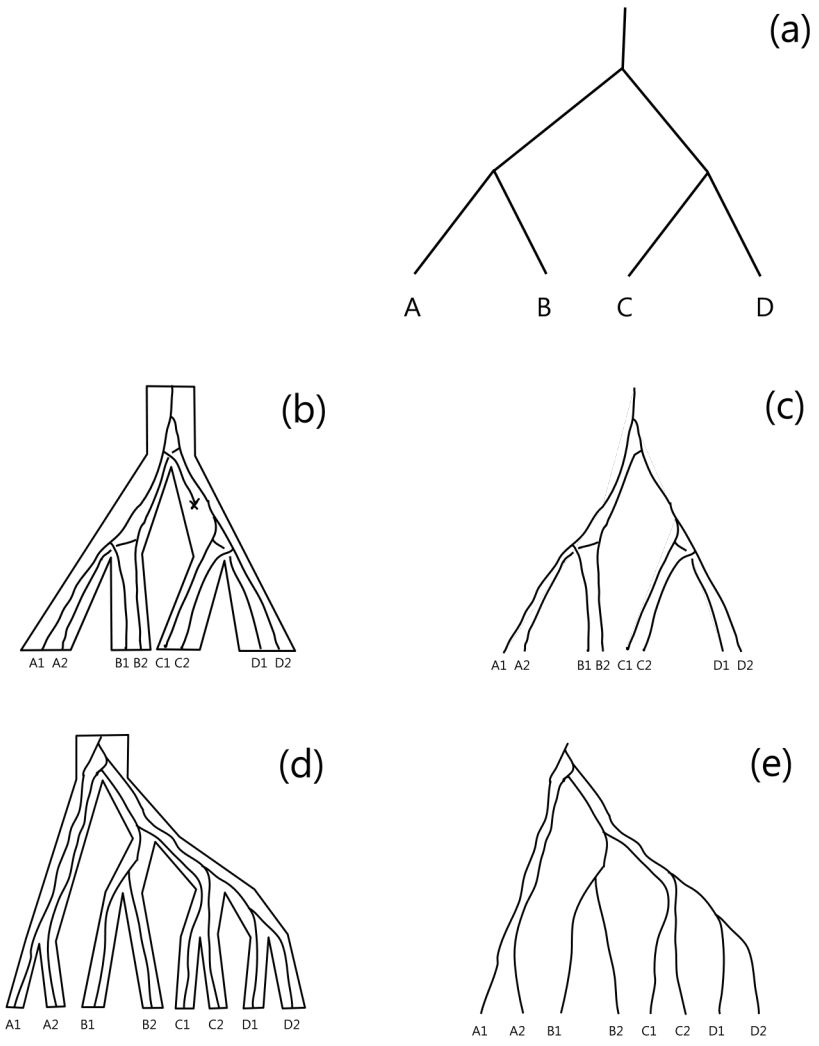


Figure 3.1: A species tree with balanced topology and leaf species A, B, C, D is given in (a). From (a), we obtain a realization of the gene duplication and loss (GDL) process with deaths not yet pruned. The pre-pruned locus tree and its generating species tree are shown in (b). Non-extant lineages in the tree are then pruned to obtain the locus tree, shown in (c). The gene tree obtained from the multi-species coalescent (MSC) process and its generating locus tree are shown in (d). The gene tree we actually observe is shown in (e).

to $\mathcal{T}$ is

$$Q_k\left(\widetilde{T}\right) = \sum_{i=1}^{k} \sum_{\mathcal{J}=\{a,b,c,d\}\subset\Sigma_i} \mathbf{1}(\widetilde{T}_{ext}^{\mathcal{J}}, t_i^{\mathcal{J}}).$$

The candidate tree $\widetilde{T}$ that maximizes this score is chosen by ASTRAL-multi.

Both procedures can be performed in exact mode or in a default constrained mode, which restricts the number of candidate species trees to those displaying the bipartitions in the given gene trees.

3.2.2 Statement of main results

We present a theoretical bound on the number of gene trees needed for ASTRAL-one to reconstruct the model species tree with high probability under the DLCoal model. Similarly to the case of the MSC model [138], the sample complexity depends on the length of the shortest species tree branch in coalescent time units. We denote this length by f. However we also highlight the influence of other relevant parameters in the sample complexity: the depth of the species tree, Δ; the duplication and loss rates, λ and μ. For simplicity, we assume throughout that $\mu \neq \lambda$.

Theorem 2 (Main result: Sample complexity of ASTRAL-one)**.** *Consider a model species tree whose minimum branch length f is finite and assume gene trees are generated under the DLCoal model. Then, for any $\epsilon > 0$, there are universal positive constants C, C' such that the exact version of ASTRAL-one returns the true species tree with probability at least $1-\epsilon$ if the number of input error-free gene trees satisfies:*

$$k \geq C' \frac{1}{f^2} \frac{e^{C|\mu-\lambda|\Delta}}{\left(1-\frac{\lambda}{\mu}\wedge\frac{\mu}{\lambda}\right)^C} \log\frac{n}{\epsilon}. \tag{3.1}$$

Somewhat surprisingly the subcritical ($\mu > \lambda$) and supercritical ($\mu < \lambda$) regimes exhibit a similar behavior. Indeed one naturally expects a higher sample complexity in the subcritical case as μ/λ becomes large because the absence of any gene copy in a species

becomes more likely and leads to the need for more gene trees in order to extract signal. That prediction is borne out in (3.1). However the sample complexity similarly increases in the supercritical regime as λ/μ becomes large. As the proof shows, the reasons for this behavior are different in that regime. They have to do with the fact that a large number of copies at the most recent common ancestor of a species quartet tends to produce large numbers of conflicting gene tree quartet topologies, thereby obscuring the signal. It is an open problem whether other inference methods (perhaps not based on quartets) are less sensitive to this last phenomenon.

Our proof of Theorem 2 involves a delicate probabilistic analysis of the DLCoal model. Along the way, we prove other results of interest. First, we show that the unrooted species tree is identifiable from the distribution of multi-labeled gene trees $\mathcal{T}$ under the DLCoal model over T. Formally, we show that two distinct unrooted species trees produce different gene tree distributions. The result is a generalization of [67, Theorem 1], where only GDL is considered. Theorem 3 was first claimed in [126]. Our novel contribution here lies in the proofs of Propositions 1 and 2 below, which give a quantitative version of the identifiability result and play a role in deriving the sample complexity of ASTRAL-one.

In fact, even in the absence of ILS (i.e., in the presence of GDL alone), no sample complexity results were previously available. While identifiability is established through symmetry arguments (see, e.g., Lemma 11 below, whose simple proof closely mirrors the core argument in [67]), sample complexity bounds require a quantitative analysis that is significantly more involved. Our arguments here (see Sections 3.3 and 3.4) combine symmetry arguments, explicit computations and a detailed case analysis of well-chosen events. In addition we incorporate ILS, which further complicates the analysis, in part because of the forward-in-time/backward-in-time nature of the combined process. We expect that the types of arguments developed here will prove useful in analyzing other reconstruction methods under related complex models of genome evolution.

Our main identifiability result is stated next.

Theorem 3 (Identifiability of species tree)**.** *Let T be a model species tree with at least*

$n \geq 4$ leaves. Then T, without its root, is identifiable from the distribution of gene trees $\mathcal{T}$ under the DLCoal model over T.

This identifiability result is established by showing that, for each quartet in the species tree, the most likely gene tree matches the species tree. As in [67, 126], a direct consequence of this proof is the statistical consistency of the ASTRAL-one.

Theorem 4 (Consistency of ASTRAL-one)**.** *As the number of input gene trees tends toward infinity, the output of ASTRAL-one converges to T almost surely, when run in exact mode or in its default constrained version.*

We use a similar reasoning to prove the consistency of ASTRAL-multi. This result is new.

Theorem 5 (Consistency of ASTRAL-multi)**.** *As the number of input gene trees tends toward infinity, the output of ASTRAL-multi converges to T almost surely, when run in exact mode or in its default constrained version.*

The first step is a proof of the identifiability of the species tree under the DLCoal model.

3.3 A proof of identifiability of the species tree

Let $\mathcal{Q} = \{A, B, C, D\}$ and assume, without loss of generality, that $T^{\mathcal{Q}}$ has unrooted quartet topology $AB|CD$, that is, it has the topology depicted in the top left of Figure 3.2 (ignoring the root). Let t be a gene tree generated under the DLCoal model on T and let $t^{\mathcal{Q}}$ be its restriction to the gene copies from the species in $\mathcal{Q}$. The high-level idea behind our proof of Theorem 3 is the following:

> *Conditioning on the number of copies in species A, B, C, D in the species in $\mathcal{Q}$, independently pick a uniformly random gene copy a, b, c, d in species A, B, C, D and let q be the corresponding quartet topology under $t^{\mathcal{Q}}$. We show that the most likely outcome is $q = ab|cd$.*

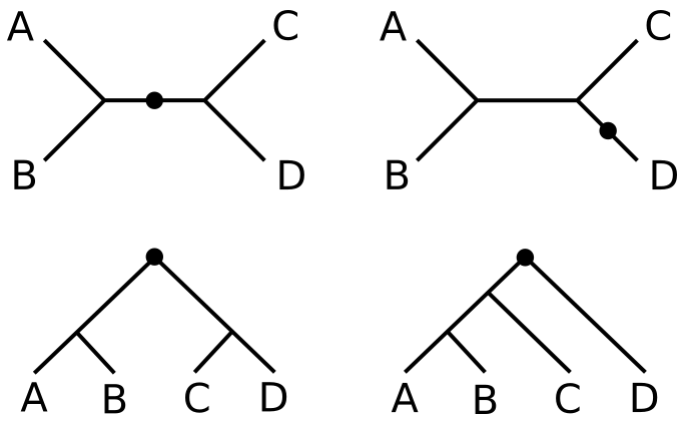


Figure 3.2: The top row shows the two cases of root location on $T^{\mathcal{Q}}$: the root of $T^{\mathcal{Q}}$ may occur either on the internal edge (left) or the pendant edge adjacent to D (right). In both cases the unrooted quartet topology of $T^{\mathcal{Q}}$ is $AB|CD$. However, as shown in the bottom row, the two choices of root location leads to two different rooted topologies: the balanced case (left) and the caterpillar case (right), which we consider separately.

This is the same approach as that used in [67], but the analysis of the model is more involved.

Define $\mathcal{X} = (\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ to be the number of copies in species A, B, C, D, respectively. For example, Figure 3.1 depicts a realization in which $\mathcal{X} = (2, 2, 2, 2)$ since the locus tree (and hence also the gene tree) has exactly two leaves in each of the species A, B, C and D. We will let $\mathbb{P}'$ be the probability measure subject to conditioning on the random vector $\mathcal{X}$.

Although we seek to reconstruct an *unrooted* species tree, under the DLCoal model, the locus trees and gene trees are in fact generated from a *rooted* species tree. Therefore when restricting the species tree to four species, there are two cases of root location to consider: when the root of the species quartet $T^{\mathcal{Q}}$ is located on the internal quartet edge of $T^{\mathcal{Q}}$ (the "balanced case") or on a pendant edge of $T^{\mathcal{Q}}$ (the "caterpillar case"). In the caterpillar case we may take D to be the pendant edge incident with the root. See Figure 3.2.

For gene copies a, b, c, d from A, B, C, D, define the following events:

$$Q_1 = \{q = ab|cd\}, \quad Q_2 = \{q = ac|bd\}, \quad Q_3 = \{q = ad|bc\}.$$

3.3.1 Balanced case

We first consider the balanced case. Without loss of generality, assume the species tree restricted to $\mathcal{Q}$ has rooted topology as in the bottom left of Figure 3.2. Let R be the most recent common ancestor of $\mathcal{Q}$ in the species quartet $T^{\mathcal{Q}}$ and I be the number of locus copies exiting R forward in time. Let $\mathbb{P}''$ be the probability measure indicating conditioning on I as well as on $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$. For any selection of copies (a, b, c, d) from each species in the quartet, let $i_x \in \{1, ..., I\}$ be the ancestral lineage on the locus tree of $x \in \{a, b, c, d\}$ in R. By the law of total probability, we have

$$\mathbb{P}[Q_i] = \mathbb{E}[\mathbb{P}''[Q_i]], \quad i = 1, 2, 3. \tag{3.2}$$

Hence, in order to show identifiability of the species quartet, it is sufficient to show that

$$\mathbb{P}''[Q_1] > \max\{\mathbb{P}''[Q_2], \mathbb{P}''[Q_3]\}$$

when the copies of (A, B, C, D) are chosen uniformly at random.

We let $\mathcal{X} \geq \vec{1}$ be the event that each of (A, B, C, D) has at least one copy to select from. On the complement of $\mathcal{X} \geq \vec{1}$ (that is, at least one of (A, B, C, D) fails to have a copy to select), then ASTRAL-one selects Q_1, Q_2, Q_3 each with probability 0. So we consider the case $\mathcal{X} \geq \vec{1}$. We will use the notation $z_1 \wedge z_2 = \min\{z_1, z_2\}$.

Proposition 1 (Quartet identifiability: Balanced case)**.** *Let* $x = \mathbb{P}''[i_a = i_b]$ *and* $y = \mathbb{P}''[i_c = i_d]$*. On the events* $\mathcal{X} \geq \vec{1}$ *and* $I \geq 1$*, we have almost surely*

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] > \frac{1}{12}\left(x - \frac{1}{I}\right) \wedge \left(y - \frac{1}{I}\right).$$

The proof of Proposition 1 is in Section 3.3.1. We first establish a series of lemmas.

The next lemma shows that $x, y \geq 1/I$, similarly to [67, Lemma 1]. In that work, it is proved that the probabilities of $\{i_a = i_b\}$ and $\{i_c = i_d\}$ are each at least $1/I$ under a different conditional probability measure. Our proof is otherwise identical.

Lemma 11. *Let* $x = \mathbb{P}''[i_a = i_b]$ *and* $y = \mathbb{P}''[i_c = i_d]$. *On the events* $\mathcal{X} \geq \vec{1}$ *and* $I \geq 1$, *we have almost surely*

$$x \wedge y \geq \frac{1}{I}.$$

Proof. For $j \in \{1, ..., I\}$, let N_j be the number of gene copies descending from j in R that survive to the most recent common ancestor of A and B. Conditioning on $(N_j)_j$, the choice of a and b in the locus tree is independent as in [67]. So i_a and i_b are picked proportionally to the N_j's by symmetry. Then

$$\mathbb{P}''[i_a = i_b] = \mathbb{E}''[\mathbb{P}''[i_a = i_b|(N_j)_j]] = \mathbb{E}''\left[\frac{\sum_{j=1}^{I} N_j^2}{(\sum_{j=1}^{I} N_j)^2}\right] \geq \frac{1}{I},$$

as in [67, Lemma 1]. The same holds for y, completing the proof of the lemma. □

Ancestral locus configurations

Conditioned on $\mathcal{X}$ and I, we will characterize the occurrence of Q_1, Q_2, Q_3 based on how $i_x, x \in \{a, b, c, d\}$, are picked at the root R, that is, based on which pairs i_x, i_y are equal. Then, in a worst-case scenario, we will analyze events of the coalescent process above R. For an arbitrary quartet (a, b, c, d), we relate the likelihood of Q_1, Q_2, Q_3 under each of the following events:

$$\begin{aligned}
E &= a - b - c - d \\
F_{ab} &= ab - c - d \qquad F_{ac} = ac - b - d \qquad F_{ad} = ad - b - c \\
F_{bc} &= bc - a - d \qquad F_{bd} = bd - a - c \qquad F_{cd} = cd - a - b \\
G_{ab} &= ab - cd \qquad G_{ac} = ac - bd \qquad G_{ad} = ad - bc \\
H_{abc} &= abc - d \qquad H_{abd} = abd - c \qquad H_{acd} = acd - b \qquad H_{bcd} = bcd - a \\
K &= abcd,
\end{aligned} \tag{3.3}$$

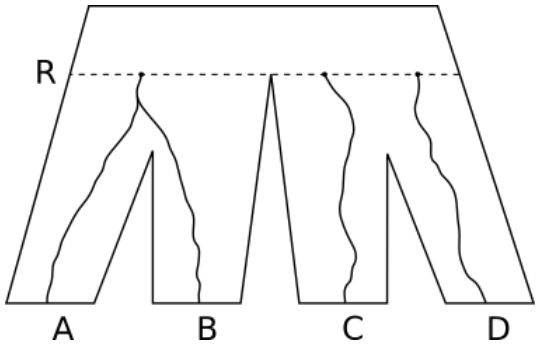


Figure 3.3: An example of a locus tree L satisfying the event F_{ab} for selected gene copies a, b, c, d. In particular, the edges of L are depicted as residing within the fat edges of $T^{\mathcal{Q}}$, however only the portion of L coinciding with the edges of $T^{\mathcal{Q}}$ is shown; the portion of L above R is omitted.

where $-$ indicates separate lineages at R for the chosen copies from A, B, C, D. For example, the event F_{ab} indicates that the i_c and i_d are different and are different from the common ancestral lineage for i_a and i_b. See Figure 3.3. These events are disjoint and mutually exhaustive. Letting $\mathcal{E}$ run across all the above events, the law of total probability implies

$$\mathbb{P}''[Q_i] = \sum_{\mathcal{E}} \mathbb{P}''[Q_i|\mathcal{E}]\mathbb{P}''[\mathcal{E}]. \tag{3.4}$$

Reduction to coalescence above R

For the rooted locus quartet implied by the four copies a, b, c, d, let $\mathcal{NC}$ be the event that no coalescence event occurs beneath R among the four corresponding lineages. The following lemma shows that conditioning on $\mathcal{NC}$ reduces the probability of Q_1 while increasing that of Q_2 and Q_3.

Lemma 12. *For any I and any $\mathcal{X} \geq \vec{1}$ and any event*

$$\mathcal{E} \in \{E, F_{ab}, \ldots, G_{ab}, \ldots, H_{abc}, \ldots\},$$

we have

$$\mathbb{P}''[Q_1|\mathcal{E}] \geq \mathbb{P}''[Q_1|\mathcal{E} \cap \mathcal{NC}] \quad \textit{and} \quad \mathbb{P}''[Q_i|\mathcal{E}] \leq \mathbb{P}''[Q_i|\mathcal{E} \cap \mathcal{NC}], \quad i \in \{2,3\},$$

almost surely.

Proof. Under $\mathcal{NC}^c$, by definition, at least one of the pairs $\{a,b\}$ or $\{c,d\}$ coalesces below R. Both cases immediately lead to the same quartet topology, so Q_1 is guaranteed. See Figure 3.5 below for an illustration when $\mathcal{E} = K$. Then the law of total probability implies

$$\mathbb{P}''[Q_1|\mathcal{E}] = \mathbb{P}''[\mathcal{NC}^c|\mathcal{E}] + \mathbb{P}''[Q_1|\mathcal{E} \cap \mathcal{NC}]\,\mathbb{P}''[\mathcal{NC}|\mathcal{E}] \geq \mathbb{P}''[Q_1|\mathcal{E} \cap \mathcal{NC}].$$

Similarly

$$\mathbb{P}''[Q_i|\mathcal{E}] = \mathbb{P}''[Q_i|\mathcal{E} \cap \mathcal{NC}]\,\mathbb{P}''[\mathcal{NC}|\mathcal{E}] \leq \mathbb{P}''[Q_i|\mathcal{E} \cap \mathcal{NC}],$$

for $i \in \{2,3\}$. □

The event K will play a special role in the proof and we treat it separately. For the other terms, combining (3.4) and Lemma 12, we have

$$\begin{aligned}
\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] &= (\mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K])\mathbb{P}''[K] \\
&\quad + \sum_{\mathcal{E} \neq K} \left(\mathbb{P}''[Q_1|\mathcal{E}] - \mathbb{P}''[Q_2|\mathcal{E}]\right) \mathbb{P}''[\mathcal{E}] \\
&\geq (\mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K])\mathbb{P}''[K] \\
&\quad + \sum_{\mathcal{E} \neq K} \left(\mathbb{P}''[Q_1|\mathcal{E} \cap \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{E} \cap \mathcal{NC}]\right) \mathbb{P}''[\mathcal{E}].
\end{aligned}$$

To prove Proposition 1, we derive an explicit bound on this last sum.

Under $\mathbb{P}''$, the events $E, H_{abc}, H_{abd}, H_{bcd}, H_{acd}$ are symmetric in the sense that switching the roles of a and c or the roles of a and d does not change the conditional probability of

Q_1 and Q_2. Hence

$$\mathbb{P}''[Q_1|\mathcal{E}\cap\mathcal{NC}] = \mathbb{P}''[Q_2|\mathcal{E}\cap\mathcal{NC}], \quad \forall\mathcal{E}\in\{E, H_{abc}, H_{abd}, H_{bcd}, H_{acd}\}$$

and using this above we get

$$\begin{aligned}\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] \geq{}& (\mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K])\mathbb{P}''[K] \\ &+ \sum_{j\in\{ab,ac,ad\}} \left(\mathbb{P}''[Q_1|G_j\cap\mathcal{NC}] - \mathbb{P}''[Q_2|G_j\cap\mathcal{NC}]\right)\mathbb{P}''[G_j] \\ &+ \sum_{j\in\{ab,\ldots,cd\}} \left(\mathbb{P}''[Q_1|F_j\cap\mathcal{NC}] - \mathbb{P}''[Q_2|F_j\cap\mathcal{NC}]\right)\mathbb{P}''[F_j]. \qquad (3.5)\end{aligned}$$

The F and G events

We now consider the events $\{F_{ab}, F_{cd}, F_{ac}, F_{bd}\}$ and $\{G_{ab}, G_{ac}\}$.

Event probabilities In the next lemma, we compute the probabilities of a given locus tree quartet satisfying the events in $\{F_{ab}, ..., F_{cd}, G_{ab}, G_{ac}\}$.

Lemma 13. *Let $x = \mathbb{P}''[i_a = i_b]$ and $y = \mathbb{P}''[i_c = i_d]$. For $I \geq 2$ and any $\mathcal{X} \geq \vec{1}$, the following hold almost surely:*

$$\begin{aligned}\mathbb{P}''[F_{ab}] &= \frac{I-2}{I}x(1-y) \\ \mathbb{P}''[F_{cd}] &= \frac{I-2}{I}(1-x)y \\ \mathbb{P}''[F_{ac}] = \mathbb{P}''[F_{bd}] &= \frac{I-2}{I(I-1)}(1-x)(1-y) \\ \mathbb{P}''[G_{ab}] &= \frac{I-1}{I}xy \\ \mathbb{P}''[G_{ac}] &= \frac{1}{I(I-1)}(1-x)(1-y)\end{aligned}$$

Proof. The calculations for F_{ab} and F_{cd} are similar, except that we condition on different

events. Indeed, note that

$$\mathbb{P}''[F_{ab}] = \mathbb{P}''[F_{ab}|i_a = i_b, i_c \neq i_d]\mathbb{P}''[i_a = i_b]\mathbb{P}''[i_c \neq i_d]$$
$$\mathbb{P}''[F_{cd}] = \mathbb{P}''[F_{cd}[i_a \neq i_b, i_c = i_d]\mathbb{P}''[i_a \neq i_b]\mathbb{P}''[i_c = i_d].$$

The conditional probability of F_{ab} is then obtained by considering that given the placement of the pair (i_c, i_d) among the I ancestral lineages, the shared lineage $i_a = i_b$ has $I - 2$ choices where they do not intersect $\{i_c, i_d\}$. The result in the statement follows. Similarly, for F_ι with $\iota \in \{ac, bd\}$, we have

$$\mathbb{P}''[F_\iota] = \mathbb{P}''[F_\iota|i_a \neq i_b, i_c \neq i_d]\mathbb{P}''[i_a \neq i_b]\mathbb{P}''[i_c \neq i_d].$$

In this case, out of $I(I-1)$ choices for i_a and i_b, the choice of i_c is determined and there are $I - 2$ remaining choices for i_d, implying the result.

We use the same principle for G_{ab} and G_{ac}. Keeping this in mind, we have

$$\mathbb{P}''[G_{ab}] = \mathbb{P}''[G_{ab}|i_a = i_b, i_c = i_d]\mathbb{P}''[i_a = i_b]\mathbb{P}''[i_c = i_d]$$
$$\mathbb{P}''[G_{ac}] = \mathbb{P}''[G_{ac}|i_a \neq i_b, i_c \neq i_d]\mathbb{P}''[i_a \neq i_b]\mathbb{P}''[i_c \neq i_d],$$

and we proceed as before to get the result. □

Using the previous lemma, we collect further bounds on the probabilities of events at the root of the locus tree.

Lemma 14. *Letting again $x = \mathbb{P}''[i_a = i_b]$ and $y = \mathbb{P}''[i_c = i_d]$, the following statements hold.*

(a) If $I = 2$ then

$$\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] \geq \left(x - \frac{1}{2}\right) \wedge \left(y - \frac{1}{2}\right)$$

(b) If $I \geq 3$ and $x \wedge y \geq 1/2$, then

$$\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] \geq 1/8.$$

(c) If $I = 2$,

$$\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}] = 0.$$

(d) If $I \geq 3$ and $x \wedge y \leq 1/2$, then

$$\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}] \geq \frac{1}{4}\left(x - \frac{1}{I}\right) \wedge \left(y - \frac{1}{I}\right).$$

Proof. By Lemma 13, $\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] = \frac{1}{2}(x + y - 1)$ which implies part (a).

To prove (b), observe that by Lemma 13 again,

$$\begin{aligned}
\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] &= \frac{I-1}{I}xy - \frac{1}{I(I-1)}(1-x)(1-y) \\
&\geq \frac{1}{2}\left(\frac{I-1}{I}y - \frac{1}{I(I-1)}(1-y)\right) \\
&\geq \frac{1}{4}\left(\frac{I-1}{I} - \frac{1}{I(I-1)}\right) \\
&= \frac{1}{4}\left(\frac{I-2}{I-1}\right)
\end{aligned}$$

where the inequalities are justified by the assumption $x \wedge y \geq 1/2$. Since $I \geq 3$, it follows that $\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] \geq 1/8$.

To prove (c) and (d), observe that by Lemma 13,

$$\begin{aligned}
&\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}] \\
&= \frac{I-2}{I}\left(x(1-y) + y(1-x)\right) - 2\frac{I-2}{I(I-1)}(1-x)(1-y) \\
&= \frac{I-2}{I(I-1)}\left((1-y)\left((I-1)x - (1-x)\right) + (1-x)\left((I-1)y - (1-y)\right)\right) \\
&= \frac{I-2}{I(I-1)}\left((1-y)(Ix-1) + (1-x)(Iy-1)\right).
\end{aligned}$$

Clearly if $I = 2$, the right-hand side is zero, which proves (c). Furthermore, since $x, y \geq 1/I$ by Lemma 11, it follows that both $(1-y)(Ix-1) \geq 0$ and $(1-x)(Iy-1) \geq 0$, and therefore

$$\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}] \geq \frac{I-2}{I(I-1)}\,(1-u)\,(Iv-1)$$

for $(u, v) \in \{(x, y), (y, x)\}$. Taking $u = \min(x, y)$ and $v = \max(x, y)$ gives

$$\begin{aligned}
\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}] &\geq \frac{I-2}{I(I-1)}\,(1-\min(x,y))\,(I\max(x,y)-1) \\
&\geq \frac{I-2}{I-1}\,(1-\min(x,y))\left(\max(x,y)-\frac{1}{I}\right) \\
&\geq \frac{1}{2}\,(1-\min(x,y))\left(\max(x,y)-\frac{1}{I}\right) \\
&\geq \frac{1}{4}\left(\max(x,y)-\frac{1}{I}\right)
\end{aligned}$$

which implies (d). □

Conditional probabilities of quartet topologies In the following lemma, we give expressions for $\mathbb{P}''[Q_i|\mathcal{E} \cap \mathcal{NC}]$ across the events $\{F_{ab}, F_{cd}, F_{ac}, F_{bd}\}$ and $\{G_{ab}, G_{ac}\}$.

Lemma 15. *(a) For any I and any $\mathcal{X} \geq \vec{1}$, we have*

$$\mathbb{P}''[Q_1|F_{ab} \cap \mathcal{NC}] = \mathbb{P}''[Q_1|F_{cd} \cap \mathcal{NC}] = \mathbb{P}''[Q_2|F_{ac} \cap \mathcal{NC}] = \mathbb{P}''[Q_2|F_{bd} \cap \mathcal{NC}] =: \phi''_+$$

and

$$\mathbb{P}''[Q_2|F_{ab} \cap \mathcal{NC}] = \mathbb{P}''[Q_2|F_{cd} \cap \mathcal{NC}] = \mathbb{P}''[Q_1|F_{ac} \cap \mathcal{NC}] = \mathbb{P}''[Q_1|F_{bd} \cap \mathcal{NC}] =: \phi''_-.$$

(b) For any I and any $\mathcal{X} \geq \vec{1}$, we have

$$\mathbb{P}''[Q_1|G_{ab} \cap \mathcal{NC}] = 1$$

and

$$\mathbb{P}''[Q_2|G_{ac} \cap \mathcal{NC}] = 1.$$

Proof. (a) The quantities ϕ''_+ and ϕ''_- are indeed well-defined as above by symmetry. (b) By switching the roles of b and c, we observe that $\mathbb{P}''[Q_1|G_{ab} \cap \mathcal{NC}] = \mathbb{P}''[Q_2|G_{ac} \cap \mathcal{NC}]$. To see why $\mathbb{P}''[Q_1|G_{ab} \cap \mathcal{NC}] = 1$, we again examine the topology above the root with leaves ab and cd. At least one of these leaves descends from a daughter edge, which implies Q_1 is constructed with probability 1. This completes the proof of the lemma. □

The following lemma establishes that, conditioned on F_{ab} and $\mathcal{NC}$, the difference in probability between Q_1 and Q_2 is at least 1/3.

Lemma 16. *For $I \geq 1$ and any $\mathcal{X} \geq \vec{1}$, we have*

$$\phi''_+ - \phi''_- \geq \frac{1}{3}.$$

Proof. By definition of ϕ''_+ and ϕ''_-, it suffices to show $\mathbb{P}''[Q_1|F_{ab} \cap \mathcal{NC}] - \mathbb{P}''[Q_2|F_{ab} \cap \mathcal{NC}] \geq 1/3$. Conditioned on $F_{ab} \cap \mathcal{NC}$, no coalescence event between the chosen lineages occurs beneath R. So, we examine the topology of the locus tree above the root with leaf set being the three leaves implied by F_{ab}. Using the law of total probability, we condition further across the three possible rooted locus topologies on the three leaves ab, c, and d. For each $i \in \{ab, c, d\}$, let τ_i be the rooted topology on three leaves in which the outgroup is labeled by i and the other two leaves are labelled from the remaining letters in $\{a, b, c, d\}$. For example, τ_{ab} has c and d as siblings with ab as the outgroup. Then

$$\begin{aligned}
&\mathbb{P}''[Q_1|F_{ab} \cap \mathcal{NC}] - \mathbb{P}''[Q_2|F_{ab} \cap \mathcal{NC}] \\
&\quad = \frac{1}{3} \sum_i \left(\mathbb{P}''[Q_1|\tau_i, F_{ab} \cap \mathcal{NC}] - \mathbb{P}''[Q_2|\tau_i, F_{ab} \cap \mathcal{NC}] \right),
\end{aligned}$$

where we used the fact that $\mathbb{P}''[\tau_i|F_{ab} \cap \mathcal{NC}] = 1/3$ for each i. Now we compute the summands. If $i = ab$, then either ab descends from a daughter lineage or the pair (c, d)

descends from a daughter lineage, meaning we observe Q_1 with probability 1 and Q_2 with probability 0. In the other two cases, let $p > 0$ be the probability that a and b coalesce along the pendant edge for ab. If they do not coalesce along the pendant edge, then the lineages from a and b live in the same population as that of, say, c. Then there is probability $1/3$ that the first coalescing pair among a, b, c is a, b. So the probability of observing Q_1 is $p + \frac{1}{3}(1-p)$. There is probability $1/3$ that the first coalescing pair among a, b, c is a, c, so the probability of observing Q_2 is $\frac{1}{3}(1-p)$. Then

$$\phi''_+ - \phi''_- = \frac{1}{3}\left(1 - 0 + 2\left(p + \frac{1}{3}(1-p) - \frac{1}{3}(1-p)\right)\right) = \frac{1}{3}(1+2p) \geq \frac{1}{3}.$$

□

Proof of Proposition 1

With that we can prove Proposition 1.

Proof of Proposition 1. In the $I = 1$ case, $\mathbb{P}''[K] = 1$ so

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] = \mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K] > 0, \tag{3.6}$$

where we used that, under $K \cap \mathcal{NC}$, the quartets Q_1 and Q_2 occur with equal probability under $\mathbb{P}''$. Since $x - 1/I = y - 1/I = 0$, the claim follows.

For $I \geq 2$, (3.5) and Lemma 15 implies that

$$\begin{aligned}
\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] \geq\ & (\mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K])\mathbb{P}''[K] \\
& + \mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] + \left(\phi''_+ - \phi''_-\right)\left(\mathbb{P}''[F_{ab}] + \mathbb{P}''[F_{cd}]\right) \\
& - \left(\phi''_+ - \phi''_-\right)\left(\mathbb{P}''[F_{ac}] + \mathbb{P}''[F_{bd}]\right) \\
>\ & \mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] \\
& + \left(\phi''_+ - \phi''_-\right)\left(\mathbb{P}''[F_{ab}] - \mathbb{P}''[F_{ac}] - \mathbb{P}''[F_{bd}] + \mathbb{P}''[F_{cd}]\right),
\end{aligned}$$

where again we used that, under $K \cap \mathcal{NC}$, the quartets Q_1 and Q_2 occur with equal

probability. If $I = 2$, then by Lemma 16 and Lemma 14 parts (a) and (c), this leads to

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] > \left(x - \frac{1}{I}\right) \wedge \left(y - \frac{1}{I}\right). \tag{3.7}$$

If $I \geq 3$, then by Lemma 14 parts (b) and (d),

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] > \begin{cases} 1/8 & \text{if } x \wedge y \geq 1/2 \\ \frac{1}{12}\left(x - \frac{1}{I}\right) \wedge \left(y - \frac{1}{I}\right) & \text{if } x \wedge y \leq 1/2 \end{cases} \tag{3.8}$$

It follows that $\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] > \frac{1}{12}\left(x - \frac{1}{I}\right) \wedge \left(y - \frac{1}{I}\right)$, finishing the proof of the main claim in the balanced case. □

3.3.2 Caterpillar case

We now consider the caterpillar case. Without loss of generality, assume the species tree restricted to $\mathcal{Q}$ has rooted topology as in the bottom right of Figure 3.2. Let R be the most recent common ancestor of A, B, C and let I be the number of locus copies exiting R (forward in time). Let $\mathbb{P}''$ be the probability measure indicating conditioning on I and $\mathcal{X}$. Let $i_x \in \{1, ..., I\}$ be the ancestral lineage of $x \in \{a, b, c\}$ in R. As with the balanced case, on the complement of $\mathcal{X} \geq \vec{1}$, ASTRAL-one selects Q_1, Q_2, Q_3 each with probability 0. To prove

$$\mathbb{P}[Q_1] > \max\{\mathbb{P}[Q_2], \mathbb{P}[Q_3]\},$$

it is sufficient to prove Proposition 2 below for $\mathcal{X} \geq \vec{1}$.

Proposition 2 (Quartet identifiability: Caterpillar case)**.** *Let $x = \mathbb{P}''[i_a = i_b]$. On the events $I \geq 1$ and $\mathcal{X} \geq \vec{1}$, we have almost surely*

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] > \frac{1}{3}\left(x - \frac{1}{I}\right).$$

Similarly to the balanced case, in order to prove this proposition we consider the

following events:

$$E = a - b - c$$
$$G_{ab} = ab - c \qquad G_{ac} = ac - b \qquad G_{bc} = bc - a$$
$$K = abc,$$

where $-$ indicates separation of lineages in R of the selected copies of A, B, C. Letting $\mathcal{E}$ run across all events, the law of total probability implies

$$\mathbb{P}''[Q_i] = \sum_{\mathcal{E}} \mathbb{P}''[Q_i|\mathcal{E}]\mathbb{P}''[\mathcal{E}]. \tag{3.9}$$

Let $\mathcal{NC}$ be the event that no coalescent event occurs beneath R between the three lineage corresponding to a, b, c.

Analogues to Lemmas 11, 12, 13, 14, and 15 hold with similar proofs.

Lemma 17. *Let $x = \mathbb{P}''[i_a = i_b]$. On the events $\mathcal{X} \geq \vec{1}$ and $I \geq 1$, we have almost surely*

$$x \geq \frac{1}{I}.$$

Lemma 18. *Let the species tree be a rooted caterpillar on four leaves A, B, C, D. For all $I \geq 1$ and $\mathcal{X} \geq \vec{1}$, and any event $\mathcal{E} \in \{E, G_{ab}, G_{ac}, G_{bc}\}$,*

$$\mathbb{P}''[Q_1|\mathcal{E}] \geq \mathbb{P}''[Q_1|\mathcal{E} \cap \mathcal{NC}] \quad \textit{and} \quad \mathbb{P}''[Q_i|\mathcal{E}] \leq \mathbb{P}''[Q_i|\mathcal{E} \cap \mathcal{NC}], \quad i \in \{2, 3\}.$$

Lemma 19. *For $I \geq 2$ and any $\mathcal{X} \geq \vec{1}$, let $x = \mathbb{P}''[i_a = i_b]$. Then the following hold:*

$$\mathbb{P}''[G_{ab}] = \frac{I-1}{I}x$$
$$\mathbb{P}''[G_{ac}] = \mathbb{P}''[G_{bc}] = \frac{1}{I}(1-x).$$

Lemma 20. *On $I \geq 1$, almost surely*

$$\mathbb{P}''[G_{ab}] - \mathbb{P}''[G_{ac}] = x - \frac{1}{I}.$$

Lemma 21. *For any I and any $\mathcal{X} \geq \vec{1}$, we have*

$$\mathbb{P}''[Q_1 | G_{ab} \cap \mathcal{NC}] = \mathbb{P}''[Q_2 | G_{ac} \cap \mathcal{NC}] =: \psi''_+$$

and

$$\mathbb{P}''[Q_2 | G_{ab} \cap \mathcal{NC}] = \mathbb{P}''[Q_1 | G_{ac} \cap \mathcal{NC}] =: \psi''_-.$$

The G events

The following lemma bounds the conditional probability difference for the G events.

Lemma 22. *On the events $I \geq 1$ and $\mathcal{X} \geq \vec{1}$, we have almost surely*

$$\psi''_+ - \psi''_- \geq \frac{1}{3}.$$

Proof. By definition of ψ''_+ and ψ''_-, it suffices to show that $\mathbb{P}''[Q_1 | G_{ab} \cap \mathcal{NC}] - \mathbb{P}''[Q_2 | G_{ab} \cap \mathcal{NC}] \geq 1/3$. The proof of this inequality involves decomposing $G_{ab} \cap \mathcal{NC}$ into a number of subcases, depicted in Figure 3.4, and computing the probabilities of Q_1 and Q_2 in each subcase. Let S be the most recent common ancestor of $\mathcal{Q}$ in the species tree. Let Λ be the event that the ab and c individuals in R descend from a common ancestor in S and let $q = \mathbb{P}''[\Lambda | G_{ab} \cap \mathcal{NC}]$. There are two cases:

1. *(Condition on Λ)* Let $\mathcal{C}$ be the event that gene copies a and b coalesce above R and below the MRCA of loci $i_{ab} = i_a = i_b$ and i_c. Let $q' = \mathbb{P}''[\mathcal{C} | \Lambda, G_{ab}, \mathcal{NC}]$. We claim that $q' \geq 1/2$. To see this, observe that conditional on $G_{ab} \cap \mathcal{NC}$ the loci i_{ab} and i_c share the same ancestral locus at S only if there occurred a duplication event between S and R which is ancestral to both of them. Therefore with probability at

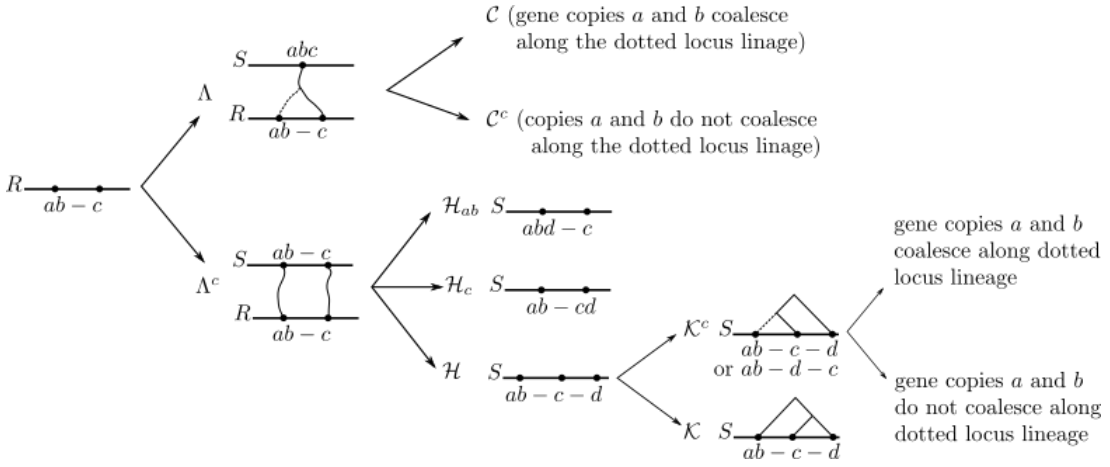


Figure 3.4: Flowchart for case analysis in Lemma 22.

least $1/2$, the gene copies a, b coalesce along their shared pendant edge in the rooted topology between R and S, proving the claim. Furthermore, it is obvious that

$$\mathbb{P}''[Q_1|\mathcal{C}, \Lambda, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{C}, \Lambda, G_{ab}, \mathcal{NC}] = 1. \tag{3.10}$$

On the other hand, conditional on $\mathcal{C}^c$, the copies of a, b and c enter the same population and are then symmetric, and hence

$$\mathbb{P}''[Q_1|\mathcal{C}^c, \Lambda, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{C}^c, \Lambda, G_{ab}, \mathcal{NC}] = 0. \tag{3.11}$$

2. *(Condition on Λ^c)* Let $\mathcal{H}_j$ be the event that copies d and j share the same ancestor in the locus tree at S, and define $\mathcal{H} = (\mathcal{H}_{ab} \cup \mathcal{H}_c)^c$ and $r = \mathbb{P}''[\mathcal{H}|\Lambda^c, G_{ab}, \mathcal{NC}]$. Then by symmetry, $\mathbb{P}''[\mathcal{H}_{ab}|\Lambda^c, G_{ab}, \mathcal{NC}] = \mathbb{P}''[\mathcal{H}_c|\Lambda^c, G_{ab}, \mathcal{NC}] = \frac{1-r}{2}$. By a further symmetry argument similar to that made in Case 1 we have

$$\mathbb{P}''[Q_1|\mathcal{H}_{ab}, \Lambda^c, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{H}_{ab}, \Lambda^c, G_{ab}, \mathcal{NC}] \geq 0, \tag{3.12}$$

where the inequality accounts for the possibility that the lineages from a and b coalesce between R and S. Let τ be the topology of the locus tree restricted to the copies ab, c, d and restricted to the portion above S (and suppressing nodes of degree

2). Conditioned on $\mathcal{H}_c$, we have $\tau = (ab, cd)$, so it must be the case that either ab descends from a daughter lineage or cd descends from a daughter lineage, meaning we observe Q_1 with probability 1 and Q_2 with probability 0. Therefore

$$\mathbb{P}''[Q_1|\mathcal{H}_c, \Lambda^c, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{H}_c, \Lambda^c, G_{ab}, \mathcal{NC}] = 1. \tag{3.13}$$

It remains to consider the case $\mathcal{H} = (\mathcal{H}_{ab} \cup \mathcal{H}_c)^c$. Let $\mathcal{K}$ be the event that $\tau = (ab, (c, d))$. By symmetry, the three possible topologies are equally likely, and so $\mathbb{P}''[\mathcal{K}|\mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] = 1/3$. Conditioned on $\mathcal{K}$, either ab descends from a daughter lineage or the pair (c, d) descends from a daughter lineage, and therefore

$$\mathbb{P}''[Q_1|\mathcal{K}, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{K}, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] = 1. \tag{3.14}$$

Conditioned on $\mathcal{K}^c$, let p denote the probability that gene copies a and b coalesce along the pendant edge for ab in the rooted triple above S. Then

$$\mathbb{P}''[Q_1|\mathcal{K}^c, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] = p + \frac{1}{3}(1 - p),$$

and

$$\mathbb{P}''[Q_2|\mathcal{K}^c, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] = \frac{1}{3}(1 - p),$$

and hence

$$\mathbb{P}''[Q_1|\mathcal{K}^c, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] - \mathbb{P}''[Q_2|\mathcal{K}^c, \mathcal{H}, \Lambda^c, G_{ab}, \mathcal{NC}] \geq p, \tag{3.15}$$

where again the inequality accounts for the possibility that the lineages from a and b coalesce between R and S.

Finally, applying the law of total probability and using equations (3.10)-(3.15) gives

$$\begin{aligned}\mathbb{P}''[Q_1|G_{ab},\mathcal{NC}] - \mathbb{P}''[Q_2|G_{ab},\mathcal{NC}] &\geq q'q + \left(\frac{1-r}{2} + \frac{1}{3}r + \frac{2}{3}pr\right)(1-q)\\ &\geq \frac{1}{2}q + \frac{1}{3}(1-q)\\ &\geq \frac{1}{3},\end{aligned}$$

where the second inequality follows from $q' \geq 1/2$ and $r \geq 0$. □

Proofs of Proposition 2 and Theorems 3 and 4

With that, the proof of Proposition 2 is similar to that of Proposition 1.

In both the balanced and caterpillar cases, observe that $\mathbb{P}''[Q_1] - \mathbb{P}''[Q_3] = \mathbb{P}''[Q_1] - \mathbb{P}''[Q_2]$ by switching the roles of c and d. By Propositions 1 and 2, all species quartet topologies are identifiable and hence we have verified Theorem 3.

Theorem 4 then follows, along similar lines as [67, Theorem 2], from the law of large numbers.

3.3.3 Proof of consistency for ASTRAL-multi

Before finishing the proof of Theorem 2, we give a proof of Theorem 5.

Proof of Theorem 5. Let $\mathcal{N}_{AB|CD}$ (respectively $\mathcal{N}_{AC|BD}, \mathcal{N}_{AD|BC}$) be the number of choices consisting of one gene copy in the gene tree from each species in $\mathcal{Q} = \{A, B, C, D\}$ whose corresponding restriction in t_1 agrees with $AB|CD$ (respectively $AC|BD$, $AD|BC$). Similarly to [67, Theorem 3], it suffices to show that

$$\mathbb{E}[\mathcal{N}_{AB|CD}] > \max\left\{\mathbb{E}[\mathcal{N}_{AC|BD}], \mathbb{E}[\mathcal{N}_{AD|BC}]\right\}. \tag{3.16}$$

Letting again $\mathcal{X} = (\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$, by taking expectation with respect to I in Propositions

1 and 2, we have on the event $\mathcal{X} \geq \vec{1}$ that

$$\mathbb{P}[q = AB|CD \,|\, \mathcal{X}] > \max\{\mathbb{P}[q = AC|BD \,|\, \mathcal{X}], \mathbb{P}[q = AD|BC \,|\, \mathcal{X}]\}, \tag{3.17}$$

where q is the topology of a uniformly chosen quartet among A, B, C, D. Let $\mathcal{M} = \mathcal{ABCD}$ be the number of quartet choices and let $q_i, i = 1, \ldots, \mathcal{M}$ be the corresponding topologies ordered arbitrarily. Because q is a uniform choice, we have

$$\mathbb{P}[q = AB|CD \,|\, \mathcal{X}] = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} \mathbb{P}[q_i = AB|CD \,|\, \mathcal{X}], \tag{3.18}$$

and similarly for the other topologies. Since

$$\mathcal{N}_{AB|CD} = \sum_{i=1}^{\mathcal{M}} \mathbf{1}\{q_i = AB|CD\},$$

and similarly for the other topologies, taking expectations and using (3.17) and (3.18) gives (3.16) as claimed. □

3.4 Proof of sample complexity bound

To prove Theorem 2, our sample complexity result for ASTRAL-one, we use a union bound over all quartets and build on the analysis of Section 3.3. In particular, the key step of the proof is a more careful analysis of the event K that appeared in the proof of Theorem 3. We first discuss a number of quantities that play an important role in the analysis.

3.4.1 Bounds on branching process quantities

We highlight the role of a number of parameters in the sample complexity: the shortest branch length in the species tree, f; the depth of the species tree, Δ; and the duplication and loss rates, λ and μ. These parameters enter the analysis through three quantities of significance:

- *Coalescence of a pair of lineages on an edge:* In the standard coalescent, the proba-

bility that a pair of lineages has coalesced by time f is

$$\gamma = 1 - e^{-f}.$$

- *Survival probability of a quartet:* For a quartet $\mathcal{Q} = \{A, B, C, D\}$, let $\mathcal{X}_{\mathcal{Q}} = (\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ be the number of gene copies in the corresponding species. The smallest probability over all quartets that a gene family contains a copy in each species will be denoted by

$$\sigma := \min_{\mathcal{Q}} \mathbb{P}\left[\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\right].$$

- *Expected number of lineages at a vertex:* For any vertex R in the species tree, let I_R be the number of copies at R in a single gene family. The largest expectation of I_R over all vertices will be denoted by

$$\alpha = \max_{R} \mathbb{E}[I_R].$$

These last two quantities can be controlled using branching process theory. See e.g. [94, Section 9.2] for the relevant results in the phylogenetic context. We use the notation $z_1 \vee z_2 = \max\{z_1, z_2\}$.

Lemma 23. *The following hold:*

- *When* $\mu > \lambda$,

$$\sigma \geq \left[\frac{1}{e^{(\mu-\lambda)\Delta}}\left(1 - \frac{\lambda}{\mu}\right)\right]^4.$$

- *When* $\lambda > \mu$,

$$\sigma \geq \left[1 - \frac{\mu}{\lambda}\right]^4.$$

Proof. Let $\mathcal{Q} = \{A, B, C, D\}$ be a quartet and let as before $\mathcal{X}_\mathcal{Q} = (\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ be the number of gene copies in the corresponding species. Assume the species tree topology on $\mathcal{Q}$ is balanced (the argument in the caterpillar case being similar). The probability that $\mathcal{A} \geq 1$ is given by

$$\mathbb{P}[\mathcal{A} \geq 1] = 1 - \frac{\mu}{\lambda} q(\Delta),$$

where

$$q(t) = \lambda \frac{1 - e^{-(\lambda-\mu)t}}{\lambda - \mu e^{-(\lambda-\mu)t}}.$$

It can be checked that, whether $\mu > \lambda$ or $\mu < \lambda$, the function $q(t)$ is increasing in t. Conditioned on $\{\mathcal{A} \geq 1\}$, there is at least one copy in each vertex along the path between the root and A. Hence

$$\mathbb{P}[\mathcal{D} \geq 1 \,|\, \mathcal{A} \geq 1] \geq 1 - \frac{\mu}{\lambda} q(\Delta).$$

Repeating this argument for B and C gives

$$\mathbb{P}\left[\mathcal{X}_\mathcal{Q} \geq \vec{1}\right] \geq \left(1 - \frac{\mu}{\lambda} q(\Delta)\right)^4.$$

It remains to bound the right-hand side. When $\lambda > \mu$, $q(t) \to 1$ as $t \to +\infty$, which implies $q(t) \leq 1$ by monotonicity. So $1 - \frac{\mu}{\lambda} q(\Delta) \geq 1 - \frac{\mu}{\lambda}$. On the other hand, when $\mu > \lambda$,

$$\begin{aligned} 1 - \frac{\mu}{\lambda} q(\Delta) &= \lambda \frac{1 - e^{-(\lambda-\mu)\Delta}}{\lambda - \mu e^{-(\lambda-\mu)\Delta}} \\ &= \frac{\mu - \lambda}{\mu e^{(\mu-\lambda)\Delta} - \lambda} \\ &\geq \frac{1}{e^{(\mu-\lambda)\Delta}} \left(1 - \frac{\lambda}{\mu}\right). \end{aligned}$$

□

Lemma 24. *We have*

$$\alpha \leq 1 \vee e^{(\lambda-\mu)\Delta}.$$

Proof. Recall that we assume there is a single lineage at the top pendant vertex of the species tree. If the time elapsed between this vertex and another vertex U is d, then the expectation number of lineages at U is $e^{(\lambda-\mu)d}$. The result follows from the fact that $d \leq \Delta$ by considering separately the cases $\mu > \lambda$ and $\lambda > \mu$. □

3.4.2 Sufficient effective number of samples

For a quartet $\mathcal{Q}$, let $\mathcal{K}_{\mathcal{Q}}$ be the set of gene trees such that each species in $\mathcal{Q}$ has at least one gene copy. For $k^* \geq 1$, let

$$\mathcal{S}_{k^*} = \{|\mathcal{K}_{\mathcal{Q}}| \geq k^* : \forall \mathcal{Q}\}.$$

Lemma 25. *For any $k^* \geq 1$ and $\epsilon \in (0,1)$, it holds that $\mathbb{P}[\mathcal{S}_{k^*}] \geq 1-\epsilon$ provided*

$$k \geq \left\{\frac{2k^*}{\sigma}\right\} \vee \left\{\frac{8}{\sigma^2}\log\frac{n}{\epsilon}\right\}.$$

Proof. For any $\mathcal{Q}$, by the definition of $\mathcal{K}_{\mathcal{Q}}$, if k is the number of loci then

$$\mathbb{E}|\mathcal{K}_{\mathcal{Q}}| \geq k\sigma.$$

Assume k is large enough that $\frac{1}{2}k\sigma \geq k^*$. Then by the union bound and Hoeffding's

inequality (see, e.g. [140]), using the fact that the gene trees are independent,

$$\begin{aligned}
\mathbb{P}\left[\mathcal{S}_{k^*}^c\right] &\leq \sum_{\mathcal{Q}} \mathbb{P}\left[|\mathcal{K}_{\mathcal{Q}}| < k^*\right] \\
&\leq \sum_{\mathcal{Q}} \mathbb{P}\left[|\mathcal{K}_{\mathcal{Q}}| < \frac{1}{2}k\sigma\right] \\
&\leq \sum_{\mathcal{Q}} \mathbb{P}\left[\mathbb{E}|\mathcal{K}_{\mathcal{Q}}| - |\mathcal{K}_{\mathcal{Q}}| > \frac{1}{2}k\sigma\right] \\
&\leq n^4 \exp\left(-2\frac{(k\sigma/2)^2}{k}\right) \\
&\leq \epsilon,
\end{aligned}$$

if

$$k \geq \frac{2}{\sigma^2}\log\frac{n^4}{\epsilon},$$

and since $\epsilon < 1$ this inequality holds whenever

$$k \geq \frac{8}{\sigma^2}\log\frac{n}{\epsilon}.$$

That proves the claim. □

3.4.3 The event K

Using the notation of Section 3.3, fix a quartet of species $\mathcal{Q} = \{A, B, C, D\}$, let $\mathbb{P}'$ denote the conditional probability given the event $\{\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\}$, and define $\delta' = \mathbb{P}'[Q_1] - 1/3$. Since $\mathbb{P}'[Q_2] = \mathbb{P}'[Q_3]$, we have

$$\delta' = \mathbb{P}'[Q_1] - \frac{\mathbb{P}'[Q_1] + \mathbb{P}'[Q_2] + \mathbb{P}'[Q_3]}{3} = \frac{2}{3}\left(\mathbb{P}'[Q_1] - \mathbb{P}'[Q_2]\right) \tag{3.19}$$

We seek to bound the right-hand side. Assume first that $T^{\mathcal{Q}}$ is balanced. Let $\mathbb{P}'_i$ indicate $\mathbb{P}'$ conditioned on $\{I = i\}$. By the proof of Proposition 1, specifically the argument leading

up to (3.6), (3.7) and (3.8), we have

$$\mathbb{P}''[Q_1] - \mathbb{P}''[Q_2] \geq (\mathbb{P}''[Q_1|K] - \mathbb{P}''[Q_2|K])\mathbb{P}''[K].$$

By further conditioning on the event $\{\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\}$ (which has positive probability when $I \geq 1$), we get for all $i \geq 1$

$$\mathbb{P}'_i[Q_1] - \mathbb{P}'_i[Q_2] \geq (\mathbb{P}'_i[Q_1|K] - \mathbb{P}'_i[Q_2|K])\mathbb{P}'_i[K].$$

On the event $K \cap \mathcal{NC}$, Q_1 and Q_2 are equally likely by symmetry. On $K \cap \mathcal{NC}^c$, at least one of the pairs $\{a, b\}$ or $\{c, d\}$ coalesces below R, guaranteeing Q_1 (similarly to the proof of Lemma 12). See Figure 3.5 for an illustration. Hence, we have

$$\begin{aligned}
\mathbb{P}'_i[Q_1] - \mathbb{P}'_i[Q_2] &\geq \mathbb{P}'_i[Q_1 \cap K] - \mathbb{P}'_i[Q_2 \cap K] \\
&= \mathbb{P}'_i[Q_1 \cap K \cap \mathcal{NC}] - \mathbb{P}'_i[Q_2 \cap K \cap \mathcal{NC}] \\
&\quad + \mathbb{P}'_i[Q_1 \cap K \cap \mathcal{NC}^c] - \mathbb{P}'_i[Q_2 \cap K \cap \mathcal{NC}^c] \\
&\geq \mathbb{P}'_i[K \cap \mathcal{NC}^c].
\end{aligned}$$

To bound the right-hand side, we consider the event $\mathcal{C}_{ab}$ that the lineages picked from A and B coalesce below R. Notice in particular that $\mathcal{C}_{ab}$ implies $\{i_a = i_b\}$. The event $K \cap \mathcal{NC}^c$ is implied by $\mathcal{C}_{ab}$ together with $\{i_c = i_d = i_a\}$, which are conditionally independent. By Lemma 11, the latter has probability at least $\mathbb{P}'_i[i_c = i_d]\frac{1}{i} \geq \frac{1}{i^2}$. Hence,

$$\mathbb{P}'_i[Q_1] - \mathbb{P}'_i[Q_2] \geq \mathbb{P}'_i[\mathcal{C}_{ab}]\,\frac{1}{i^2}. \tag{3.20}$$

By a similar argument in the caterpillar case, we have

$$\begin{aligned}\mathbb{P}_i'[Q_1] - \mathbb{P}_i'[Q_2] &\geq \mathbb{P}_i'[K \cap \mathcal{NC}^c] \\ &\geq \mathbb{P}_i'[\mathcal{C}_{ab}]\,\frac{1}{i} \\ &\geq \mathbb{P}_i'[\mathcal{C}_{ab}]\,\frac{1}{i^2}. \end{aligned} \tag{3.21}$$

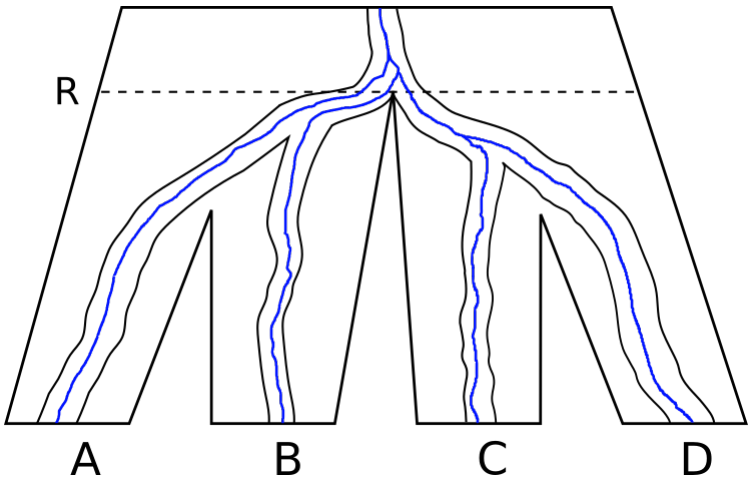


Figure 3.5: An example realization of $K \cap \mathcal{NC}^c$. In particular, a locus tree satisfying event K for copies a, b, c, d is depicted in the case where $T^{\mathcal{Q}}$ is balanced; additionally, a realization of the gene tree is shown (in blue) within the locus tree. This realization satisfies $K \cap \mathcal{NC}^c$ since at least one coalescence event—in this case between the ancestors of gene copies c and d—has occurred on the gene tree below R. Since copies c and d are therefore not separated by an internal edge on the gene tree, the event Q_1 occurs.

It remains to bound $\mathbb{P}_i'[\mathcal{C}_{ab}]$.

Lemma 26. *We have*

$$\mathbb{P}_i'[\mathcal{C}_{ab}] \geq \left\{\gamma \wedge \frac{1}{8}\right\}\frac{1}{i}.$$

Proof. Similarly to the proof of Lemma 11, for copy ℓ at R, let N_ℓ be the number of its descendant copies at R', the most recent common ancestor of A and B, and let $J = \sum_{\ell=1}^{i} N_\ell$. We consider two cases for N_ℓ:

1. In the case $N_\ell = 1$ and $\{i_a = i_b = \ell\}$, let Y_ℓ be the indicator function of the event that

the lineages from a and b coalesce before R. So $Y_\ell = 1$ if the lineages from a and b coalesce before R, and 0 otherwise. Under the standard coalescent, the probability of that coalescent event is at least γ. Here, though, we are working under the bounded coalescent. In the case that a daughter edge is ancestral to the lineages from a and b, the additional conditioning on complete coalescence only increases the probability that a and b coalesce before R. So γ remains a lower bound.

2. In the case $N_\ell \geq 2$ and $\{i_a = i_b = \ell\}$, let Z_ℓ be the indicator function of the event that the lineages from a and b coalesce before R. So $Z_\ell = 1$ if the lineages from a and b coalesce before R, and 0 otherwise. Since $N_\ell \geq 2$, there is at least one duplication below ℓ before R'. By symmetry, there is probability at least $1/2$ that the first duplication produces a daughter edge with at least half of the descendants of ℓ below it at R'. Under $\{i_a = i_b = \ell\}$, there is then a probability at least $1/4$ that a and b descend from copies at R' below that daughter edge. So overall there is probability at least $1/8$ that $Z_\ell = 1$ in that case.

Putting these two cases together , we get

$$\begin{aligned}
\mathbb{P}'_i[\mathcal{C}_{ab}] &= \mathbb{E}'_i\left[\mathbb{E}'_i\left[\sum_{\ell:N_\ell=1}\left(\frac{N_\ell}{J}\right)^2 Y_\ell + \sum_{\ell:N_\ell>1}\left(\frac{N_\ell}{J}\right)^2 Z_\ell \,\middle|\, (N_\ell)_{\ell=1}^i\right]\right] \\
&= \mathbb{E}'_i\left[\sum_{\ell:N_\ell=1}\left(\frac{N_\ell}{J}\right)^2 \mathbb{E}'_i\left[Y_\ell \,|\, N_\ell\right] + \sum_{\ell:N_\ell>1}\left(\frac{N_\ell}{J}\right)^2 \mathbb{E}'_i\left[Z_\ell \,|\, N_\ell\right]\right] \\
&\geq \left\{\gamma \wedge \frac{1}{8}\right\} \mathbb{E}'_i\left[\sum_{\ell:N_\ell=1}\left(\frac{N_\ell}{J}\right)^2 + \sum_{\ell:N_\ell>1}\left(\frac{N_\ell}{J}\right)^2\right] \\
&\geq \left\{\gamma \wedge \frac{1}{8}\right\} \frac{1}{i},
\end{aligned}$$

as in [67, Lemma 1], proving the claim. □

Lemma 27. *We have*

$$\delta' \geq \frac{2}{3}\left\{\gamma \wedge \frac{1}{8}\right\}\frac{\sigma^3}{\alpha^3}.$$

Proof. By (3.19), (3.20), (3.21), and Lemma 26,

$$\begin{aligned}\delta' &= \frac{2}{3}\left(\mathbb{P}'[Q_1] - \mathbb{P}'[Q_2]\right) \\ &\geq \frac{2}{3}\left\{\gamma \wedge \frac{1}{8}\right\}\mathbb{E}'\left[\frac{1}{I^3}\right] \\ &\geq \frac{2}{3}\left\{\gamma \wedge \frac{1}{8}\right\}\frac{1}{\mathbb{E}'\left[I\right]^3},\end{aligned} \tag{3.22}$$

where the last line follows from Jensen's inequality. Moreover

$$\begin{aligned}\alpha &\geq \mathbb{E}[I] \\ &= \mathbb{E}\left[I\middle|\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\right]\mathbb{P}\left[\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\right] + \mathbb{E}\left[I\middle|\{\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\}^c\right]\mathbb{P}\left[\{\mathcal{X}_{\mathcal{Q}} \geq \vec{1}\}^c\right] \\ &\geq \mathbb{E}'[I]\,\sigma.\end{aligned}$$

Plugging back into (3.22) gives the claim. □

3.4.4 Final analysis

Proof of Theorem 2. The following, adapted from [138, Lemmas A.1 and A.2], gives a bound on the k^* required to reconstruct the correct species tree with probability $1-\epsilon$ in terms of δ'

$$k^* > 2\log\left(\frac{n^4}{\epsilon}\right)\frac{1}{(\delta')^2}.$$

In particular, for $\epsilon < 1$ this inequality holds whenever

$$k^* > 8\log\left(\frac{n}{\epsilon}\right)\frac{1}{(\delta')^2}.$$

By Lemmas 25 and 27, it suffices to have

$$\begin{aligned} k &\geq \left\{ \frac{16}{\sigma} \log\left(\frac{n}{\epsilon}\right) \frac{1}{(\frac{2}{3}\{\gamma \wedge \frac{1}{8}\}\frac{\sigma^3}{\alpha^3})^2} \right\} \vee \left\{ \frac{8}{\sigma^2} \log \frac{n}{\epsilon} \right\} \\ &\geq \frac{2304\alpha^6}{\sigma^7\gamma^2} \log \frac{n}{\epsilon}. \end{aligned}$$

The claim follows from Lemmas 23 and 24. □

3.5 Concluding remarks

Through a probabilistic analysis of the DLCoal model, we established identifiability of the model species tree and statistical consistency of quartet-based species tree estimation methods ASTRAL-one and ASTRAL-multi. In our main new result, we derived an upper bound on the required number of gene trees to reconstruct the species tree with high probability. In particular, we highlighted the roles of the branching process parameters λ and μ as well as the tree depth Δ. These parameters enter naturally through two relevant quantities: the minimum survival probability of a quartet (σ) and the maximum expected number of lineages at a vertex (α).

Our results suggest many open problems. First, can we derive a lower bound (and matching upper bound) on the required number of gene trees for ASTRAL-one and similar methods (including ASTRAL-multi)? In particular, is our dependence on α (in the supercritical regime) and σ (in the subcritical regime) optimal? An improvement on the polynomial dependence on α (see Lemma 27) is likely possible with a more detailed analysis of the events in Section 3.3. But, perhaps more importantly, are there alternative ways of processing multi-labeled gene trees (not necessarily quartet-based) that dampen or even exclude the effect of α?

More generally, it would be interesting to obtain statistical consistency and sample complexity results for models also including LGT, under which at low enough rates quartet-based methods have also been shown to be consistent [118]. One such more general model was recently introduced in [125].

Chapter 4

Some Lower Bounds on the Sample Complexity of Species Tree Estimation with Variable Mutation Rates

Abstract: In this chapter we consider the effect of intergenic mutation rate heterogenity on the sample complexity of species tree estimation. We prove two main results. The first main result is a lower bound for distinguishing the 2-leaf trees under the assumption that mutation rates are gamma distributed, possibly with an atom at zero. In particular we show that to distinguish two edges differing by length f with high probability, one requires $O(f^{-2})$ samples. The second main result is a theorem which establishes similar lower bounds under a general class of continuous distributions.

4.1 Introduction

This chapter is concerned with the sample complexity of species tree estimation when genes exhibit rate variation across the genome, that is, when mutation rates may vary randomly between genes. Broadly speaking, the term *sample complexity* refers to the amount of data needed to achieve high probability of correct estimation. The topic of sample complexity includes *lower bounds*, which pertain to how much data is *necessary* for *any* inference method (or any method in specified class of methods) to succeed with high probability; it also includes *upper bounds*, which pertain to how much data is *sufficient* for a *particular* inference method.

This chapter is structured as follows. In Section 4.1, after discussing related work, we introduce our model of gene evolution and describe the general approach for proving lower bounds that will be used. In Section 4.2 we prove the first main result, which considers the special case of distinguishing edges when mutation rates are gamma-distributed with an atom at zero. In Section 4.3, we describe a complication which arises when attempting to generalize the first main result to more general probability distributions, and prove another lower bound in this setting.

4.1.1 Related work

Before introducing our model and results, we briefly review previous work on sample complexity for mixture models of rate variation as well as in the case of the multispecies coalescent (MSC).

RAS rate variation models

Rate variation in the genome has been studied extensively using mixture models, such as the rates-across-sites (RAS) model, which assumes that branch lengths vary between sites, but that all sites share a common tree topology. In the RAS model, branch length variation between sites is obtained by multiplying the mutation rate for each site by a random iid scaling factor; here, a default formulation is the *general time-reversible model*

with gamma-distributed rates and invariable sites, abbreviated GTR + Γ + I. Under this model, each site is either invariant (I) or undergoes a mutational process governed by a continuous-time Markov process with rate matrix rQ, where Q is a general time reversible (GTR) instantaneous rate matrix, and r is an independent gamma-distributed random variable (Γ). A common gamma distribution is used for all sites; typically it is assumed to be normalized to have rate 1, and so depends only on an unknown shape parameter α, which may be regarded as an unknown model parameter [141, 70].

When α is unknown, the question of identifying model parameters (which include Q, α, the stationary distribution, and all features of the species phylogeny) is somewhat subtle, as pairwise sequence comparisons are not sufficient to recover the model parameters [141]. Nonetheless, it has been shown that all parameters are generically identifiable from sequence data under both the GTR + Γ model [70], as well as the GTR + Γ + I model under reasonable assumptions about the species tree branch lengths and rate matrix Q [142]. More general RAS models than GTR + Γ + I were considered by [143] in the setting of the *large-tree limit* (i.e. as the number of leaves n goes to infinity), who showed that under certain regularity assumptions about the species tree and the distribution of mutation rates, and provided that the number of leaves n is sufficiently large, the species tree topology is both identifiable from the distribution at the leaves and—regarding the sample complexity—can be recovered with high probability from DNA sequence data of polynomial length in n.

Sample complexity for models involving the MSC

The model of rate variation considered in this chapter will differ from the RAS models in an important way: rather than requiring that all gene trees share a common topology, it will instead be assumed that gene trees are generated according to the MSC. For data generated in a manner based on the MSC, questions of the sample complexity have been studied from two perspectives: (1) using error-free gene trees as data, and (2) using sequences as data. In the latter case, a 2-step model is typically employed, with the

MSC used together with a site substitution model (e.g., the Jukes-Cantor model), which is utilized to generate sequence data using the gene trees obtained from the MSC as input.

The perspective usually taken here is how the data requirement for consistent inference grows as a function of certain key model parameters. In case (1), the data requirement can be expressed by a single quantity, m, the number of gene trees. In case (2) by contrast, the total data amount of data is the product mk, where k is the length of each gene sequence k. In this latter case, m and k do not play equivalent roles. Increasing one of these variables is not equivalent to increasing the other. Moreover, if the total sequence length mk is fixed, there is in fact a tradeoff between the two.

Much of the work on sample complexity for data generated under the MSC has been aimed at better understanding how inference is affected by this tradeoff between m and k. On one hand, when k is large, the accuracy of estimated gene trees may increase, so that one needs fewer gene trees to estimate the species tree [129]. On the other hand, it *is* possible to have too few gene trees, irrespective of how large k may be. To see this, observe that in both settings (1) and (2), the amount of data needed to recover the species tree with high probability depends critically on the minimal branch length f of the species tree, since under the MSC this parameter has an important effect on the probability of ILS, which can result in the failure to detect internal branches as well as topological discordance between gene trees and species tree. As observed in [144], statistically consistent estimation of the species tree thus requires that *at least one gene tree does not exhibit ILS on the shortest branch of the species tree;* elementary coalescent calculations show this event has probability

$$1 - e^{-mf}. \tag{4.1}$$

As f gets smaller, the estimation problem becomes harder, and so the amount of data m required must correspondingly grow, and the relevant question is:

As f shrinks to zero, how quickly must m grow to ensure high probability of

correct inference?

If $m = o(f^{-1})$, then Eq. (4.1) tends to zero as $f \to 0$. In that case, the probability of correctly estimating the species tree is bounded away from 1, so correct estimation *with high probability* is not possible, regardless of the inference method used. This implies the following lower bound on the required number of genes m: as $f \to 0$, reconstruction of the species tree with high probability requires at the very least that

$$m = \Omega\left(\frac{1}{f}\right), \tag{4.2}$$

and, moreover, this lower bound does not depend on k at all (for further discussion, see [144, 129]).

Another lower bound for species tree estimation from sequence data, established by [145] is that the total sequence length must grow quadratically with f^{-1}. This lower bound was proved under the assumption that data be generated according to a substitution model using the species tree directly as input, a model which does not incorporate the MSC. Nonetheless, this result still has implications for the setting where sequence data is generated according to a 2-step model combining the MSC with substitution model; intuitively, it says that even if all the gene trees were exact copies of the species tree (so that e.g., there is no estimation error due to ILS, variable branch lengths, etc), the total data data required as $f \to 0$ would still be at least

$$mk = O\left(\frac{1}{f^2}\right) \tag{4.3}$$

in order to achieve a high probability of correct inference.

For the setting in which data takes the form of error-free gene trees generated according to the MSC, a number of results have been proved for quartet-based methods. The most popular of these is a suite of methods known as ASTRAL [31], which estimates the species tree from a collection of gene trees by scoring trees based on how many unrooted topological quartets they share with the gene trees. In [146], it was shown that when f is small,

ASTRAL achieves high probability of correctly estimating the species tree topology if and only if that the number of samples is at least $m = \Omega\left(f^2 \log n\right)$, where n is the number of leaves. A similar approach to the upper bound of [146] was used in [147] (i.e., Chapter 3 of this thesis) to show that this upper bound also holds for gene trees generated under the DLCoal model. A different asymptotic regime is considered in [148], which focuses on the probability of incorrect tree estimation as $m \to \infty$ for a fixed species tree S; in that paper it is shown that for general models of evolution (which includes the MSC and may include DLCoal), the error probability of quartet-based methods decays exponentially as $m \to \infty$, and upper bounds are numerically obtained which improve on those in [146] when m is large.

Although mathematically convenient, the use of error-free gene trees as data has important limitations. It is unrealistic to assume that gene trees can be estimated from DNA sequences without error. Moreover, the errors in estimating branch lengths may be large relative to species tree branch lengths, and inference methods which rely on branch lengths may perform markedly worse when this error is introduced [149].

The study of sample complexity from sequence data has seen significant advances over the last decade. In particular, a number of papers [144, 129, 150] have considered the sample complexity in the context of the 2-step MSC+JC model, whereby gene trees are generated according to the MSC, and then the JC model is utilized to generate sequence data using the gene trees as input. The first of these, [144], considered the general setting in which edges of the species tree may be assigned distinct mutation rates and population sizes; in this setting it was shown that there exist distance-based algorithms which can achieve high probability of correct inference for any fixed $k \geq 1$ provided that $m = O(f^{-2})$, thus demonstrating that the lower bound in Eq. (4.3) can be achieved. Under the more restrictive *molecular clock* assumption, which assumes no variation in mutation rates or population sizes between edges of the species tree, [129] showed that for distance-based

methods, the data tradeoff between k and m takes the form

$$m = O\left(\frac{1}{f^2\sqrt{k}}\right) \tag{4.4}$$

which was shown to be the number of genes which, as $f \to 0$, is both necessary and sufficient to achieve high probability of correct inference using distance-based inference methods. In [150], the data requirement tradeoff in Eq. (4.4) was extended to more general species trees using a novel reduction to the molecular clock case.

4.1.2 Key Definitions and Model Description

We begin by introducing key definitions and describing the model of gene evolution used in this chapter to account for variable mutation rates.

A *species tree* S is defined as a rooted edge-weighted tree with leaves $\mathcal{L}_S$, labeled $1, \ldots, n$, where $\mathcal{L}_S$ is regarded as a set of distinct contemporary species (other taxonomic grouping). A species tree represents a hypothesis about the evolutionary history of $\mathcal{L}_S$ and their ancestors, with edges regarded as *populations* and vertices as speciation events. The edge weights, or *branch lengths*, represent some measure of evolutionary distance and are therefore nonnegative.

We use the term *gene* to refer to a non-recombining locus on the genome; that is, a segment of DNA which is passed down in its entirety from parent to offspring. Any gene shared by one or more of the species in $\mathcal{L}$ can be associated with a corresponding *gene tree*, a rooted edge-weighted tree T with leaf labels $\mathcal{L}_T$, along with a function $\phi_T : \mathcal{L}_T \to \mathcal{L}_S$. A gene tree represents the genealogical history of the gene, with edges regarded as lines of descent, or *ancestral lineages*, and vertices corresponding to most-recent common ancestors. The leaves of a gene tree are regarded as homologous copies of the gene, each sampled from a species in $\mathcal{L}$, and ϕ_T simply associates each leaf of the gene tree with the corresponding leaves of S from which they were sampled.

The most commonly-used model of gene evolution is the *multi-species coalescent* (MSC), due to [25], which describes a probability distribution of a gene tree T as a function of a

species tree parameter S [24], and which may be regarded as an algorithm for generating gene tree samples from a species tree [21]. Informally, the algorithm proceeds as follows: one starts with some number of gene copies at the leaves of S (e.g. one gene copy in each species of $\mathcal{L}$), and then traces each of these copies upwards within the edges of S, forming lineages going backwards in time toward the root of S. Any pair of lineages located in the same edge of S are regarded as being in the same population, and pairs of lineages in the same population *coalesce*, or join together into a single lineage, independently at a pre-specified rate (which may depend on which edge of S the lineages are in). The process continues until all lineages have coalesced and only a single lineage remains, an event which occurs with probability one in the root population of S. The output is a gene tree representing the genealogy of the gene copies initially sampled from the leaves of S. A fuller description of this model can be found in [29, 21].

Given a gene which undergoes spontaneous neutral mutation at a rate of μ mutations per site per generation, the corresponding population mutation parameter is

$$\theta := 4N_{\mathrm{e}}\mu,$$

where N_{e} is the effective population size, assuming the gene is autosomal and in a diploid population. We refer to θ as the *mutation rate*. Originally due to [151], θ is also known as the heterozygosity, a standard measure of population genetic diversity, and is the usual mutation rate parameter used in the coalescent since the parameters N_{e} and μ are not separately identifiable from DNA sequence data alone [21, 24]. In this chapter we assume that the mutation rate θ is constant in time and across populations, but is chosen independently for each gene according to a pre-specified probability distribution.

In the MSC, it is usually assumed that branch lengths of the species tree are measured in units of expected number of mutations per site [24]. Under assumption that all genes share a common mutation rate, it is reasonable to use this unit of measurement for branch lengths of the species tree. In this chapter however, due to the assumption of *variable* mutation rates, there is not a unique mutation rate with which to measure the branch

lengths of the species tree. Therefore, we assume instead that the edges of the species tree are measured in *coalescent units*, i.e. in units of $2N_{\mathrm{e}}$ generations, as this does not depend on the rate of mutation of any given gene.

Given a fixed species tree S, the procedure we adopt for modeling gene evolution (i.e. generating a gene tree) has two steps:

1. Sample a gene tree T according to the MSC on S, with the caveat that mutation rates are not taken into account; instead it is assumed that each pair of gene copies in the same population coalesce independently at rate 1, and hence the edge lengths of T are assumed to be in coalescent units. We denote the distribution of T generated in this manner by $\mathcal{M}(S)$.

2. Next, to model unknown mutation-rate heterogeneity between genes, randomly scale T by multiplying each edge of T by a random *scaling factor* $\theta/2$, where θ is a random variable drawn independently from a fixed probability distribution μ_θ. The output is a *mutation-scaled gene tree*, or more simply, *scaled gene tree* $\widetilde{T}$. Since $\theta/2$ is the expected number of mutations per site per coalescent unit, the edge lengths of $\widetilde{T}$ are measured in *expected number of mutations per site*.

For $\lambda \geq 0$, it is convenient to denote by $\lambda \cdot \mathcal{T}$ the tree $\mathcal{T}$ with all edge lengths multiplied by λ. Using this notation a scaled gene tree $\widetilde{T}$ can be written

$$\widetilde{T} = \frac{\theta}{2} \cdot T$$

where $\theta \sim G$ and $T \sim \mathcal{M}(S)$, with $\theta \perp\!\!\!\perp T$. As noted, the appropriate unit with which to measure scaled gene tree edge lengths is in *expected number of mutations per site*; henceforth any quantity referred to as an *evolutionary distance* is to be regarded as using this unit of measurement.

4.1.3 The Problem: Phylogenetic Sample Complexity

Description of the Problem

The primary question we are interested in is understanding how many genes need to be observed to distinguish two distinct n-leaf species trees with high probability. This question of how many samples are *needed* corresponds to establishing an information-theoretic *lower bound*; the complementary question of how many samples are *sufficient* corresponds to establishing an *upper bound*. The former question is the primary focus of this chapter.

In the context of estimating phylogenetic trees, two key quantities related to this question are (1) the *depth* of the species tree, defined as the length of the longest path from the root to one of the leaves, and (2) the *minimal branch length*, defined as the length of the shortest edge on the species tree. Throughout, we shall always use the letters d and f to represent the species tree depth and minimal branch length respectively. Specifically, we ask how many scaled gene trees are needed—as a function of f and d—in order to have a high probability of correctly identifying the latent (unknown) species tree parameter.

To make this precise, let $\tilde{\mathbb{P}}$ and $\tilde{\mathbb{Q}}$ be the distribution of a single scaled gene tree $T \in \mathbb{T}$ generated in the manner described in Section 4.1.2 with species tree parameters $S_{\tilde{\mathbb{P}}}$ and $S_{\tilde{\mathbb{Q}}}$ respectively. Then the joint distribution of $m \geq 1$ independently sampled scaled gene trees $T_1, \ldots, T_m$ is given by the probability measures

$$\tilde{\mathbb{P}}^{\otimes m} \quad \text{and} \quad \tilde{\mathbb{Q}}^{\otimes m}$$

on the product space $\prod_{i=1}^{m} \mathbb{T}$. To establishing a lower bound in this context, it will be enough to estimate the *total variation distance*

$$\|\tilde{\mathbb{P}}^{\otimes m} - \tilde{\mathbb{Q}}^{\otimes m}\|_{\mathrm{TV}} := \sup \left\{ \left| \tilde{\mathbb{P}}^{\otimes m}(A) - \tilde{\mathbb{Q}}^{\otimes}(A) \right| : A \subseteq \prod_{i=1}^{m} \mathbb{T}, A \text{ is measurable} \right\}. \tag{4.5}$$

To see why this is the case, observe that any statistical test to distinguish hypotheses $S_{\tilde{\mathbb{Q}}}$

and $S_{\tilde{\mathbb{Q}}}$ on the basis of m independent samples all drawn from one of the distributions $\tilde{\mathbb{P}}$ or $\tilde{\mathbb{Q}}$ can be identified with its *rejection region*, a measurable set $A \subseteq \prod_{i=1}^{m} \mathbb{T}$ defined as the set of sample points for which $S_{\tilde{\mathbb{P}}}$ is rejected [152]. Given any test A, the maximum of its probabilities of type-I and type-II errors is

$$\max\left\{\tilde{\mathbb{P}}^{\otimes m}(A), \tilde{\mathbb{Q}}^{\otimes}(A^c)\right\}.$$

Taking the infimum over all tests,

$$\begin{aligned}\inf_A \max\left\{\tilde{\mathbb{P}}^{\otimes m}(A), \tilde{\mathbb{Q}}^{\otimes}(A^c)\right\} &\geq \frac{1}{2}\inf_A \left\{\tilde{\mathbb{P}}^{\otimes m}(A) + \tilde{\mathbb{Q}}^{\otimes m}(A^c)\right\} \\ &= \frac{1}{2}\left(1 - \left\|\tilde{\mathbb{P}}^{\otimes m} - \tilde{\mathbb{Q}}^{\otimes m}\right\|_{\mathrm{TV}}\right).\end{aligned}$$

Hence if $\|\tilde{\mathbb{P}}^{\otimes m} - \tilde{\mathbb{Q}}^{\otimes m}\|_{\mathrm{TV}} \leq \epsilon$, then any statistical test to identify tree parameter from m samples drawn from one of the distributions $\tilde{\mathbb{P}}$ or $\tilde{\mathbb{Q}}$ cannot have type-I and type-II error probabilities both less than $\frac{1}{2} - \frac{\epsilon}{2}$. (A fuller, more general version of this argument can be found in Theorem 2.2 in [153].)

Parameter Regime of Interest: $(f, d) \to (0, \infty)$

We are interested in the asymptotic regime in which $(f, d) \to (0, \infty)$, as it is in this case that distinguishing species trees becomes more difficult. Throughout this chapter, the following assumptions about f and d will always be made:

- there exists a constant $d_{\min} > 0$ such that $d \geq d_{min}$
- $0 < f < \frac{1}{2}$
- $\frac{f}{d} < \frac{1}{2}$

These assumptions do little harm given the asymptotic regime of interest. Notably, the assumption that $\frac{f}{d} < \frac{1}{2}$ is automatically satisfied by definition of f and d for any species tree with at least one internal edge. Later in the chapter, for technical reasons—namely,

for proving Lemmas 34 and 36—it will not be enough to assume $\frac{f}{d} < \frac{1}{2}$, but instead will also be necessary to assume that $\frac{f}{d}$ be bounded away from $\frac{1}{2}$.

Hellinger distance

Given two probability measures P and Q defined on the same measure space $\mathcal{X}$, the *Hellinger distance*, denote $H(P,Q)$ is defined the L^2-distance between the square roots of their densities; that is,

$$H^2(P,Q) := \int_{\mathcal{X}} \left(\sqrt{f_P} - \sqrt{f_Q}\right)^2 d\nu \tag{4.6}$$

where ν is any measure such that the densities f_P and f_Q are absolutely continuous with respect to ν.

Using the observation that

$$\|P - Q\|_{TV} = \int_{[f_P > f_Q]} (f_P - f_Q) d\nu = \frac{1}{2} \int_{\mathcal{X}} |f_P - f_Q| \, d\nu$$

along with the Cauchy-Schwarz inequality, it is not difficult to show the following inequality (see, e.g., [154]):

$$\begin{aligned} \|P - Q\|_{\mathrm{TV}} &\leq \frac{1}{2} H\left(P,Q\right)\left(4 - H^2\left(P,Q\right)\right)^{\frac{1}{2}} \\ &\leq H(P,Q), \end{aligned} \tag{4.7}$$

so that $H(\tilde{\mathbb{P}}^{\otimes}, \tilde{\mathbb{Q}}^{\otimes})$ can be used to bound Eq. (4.5). The advantage of considering the Hellinger distance rather than the total variation distance directly is that it satisfies the following *tensorization property:*

$$H^2\left(P^{\otimes m}, Q^{\otimes m}\right) = 2 - 2\left(1 - \frac{1}{2} H^2(P,Q)\right)^m \tag{4.8}$$

which makes it useful for dealing with product measures [154, 155], as is the case in Eq. (4.5).

The approach we take to establish lower bounds in this chapter is the same as that used in the proof of Theorem 1 in [73]; namely, we will use an estimate of the squared Hellinger distance $H^2(\tilde{\mathbb{P}}, \tilde{\mathbb{Q}})$ along with Eqs. (4.7) and (4.8) to bound Eq. (4.5). In particular, the argument from Theorem 1 in [73] is summarized in the following lemma.

Lemma 28 (Relating Hellinger Distance to Sample Lower Bound)**.** *Let P, Q be two probability distributions depending on parameters f and d. Then*

$$\left\|P^{\otimes m} - Q^{\otimes m}\right\|_{\mathrm{TV}}^2 \leq mH^2(P, Q). \tag{4.9}$$

for all $m \geq 1$.

Proof of Lemma 28. By Eqs. (4.7) and (4.8)

$$\begin{aligned}\frac{1}{2}\left\|P^{\otimes m} - Q^{\otimes m}\right\|_{\mathrm{TV}}^2 &\leq \frac{1}{2}H^2(P^{\otimes m}, Q^{\otimes m}) \\ &= 1 - \left(1 - \frac{1}{2}H^2(P, Q)\right)^m\end{aligned} \tag{4.10}$$

Moreover, by the mean value theorem

$$1 - (1 - x)^m = m(1 - \xi)^{m-1}x$$

for some $\xi \in (0, x)$. Therefore since $m \geq 1$, it holds that

$$1 - (1 - x)^m \leq mx \tag{4.11}$$

whenever $0 \leq x \leq 1$. It follows trivially from Eq. (4.6) that $H^2(P, Q) \leq 2$. Therefore Eq. (4.11) implies

$$1 - \left(1 - \frac{1}{2}H^2(P, Q)\right)^m \leq \frac{m}{2}H^2(P, Q)$$

Combining this inequality with Eq. (4.10) implies Eq. (4.9). □

Remark 1. *In particular, if $\Psi(f,d)$ is any positive function satisfying*

$$H^2(P,Q) = O(\Psi(f,d))$$

as $(f,d) \to (0,\infty)$, then distinguishing P and Q with high probability requires at least $O(\frac{1}{\Psi(f,d)})$ samples; to see why this is the case, observe that if $m = o(\frac{1}{\Psi(f,d)})$ then by Lemma 28,

$$\left\|P^{\otimes m} - Q^{\otimes m}\right\|_{\mathrm{TV}} = o(1)$$

as $(f,d) \to (0,\infty)$. That is, for any $\delta > 0$ there exists a constant c_δ such that

$$\left\|P^{\otimes m} - \mathbb{Q}^{\otimes m}\right\|_{\mathrm{TV}} \leq \delta$$

whenever

$$m \leq c_\delta \frac{1}{\Psi(f,d)}.$$

4.2 Distinguishing 2-Leaf Trees

4.2.1 The Distributions $\mathbb{Q}$ and $\mathbb{P}_0$

We first consider the simplest case, that of a species tree with only two leaves. Fix $d_{\min} > 0$, and let S_d be a species tree of height $d \geq d_{\min}$ with exactly two leaves, labeled 1 and 2. Assume the leaves are equidistant from the root. Let $f \in (0, d/2)$, and let S_{d-f} be a species tree of height $d - f$, again with two equidistant leaves labeled 1 and 2. As mentioned, both f and d are measured in coalescent units; however, since $\frac{\theta}{2}$ is the expected number of mutations per site per coalescent unit, we may apply a unit conversion to d and f into expected number of mutations per site by scaling them by $\theta/2$:

$$\widetilde{d} := \frac{\theta}{2} \cdot d \quad \text{and} \quad \widetilde{f} := \frac{\theta}{2} \cdot f. \tag{4.12}$$

We will be interested in understanding how difficult it is to distinguish S_d and S_{d-f}

using independent scaled gene trees generated in the manner described in Section 4.1.2. The problem of distinguishing S_d and S_{d-f} involves consideration of the following two probability distributions:

- **The multi-species coalescent on S_{d-f}:** Let $\mathbb{Q}$ be the distribution of a scaled gene tree $\widetilde{T} = \frac{\theta}{2} \cdot T$, where T is drawn according to the multi-species coalescent on S_{d-f} and θ is drawn independently according to some distribution G.
- **Delayed coalescence on S_{d-f}:** Let $\mathbb{P}_0$ be the distribution of $\widetilde{T}$ conditional gene copies 1 and 2 on T coalescing at least d coalescent units in the past. (Note that $\mathbb{P}_0$ is also the distribution of $\widetilde{D}_{12}$ had the gene tree T been drawn according to the MSC with tree parameter S_d rather than S_{d-f}. As such, distinguishing the probability distributions $\mathbb{P}_0$ and $\mathbb{Q}$ is equivalent to distinguishing the species trees S_d and S_{d-f}.)

As shown in [73], $\mathbb{Q}$ can be expressed as a mixture of P_0 and another independent distribution $\mathbb{P}_1$

$$\mathbb{Q} = (1 - \sigma_f)\mathbb{P}_0 + \sigma_f \mathbb{P}_1$$

where $\mathbb{P}_1$ is the distribution of D_{12} conditional on gene copies 1 and 2 coalescing in the time interval $(d - f, d)$, and where $\sigma_f = O(f)$. Consequently, $\mathbb{P}_0$ and $\mathbb{Q}$ are very similar when f is small, and become harder to distinguish as $f \to 0$, as the admixed sparse signal $\mathbb{P}_1$ becomes less noticeable. Viewed in this light, the essential difference between $\mathbb{Q}$ and $\mathbb{P}_0$ is that under $\mathbb{Q}$ the gene copies from leaves 1 and 2 may coalesce at any time after $d - f$, whereas under $\mathbb{P}_0$ coalescence may occur only after time d; that is, under $\mathbb{P}_0$ the gene copies must wait a little longer before coalescence can occur, and hence the gene trees under $\mathbb{P}_0$ will tend to be a little taller. More precisely, given our assumption that any two gene copies in the same population coalesce at rate 1 per coalescent unit. Therefore under $\mathbb{Q}$, the age (in coalescent units) at which gene copies 1 and 2 coalesce has distribution $d - f + Z$ where $Z \sim \exp(1)$. Converted to evolutionary time, the age of coalescence is $\widetilde{d} - \widetilde{f} + \widetilde{Z}_1$ where $\widetilde{Z}_1 := \frac{\theta}{2} \cdot Z_1 \sim \exp(\frac{2}{\theta})$. On the other hand, by the Markov property the age of coalescence in evolutionary time under $\mathbb{P}_0$ is $\widetilde{d} + \widetilde{Z}_2$ where $Z_2 \sim \exp(\frac{2}{\theta})$ is independent.

The distribution of a two-leaf scaled gene tree $\widetilde{T}$ with leaves equidistant from the root is entirely determined by one quantity: the distance between leaves 1 and 2 on the tree. Let $\widetilde{D}_{12}$ the length of the path between leaves 1 and 2 on $\widetilde{T}$. Since $\widetilde{D}_{12}$ is twice the age of coalescence, it follows that

$$\begin{aligned} \widetilde{D}_{12} &\stackrel{d}{=} \begin{cases} 2(\widetilde{d} - \widetilde{f} + \widetilde{Z}_1) & \text{under } \mathbb{Q} \\ 2(\widetilde{d} + \widetilde{Z}_2) & \text{under } \mathbb{P}_0 \end{cases} \\ &\stackrel{d}{=} \begin{cases} \theta(d - f + Z_1) & \text{under } \mathbb{Q} \\ \theta(d + Z_2) & \text{under } \mathbb{P}_0 \end{cases} \end{aligned} \tag{4.13}$$

where $Z_i \sim \exp(1)$ are independent, $i = 1, 2$.

4.2.2 Constant Mutation Rates

To provide a baseline for comparison with the main result of this section, it is helpful to first consider the case in which genes do *not* exhibit random mutation rates, i.e. the case in which the mutation rate is constant and shared by all genes. In that case, we will show that the number of samples required to distinguish S_d and S_{d-f} with high probability is on the order of f^{-1}.

Our goal is to bound the square of the Hellinger distance between $\mathbb{P}_0$ and $\mathbb{Q}$:

$$H^2(\mathbb{P}_0, \mathbb{Q}) = \frac{1}{2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt$$

Proposition 1 (Nonrandom Mutation Rates)**.** *If $\theta = \theta_0 > 0$ is a fixed constant not varying between genes, then*

$$H^2(\mathbb{P}_0, \mathbb{Q}) = O(f).$$

as $(f, d) \to (0, \infty)$.

Proof of Proposition 1. By Eq. (4.13), it holds that under $\mathbb{Q}$,

$$\widetilde{D}_{12} \stackrel{d}{=} \theta_0(d - f + Z)$$

where $Z \sim \exp(1)$. Let F_Z and f_Z denote the distribution function and density function of Z respectively. Let $t \geq 0$. Then

$$\begin{aligned}
\mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] &= \mathbb{Q}\left[\theta_0(d - f + Z) \leq t\right] \\
&= \mathbb{Q}\left[Z \leq \frac{t}{\theta_0} - d + f\right] \\
&= F_Z\left(\frac{t}{\theta_0} - d + f\right)
\end{aligned}$$

Differentiating both sides with respect to t, it follows that the density function of $\widetilde{D}_{12}$ under $\mathbb{Q}$ is

$$f_{\widetilde{D}_{12};\mathbb{Q}}(t) := f_Z\left(\frac{t}{\theta_0} - d + f\right) = \frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d-f} \cdot \mathbf{1}_{[t>\theta_0(d-f)]}.$$

By a similar calculation, the density function of $\widetilde{D}_{12}$ under $\mathbb{P}_0$ is

$$f_{\widetilde{D}_{12};\mathbb{P}_0}(t) = \frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d} \cdot \mathbf{1}_{[t>\theta_0 d]}.$$

Noting that the intersection of the supports of these two densities is the interval $[\theta_0 d, \infty)$,

$$\begin{aligned}
2H^2\left(\mathbb{P}_0, \mathbb{Q}\right) &= \int_0^\infty \left(\sqrt{f_{\widetilde{D}_{12};\mathbb{P}_0}(t)} - \sqrt{f_{\widetilde{D}_{12};\mathbb{Q}}(t)}\right)^2 dt \\
&= \int_{\theta_0(d-f)}^{\theta_0 d} \frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d-f} dt + \int_{\theta_0 d}^\infty \left(\sqrt{\frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d}} - \sqrt{\frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d-f}}\right)^2 dt \\
&= \left(1 - e^{-f}\right) + \int_{\theta_0 d}^\infty \frac{1}{\theta_0} e^{-\frac{t}{\theta_0}+d} \left(1 - \sqrt{e^{-f}}\right)^2 dt \\
&= \left(1 - e^{-f}\right) + \left(1 - e^{-\frac{f}{2}}\right)^2 \\
&= O(f)
\end{aligned}$$

as $(f, d) \to (0, \infty)$. □

In the proof of Proposition 1, the $O(f)$ term arises in the region $(\theta_0(d-f), \theta_0 d)$, which is in the support of $f_{D_{12};\mathbb{Q}}$ but not $f_{D_{12};\mathbb{P}_0}$. This suggests that in order to distinguish S_d and S_{d-f}, one possible strategy is to look for genes with $\widetilde{D}_{12} \in (\theta_0(d-f), \theta_0 d)$. Since this cannot occur under S_d, observing a scaled gene tree exhibiting such a value of $\widetilde{D}_{12}$ would

immediately imply that the gene tree distribution is parameterized by S_{d-f} rather than S_d, and under $\mathbb{Q}$ the probability of observing such a gene tree is $O(f)$.

4.2.3 Random Mutation Rates

Overview

In this section, we will establish an information-theoretic lower bound for distinguishing two-leaf species trees when genes exhibit random mutation rates drawn from a fixed distribution. As noted in the introduction, the parameters of interest will be the minimal branch length f and the height of the tree d. We will show in this setting,

$$H^2\left(\mathbb{P}_0, \mathbb{Q}\right) = O(f^2)$$

given reasonable assumptions about the distribution of θ. That is, when genes exhibit random independent mutation rates, the number of samples required to distinguish S_d and S_{d-f} with high probability goes from order f^{-1} to order f^{-2}, when f is small. The height d has no effect on the number of samples required, as $d \to \infty$.

Assumptions about the distribution of θ

Throughout we will take θ to be a nonnegative random variable. A commonly used distribution for modeling mutation rates is the gamma distribution, or a mixture of a gamma distribution with a Dirac point mass at zero (i.e. an "atom" at zero) [69, 70, 71]. While there is not much evidence to justify the use of the gamma distribution (aside from mathematical tractability), the consideration of a point mass at zero is important to model genes which are invariant, for example genes which are sufficiently conserved that they do not change at all over long periods of time [69].

Motivation for guessing $H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$

Suppose that there exists some $a > 0$ such that $\theta \geq a$ almost surely. In this case, the motivation for why $H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$ rather than $O(f)$ is as follows. Given two taxa separated by D_{12} coalescent units, and having mutation rate $\theta/2$, the evolutionary distance is then $\widetilde{D}_{12} := (\theta/2) \cdot D_{12}$. By Eq. (4.13), $\widetilde{D}_{12} > a(d-f)$ with probability one under $\mathbb{Q}$, and under $\mathbb{P}_0$ it holds that $\widetilde{D}_{12} > ad$ with probability one. Therefore, if given a number of scaled gene trees all sampled from an unknown distribution either $\mathbb{Q}$ or $\mathbb{P}_0$, one strategy is to look for samples exhibiting $\widetilde{D}_{12} < ad$. This event cannot happen for gene trees sampled from $\mathbb{P}_0$, so if it happens at least once, then the samples must be distributed according to $\mathbb{Q}$ rather than $\mathbb{P}_0$. On the other hand, if one observes a large number of samples and *none* of them exhibit $\widetilde{D}_{12} < ad$, then it is unlikely that they are distributed according to $\mathbb{Q}$.

How many gene trees must one observe for this event to have occurred with high probability? To answer this, observe that in order for a gene tree drawn from $\mathbb{Q}$ to have $\widetilde{D}_{12} < ad$, the following two conditions are necessary:

- $a \leq \theta < \frac{ad}{d-f}$
- $d - f < D_{12} < d$,

i.e., both θ and D_{12} must be small. Moreover, both of these events are of probability $O(f)$, and since they are by assumption independent, this event has $O(f^2)$ probability. This event occurs with high probability provided one has sampled a number of gene trees on the order of f^{-2}.

The takeaway is that with $O(f^{-2})$ samples, one can distinguish S_d and S_{d-f} with high probability. It is reasonable to guess that one *requires* $O(f^{-2})$ samples to distinguish the two trees. As we will show, this guess turns out to be mostly–but not entirely–correct.

The scaled densities p and q

Let p and q be the probability density functions of the $\widetilde{D}_{12}$ under the probability measures $\mathbb{P}_0$ and $\mathbb{Q}$ respectively. We refer to p and q as the *scaled densities*, as they represent the

density of the random variable $\widetilde{D}_{12}$, which is the distance D_{12} in coalescent units after being scaled by the mutation rate $\theta/2$. Formulas for p and q are derived in the following lemma.

Lemma 29 (Scaled Density Formulas)**.** *If θ is a nonnegative continuous random variable with density g such that $a := \inf \operatorname{supp}(g)$, then*

$$q(t) = e^{d-f} \int_a^{t/(d-f)} \frac{1}{x} e^{-\frac{t}{x}} g(x)dx \cdot \mathbf{1}_{[t \geq a(d-f)]} \tag{4.14}$$

$$p(t) = e^{d} \int_a^{t/d} \frac{1}{x} e^{-\frac{t}{x}} g(x)dx \cdot \mathbf{1}_{[t \geq ad]} \tag{4.15}$$

and in particular, $\inf \operatorname{supp}(q) = a(d-f)$ *and* $\inf \operatorname{supp}(p) = ad$.

Proof of Lemma 29. Let $Z \sim \exp(1)$ be independent from θ. Then by Eq. (4.13),

$$\begin{aligned}
\mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] &= \mathbb{Q}\left[2\left(\widetilde{d} - \widetilde{f} + \widetilde{Z}_1\right) \leq t\right] \\
&= \mathbb{Q}\left[\widetilde{Z}_1 \leq \frac{t}{2} - \widetilde{d} + \widetilde{f}\right] \\
&= \mathbb{Q}\left[\frac{\theta}{2} Z \leq \frac{t}{2} - \frac{\theta}{2}(d-f)\right],
\end{aligned}$$

where $Z \sim \exp(1)$ is independent from θ. Let F_Z and f_Z denote the distribution function of Z and density of Z respectively. Since θ is $\mathbb{Q}$-a.s. positive,

$$\mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] = \mathbb{Q}\left[Z \leq \frac{t}{\theta} - d + f\right].$$

Therefore by the law of total expectation, and taking $\mathbb{E}_{\mathbb{Q}}$ to be the expectation operator with respect to the measure $\mathbb{Q}$,

$$\begin{aligned}
\mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] &= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{Q}\left[Z \leq \frac{t}{\theta} - d + f \mid \theta\right]\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[F_Z\left(\frac{t}{\theta} - d + f\right)\right] \\
&= \int_a^{\infty} F_Z\left(\frac{t}{x} - d + f\right) g(x)dx.
\end{aligned}$$

By the Leibniz rule, differentiating both sides with respect to t implies

$$\begin{aligned} q(t) &= \int_a^\infty f_Z\left(\frac{t}{x} - d + f\right)\frac{1}{x}g(x)dx \\ &= \int_a^\infty \mathbf{1}_{[t\geq x(d-f)]}\frac{1}{x}e^{-\frac{t}{x}+d-f}g(x)dx. \end{aligned}$$

This proves Eq. (4.14). Replacing $\mathbb{Q}$ by $\mathbb{P}_0$, the same argument with $f = 0$ implies Eq. (4.15). □

4.2.4 First main result: gamma-distributed rates with atom at zero

Let $\delta_{\{0\}}$ denote the *Dirac measure* at $\{0\}$ on $\mathbb{R}$, defined as the measure on $\mathbb{R}$ for which

$$\delta_{\{0\}}(A) = \begin{cases} 1 & : \quad 0 \in A \\ 0 & : \quad 0 \notin A \end{cases}$$

for every Borel set $A \subset \mathbb{R}$.

Let μ_θ be the distribution of θ. We say that θ is *Gamma distributed with an atom at zero* if there exists a constant $\lambda \in (0,1)$ such that

$$\mu_\theta = \lambda\delta_{\{0\}} + (1-\lambda)g(x)dx$$

where dx is the Lebesgue measure and g is the probability density function of a Gamma distribution (i.e. $g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ for some $\alpha, \beta > 0$).

In this section we show that if θ is either Gamma distributed or Gamma distributed with an atom at zero then

$$H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$$

as $(f, d) \to (0, \infty)$.

To prove this we prove a slightly stronger result, which will rely on the following growth condition:

A.1 (Growth condition). There exists a constant $C_{\text{gc}} > 0$ such that for all $s > 0$,

$$\sup_{x \in (s,2s)} g(x) \leq C_{\text{gc}} \inf_{x \in (\frac{s}{2},s)} g(x). \tag{4.16}$$

Remark 2. *The assumption **A.1**, while satisfied by a reasonably large class of distribution functions (see Corollary 2), is nonethelss more restrictive than necessary; for example, if $g(x) = 0$ for some $x > 0$, then **A.1** implies that $g \equiv 0$ on $[x, \infty)$, a seemingly arbitrary restriction which causes **A.1** to be violated by any distribution g with $a > 0$. In Section 4.3, we will show that **A.1** can be replaced by a substantially less restrictive set of assumptions about g.*

The precise statement of the main result is as follows:

Theorem 6 (First Main Result: Lower Bound for Distinguishing 2-Leaf Trees)**.** *Let $\lambda \in [0, 1)$. Assume θ has distribution*

$$\mu_\theta = \lambda \delta_{\{0\}} + (1 - \lambda) g(x) dx$$

*where g is a probability density function with $g(x) > 0$ for all $x > 0$. Further assume that g satisfies **A.1**. Then there exists a constant $C > 0$ not depending on f or d such that*

$$H^2(\mathbb{P}_0, \mathbb{Q}) \leq C f^2$$

for all $d > d_{\min}$ and all $f < \frac{1}{2} \wedge \frac{d}{2}$.

Before digging into the proof of Theorem 6, we first state the following important corollary.

Corollary 2. *If θ is Gamma distributed, or is Gamma distributed with an atom at zero, then*

$$H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$$

as $(f, d) \to (0, \infty)$.

Proof of Corollary 2. If θ has density $g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ with $\alpha, \beta > 0$, then clearly $g(x) > 0$ for all $x > 0$. It remains only to show that g satisfies **A.1**, in which case the assumptions of Theorem 6 will be satisfied. To see this, observe that

$$\begin{aligned}\frac{\sup\{g(y) : s \le y \le 2s\}}{\inf\left\{g(x) : \frac{s}{2} \le x \le s\right\}} &= \frac{\sup\left\{y^{\alpha-1} : s \le y \le 2s\right\}}{\inf\left\{x^{\alpha-1} : \frac{s}{2} \le x \le s\right\}} \\ &\le 4^{\alpha-1} \vee 1.\end{aligned}$$

This implies that Eq. (4.16) is satisfied with $C_{\text{gc}} = 4^{\alpha-1} \vee 1$. □

The proof of Theorem 6 consists of four lemmas. The first of these captures an intermediate estimate commonly used in this and subsequent sections.

Lemma 30 (Integral Calculation)**.** *If $\gamma > 1$ then there exists a constant $C_\gamma^{(j)} > 0$ not depending on f such that for all $f < \frac{1}{2} \wedge \frac{d}{2}$ and $d \ge d_{\min}$, the following inequalities hold:*

$$\frac{\left(\int_{d-f}^{d} \frac{1}{u} e^{-u} du\right)}{\int_{d}^{\gamma d} \frac{1}{u} e^{-u} du} \le C_\gamma^{(1)} f \quad \textit{and} \quad \frac{\left(\int_{d-f}^{d} \frac{1}{u} e^{-u} du\right)^2}{\int_{d}^{\gamma d} \frac{1}{u} e^{-u} du} \le \frac{C_\gamma^{(2)} f^2}{d e^d},$$

where

$$C_\gamma^{(j)} = \frac{\gamma \left(2\sqrt{e}\right)^j}{1 - e^{-(\gamma-1)d_{\min}}}, \quad j = 1, 2.$$

Proof of Lemma 30. First observe that

$$\begin{aligned}\int_{d-f}^{d} \frac{1}{u} e^{-u} du &\le \frac{1}{d-f}\left(e^{-(d-f)} - e^{-d}\right) \\ &= \frac{e^{-d}}{d-f}\left(e^{f} - 1\right) \\ &\le e^{-(d-f)}\left(\frac{f}{d-f}\right) \\ &< \frac{2\sqrt{e} e^{-d} f}{d}, \end{aligned} \tag{4.17}$$

where the final inequality follows from the assumptions that $f < 1/2$ and $f \le d/2$.

Second, since $\gamma > 1$,

$$\int_d^{\gamma d} \frac{1}{u} e^{-u} du \geq \frac{1}{\gamma d} \int_d^{\gamma d} e^{-u} du = \frac{1}{\gamma d} \left(e^{-d} - e^{-\gamma d} \right) = \frac{1 - e^{-(\gamma-1)d}}{\gamma d e^d} > 0. \tag{4.18}$$

Combining bounds in Eqs. (4.17) and (4.18) gives the inequality

$$\frac{\left(\int_{d-f}^{d} \frac{1}{u} e^{-u} du \right)^j}{\int_d^{\gamma d} \frac{1}{u} e^{-u} du} \leq \frac{\gamma \left(2\sqrt{e}\right)^j}{1 - e^{-(\gamma-1)d}} \left(d e^d \right)^{1-j} f^j \leq \frac{\gamma \left(2\sqrt{e}\right)^j}{1 - e^{-(\gamma-1)d_{\min}}} \left(d e^d \right)^{1-j} f^j$$

for $j = 1, 2$, which implies both inequalities in the statement of the lemma. □

The next lemma utilizes a Taylor expansion to give an estimate of an integral of form

$$\int_S \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt, \quad S \subseteq (ad, \infty) \text{ a bounded interval}$$

by means of a more convenient error term. As with the previous lemma, this lemma will also be used in a subsequent section.

Lemma 31 (Error Term Lemma)**.** *Assume that $g(x) > 0$ for a.e. $x \in (a, a + \delta_+)$, for some $\delta_+ > 0$. Let $S \subseteq (ad, \infty)$ be a measurable set and for each $t \in S$, let*

$$E(t, f) := \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx}{\int_a^{t/d} \frac{1}{x} e^{-t/x} g(x) dx}. \tag{4.19}$$

If $C_E \geq 0$ does not depend on t and $E(t, f) \leq C_E f$ for all $t \in S$, then

$$\int_S \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt \leq \left(\frac{1}{4} C_E^2 \vee 1 \right) f^2$$

for all $f \leq 1/2$.

Proof of Lemma 31. By Lemma 29 and the assumption that $g(t) > 0$ for all $t \in (a, a+\delta_+)$,

it holds that $p(t) > 0$ for all $t \in S$. Therefore

$$\int_S \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt = \int_S p(t) \left(1 - \sqrt{\frac{q(t)}{p(t)}}\right)^2 dt, \tag{4.20}$$

Assume $t \in S$, so that $t > ad$, and let $r := \frac{q(t)}{p(t)}$, it follows by Lemma 29 that

$$\begin{aligned} r &= \frac{e^{-f} \int_a^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx}{\int_a^{t/d} \frac{1}{x} e^{-t/x} g(x) dx} \\ &= e^{-f} + e^{-f} \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx}{\int_a^{t/d} \frac{1}{x} e^{-t/x} g(x) dx}. \\ &= e^{-f} + e^{-f} E(t, f). \end{aligned}$$

Since $0 \leq E(t, f) \leq C_E f$ and $1 - f \leq e^{-f}$, it follows that

$$1 - f \leq r \leq 1 + C_E f$$

for all $t \in S$. By Taylor's theorem,

$$(1 - \sqrt{r})^2 = \frac{1}{4}(r-1)^2 - \frac{1}{8\xi^{5/2}}(r-1)^3 \tag{4.21}$$

for some ξ between r and 1. Considering the cases when $r < 1$ and $r > 1$ separately, Eq. (4.21) implies

$$(1 - \sqrt{r})^2 \leq \begin{cases} \frac{1}{4} f^2 + \frac{1}{\sqrt{2}} f^3 & : \quad 1 - f < r < 1 \\ \frac{1}{4} C_E^2 f^2 & : \quad 1 < r < 1 + C_E f \end{cases}$$

where the first bound is easily obtained from Eq. (4.21) using the fact that $\xi \geq 1 - f \geq 1/2$, and the second bound follows from the fact that the remainder term in Eq. (4.21) is negative when $r > 1$.

Plugging these bounds into Eq. (4.20) and using the fact that $\int_0^\infty p(t) dt = 1$ yields the

estimate

$$\int_S \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \max\left\{\frac{1}{4}C_E^2 f^2, \frac{1}{4}f^2 + \frac{1}{\sqrt{2}}f^3\right\}.$$

Since $f \leq 1/2$, this estimate implies the inequality in the statement of the lemma. □

The next lemma utilizes Lemmas 30 and 31 to prove $H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$ in the case where θ is positive and has a continuous distribution satisfying **A.1**.

Lemma 32. *Let θ be a nonnegative random variable with probability density function g such that $g(x) > 0$ for all $x > 0$. Further assume that g satisfies **A.1**. Then there exists a constant $C > 0$ such that*

$$H^2\left(\mathbb{P}_0, \mathbb{Q}\right) := \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq Cf^2$$

for all $d > d_{\min} > 0$ and $f < \frac{d}{2} \wedge \frac{1}{2}$.

Proof of Lemma 32. By assumption $a = 0$. Therefore taking $S = (0, \infty)$, it suffices by Lemma 31 to show that there exists a constant $C_1 > 0$ such that for all $t > 0$,

$$E(t, f) := \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx}{\int_0^{t/d} \frac{1}{x} e^{-t/x} g(x) dx} \leq C_1 f \tag{4.22}$$

for all $d \geq d_{\min}$ and $f < \frac{d}{2} \wedge \frac{1}{2}$.

By making the domain of integration on the denominator smaller,

$$E(t, f) < \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx}{\int_{t/2d}^{t/d} \frac{1}{x} e^{-t/x} g(x) dx}. \tag{4.23}$$

Here we have used the assumption that $g(x) > 0$ for all $x > 0$ to ensure that the denominator on the right-hand side of Eq. (4.23) is nonzero. Next, since $f < d/2$, it follows that

$$\frac{t}{d-f} < \frac{2t}{d}.$$

Using this inequality along with Eq. (4.16) (taking $s = t/d$), it follows that

$$\sup_{s\in\left(\frac{t}{d},\frac{t}{d-f}\right)} g(x) \leq \sup_{s\in\left(\frac{t}{d},\frac{2t}{d}\right)} g(x) \leq C_{\mathrm{gc}} \inf_{x\in\left(\frac{t}{2d},\frac{t}{d}\right)} g(x).$$

Therefore

$$\begin{aligned}(4.23) &\leq C_{\mathrm{gc}} \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} dx}{\int_{t/2d}^{t/d} \frac{1}{x} e^{-t/x} dx} \\ &= C_{\mathrm{gc}} \frac{\int_{d-f}^{d} \frac{1}{u} e^{-u} du}{\int_{d}^{2d} \frac{1}{u} e^{-u} du} \qquad (4.24)\end{aligned}$$

where the second step follows by the substitution $u = t/x$, $\frac{1}{u}du = -\frac{1}{x}dx$. Therefore by Lemma 30 there exists a constant $C_2^{(1)} > 0$ such that

$$(4.24) \leq C_{\mathrm{gc}} C_2^{(1)} f$$

for all $d > d_{\min}$ and all $f < d/2$. This establishes Eq. (4.22), completing the proof.

□

The previous lemma suffices to show the result of Theorem 6 when θ is everywhere positive and has a continuous distribution satisfying **A.1**, for example when θ has a gamma distribution; the next lemma extends this result by allowing for the distribution of θ to have an atom at zero.

Lemma 33 (Addition of an atom at zero)**.** *Assume θ has distribution*

$$\mu_\theta = \lambda\delta_{\{0\}} + (1-\lambda)g(x)dx \qquad (4.25)$$

where $\lambda \in (0,1)$ and g is any probability density function with $a = \inf \mathrm{supp}(g) \geq 0$. Then

$$2H^2(\mathbb{P}_0, \mathbb{Q}) = (1-\lambda)\int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt$$

provided that $f < d$.

Proof of Lemma 33. Let μ_p and μ_q be the distributions induced by $\widetilde{D}_{12} := \frac{\theta}{2} \cdot D_{12}$ under $\mathbb{P}_0$ and $\mathbb{Q}$ respectively; that is, for all $t \in \mathbb{R}$

$$\mu_q(-\infty, t] := \mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] \quad \text{and} \quad \mu_p(-\infty, t] := \mathbb{P}_0\left[\widetilde{D}_{12} \leq t\right].$$

Claim 6 (The form of μ_q and μ_p)**.** *For every Borel set $A \subset \mathbb{R}$,*

$$\mu_q(A) = \lambda\delta_{\{0\}}(A) + (1-\lambda)\int_A q(t)dt, \tag{4.26}$$

and

$$\mu_p(A) = \lambda\delta_{\{0\}}(A) + (1-\lambda)\int_A p(t)dt. \tag{4.27}$$

Proof of Claim 6. We will prove only Eq. (4.26), as the proof of Eq. (4.27) is similar but with $f = 0$. By a standard argument involving the $\pi - \lambda$ theorem, to prove Eq. (4.26) it suffices to show that

$$\mu_q((-\infty, t]) = \lambda\delta_{\{0\}}((-\infty, t]) + (1-\lambda)\int_{(-\infty,t]} q(s)ds \tag{4.28}$$

for all $t \in \mathbb{R}$. There are two cases, depending on whether $t < 0$ or $t \geq 0$.

Case 1. Suppose $t < 0$. By Eq. (4.13), $\mathbb{Q}\left[\widetilde{D}_{12} \geq 0\right] = 1$. Therefore by definition of μ_q the left-hand side of Eq. (4.28) is zero. On the other hand, $\delta_{\{0\}}((-\infty, t]) = 0$ since $t < 0$, and moreover by Lemma 29, $q(s) = 0$ whenever $s < 0$. Therefore the right hand side of Eq. (4.28) is also zero. Since both side vanish, Eq. (4.28) holds for all $t < 0$.

Case 2. Suppose $t \geq 0$. Then by Eq. (4.13), $\widetilde{D}_{12} \stackrel{d}{=} \theta(d - f + Z)$ under $\mathbb{Q}$, where

$Z \sim \exp(1)$ and $Z \perp\!\!\!\perp \theta$. Therefore by the law of total expectation

$$\begin{aligned}\mu_q((-\infty, t]) &= \mathbb{Q}\left[\widetilde{D}_{12} \leq t\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{Q}\left[\widetilde{D}_{12} \leq t \mid \theta\right]\right] \\ &= \int_{[0,\infty)} \mathbb{Q}\left[x(d - f + Z) \leq t\right] \mu_\theta(dx). \end{aligned} \tag{4.29}$$

Using the assumption from Eq. (4.25),

$$\begin{aligned}(4.29) &= \lambda \mathbb{Q}\left[0 \leq t\right] + (1 - \lambda) \int_{(a,\infty)} \mathbb{Q}\left[x(d - f + Z) \leq t\right] g(x)dx. \\ &= \lambda + (1 - \lambda) \int_{(a,\infty)} \mathbb{Q}\left[Z \leq \frac{t}{a} - d + f\right] g(x)dx.\end{aligned}$$

Therefore since $\mathbb{Q}\left[0 \leq t\right] = 1$ and x is nonzero in the domain of integration, it follows that

$$\mu_q((-\infty, t]) = \lambda \delta_{\{0\}}((-\infty, t]) + (1 - \lambda) \int_{(a,\infty)} \mathbb{Q}\left[Z \leq \frac{t}{a} - d + f\right] g(x)dx. \tag{4.30}$$

Next, consider the integral on the right-hand side of Eq. (4.30). Letting F_Z denote the distribution function of Z,

$$\begin{aligned}\int_{(a,\infty)} \mathbb{Q}\left[Z \leq \frac{t}{x} - d + f\right] g(x)dx &= \int_{(a,\infty)} F_Z\left(Z \leq \frac{t}{x} - d + f\right) g(x)dx \\ &= \int_{(a,\infty)} \mathbf{1}_{[t \geq x(d-f)]} \left(1 - e^{-\frac{t}{x} + d - f}\right) g(x)dx \\ &= \int_a^{\frac{t}{d-f}} \left(1 - e^{-\frac{t}{x} + d - f}\right) g(x)dx. \end{aligned} \tag{4.31}$$

Moreover, by the Leibniz integral rule,

$$\begin{aligned}
\frac{d}{dt}\left[\int_a^{\frac{t}{d-f}}\left(1-e^{-\frac{t}{x}+d-f}\right)g(x)dx\right] &= \int_a^{\frac{t}{d-f}}\frac{\partial}{\partial t}\left[\left(1-e^{-\frac{t}{x}+d-f}\right)g(x)\right]dx \\
&\quad + \left[\left(1-e^{-\frac{t}{x}+d-f}\right)g(x)\frac{1}{d-f}\right]_{x=\frac{t}{d-f}} \\
&= e^{d-f}\int_a^{\frac{t}{d-f}}\frac{1}{x}e^{-\frac{t}{x}}g(x)dx \\
&\quad + \left(1-e^{-(d-f)+d-f}\right)g\left(\frac{t}{d-f}\right)\frac{1}{d-f} \\
&= e^{d-f}\int_a^{\frac{t}{d-f}}\frac{1}{x}e^{-\frac{t}{x}}g(x)dx. \qquad (4.32)
\end{aligned}$$

By the Fundamental Theorem of Calculus,

$$\begin{aligned}
(4.31) &= \int_{-\infty}^{t}\frac{d}{dt}\left[\int_a^{\frac{t}{d-f}}\left(1-e^{-\frac{t}{x}+d-f}\right)g(x)dx\right]_{t=s}ds \\
&= \int_{-\infty}^{t}\left[e^{d-f}\int_a^{\frac{s}{d-f}}\frac{1}{x}e^{-\frac{s}{x}}g(x)dx\right]ds && \text{by Eq. (4.32)} \\
&= \int_{-\infty}^{t}q(s)ds && \text{by Lemma 29.}
\end{aligned}$$

Taken together with Eq. (4.30), it follows that Eq. (4.28) holds for all $t \geq 0$. This proves the claim. □ Claim

Let $\nu := \mu + \delta_{\{0\}}$ where μ is the Lebesgue measure on $\mathbb{R}$. Then μ_p, μ_q are both absolutely continuous with respect to ν. Therefore by the Radon-Nikodyn Theorem, there exists μ-a.e. unique functions $\frac{d\mu_p}{d\nu}$ and $\frac{d\mu_q}{d\nu}$ such that

$$\mu_q(A) = \int_A \frac{d\mu_q}{d\nu}(t)\nu(dt) \qquad (4.33)$$

for every Borel set $A \subset \mathbb{R}$. Two conclusions can be drawn:

- Taking $A = \{0\}$ in Eq. (4.33) implies

$$\mu_q(\{0\}) = \frac{d\mu_q}{d\nu}(0).$$

Therefore since $\mu_q(\{0\}) = \lambda$ by Eq. (4.26), it follows that

$$\frac{d\mu_q}{d\nu}(0) = \lambda. \tag{4.34}$$

- For any Borel set $A \subseteq \mathbb{R}\backslash\{0\}$,

$$\begin{aligned}(1-\lambda)\int_A q(t)d\mu(t) &= \mu_q(A) && \text{by Eq. (4.26)}\\ &= \int_A \frac{d\mu_q}{d\nu}(t)d\nu(t) && \text{by Eq. (4.33)}\\ &= \int_A \frac{d\mu_q}{d\nu}(t)d\mu(t) && \text{since } 0 \notin A.\end{aligned}$$

Since $A \subset \mathbb{R}\backslash\{0\}$ is arbitrary, it follows that

$$\frac{d\mu_q}{d\nu}(t) = (1-\lambda)q(t) \tag{4.35}$$

for μ-a.e. $t \in \mathbb{R}\backslash\{0\}$.

A similar argument with $\frac{d\mu_p}{d\nu}$ yields the following analogues to Eqs. (4.34) and (4.35):

$$\frac{d\mu_p}{d\nu}(0) = \lambda \quad \text{and} \quad \frac{d\mu_p}{d\nu}(t) = (1-\lambda)p(t)$$

for μ-a.e. $t \in \mathbb{R}\backslash\{0\}$. Therefore

$$\begin{aligned}2H^2(\mathbb{P}_0, \mathbb{Q}) &= \int_{-\infty}^{\infty}\left(\sqrt{\frac{d\mu_p}{d\nu}(t)} - \sqrt{\frac{d\mu_q}{d\nu}(t)}\right)^2 d\nu(t)\\ &= \left(\sqrt{\frac{d\mu_p}{d\nu}(0)} - \sqrt{\frac{d\mu_q}{d\nu}(0)}\right)^2 + \int_{\infty}^{\infty}\left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt\\ &= (\lambda-\lambda)^2 + (1-\lambda)\int_0^{\infty}\left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt\\ &= (1-\lambda)\int_0^{\infty}\left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt.\end{aligned}$$

This completes the proof of the lemma. □

Taken together Lemmas 32 and 33 immediately imply Theorem 6.

4.3 More general mutation distributions

One question we might ask as is whether Theorem 6 holds for more general assumptions about the distribution of θ. For example, recall that Theorem 6 assumes that a, the infimum of the support of g, is zero. Does the $O(f^2)$ lower bound still hold if a is positive? The next example shows that in general it does not.

4.3.1 A motivating example

The next example gives a concrete example showing that when $a > 0$, we are not guaranteed to have $H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$.

An example where $H^2(\mathbb{P}_0, \mathbb{Q}) \neq O(f^2)$ if $a > 0$

Example 1. For simplicity take $d = 1$, and suppose θ is uniformly distributed on the interval $[1, 2]$, so that $g(t) = \mathbf{1}_{[1,2]}(t)$ and $a = 1$. See Figs. 4.1 and 4.2.

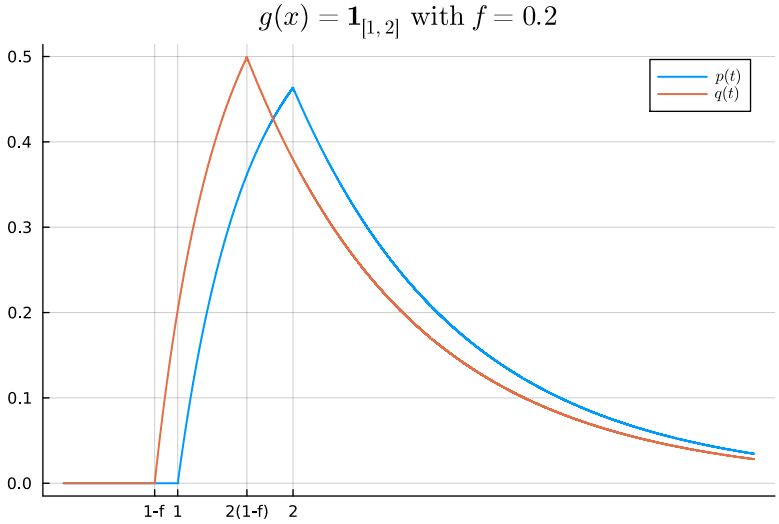


Figure 4.1: A plot of p and q when $g = \mathbf{1}_{[1,2]}$

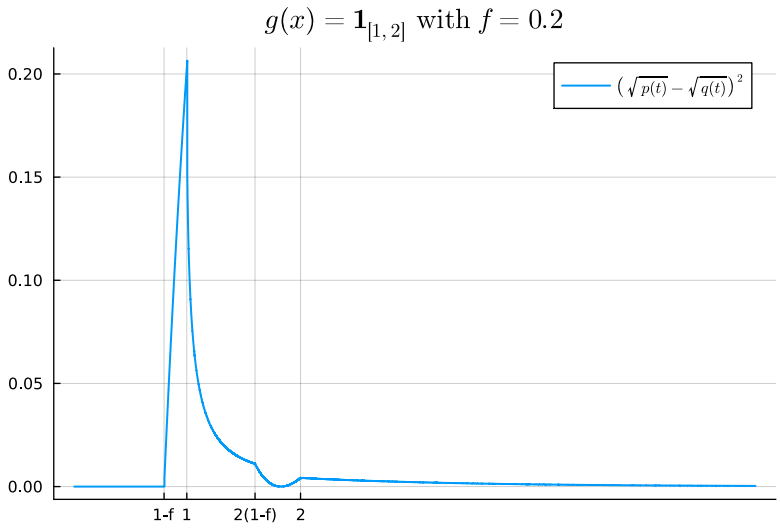


Figure 4.2: The mass of the integral $\int_0^\infty \left(\sqrt{q(t)} - \sqrt{p(t)}\right)^2 dt$ is concentrated near $a = 1$.

We claim that for this choice of g,

$$\lim_{f \to 0+} \frac{1}{f^2} H^2\left(\mathbb{P}_0, \mathbb{Q}\right) = +\infty \tag{4.36}$$

so that $H^2\left(\mathbb{P}_0, \mathbb{Q}\right) \neq O(f^2)$.

This result is surprising since the conclusion $H^2(\mathbb{P}_0, \mathbb{Q}) = O(f^2)$ holds in the case where $g(x) = \mathbf{1}_{[0,1]}$ by Theorem 6 because the function $\mathbf{1}_{[0,1]}$ satisfies **A.1**.

Before proving Eq. (4.36), we sketch the approach we will take, which can be broken down into three steps:

- First we will show that there exists a constant $c > 0$ such that

$$q(t) \geq p(t) + cf$$

  so that when f is small,

$$\begin{aligned} \frac{1}{\sqrt{p(t)}} &\lesssim \frac{\sqrt{p(t) + cf} - \sqrt{p(t)}}{f} \\ &\leq \frac{\sqrt{q(t)} - \sqrt{p(t)}}{f}. \end{aligned}$$

- Using the above, along with Fatou's lemma, we will then prove that

$$\int_{1+\eta}^{3/2} \frac{1}{p(t)} dt \lesssim \liminf_{f \to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{q(t)} - \sqrt{p(t)} \right)^2 dt.$$

- Finally we will show that the integral on the left-hand side blows up as $\eta \to 0^+$.

Proof of Eq. (4.36). Let $t \in [1+\eta, 2(1-f)]$ and $f \in (0,1)$. Define

$$\phi_t(f) := e^{1-f} \int_1^{t/(1-f)} \frac{1}{x} e^{-t/x} dx.$$

By Lemma 29, $\phi_t(f) = q(t)$ and $\phi_t(0) = p(t)$. Using the product rule and the Leibniz integral rule to differentiate ϕ_t with respect to f gives

$$\begin{aligned}
\phi_t'(f) &= -\phi_t(f) + e^{1-f} \frac{d}{df} \left[\int_1^{\frac{t}{1-f}} \frac{1}{x} e^{-\frac{t}{x}} dx \right] \\
&= -\phi_t(f) + e^{1-f} \left(\frac{1-f}{t} e^{-(1-f)} \frac{d}{dt} \left[\frac{t}{1-f} \right] \right) \\
&= -\phi_t(f) + (1-f) \frac{1}{(1-f)^2} \\
&= -\phi_t(f) + \frac{1}{1-f}. \qquad (4.37)
\end{aligned}$$

Therefore since $\phi_t(0) = p(t)$,

$$\phi_t'(0) = 1 - p(t). \qquad (4.38)$$

Differentiating both sides of Eq. (4.37) with respect to f,

$$\begin{aligned}
\phi_t''(f) &= -\phi_t'(f) + \frac{1}{(1-f)^2} \\
&= -\left[-\phi_t(f) + \frac{1}{1-f} \right] + \frac{1}{(1-f)^2} \qquad \text{by Eq. (4.37)} \\
&= \phi_t(f) - \frac{1}{1-f} + \frac{1}{(1-f)^2}. \qquad (4.39)
\end{aligned}$$

In particular, this implies that $\phi_t'' > 0$ on the interval $(0,1)$. Using this fact along with

Taylor's theorem, there exists $\xi_t \in (0, f)$ such that

$$\begin{aligned}\phi_t(f) &= \phi_t(0) + \phi_t'(0)f + \frac{\phi_t''(\xi_t)}{2}f^2 \\ &\geq \phi_t(0) + \phi_t'(0)f.\end{aligned}$$

This inequality, along with Eq. (4.38), $\phi_t(f) = q(t)$, and $\phi_t(0) = p(t)$, together imply that

$$q(t) \geq p(t) + (1 - p(t))\, f \tag{4.40}$$

for all $t \in [1 + \eta, 2(1 - f)]$ and all $f \in (0, 1)$, Assume $0 < f < 1/4$ and $0 < \eta < 1/2$. Making the substitution $u = \frac{t}{x}$, $-\frac{1}{u}du = \frac{1}{x}dx$ in the formula for p given by Lemma 29 implies

$$p(t) = e\int_1^t \frac{1}{u}e^{-u}du. \tag{4.41}$$

Since $p(t)$ is the integral of a nonnegative function, it is increasing on $(0, \infty)$, and hence for all $t \in (1 + \eta, 2(1 - f))$,

$$\begin{aligned}p(t) &\leq p(2(1-f)) \\ &= e\int_1^{2(1-f)} \frac{1}{x}e^{-2(1-f)/x}dx \\ &\leq \int_1^{2(1-f)} \frac{1}{x}dx \\ &= \log(2 - 2f) \\ &\leq \log 2 \\ &< 0.7.\end{aligned}$$

Plugging this inequality into Eq. (4.40),

$$\begin{aligned}q(t) &\geq p(t) + (1 - 0.7)\, f \\ &= p(t) + 0.3f\end{aligned}$$

Therefore, since $q(t) \geq p(t)$,

$$\begin{aligned}\int_{1+\eta}^{2(1-f)} \left(\sqrt{q(t)} - \sqrt{p(t)}\right)^2 dt &\geq \int_{1+\eta}^{2(1-f)} \left(\sqrt{p(t)+0.3f} - \sqrt{p(t)}\right)^2 dt \\ &\geq \int_{1+\eta}^{3/2} \left(\sqrt{p(t)+0.3f} - \sqrt{p(t)}\right)^2 dt,\end{aligned}$$

where the second inequality is justified by $f < 1/4$. Therefore

$$\liminf_{f\to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \geq \liminf_{f\to 0^+} \frac{1}{f^2} \int_{1+\eta}^{3/2} \left(\sqrt{p(t)+0.3f} - \sqrt{p(t)}\right)^2 dt. \tag{4.42}$$

Therefore by Fatou's lemma,

$$\begin{aligned}\liminf_{f\to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt &\geq \int_{1+\eta}^{3/2} \liminf_{f\to 0^+} \frac{1}{f^2} \left(\sqrt{p(t)+0.3f} - \sqrt{p(t)}\right)^2 dt \\ &= \int_{1+\eta}^{3/2} \left(\liminf_{f\to 0^+} \frac{\sqrt{p(t)+0.3f} - \sqrt{p(t)}}{f}\right)^2 dt \\ &= \int_{1+\eta}^{3/2} \left(\frac{3}{20\sqrt{p(t)}}\right)^2 dt, \end{aligned} \tag{4.43}$$

where the last step follows by definition of derivative since $p(t) > 0$ for all $t \in [1+\eta, 3/2]$ (and hence the limit exist and equals its limit infimum). Simplifying the right hand side of Eq. (4.43), we obtain

$$\liminf_{f\to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \geq \frac{9}{400} \int_{1+\eta}^{3/2} \frac{1}{p(t)} dt. \tag{4.44}$$

Next, we claim that

$$\frac{1}{p(t)} \geq \frac{1}{t-1}. \tag{4.45}$$

To see this, observe that by Eq. (4.41) and the trivial inequality $\frac{1}{u}e^{-u} \leq \frac{1}{e}$ for all $u \geq 1$,

it holds that

$$p(t) \leq e \int_1^t \frac{1}{e} dt = t - 1$$

which implies Eq. (4.45). Together, Eqs. (4.44) and (4.45) imply

$$\begin{aligned} \liminf_{f \to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt &\geq \frac{9}{400} \int_{1+\eta}^{3/2} \frac{1}{t-1} dt \\ &= \frac{9}{400} \log \left(\frac{1}{2\eta} \right) \to +\infty \end{aligned}$$

as $\eta \to 0^+$.

Since $\eta \in (0, \frac{1}{2})$ can be chosen arbitrarily small, sending $\eta \to 0^+$ implies

$$\liminf_{f \to 0^+} \frac{1}{f^2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt = +\infty,$$

and hence

$$H^2(\mathbb{P}_0, \mathbb{Q}) = \frac{1}{2} \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt \neq O(f^2).$$

□

Discussion of Example 1: a different lower bound

Example 1 shows that the case with $a > 0$ exhibits somewhat different behavior compared to the case with $a = 0$. The reason for the difference concerns the magnitude of the integral

$$\int_{ad}^{ad+\delta} \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt$$

when $\delta > 0$ is fixed and f is very small. Recall that here we have $p(ad) = 0$ but $q(ad) > 0$ and this can cause the size of the integrand to spike when t is just slightly larger than ad (for an example of this, see Fig. 4.2), and this spike may not decrease quickly enough for the above integral to be $O(f^2)$. Instead, it appears that for $a > 0$, $H^2(\mathbb{P}_0, \mathbb{Q}) = O(-f^2 \log f)$ as $f \to 0$ for fixed d.

But why is there a factor of $-\log f$? The beginnings of an answer to this question is captured in the following formal calculation. For simplicity assume $d = 1$, and assume that

$$p(t) \approx t - a \quad \text{and} \quad q(t) \approx t - a + f$$

for all $t \in (a, a + \delta)$. (In Example 1, these assumptions can be shown to hold for δ sufficiently small by regarding q and p as functions of both t and f and taking the first order Taylor approximation at the point $(t, f) = (1, 0)$.)

Then

$$\begin{aligned} H^2(\mathbb{P}_0, \mathbb{Q}) &\geq \int_a^{a+\delta} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \\ &\approx \int_a^{a+\delta} \left(\sqrt{t-a+f} - \sqrt{t-a}\right)^2 dt. \end{aligned}$$

By the substitution $t \mapsto t + a$,

$$\begin{aligned} H^2(\mathbb{P}_0, \mathbb{Q}) &\geq \int_0^{\delta} \left(\sqrt{t+f} - \sqrt{t}\right)^2 dt \\ &= f^2 \int_0^{\delta} \left(\frac{1}{\sqrt{t+f} + \sqrt{t}}\right)^2 dt \\ &\geq f^2 \int_0^{\delta} \left(\frac{1}{2\sqrt{t+f}}\right)^2 dt. \end{aligned}$$

Evaluating the integral on the right hand side gives

$$\begin{aligned} H^2(\mathbb{P}_0, \mathbb{Q}) &\geq \frac{f^2}{4} \log\left(1 + \frac{\delta}{f}\right) \\ &= O(-f^2 \log f). \end{aligned}$$

The key point of this calculation is to show how the $-\log f$ factor arises due to the denominator becoming close to t as $f \to 0^+$. Since $-\log f$ grows so slowly, the lower bound of $O(-f^2 \log f)$ is $o(f^{1-\epsilon})$ for all $\epsilon > 0$, and thus for practical purposes, this bound may not be meaningfully different from $O(f^2)$.

4.3.2 Establishing more general lower bound

In the remainder of this section will be devoted to considering more general conditions for the distribution of θ than those required by Theorem 6. By replacing **A.1** with substantially more general (albeit more complicated) assumptions, the main result of this section will help to answer the question raised in Section 4.3 by characterizing when the $O(f^2)$ lower bound can be obtained if $a > 0$, and when $O(-f^2 \log f)$ is obtained instead.

Assumptions

At a high level, the techniques used to establish estimates will rely on two types of assumptions about g: first, that g is both "mostly" bounded and "mostly" bounded away from zero, and second, that g satisfies certain growth/decay conditions at points where it is not. Making these notions precise will require the following definitions.

Definition 1 (Balanced polynomial growth/decay)**.** *Let $x < y$, and let ψ be a real-valued function defined on $[x, y]$, and let $A, B > 0$ such that $A \leq B$.*

- *If $\psi(x) = 0$, we say that ψ has **balanced polynomial growth** on the interval $[x, y]$ with constants A, B if there exist constants $k \geq k' > 0$ satisfying $2k' - k > 0$ such that*

$$A(\xi - x)^k \leq \psi(\xi) \leq B(\xi - x)^{k'} \tag{4.46}$$

  for all $\xi \in [x, y]$.

- *If $\psi(y) = 0$, we say that ψ has **balanced polynomial decay** on the interval $[x, y]$ with constants A, B if there exist constants $k \geq k' > 0$ satisfying $2k' - k > 0$ such that*

$$A(y - \xi)^k \leq \psi(\xi) \leq B(y - \xi)^{k'} \tag{4.47}$$

  for all $\xi \in [x, y]$.

The above definition can be generalized by loosening the restrictions on the exponents k and k'.

Definition 2 (Almost-polynomial growth/decay)**.** *Let $x < y$. Assume that $A, B > 0$ and $k \geq k' > -1$. A function ψ is said to have* ***almost-polynomial growth*** *on the inteval $(x, y]$ with constants A, B and exponents k, k' if*

$$A(\xi - x)^k \leq \psi(\xi) \leq B(\xi - x)^{k'} \tag{4.48}$$

for all $\xi \in (x, y]$. Similarly, we say that ψ has ***almost-polynomial decay*** *on $[x, y)$ if*

$$A(y - \xi)^k \leq \psi(\xi) \leq B(y - \xi)^{k'} \tag{4.49}$$

for all $\xi \in [x, y)$.

Informally, the almost-polynomial growth condition is satisfied on the interval $[w, w+\delta]$ for some $\delta > 0$ if ψ exhibits essentially one of three following behaviors locally to the right of w:

- $\limsup_{x \to w^+} \psi(x) = 0$ and ψ grows at least as fast a nonconstant polynomial (i.e. this is the case when $k, k' > 0$),
- ψ is bounded and bounded away from zero locally to the right of w (i.e. when $k = k' = 0$); or,
- $\psi(w+) = \infty$ and ψ decays sufficiently fast that the singularity is integrable (i.e. when $k, k' < 0$).

The almost-polynomial decay condition can be interpreted in a similar manner with respect to the behavior of ψ locally to the left of z.

With these definitions, the assumptions that will be used are as follows.

A.2 (Beahaviour near a.). There exists a constant $\delta_a > 0$ such that g has almost-polynomial growth on $[a, a + \delta_a]$ for some choice of constants A, B and exponents k_a, k'_a; i.e., g has almost-polynomial growth at a. In addition, letting $\eta := 2k'_a - k_a$, assume $\eta > -1$ if $a = 0$, or $\eta \geq 0$ if $a > 0$.

A.3 (Eventually pseudo-decreasing). There exists constants $K > a$ and $C_K > 0$ such that

$$g(y) \leq C_K g(x) \tag{4.50}$$

whenever $K \leq x < y$.

A.4 (Behavior on $(a + \delta_a, K)$). There exist positive constants $\tilde{A}, \tilde{B}$ and δ_z with $\delta_z \leq \delta_a$ such that the following two conditions hold:

(i) If $z \in Z := \{x \in (a + \delta_a, K) : g(x) = 0\}$, then

- g has balanced polynomial decay on $[z - \delta_z, z]$ with constants $\tilde{A}, \tilde{B}$; and,
- g has balanced polynomial growth on $[z, z + \delta_z]$ with constants $\tilde{A}, \tilde{B}$.

(ii) If $x \in (a + \delta_a, K) \backslash \bigcup_{z \in Z} (z - \delta_z, z + \delta_z)$ then $\tilde{A} \leq g(x) \leq \tilde{B}$.

Remark 3 (Explanation of Assumptions **A.2**, **A.3**, and **A.4**)**.**

- *Consideration of the special case $k_a = k'_a$ is sufficient to show that **A.2** is quite general, allowing for the following situations:*
  - *If, on the interval $(a, a + \delta_a)$, the function g is both bounded and bounded away from zero then **A.2** is satisfied with $k_a = k'_a = 0$. This is the case, for example, if $g(a+)$ exists and is positive.*
  - *If $a = 0$ and $g(x) = O(x^{-\kappa})$ as $x \to 0^+$ for some $\kappa \in (0, 1)$, then **A.2** is satisfied with $k_a = k'_a = \kappa$. This allows for certain integrable singularities.*
  - *We claim that if $g(a+) = 0$ then **A.2** is satisfied provided that there exists $k \geq 1$ such that g is k-times differentiable at a, $g^{(1)}(a) = \ldots = g^{(k-1)}(a) = 0$, and $g^{(k-1)}$ is right-differentiable at a with $g_+^{(k)}(a) \neq 0$. To see why this is the case, observe that by Taylor's theorem,*

$$g(x) = \frac{g_+^{(k)}(a)(x - a)^k}{k!} + o(|x - a|^k)$$

as $x \to a^+$. *Therefore*

$$\lim_{x \to a^+} \frac{g(x)}{(x-a)^k} = \frac{g_+^{(k)}(a)}{k!} \neq 0.$$

Since g *is nonnegative, the above expression implies that* $g_+^{(k)}(a) > 0$. *Choosing* $A, B > 0$ *such that* $A < \frac{g_+^{(k)}(a)}{k!} < B$, *it follows by definition of limit that there exists* $\delta_a > 0$ *such that*

$$A \leq \frac{g(x)}{(x-a)^k} \leq B$$

for all $x \in (a, a+\delta_a)$. *Therefore* ***A.2*** *is satisfied at* a *with* $k_a = k'_a = k$. *Incidentally, this argument also shows that in this case* δ_a *can be chosen sufficiently small that the ratio* $B/A > 1$ *can be made arbitrarily close to* 1.

- *Assumption* ***A.3*** *generalizes the notion that* g *be eventualy nonincreasing. For example, Eq.* (4.50) *is satisfied with* $C_K = 1$ *if* g *is decreasing on* $[K, \infty)$, *and hence is also trivially satisfied with* $K = \sup\{x : g(x) > 0\}$ *if* g *is compactly supported.*

- *Assumption* ***A.4*** *captures the idea that the function* g *"looks like" a polynomial locally near any zeros located in the interior of its support (and that between those zeros,* g *is both bounded and bounded away from zero).*

  - *If* g *has no zeros on the interval* $(a+\delta_a, K)$, *as is typical for standard continuous distributions used in modeling mutation rates, then* ***A.4*** *reduces simply to the assumption that there exists* $\tilde{A}, \tilde{B} > 0$ *such that* $\tilde{A} \leq g(x) \leq \tilde{B}$ *for all* $x \in (a+\delta_a, K)$, *which in most cases will be trivially satisfied.*

  - *On the other hand,* ***A.4*** *allows for* g *to have up to finitely many zeros on* $(a+\delta_a, K)$; *to see why this is the case, let* $z, z' \in Z$ *such that* $z \neq z'$, *and suppose that* $|z - z'| < \delta_z$. *Then since* $g(z) = 0$ *and* $|z - z'| < \delta_z$, ***A.4*** *(ii) implies that* $g(z') \geq \tilde{A} > 0$, *contradicting* $z' \in Z$. *Therefore* $|z - z'| \geq \delta_z$, *i.e., the elements are of* Z *are isolated, and hence* $|Z| < \infty$.

  - *Condition* ***A.4*** *(i) is satisfied under mild regularity conditions, as we now ex-*

plain. Let $z \in Z$. Suppose there exist constants $\kappa, \kappa' > 0$ such that both

$$\lim_{\xi \to z^-} \frac{g(\xi)}{(z-\xi)^\kappa} > 0 \quad \textit{and} \quad \lim_{\xi \to z^+} \frac{g(\xi)}{(\xi - z)^{\kappa'}} > 0. \tag{4.51}$$

We claim that Eq. (4.51) *is sufficient condition for g to have both balanced polynomial growth on $[z, z+\delta]$ and balanced polynomial decay on $[z-\delta, z]$ for some δ. Indeed, let*

$$C_2 := \lim_{\xi \to z^-} \frac{g(\xi)}{(z-\xi)^\kappa} > 0.$$

and let $A, B > 0$ such that $A < C_2 < B$. Then there exists a constant $\delta > 0$ such that

$$A(z-\xi)^\kappa < g(\xi) < B(z-\xi)^\kappa$$

for all $\xi \in [z-\delta, z]$. That is, g has balanced polynomial decay on $[z-\delta, z]$. The proof of balanced polynomial growth is similar.

Theorem statement and proof sketch

The main result in this section is the following theorem.

Theorem 7 (Lower bound for distinguishing 2-leaf trees)**.** *Assume $\theta/2$ has probability density function g with $a := \inf \operatorname{supp}(g) \geq 0$, and that assumptions **A.2**, **A.3**, and **A.4** hold, with η defined as in **A.2**. Let $d_{\min} > 0$. Then there exist constants $C, r_0 > 0$ not depending on f or d such that*

$$H^2(\mathbb{P}_0, \mathbb{Q}) \leq \begin{cases} Cf^2 & : \quad a, \eta > 0 \text{ or } a = 0 \text{ and } \eta > -1 \\ Cf^2 \log \frac{1}{f} & : \quad a > 0 \text{ and } \eta = 0 \end{cases} \tag{4.52}$$

for all $f \in (0, \frac{1}{2})$ and $d \geq d_{\min}$ satisfying $\frac{f}{d} \leq r_0$.

By Eq. (4.6), the proof will consist of estimating the integral

$$H^2(\mathbb{P}_0, \mathbb{Q}) = \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt. \tag{4.53}$$

This will be done by breaking the domain of integration into four regions:

- $0 \leq t \leq ad$
- $ad \leq t \leq (a+\delta_a)d$
- $t \geq Kd$
- $(a+\delta_a)d \leq t \leq Kd$

and obtaining estimates for each region. We start by establishing some key lemmas which will be used to do that.

Key lemmas

The first lemma used in estimating Eq. (4.53) provides an estimate for those t satisfying $0 \leq t \leq ad$, in the case where $a > 0$.

Lemma 34 (Estimate for $0 \leq t \leq ad$)**.** *Let $a > 0$ and $d > d_{\min}$. Assume that $g(t) \leq C_\delta$ for a.e. $t \in (a, a+\delta)$. Then*

$$\int_0^{ad} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \left(\frac{C_\delta a}{d-f}\right) f^2$$

for all $f < \frac{1}{2}$ with $\frac{f}{d} < \frac{\delta}{\delta+a}$.

Proof of Lemma 34. By Lemma 29, $p(t) = 0$ for all $t \leq ad$ and $q(t) = 0$ for all $t \leq a(d-f)$. Therefore

$$\begin{aligned}\int_0^{ad} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt &= \int_{a(d-f)}^{ad} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \\ &= \int_{a(d-f)}^{ad} q(t) dt.\end{aligned}$$

Then using Lemma 29 and the Fubini-Tonelli theorem,

$$\begin{aligned}
\int_{a(d-f)}^{ad} q(t)dt &= e^{d-f} \int_{a(d-f)}^{ad} \left[\int_{a}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx \right] dt \\
&= e^{d-f} \int_{a}^{ad/(d-f)} g(x) \left[\int_{(d-f)x}^{ad} \frac{1}{x} e^{-t/x} dt \right] dx \\
&= e^{d-f} \int_{a}^{ad/(d-f)} g(x) \left[e^{-t/x} \right]_{t=ad}^{t=(d-f)x} dx \\
&= e^{d-f} \int_{a}^{ad/(d-f)} g(x) \left[e^{-(d-f)} - e^{-ad/x} \right] dx \\
&= \int_{a}^{ad/(d-f)} g(x) \left[1 - e^{d-f-ad/x} \right] dx.
\end{aligned}$$

Since $x \geq a$, it holds that $1 - e^{d-f-ad/x} \leq 1 - e^{-f} \leq f$, and therefore

$$\int_{a(d-f)}^{ad} q(t)dt \leq \int_{a}^{ad/(d-f)} g(x)dx \cdot f \tag{4.54}$$

Next observe that $\frac{ad}{d-f} \leq a + \delta$ by the assumption that $\frac{f}{d} \leq \frac{\delta}{\delta+a}$. Therefore $g(x) \leq C_\delta$ for a.e. $x \in \left(a, \frac{ad}{d-f}\right)$. Using this to bound $g(x)$ in Eq. (4.54) gives:

$$\begin{aligned}
\int_{a(d-f)}^{ad} q(t)dt &\leq C_\delta \left(\frac{ad}{d-f} - a \right) f \\
&= \left(\frac{C_\delta a}{d-f} \right) f^2 \\
&= O\left(\frac{f^2}{d} \right)
\end{aligned}$$

as $d \to \infty$ and $f \to 0$. □

The next two lemmas will be used to establish that there exists $\delta > 0$ such that

$$\int_{ad}^{(a+\frac{\delta}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt \lesssim f^2. \tag{4.55}$$

Both estimates will later be used in other parts of the proof as well. The first of these

lemmas, Lemma 35, establishes an intermediate estimate for

$$\int_I \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt$$

when $I \subseteq [ad, \infty)$ is any bounded interval.

Lemma 35 (Bounded Intervals)**.** *Assume there exists some $\delta_+ > 0$ such that $g(x) > 0$ for all $x \in (a, a+\delta_+)$. If $a(d-f) < w < z$ then*

$$\int_w^z \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq ef^2 + de^{d+f} \int_{w/d}^{z/d} \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} g(x) dx\right)^2}{\int_a^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} g(x) dx} dt$$

for all $f \leq \frac{1}{2} \wedge \frac{d}{2}$.

Proof of Lemma 35. Using the numerical identity $u - v = \frac{u^2 - v^2}{u+v}$,

$$\begin{aligned}\int_w^z \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt &= \int_w^z \frac{(q(t) - p(t))^2}{\left(\sqrt{p(t)} + \sqrt{q(t)}\right)^2} dt \\ &\leq \int_w^z \frac{(q(t) - p(t))^2}{q(t)} dt\end{aligned}$$

The denominator is nonzero by the assumption that $g(x) > 0$ for all $x \in (a, a+\delta_+)$, which along with Lemma 29, implies that $q(t) > 0$ for all $t \in (w, z)$. Therefore

$$\begin{aligned}\int_w^z \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq &\int_{(w,z) \cap \{t : p(t) < q(t)\}} \frac{(q(t) - p(t))^2}{q(t)} dt \\ &+ \int_{(w,z) \cap \{t : p(t) > q(t)\}} \frac{(q(t) - p(t))^2}{q(t)} dt \qquad (4.56)\end{aligned}$$

Consider the two integrals on the right-hand side separately. If $p(t) < q(t)$ then $p(t) < e^f q(t)$ and therefore, using the formulas from Lemma 29,

$$(q(t) - p(t))^2 \leq \left(e^f q(t) - p(t)\right)^2 = \left(e^d \int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x) dx\right)^2.$$

By this inequality, the definition of $q(t)$, and nonnegativity of the integrand, it follows that

$$\begin{aligned}\int_{(w,z)\cap\{t:q(t)>p(t)\}} \frac{(q(t)-p(t))^2}{q(t)}dt &\le \int_w^z \frac{\left(e^d \int_{t/d}^{t/(d-f)} \frac{1}{x}e^{-t/x}g(x)dx\right)^2}{e^{d-f}\int_a^{t/(d-f)} \frac{1}{x}e^{-t/x}g(x)dx}dt \\ &= e^{d+f}\int_w^z \frac{\left(\int_{t/d}^{t/(d-f)} \frac{1}{x}e^{-t/x}g(x)dx\right)^2}{\int_a^{t/(d-f)} \frac{1}{x}e^{-t/x}g(x)dx}dt \\ &= de^{d+f}\int_{w/d}^{z/d} \frac{\left(\int_t^{td/(d-f)} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_a^{td/(d-f)} \frac{1}{x}e^{-td/x}g(x)dx}dt, \quad (4.57)\end{aligned}$$

where the last step follows by the substitution $t \mapsto td$.

On the other hand, if $p(t) > q(t)$ then using Lemma 29 again, it holds for all $f \le 1/2$ that

$$\begin{aligned}(p(t)-q(t))^2 &= \left(e^d\int_a^{t/d}\frac{1}{x}e^{-t/x}g(x)dx - e^{d-f}\int_a^{t/(d-f)}\frac{1}{x}e^{-t/x}g(x)dx\right)^2 \\ &\le \left(e^d\int_a^{t/(d-f)}\frac{1}{x}e^{-t/x}g(x)dx - e^{d-f}\int_a^{t/(d-f)}\frac{1}{x}e^{-t/x}g(x)dx\right)^2 \\ &= \left(1-e^{-f}\right)^2\left(e^d\int_a^{t/(d-f)}\frac{1}{x}e^{-t/x}g(x)dx\right)^2 \\ &= \left(1-e^{-f}\right)^2 e^{2f}\left(e^{d-f}\int_a^{t/(d-f)}\frac{1}{x}e^{-t/x}g(x)dx\right)^2 \\ &= \left(1-e^{-f}\right)^2 e^{2f}\left(q(t)\right)^2 \\ &\le ef^2\left(q(t)\right)^2.\end{aligned}$$

Therefore, again noting that $q(t) > 0$ for all $t > w$,

$$\begin{aligned}\int_{(w,z)\cap\{t:p(t)>q(t)\}}\frac{(q(t)-p(t))^2}{q(t)}dt &\le ef^2\int_{(w,z)\cap\{t:p(t)>q(t)\}}\frac{q(t)\cdot q(t)}{q(t)}dt \\ &= ef^2\int_{(w,z)\cap\{t:p(t)>q(t)\}} q(t)dt \\ &\le ef^2. \quad (4.58)\end{aligned}$$

where the last step follows from the fact that q is a probability density function.

Plugging Eqs. (4.57) and (4.58) into Eq. (4.56) implies the statement of the lemma. □

The next lemma, which uses Lemmas 30 and 35, will be used to establish a general "right-side estimate" i.e., an estimate of the form

$$\int_{yd}^{(y+\lambda\delta)d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt$$

at any $y \geq a$ such that g satisfies the almost-polynomial growth conditions on $[y, y+\delta]$. In particular, taking $y = a$, $\delta = \delta_a$ and $\lambda = \frac{1}{2}$ will give the necessary estimate for Eq. (4.55). As with Lemmas 30 and 35, Lemma 36 will be utilized for subsequent estimates as well.

Lemma 36 (Right-Side Estimates)**.** *Let $y \geq a$. Assume that there exists some $\delta > 0$ such that g has almost-polynomial growth on $[y, y+\delta]$ with constants $A, B > 0$ and exponents $k \geq k' > -1$, with k, k' sharing the same sign; that is,*

$$A(x-y)^k \leq g(x) \leq B(x-y)^{k'} \tag{4.59}$$

whenever $y < x \leq y + \delta$. Let $\eta := 2k' - k$ and $\lambda \in [\frac{1}{2}, 1)$. Then the following statements hold:

(i.) If $y = 0$ and $\eta > -1$ then there exists a constant $C > 0$ not depending on f or d such that

$$\int_0^{\lambda\delta d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq Cf^2$$

for all $d \geq d_{min}$ and $f < \frac{1}{2}$ such that $\frac{f}{d} < \frac{(1-\lambda)\lambda\delta}{2\lambda\delta+y}$.

(ii.) If $y > 0$ then there exists a constant C, possibly depending on y, but not depending on f or d such that

$$\int_{yd}^{(y+\lambda\delta)d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \begin{cases} Cf^2 & : \quad \eta > 0 \\ Cf^2 \log \frac{1}{f} & : \quad \eta = 0 \end{cases}$$

for all $d \geq d_{min}$ *and* $f < \frac{1}{2}$ *such that* $\frac{f}{d} < \frac{(1-\lambda)\lambda\delta}{2\lambda\delta+y}$.

Proof of Lemma 36. Fix $\lambda \in (0,1)$. Let $t \in (y, y+\lambda\delta)$ and define

$$\phi(t) := \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_a^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx}. \tag{4.60}$$

By Lemma 35 (taking $w = yd$ and $z = (y+\lambda\delta)d$), it will suffice to estimate

$$\int_y^{y+\lambda\delta} \phi(t)dt \tag{4.61}$$

for all values of f and d in the ranges specified in the statement of the lemma.

Assume that

$$\frac{f}{d} \leq \frac{(1-\lambda)\lambda\delta}{y+2\lambda\delta}. \tag{4.62}$$

It follows trivially that $\frac{f}{d} \leq \frac{(1-\lambda)\delta}{y+\delta}$. Therefore

$$\frac{td}{d-f} = \frac{t}{1-\frac{f}{d}} \leq \frac{t}{1-\frac{(1-\lambda)\delta}{y+\delta}} = \frac{(y+\delta)t}{y+\lambda\delta} < y+\delta, \tag{4.63}$$

where the last inequality is due to $t < y+\lambda\delta$. The remainder of the proof is split into two cases, corresponding to statements (i.) and (ii.) respectively.

Case 1. Assume that $y = 0$ and $\eta > -1$. Then by making the domain of integration on the denominator smaller in Eq. (4.60),

$$\phi(t) \leq \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_{t/2}^{t} \frac{1}{x}e^{-td/x}g(x)dx}.$$

By Eq. (4.63), the inequalities in Eq. (4.59) can be applied to $g(x)$ on both the numer-

ator and denominator of Eq. (4.60), which implies

$$\phi(t) \leq \frac{de^{d+f}B^2}{A} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} x^{k'} dx\right)^2}{\int_{t/2}^t \frac{1}{x} e^{-td/x} x^k dx}. \tag{4.64}$$

Next note that k and k' share the same sign, and make the following two observations:

- If $k, k' \geq 0$ then, continuing from Eq. (4.64),

$$\begin{aligned}\phi(t) &\leq \frac{de^{d+f}B^2}{A} \cdot \frac{\left(\frac{td}{d-f}\right)^{2k'} \left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx\right)^2}{\left(\frac{t}{2}\right)^k \int_{t/2}^t \frac{1}{x} e^{-td/x} dx} \\ &= \frac{de^{d+f}B^2 2^k}{A} \cdot t^{2k'-k} \left(\frac{d}{d-f}\right)^{2k'} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx\right)^2}{\int_{t/2}^t \frac{1}{x} e^{-td/x} dx}\end{aligned}$$

- Similarly, if $k, k' < 0$ then

$$\begin{aligned}\phi(t) &\leq \frac{de^{d+f}B^2}{A} \cdot \frac{t^{2k'} \left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx\right)^2}{t^k \int_{t/2}^t \frac{1}{x} e^{-td/x} dx} \\ &= \frac{de^{d+f}B^2}{A} \cdot t^{2k'-k} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx\right)^2}{\int_{t/2}^t \frac{1}{x} e^{-td/x} dx}.\end{aligned}$$

In both cases, it follows that there exists a constant $C_3 > 0$ such that

$$\phi(t) \leq C_3 de^d t^{2k'-k} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx\right)^2}{\int_{t/2}^t \frac{1}{x} e^{-td/x} dx}$$

whenever $\frac{f}{d} \leq \frac{1}{2}$ and $f \leq 1/2$. By the substitution $u = td/x$, so that $-\frac{1}{u}du = \frac{1}{x}dx$, it holds that

$$\phi(t) \leq C_3 de^d t^{2k'-k} \cdot \frac{\left(\int_{d-f}^d \frac{1}{u} e^{-u} du\right)^2}{\int_d^{2d} \frac{1}{u} e^{-u} du}. \tag{4.65}$$

Moreover by Lemma 30 (with $\gamma = 2$),

$$\frac{\left(\int_{d-f}^{d} \frac{1}{u} e^{-u} du\right)^2}{\int_{d}^{2d} \frac{1}{u} e^{-u} du} \leq \frac{8e}{1 - e^{-d}} \cdot \frac{f^2}{de^d} \leq \frac{8e}{1 - e^{-d_{\min}}} \cdot \frac{f^2}{de^d} \tag{4.66}$$

where it is assumed that $d \geq d_{\min} > 0$. By Eqs. (4.65) and (4.66), it follows that there exists a constant $C_4 > 0$ not depending on d or f such that

$$\phi(t) \leq C_4 t^{2k'-k} f^2. \tag{4.67}$$

whenever $\frac{f}{d} < \min\left\{\frac{1}{2}, 1 - \lambda\right\}$ and $d > d_{\min}$. By the assumption that $\eta = 2k' - k > -1$, we can integrate both sides of Eq. (4.67) over the interval $(0, \lambda\delta)$, which implies that there exists a constant $C_4 > 0$ such that

$$\int_0^{\lambda\delta} \phi(t) dt \leq C_4 f^2$$

for all $d \geq d_{\min}$ and $f < \frac{1}{2}$ such that $\frac{f}{d} \leq \frac{(1-\lambda)\lambda\delta}{2\lambda\delta + y}$. Having estimated the integral in Eq. (4.61), this concludes the proof of part (i.) of the lemma. It remains to prove part (ii.).

Case 2. Assume that $y > 0$ and $\eta \geq 0$. Before continuing, note that Eq. (4.62) implies that $d - 2f \geq \lambda d_{\min} > 0$, and hence any division by $d - 2f$ and $d - f$ (which will occur frequently in the rest of the proof) is not division by zero.

By Eq. (4.62), it holds that $\frac{f}{d} < \frac{\lambda\delta}{2\lambda\delta + y}$, and hence

$$y < \frac{d - f}{d - 2f} y < y + \lambda\delta. \tag{4.68}$$

By Eq. (4.61), it will suffice to estimate the integral

$$\int_y^{y+\lambda\delta} \phi(t) dt = \int_y^{\frac{d-f}{d-2f} y} \phi(t) dt + \int_{\frac{d-f}{d-2f} y}^{y+\lambda\delta} \phi(t) dt, \tag{4.69}$$

where ϕ is defined as in Eq. (4.60).

The required estimate for case 2 will follow directly from the next two claims, each of which bounds one of the two integrals on the right-hand side of Eq. (4.69).

Claim 7 (Bound on first integral in Eq. (4.69))**.** *There exists a constant $C > 0$, possibly depending on y, such that*

$$\int_y^{\frac{d-f}{d-2f}y} \phi(t)dt \leq Cf^{2+k'}d^{-(k'+1)}$$

for all $d > d_{\min}$ and $f < \frac{1}{2}$ with $\frac{f}{d} \leq \frac{(1-\lambda)\lambda\delta}{2\lambda\delta+y}$.

Proof of Claim 7. Assume $t \in \left(y, \frac{d-f}{d-2f}y\right)$. By Eq. (4.59), $g(x) > 0$ for all $x \in (y, y+\delta)$. It follows that

$$\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx > 0.$$

Therefore by Eq. (4.60),

$$\begin{aligned}
\phi(t) &= \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_a^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx} \\
&\leq \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx} \\
&= de^{d+f}\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx.
\end{aligned}$$

Using Eq. (4.59),

$$\phi(t) \leq Bde^{d+f}\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}(x-y)^{k'}dx$$

Therefore since $k' \geq 0$ and $x \geq t > y$,

$$\phi(t) \leq Bde^{d+f}\left(\frac{td}{d-f} - y\right)^{k'}\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}dx.$$

Applying the substitution $u = \frac{td}{x}$, $\frac{1}{u}du = -\frac{1}{x}dx$ to the integral on the right-hand side,

$$\begin{aligned}\phi(t) &\le Bde^{d+f}\left(\frac{td}{d-f}-y\right)^{k'}\int_{d-f}^{d}\frac{1}{u}e^{-u}du\\ &\le 2\sqrt{e}Be^{f}\left(\frac{td}{d-f}-y\right)^{k'}f \qquad \text{by Eq. (4.17).}\end{aligned}$$

Therefore

$$\int_{y}^{\frac{d-f}{d-2f}y}\phi(t)dt \le 2\sqrt{e}Be^{f}f\int_{y}^{\frac{d-f}{d-2f}y}\left(\frac{td}{d-f}-y\right)^{k'}dt,$$

and so by the substitution $u = \frac{td}{d-f} - y$, $du = \left(\frac{d}{d-f}\right)dt$,

$$\begin{aligned}\int_{y}^{\frac{d-f}{d-2f}y}\phi(t)dt &= 2y^{k'+1}\sqrt{e}Be^{f}f\left(\frac{d-f}{d}\right)\int_{y\left(\frac{f}{d-f}\right)}^{y\left(\frac{2f}{d-2f}\right)}u^{k'}du\\ &\le 2y^{k'+1}\sqrt{e}Be^{f}f\left(1-\frac{f}{d}\right)\left(\frac{2f}{d-2f}\right)^{k'}\left[\frac{2f}{d-2f}-\frac{f}{d-f}\right]\\ &= 2y^{k'+1}\sqrt{e}Be^{f}f\left(1-\frac{f}{d}\right)\left(\frac{2f}{d-2f}\right)^{k'}\left[\frac{fd}{(d-2f)(d-f)}\right]\\ &\lesssim f^{2+k'}d^{-(k'+1)}\end{aligned}$$

Notably, Eq. (4.62) implies that $d - 2f \ge \lambda d_{\min} > 0$, so that no division by zero occurs in the above estimates. □ Claim

Claim 8 (Bound on second integral in Eq. (4.69))**.** *There exists a constant $C > 0$ such that*

$$\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta}\phi(t)dt \le \begin{cases} Cf^2\log\frac{1}{f} & : \quad 2k'-k=0\\ Cf^2 & : \quad 2k'-k>0\end{cases} \tag{4.70}$$

for all $f < \frac{1}{2}$ and $d > d_{\min}$ with $\frac{f}{d} < \frac{\lambda(1-\lambda)\delta}{2\lambda\delta+y}$.

Proof of Claim 8. Assume that $t \in \left(\frac{d-f}{d-2f}y, y+\lambda\delta\right)$. Since $t > y$, it holds that

$$a \le y < \frac{y}{2} + \frac{td}{2(d-f)} < \frac{td}{d-f}.$$

Therefore, by making the domain of integration on the denominator of Eq. (4.60) smaller, it holds that

$$\phi(t) \le \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_{\frac{y}{2}+\frac{td}{2(d-f)}}^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx}.$$

By Eq. (4.63), the inequalities in Eq. (4.59) can be applied on both the numerator and denominator, which implies

$$\phi(t) \le \frac{de^{d+f}B^2}{A} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}(x-y)^{k'}dx\right)^2}{\int_{\frac{y}{2}+\frac{td}{2(d-f)}}^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}(x-y)^{k}dx}. \tag{4.71}$$

By the assumptions that $k \ge k'$ and $\eta \ge 0$, it follows that both k and k' are nonnegative (since $0 \le \eta = 2k' - k \le 2k' - k' = k' \le k$). Therefore by Eq. (4.71),

$$\begin{aligned}\phi(t) &\le \frac{de^{d+f}B^2}{A} \cdot \frac{\left(\frac{td}{d-f}-y\right)^{2k'}}{\left(\frac{td}{2(d-f)}-\frac{y}{2}\right)^{k}} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}dx\right)^2}{\int_{\frac{y}{2}+\frac{td}{2(d-f)}}^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}dx} \\ &= \frac{de^{d+f}B^2 2^k}{A}\left(\frac{td}{d-f}-y\right)^{2k'-k} \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}dx\right)^2}{\int_{\frac{y}{2}+\frac{td}{2(d-f)}}^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}dx}.\end{aligned} \tag{4.72}$$

Applying the substitution $u = \frac{td}{x}$ (so that $\frac{1}{x}dx = -\frac{1}{u}du$) to both integrals on the right-hand side, and noting that $\frac{y}{2} + \frac{td}{2(d-f)} = \frac{y(d-f)+td}{2(d-f)}$,

$$\phi(t) \le \frac{de^{d+f}B^2 2^k}{A}\left(\frac{td}{d-f}-y\right)^{2k'-k} \frac{\left(\int_{d-f}^{d} \frac{1}{u}e^{-u}du\right)^2}{\int_{d-f}^{\frac{2td(d-f)}{y(d-f)+td}} \frac{1}{u}e^{-u}du}. \tag{4.73}$$

Using Eq. (4.17) to bound the numerator on the right-hand side of Eq. (4.73),

$$\begin{aligned}\phi(t) &\le \frac{4e^{1+f}B^2 2^k f^2}{Ade^d}\left(\frac{td}{d-f}-y\right)^{2k'-k}\frac{1}{\int_{d-f}^{\frac{2td(d-f)}{y(d-f)+td}}\frac{1}{u}e^{-u}du}\\ &= \frac{4e^{1+f}B^2 2^k f^2}{Ade^d}\left(\frac{td}{d-f}-y\right)^{2k'-k}\frac{1}{\int_{d-f}^{\gamma d}\frac{1}{u}e^{-u}du}\end{aligned} \tag{4.74}$$

where

$$\gamma := \frac{2t(d-f)}{y(d-f)+td} = \frac{2t(1-\frac{f}{d})}{(1-\frac{f}{d})y+t}. \tag{4.75}$$

Differentiating both sides of Eq. (4.75) with respect to t, one obtains

$$\begin{aligned}\gamma'(t) &= \frac{2(1-\frac{f}{d})^2 y}{\left[(1-\frac{f}{d})y+t\right]^2}\\ &\ge \frac{y}{2\left[(1-\frac{f}{d})y+t\right]^2} && \text{since } \frac{f}{d}\le\frac{1}{2}\\ &\ge \frac{y}{2\left[(1-\frac{f}{d})y+2y\right]^2} && \text{since } t\le y+\lambda\delta<2y\\ &= \frac{1}{2y\left(3-\frac{f}{d}\right)}\\ &\ge \frac{1}{6y}.\end{aligned} \tag{4.76}$$

In particular, this shows that $t\mapsto\gamma(t)$ is increasing. Therefore by the claim assumption that $t > \frac{d-f}{d-2f}y$, it follows that $\gamma(t) > \gamma\left(\frac{d-f}{d-2f}y\right) = 1$.

Therefore

$$\begin{aligned}\int_{d-f}^{\gamma d}\frac{1}{u}e^{-u}du &\ge \frac{e^f}{\gamma de^d}\left(1-e^{-(\gamma-1)d-f}\right)\\ &\ge \frac{e^f}{\gamma de^d}\left(1-e^{-(\gamma-1)d_{\min}-f}\right)\end{aligned} \tag{4.77}$$

By Eq. (4.76) and the observation that $\gamma = 1$ when $t = \frac{d-f}{d-2f}y$,

$$\gamma(t) - 1 = \int_{\frac{d-f}{d-2f}y}^{t} \gamma'(s)ds \geq \frac{1}{6y}\left(t - \frac{(d-f)y}{d-2f}\right). \tag{4.78}$$

Plugging Eq. (4.78) into Eq. (4.77) implies

$$\int_{d-f}^{\gamma d} \frac{1}{u}e^{-u}du \geq \frac{e^f}{\gamma d e^d}\left(1 - e^{-x_0}\right) \tag{4.79}$$

where

$$x_0 := \frac{d_{\min}}{6y}\left(t - \frac{(d-f)y}{d-2f}\right) + f \geq 0.$$

Moreover, by Eq. (4.68) and the inequality $t \leq y + \lambda\delta < 2y$, it is not difficult to see that

$$\begin{aligned} x_0 &\leq \frac{1}{6y}\left(y + \lambda\delta + \frac{d-f}{d-2f}y\right)d_{\min} + f \\ &\leq \frac{2}{3}d_{\min} + f \\ &< d_{\min} + 1 \end{aligned} \tag{4.80}$$

Therefore, by applying the numerical inequality $1 - e^{-x} \geq \frac{x}{x+1}$ for $x \geq 0$ (which holds by rearranging $e^x \geq x + 1$) to Eq. (4.79) implies

$$\begin{aligned} \int_{d-f}^{\gamma d} \frac{1}{u}e^{-u}du &\geq \frac{e^f}{\gamma d e^d}\left(\frac{x_0}{x_0+1}\right) \\ &> \frac{e^f}{\gamma d e^d}\left(\frac{x_0}{d_{\min}+2}\right) \qquad \text{by Eq. (4.80)} \end{aligned}$$

Plugging this inequality into Eq. (4.74),

$$\begin{aligned}
\phi(t) &< \frac{4e^{1+f}B^2 2^k f^2}{Ade^d}\left(\frac{td}{d-f}-y\right)^{2k'-k}\left[\frac{\gamma de^d}{e^f}\left(\frac{d_{\min}+2}{x_0}\right)\right] \\
&= \frac{4e\gamma B^2 2^k f^2}{A}\left(\frac{td}{d-f}-y\right)^{2k'-k}\left(\frac{d_{\min}+2}{x_0}\right) \\
&\le \frac{8eB^2 2^k (d_{\min}+2) f^2}{A}\left(\frac{td}{d-f}-y\right)^{2k'-k}\frac{1}{x_0} && \text{since } \gamma \le 2 \\
&= \frac{48eyB^2 2^k f^2}{A}\left(\frac{d_{\min}+2}{d_{\min}}\right)\left(\frac{td}{d-f}-y\right)^{2k'-k}\left[t-\frac{d-f}{d-2f}y+\frac{6y}{d_{\min}}f\right]^{-1}.
\end{aligned}$$

In particular, letting $C_6 := \frac{48eyB^2 2^k f^2}{A}\left(\frac{d_{\min}+2}{d_{\min}}\right)$, we have shown that

$$\phi(t) \le C_6\left(\frac{td}{d-f}-y\right)^{\eta}\frac{1}{t-\frac{d-f}{d-2f}y+\frac{6y}{d_{\min}}f} \tag{4.81}$$

for all $\frac{d-f}{d-2f}y < t < y+\lambda\delta$, where $\eta = 2k'-k$.

We consider the cases $\eta = 0$ and $\eta > 0$ separately.

- Suppose $\eta = 0$. By Eq. (4.81),

$$\begin{aligned}
\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta}\phi(t)dt &\le C_6 f^2\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta}\frac{1}{t-\frac{d-f}{d-2f}y+\frac{6y}{d_{\min}}f}dt \\
&= C_6 f^2\log\left(1+\frac{d_{\min}}{6(d-f)}+\frac{\lambda d_{\min}}{6yf}\right) \\
&= O(-f^2\log f).
\end{aligned}$$

  Therefore there exists a constant $C_7 > 0$

$$\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta}\phi(t)dt \le C_7 f^2\log\frac{1}{f}$$

  for all $f \in (0,\frac{1}{2})$ and $d \ge d_{\min}$ such that $\frac{f}{d} < \frac{\lambda\delta}{2\lambda\delta+y}$. This prove the first half of Eq. (4.70).

- Suppose $\eta > 0$. By Eq. (4.81),

$$\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta} \phi(t)dt \le C_6 f^2 \int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta} \left(\frac{td}{d-f} - y\right)^{\eta} \frac{1}{t - \frac{d-f}{d-2f}y + \frac{6y}{d_{\min}}f} dt.$$

Therefore by the substitution $u = t - \frac{d-f}{d-2f}y$, $du = dt$,

$$\begin{aligned}\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta} \phi(t)dt &\le C_6 f^2 \int_0^{\lambda\delta - \frac{yf}{d-2f}} \left(\frac{ud}{d-f} + \frac{2fy}{d-2f}\right)^{\eta} \frac{1}{u + \frac{6y}{d_{\min}}f} du \\ &\le C_6 f^2 \int_0^{\lambda\delta} \left(\frac{ud}{d-f} + \frac{2fy}{d-2f}\right)^{\eta} \frac{1}{u + \frac{6y}{d}f} du,\end{aligned}$$

where the second inequality follows by making the domain of integration larger and replacing $d_{\min}$ by the larger quantity d. Then, rewriting the right-hand side using $\rho := \frac{f}{d}$,

$$\begin{aligned}\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta} \phi(t)dt &= C_6 f^2 \int_0^{\lambda\delta} \left(\frac{u}{1-\rho} + \frac{2y\rho}{1-2\rho}\right)^{\eta} \frac{1}{u + 6y\rho} du, \\ &\le C_6 f^2 \int_0^{\lambda\delta} \frac{(u+2y\rho)^{\eta}}{(1-2\rho)^{\eta}} \frac{1}{u+6y\rho} du.\end{aligned}$$

By Eq. (4.62), $\rho \le \frac{1-\lambda}{2} < \frac{1}{2}$, and hence $(1-2\rho)^{\eta} \ge \lambda^{\eta}$. Therefore

$$\begin{aligned}\int_{\frac{d-f}{d-2f}y}^{y+\lambda\delta} \phi(t)dt &\le C_6 f^2 \lambda^{-\eta} \int_0^{\lambda\delta} \frac{(u+2y\rho)^{\eta}}{u+6y\rho} du \\ &\le C_6 f^2 \lambda^{-\eta} \int_0^{\lambda\delta} (u+2y\rho)^{\eta-1} du \\ &\le C_6 f^2 \lambda^{-\eta} \eta^{-1} \left[(\lambda\delta + 2y\rho)^{\eta} - (2y\rho)^{\eta}\right] \\ &\le C_6 f^2 \lambda^{-\eta} \eta^{-1} (y + \lambda\delta)^{\eta}\end{aligned}$$

where the last inequality follows by $2\rho \le 1$, since it is assumed that $\frac{f}{d} \le \frac{1}{2}$. This proves the second half of Eq. (4.70).

□ Claim

This completes the proof of the lemma. □

As the counterpart to Lemma 36, the next lemma establishes an analogous "left-side" estimate, which holds when g satisfies the appropriate growth/decay conditions.

Lemma 37 (Left-Side Estimates)**.** *Let $y > a$. Suppose there exist $A, B, \delta > 0$ such that*

$$A(y-\xi)^k \le g(\xi) \le B(y-\xi)^{k'} \tag{4.82}$$

for all $\xi \in [y-\delta, y]$, where k, k' are nonnegative constants satisfying $k \ge k' \ge 0$ and $\eta := 2k' - k > -1$. Further suppose that there exists a constant $\hat{k}' \ge 0$ such that

$$g(\xi) \le B(\xi - y)^{\hat{k}'} \tag{4.83}$$

for all $\xi \in [y, y+\delta]$. Then there exists a constant C not depending on d or f such that

$$\int_{(y-\frac{\delta}{2})d}^{yd} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \le Cf^2$$

for all $f \in (0, 1/2)$ and $d \ge d_{\min}$ satisfying $\frac{f}{d} \le \frac{\delta}{2(y+\delta)}$.

Proof of Lemma 37. For each $t \in (y - \frac{\delta}{2}, y)$, define

$$\phi(t) := \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} g(x)dx\right)^2}{\int_a^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} g(x)dx}. \tag{4.84}$$

By Lemma 35, it suffices to show that there exists some constant $C_8 > 0$ not depending on f or d such that

$$\int_{y-\delta/2}^{y} \phi(t)dt \le C_8 f^2 \tag{4.85}$$

for all f, d in the ranges specified in the statement of the lemma. Assume that

$$\frac{f}{d} \le \frac{\delta}{2(y+\delta)}. \tag{4.86}$$

Then

$$y\left(1-\frac{f}{d}\right) \geq y\left(1-\frac{\delta}{2(y+\delta)}\right) \geq y\left(1-\frac{\delta}{2y}\right) = y-\frac{\delta}{2}, \tag{4.87}$$

and hence the integral in Eq. (4.85) can be written as

$$\int_{y-\delta/2}^{y} \phi(t)dt = \int_{y-\frac{\delta}{2}}^{y(1-\frac{f}{d})} \phi(t)dt + \int_{y(1-\frac{f}{d})}^{y} \phi(t)dt. \tag{4.88}$$

The remainder of the proof consists of two claims, each bounding one of these two integrals.

Claim 9 (Bound on first integral in Eq. (4.88))**.** *There exists a constant $C_9 > 0$ such that*

$$\int_{y-\frac{\delta}{2}}^{y\left(1-\frac{f}{d}\right)} \phi(t)dt \leq C_9 f^2$$

for all $d \geq d_{\min}$ and all $f < 1/2$ satisfying Eq. (4.86)*.*

Proof of Claim 9. Let t be given such that

$$y-\frac{\delta}{2} \leq t \leq y\left(1-\frac{f}{d}\right). \tag{4.89}$$

Since since making δ smaller can only weaken the assumptions Eqs. (4.82) and (4.83), assume without loss of generality that $y-\delta > a$. Moreover, $t < \frac{td}{d-f}$ trivially. Therefore, by making the domain of integration smaller in the denominator of Eq. (4.84),

$$\phi(t) \leq \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_{y-\delta}^{t} \frac{1}{x}e^{-td/x}g(x)dx}.$$

(Note that $g(x) > 0$ for all $x \in (y-\delta, y-\frac{\delta}{2})$ by Eq. (4.82), and the denominator in the above equation is nonzero since $y-\frac{\delta}{2} \leq t$ by the claim assumption).

Next observe that by Eq. (4.89),

$$\frac{td}{d-f} = \frac{t}{1-\frac{f}{d}} \leq y, \tag{4.90}$$

and therefore Eq. (4.82) may be used to estimate the term $g(x)$ on both the numerator and the denominator:

$$\phi(t) \leq \frac{B^2 d e^{d+f} \left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} (y-x)^{k'} dx \right)^2}{A \int_{y-\delta}^{t} \frac{1}{x} e^{-td/x} (y-x)^k dx}.$$

Therefore since $k, k' \geq 0$,

$$\phi(t) \leq \frac{B^2 d e^{d+f} (y-t)^{2k'-k}}{A} \cdot \frac{\left(\int_t^{\frac{td}{d-f}} \frac{1}{x} e^{-td/x} dx \right)^2}{\int_{y-\delta}^{t} \frac{1}{x} e^{-td/x} dx}.$$

Making the substitution $u = \frac{td}{x}$, so that $\frac{1}{x} dx = -\frac{1}{u} du$,

$$\phi(t) \leq \frac{B^2 d e^{d+f} (y-t)^{2k'-k}}{A} \cdot \frac{\left(\int_{d-f}^{d} \frac{1}{u} e^{-u} du \right)^2}{\int_{d}^{\frac{td}{y-\delta}} \frac{1}{u} e^{-u} du}.$$

Let $\gamma := \frac{2y-\delta}{2y-2\delta}$. Then by Eq. (4.89), $\frac{td}{y-\delta} \geq \gamma > 1$, and hence Lemma 30 implies

$$\begin{aligned} \phi(t) &\leq \frac{B^2 d e^{d+f} (y-t)^{2k'-k}}{A} \cdot \frac{C_\gamma^{(2)} f^2}{d e^d} \\ &= \frac{C_\gamma^{(2)} B^2 e^f}{A} \cdot (y-t)^{2k'-k} f^2 \end{aligned}$$

where

$$C_\gamma^{(2)} = \frac{4e\gamma}{1 - e^{-(\gamma-1)d}} \leq \frac{4e\gamma}{1 - e^{-(\gamma-1)d_{\min}}}$$

which does not depend on d or f. Therefore there exists a constant $C_9 > 0$ such that

$$\phi(t) \leq C_9 (y-t)^{2k'-k} f^2$$

for all $f < \frac{1}{2}$ and $d > d_{\min}$ satisfying Eq. (4.86). Moreover, since $\eta := 2k' - k > -1$,

$$\begin{aligned}\int_{y-\frac{\delta}{2}}^{y(1-\frac{f}{d})} \phi(t)dt &\le \int_{y-\frac{\delta}{2}}^{y} C_9 \left(y-t\right)^{\eta} f^2 dt \\ &= \frac{C_9}{\eta+1}\left(\frac{\delta}{2}\right)^{\eta+1} f^2.\end{aligned}$$

This proves the claim. □ Claim

Let $k^* := k' \wedge \hat{k}'$. Then $k^* \ge 0$.

Claim 10 (Bound on second integral in Eq. (4.88))**.** *There exists a constant $C_{10} > 0$ such that*

$$\int_{y(1-\frac{f}{d})}^{y} \phi(t)dt \le \frac{C_{10} f^{k^*+2}}{d^{k^*+1}}$$

for all $d \ge d_{\min}$ and $f < \frac{1}{2}$ satisfying Eq. (4.86)*.*

Proof of Claim 10. Let t be given such that

$$y\left(1-\frac{f}{d}\right) \le t \le y.$$

Therefore by Eq. (4.87),

$$t \ge y - \frac{\delta}{2}. \tag{4.91}$$

Therefore using Eq. (4.82),

$$\begin{aligned}\int_a^t \frac{1}{x} e^{-td/x} g(x)dx &\ge \int_{y-\delta}^{y-\frac{\delta}{2}} \frac{1}{x} e^{-t/x} g(x)dx \\ &\ge A \int_{y-\delta}^{y-\frac{\delta}{2}} \frac{1}{x} e^{-t/x} (y-x)^k dx \\ &> 0. \end{aligned} \tag{4.92}$$

For the function ϕ defined in Eq. (4.84), it holds by Eq. (4.92) that

$$\begin{aligned}\phi(t) &= \frac{de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right)^2}{\int_a^t \frac{1}{x}e^{-td/x}g(x)dx + \int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx} \\ &\leq de^{d+f}\left(\int_t^{\frac{td}{d-f}} \frac{1}{x}e^{-td/x}g(x)dx\right).\end{aligned}$$

Therefore by applying a trivial inequality,

$$\begin{aligned}\phi(t) &\leq de^{d+f} \cdot \frac{e^{-(d-f)}}{t}\left(\int_t^{\frac{td}{d-f}} g(x)dx\right) \\ &= \frac{de^{2f}}{t}\left(\int_t^{\frac{td}{d-f}} g(x)dx\right). \qquad (4.93)\end{aligned}$$

Using Eq. (4.86) and the assumption that $t \leq y$,

$$\frac{td}{d-f} \leq \frac{y}{1-\frac{f}{d}} \leq \frac{y}{1-\frac{\delta}{2(y+\delta)}} = \frac{2y(y+\delta)}{2y+\delta} < y+\delta. \qquad (4.94)$$

Eqs. (4.91) and (4.94) together imply that $(t, \frac{td}{d-f})$ is a subinterval of $(y-\frac{\delta}{2}, y+\delta)$. Therefore by Eqs. (4.82) and (4.83), there exists a positive constant $B^* \geq B$ such that

$$g(x) \leq B^*|x-y|^{k^*}$$

for all $x \in (t, \frac{td}{d-f})$. Applying this inequality to the right-hand side of Eq. (4.93),

$$\begin{aligned}\phi(t) &\leq \frac{B^*de^{2f}}{t}\int_t^{\frac{td}{d-f}} |x-y|^{k^*}dx \\ &= \frac{B^*de^{2f}}{t}\left[\int_t^y (y-x)^{k^*}dx + \int_y^{\frac{td}{d-f}} (x-y)^{k^*}\right] \\ &= \frac{B^*de^{2f}}{t(k^*+1)}\left[(y-t)^{k^*+1} + \left(\frac{td}{d-f}-y\right)^{k^*+1}\right].\end{aligned}$$

Observe that since $k^* + 1 > 0$, it holds that $w^{k^*+1} + z^{k^*+1} \leq (w+z)^{k^*+1}$ for any $w, z \geq 0$. Therefore

$$\begin{aligned}\phi(t) &\leq \frac{B^* d e^{2f}}{t(k^*+1)} \left(\frac{td}{d-f} - t \right)^{k^*+1} \\ &= \frac{B^* d e^{2f} t^{k^*}}{(k^*+1)} \left(\frac{f}{d-f} \right)^{k^*+1}.\end{aligned}$$

Therefore there exists a constant $C_{11} > 0$ not depending on f or d such that

$$\phi(t) \leq C_{11} \frac{f^{k^*+1}}{d^{k^*}} \tag{4.95}$$

for all $d \geq d_{\min}$ and $f < \frac{1}{2}$ such that $\frac{f}{d} \leq \frac{\delta}{2(y+\delta)}$. Since Eq. (4.95) holds for all $t \in (y(1-\frac{f}{d}), y)$, integrating both sides with respect to t implies the statement of the claim. $\square_{\text{Claim}}$

The statement of the lemma follows from Eq. (4.88) and the two above claims. $\square$

The next lemma utilizes **A.3** to obtain an estimate for when t is large.

Lemma 38 (Estimate for $[\alpha d K, \infty)$.)**.** *Let $a \geq 0$. Assume that there exist constants $K > a$ and $C_K > 0$ such that*

$$g(y) \leq C_K g(x) \tag{4.96}$$

whenever $K \leq x < y$. Further assume that there exists some $\delta_+ > 0$ such that $g(x) > 0$ for all $x \in (a, a+\delta_+)$. Then for any $\alpha > 1$ there exists a constant C not depending on f or d such that

$$\int_{\alpha d K}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt \leq C f^2$$

whenever $0 \leq f < \frac{1}{2} \wedge \frac{d}{2}$ and $d \geq \delta_{\min}$.

Proof of Lemma 38. By Lemma 31, it will suffice to show that there exists a constant $C_{12} > 0$ such that

$$E(t, f) \leq C_{12} f \tag{4.97}$$

for all $t \geq \alpha d K$.

Let $t \geq \alpha dK$. By Eq. (4.19),

$$E(t,f) = \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x)dx}{h_+(t) + \int_K^{t/d} \frac{1}{x} e^{-t/x} g(x)dx}, \tag{4.98}$$

where

$$h_+(t) := \int_a^K \frac{1}{x} e^{-t/x} g(x)dx.$$

In particular, $h_+(t) > 0$ for all t due to the assumption that $g(x) > 0$ for all $x \in (a, a+\delta_+)$. Next observe that by Eq. (4.96)

$$\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} g(x)dx \leq C_K g(t/d) \int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} dx.$$

In addition, Eq. (4.50) also implies

$$\int_K^{t/d} \frac{1}{x} e^{-t/x} g(x)dx \geq \frac{1}{C_K} g(t/d) \int_K^{t/d} \frac{1}{x} e^{-t/x} dx.$$

Applying these two inequalities to Eq. (4.98),

$$E(t,f) \leq \frac{C_K g(t/d) \int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} dx}{h_+(t) + \frac{1}{C_K} g(t/d) \int_K^{t/d} \frac{1}{x} e^{-t/x} dx}.$$

If $g(t/d) = 0$, then the right hand side is zero and there is nothing to show, so henceforth assume that $g(t/d) > 0$. Then

$$E(t,f) \leq C_K^2 \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} dx}{\int_K^{t/d} \frac{1}{x} e^{-t/x} dx}. \tag{4.99}$$

Therefore, since $K \leq \frac{t}{\alpha d} < \frac{t}{d}$, the domain of integration on the denominator of Eq. (4.99) can be made smaller by replacing K by $\frac{t}{\alpha d}$. This yields the estimate

$$E(t,f) \leq C_K^2 \frac{\int_{t/d}^{t/(d-f)} \frac{1}{x} e^{-t/x} dx}{\int_{t/\alpha d}^{t/d} \frac{1}{x} e^{-t/x} dx}. \tag{4.100}$$

Applying the substitution $u = t/x$ with $\frac{1}{x}dx = -\frac{1}{u}du$ gives

$$E(t,f) \leq C_K^2 \cdot \frac{\int_{d-f}^{d} \frac{1}{u} e^{-u} du}{\int_{d}^{\alpha d} \frac{1}{u} e^{-u} du}.$$

Therefore by Lemma 30,

$$\begin{aligned} E(t,f) &\leq C_K^2 \cdot \frac{2\alpha\sqrt{e}f}{1 - e^{-(\alpha-1)d}} \\ &\leq C_K^2 \cdot \frac{2\alpha\sqrt{e}f}{1 - e^{-(\alpha-1)d_{\min}}} \end{aligned}$$

for all $t \geq \alpha dK$, $d \geq d_{\min}$, and all $f < \frac{1}{2} \wedge \frac{d}{2}$. □

The next lemma will be used to strengthen **A.3**. Specifically, it gives conditions under which the interval $[K, \infty)$ (i.e., the interval on which on which g is known to be pseudo-decreasing) can be expanded to a larger interval, albeit at the cost of a possibly larger constant.

Lemma 39 (Tail bound adjustment)**.** *Let $K \geq \tilde{K} > 0$. Suppose*

(i.) there exists a constant C_K such that $g(y) \leq C_K g(x)$ whenever $K \leq x < y$; and,

(ii.) there exist constants $c_1 \geq c_0 > 0$ such that $c_0 \leq g(\xi) \leq c_1$ whenever $\tilde{K} \leq \xi \leq K$.

Letting $\tilde{C}_{\tilde{K}} := \max\left\{\frac{C_K c_1}{c_0}, \frac{c_1}{c_0}\right\}$, it follows that

$$g(y) \leq \tilde{C}_{\tilde{K}} g(x) \tag{4.101}$$

whenever $\tilde{K} \leq x < y$.

Proof of Lemma 39. Suppose $\tilde{K} \leq x < y$. The following three cases together imply Eq. (4.101).

- If $K \leq x < y$ then Eq. (4.101) follows immediately by (i.), since $C_K \leq \tilde{C}_{\tilde{K}}$.

- If $x < y \leq K$ then using (ii.) twice, along with the observation that $\frac{c_1}{c_0} \leq \tilde{C}_{\tilde{K}}$,

$$\begin{aligned} g(y) &\leq c_1 && \text{by (ii.)} \\ &= \frac{c_1}{c_0} \cdot c_0 && \text{since } c_0 \leq c_1 \\ &\leq \frac{c_1}{c_0} \cdot g(x) && \text{by (ii.)} \\ &\leq \tilde{C}_{\tilde{K}} g(x). \end{aligned}$$

- If $x \leq K < y$ then

$$\begin{aligned} g(y) &\leq C_K g(K) && \text{by (i.)} \\ &\leq C_K c_1 && \text{by (ii.)} \\ &\leq C_K \frac{c_1}{c_0} g(x) && \text{since } g(x)/c_0 \geq 1 \text{ by (ii.)} \\ &\leq \tilde{C}_{\tilde{K}} g(x). \end{aligned}$$

□

Proof of Theorem 7

Proof of Theorem 7. Assume that $0 \leq f \leq \frac{1}{2}$ and that $d \geq d_{\min} > 0$. Some additional assumptions about $\frac{f}{d}$ will be made later. The goal is to estimate the following quantity:

$$H^2(\mathbb{P}_0, \mathbb{Q}) = \int_0^\infty \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt \tag{4.102}$$

By **A.2** there exists $k_a \geq k'_a > -1$ and $A, B, \delta_a > 0$ such that

$$A(x-a)^{k_a} \leq g(x) \leq B(x-a)^{k'_a} \tag{4.103}$$

for all $x \in (a, a+\delta_a]$. In addition, under **A.2** it is assumed that exactly one of the following conditions is true:

(i.) $a = 0$ and $\eta := 2k'_a - k_a > -1$

(ii.) $a > 0$ and $\eta := 2k'_a - k_a \geq 0$

Finally, assume that

$$\frac{f}{d} \leq \frac{\delta_a}{4(a+\delta_a)}. \tag{4.104}$$

Claim 11 (Bound for small values of t)**.** *There exists a constant $C_1^* > 0$ such that*

$$\int_0^{(a+\frac{\delta_a}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \begin{cases} C_1^* f^2 & : \quad a, \eta > 0 \text{ or } a = 0 \\ C_1^* f^2 \log \frac{1}{f} & : \quad a > 0 \text{ and } \eta = 0 \end{cases} \tag{4.105}$$

Proof of Claim 11. Under assumption (ii.) it holds that both $k_a \geq k'_a$ and $2k'_a - k_a \geq 0$, and together these imply $k'_a \geq 0$. Since $k'_a \geq 0$, Eq. (4.103) implies that $g(x) \leq B\delta_a^{k'_a}$ whenever $a \leq x \leq a + \delta_a$. Therefore by Lemma 34 (taking $C_{\delta_a} = B\delta_a^{k'_a}$ in the statement of the lemma), it follows that

$$\int_0^{ad} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \left(\frac{B\delta_a^{k'_a} a}{d-f}\right) f^2. \tag{4.106}$$

On the other hand, Eq. (4.106) is trivial under assumption (i.) because in that case both sides are zero. Therefore we have shown that Eq. (4.106) holds under either assumption (i.) or (ii.).

Moreover, by **A.2** and Eq. (4.104), the assumptions of Lemma 36 are satisfied (for $y = a$ and $\lambda = \frac{1}{2}$ in the statement of the lemma). Therefore by Lemma 36 there exists a constant $C_{13} > 0$ not depending on f or d such that

$$\int_{ad}^{(a+\frac{\delta_a}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \begin{cases} C_{13} f^2 & : \quad a, \eta > 0 \text{ or } a = 0 \\ C_{13} f^2 \log \frac{1}{f} & : \quad a > 0 \text{ and } \eta = 0 \end{cases} \tag{4.107}$$

The statement of the claim follows from Eqs. (4.106) and (4.107). □ Claim

By Claim 11, it remains only to show that

$$\int_{(a+\frac{\delta_a}{2})d}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \lesssim f^2.$$

By **A.3** there exists $K > a$ and $C_K > 0$ such that

$$g(y) \leq C_K g(x) \tag{4.108}$$

whenever $K \leq x < y$. Define

$$Z := \{x \in (a + \delta_a, K) : g(x) = 0\}.$$

Note that Z has at most finitely many elements by **A.4** (see Remark 3). Therefore, one may define

$$z^* := \max Z \cup \{a\}$$

and

$$\tilde{K} := z^* + \frac{\delta_z}{3}.$$

The next claim, which uses Lemma 39, provides an inequality like that of Eq. (4.108) but with $\tilde{K}$ rather than K; this is an improvement over Eq. (4.108) since $\tilde{K}$ is guaranteed to be close to z^* whereas K is not.

Claim 12. *There exists a constant $\tilde{C}_{\tilde{K}} > 0$ such that*

$$g(y) \leq \tilde{C}_{\tilde{K}} g(x) \tag{4.109}$$

whenever $\tilde{K} \leq x < y$.

Proof of Claim 12. If $\tilde{K} \geq K$ then the claim is an immediate consequence of **A.3** in which case there is nothing to show. So henceforth assume instead that $K > \tilde{K}$.

It will suffice to show that the assumptions of Lemma 39 are satisfied for $\tilde{K} = z^* + \frac{\delta_z}{3}$,

as this will imply the claim. In fact, since assumption (i.) in Lemma 39 is nothing other than **A.3**, we need only to prove assumption (ii.). We start by making two observations:

- First, by maximality of z^*, $g(x) > 0$ for all $x \in (z^*, K)$. Therefore by part (ii) of **A.4**, there exists constants $\tilde{A}, \tilde{B}, \delta_z > 0$ with $\delta_z \leq \delta_a$ such that

$$\tilde{A} \leq g(x) \leq \tilde{B} \tag{4.110}$$

  for all $x \in (z^* + \delta_z, K)$.

- Second, it is easy to check that there exist constants $\tilde{B}_1 \geq \tilde{A}_1 > 0$ such that

$$\tilde{A}_1 \leq g(x) \leq \tilde{B}_1 \tag{4.111}$$

  for all $x \in [\tilde{K}, z^* + \delta_z]$, as we now explain: on one hand, if $z^* \neq a$ then $z^* \in Z$, in which case Eq. (4.111) is implied by part (i) of **A.4**. On the other hand, if $z^* = a$ then **A.2** implies that Eq. (4.111) holds for all $x \in [\tilde{K}, z^* + \delta_a]$, and therefore all $x \in [\tilde{K}, z^* + \delta_z]$ since $\delta_z \leq \delta_a$.

By Eqs. (4.110) and (4.111),

$$\tilde{A} \wedge \tilde{A}_1 \leq g(x) \leq \tilde{B} \vee \tilde{B}_1$$

for all $x \in [\tilde{K}, K)$, and hence

$$\tilde{A} \wedge \tilde{A}_1 \leq g(x) \leq \tilde{B} \vee \tilde{B}_1 \vee g(K)$$

for all $x \in [\tilde{K}, K]$. Thus we have shown that assumption (ii.) in Lemma 39 holds, as required. □ Claim

The rest of the proof is split into two cases, depending on whether $Z = \emptyset$ or $Z \neq \emptyset$.

Case 1. Suppose $Z = \emptyset$. Therefore $z^* = a$, and so by Claim 12,

$$g(y) \leq \tilde{C}_{\tilde{K}} g(x)$$

whenever $a + \frac{\delta_z}{3} \leq x < y$. Therefore by Lemma 38 (with K replaced by $a + \frac{\delta_z}{3}$ in the statement of the lemma), for all $\alpha > 1$ there exists a constant C_{14} not depending on f or d such that

$$\int_{\alpha(a+\frac{\delta_z}{3})d}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{14} f^2. \tag{4.112}$$

Since $\alpha > 1$ can be chosen sufficiently small that $\alpha(a + \frac{\delta_z}{3}) \leq a + \frac{\delta_a}{2}$, Eq. (4.112) implies that there exists $C_{15} > 0$ not depending on f or d such that

$$\int_{(a+\frac{\delta_a}{2})d}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{15} f^2. \tag{4.113}$$

Combining the inequalities in Eqs. (4.105) and (4.113), we get Eq. (4.52). This completes the proof in the case where $Z = \emptyset$.

Case 2. Assume $Z \neq \emptyset$. Let $z_0 := a$ and enumerate the elements of Z by $z_1, \ldots, z_\ell$ such that

$$a = z_0 < z_1 < z_2 < \cdots < z_\ell.$$

By Claim 11, it will be sufficient to show that there exists a constant $C_{16} > 0$ not depending on f or d such that

$$\int_{(a+\frac{\delta_a}{2})d}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{16} f^2. \tag{4.114}$$

for all $f < \frac{1}{2}$, $d \geq d_{\min}$ such that $\frac{f}{d}$ is sufficiently small. This will be accomplished by dividing up the interval $[(a + \frac{\delta_a}{2})d, \infty)$ into several subsets and proving estimates for each in a series of claims.

The first of these claims gives an estimate for the right-hand tail of the integral in Eq. (4.114).

Claim 13. *There exists a constant $C_2^* > 0$ such that*

$$\int_{(z_\ell + \frac{\delta_z}{2})d}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_2^* f^2$$

for all $f < \frac{1}{2}$ and $d \geq d_{\min}$ such that $\frac{f}{d} < \frac{1}{2}$.

Proof of Claim 13. By Claim 12 there exists a constant $\tilde{C}_{\tilde{K}}$ such that

$$g(y) \leq \tilde{C}_{\tilde{K}} g(x)$$

whenever $\tilde{K} \leq x < y$. Therefore by Lemma 38 (taking $K = \tilde{K}$ in the statement of the lemma), there exists a constant $C_{17} > 0$ depending on α but not on f or d such that

$$\int_{\alpha d\tilde{K}}^{\infty} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{17} f^2. \tag{4.115}$$

Since $\alpha > 1$ is arbitrary, it follows that α can be chosen sufficiently small that $\alpha\tilde{K} = \alpha(z_\ell + \frac{\delta_z}{3}) \leq z_\ell + \frac{\delta_z}{2}$, in which case the claim follows immediately from Eq. (4.115). □ Claim

Define

$$I := \bigcup_{i=1}^{\ell} \left[(z_i - \frac{\delta_z}{2})d, (z_i + \frac{\delta_z}{2})d\right]$$

In addition to the inequality in Eq. (4.104), further assume that

$$\frac{f}{d} < \min_{1 \leq i \leq \ell} \left\{ \frac{\delta_z}{4(\delta_z + z_i)} \right\} \tag{4.116}$$

for all $z \in Z$.

Claim 14. *There exists a constant $C_I > 0$ not depending on f or d such that*

$$\int_I \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_I f^2$$

for all $f < \frac{1}{2}$, $d \geq d_{\min}$ such that $\frac{f}{d}$ is sufficently small that the inequalities in Eqs. (4.104) *and* (4.116) *hold.*

Proof of Claim 14.

Let $i \in \{1, \ldots, \ell\}$. By Eq. (4.116) and **A.4**, the assumptions of Lemma 36 (ii.) are satisfied (with $y = z_i$, $\delta = \delta_z$ and $\lambda = \frac{1}{2}$ in the statement of the lemma); therefore there exists a constant $C^+_{z_i} > 0$ not depending on f or d such that

$$\int_{z_i d}^{(z_i + \frac{\delta_z}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C^+_{z_i} f^2.$$

In addition, the assumptions of Lemma 37 are similarly satisfied (with $y = z_i$, $\delta = \delta_z$); therefore there exists a contant $C^-_{z_i} > 0$ not depending on f or d such that

$$\int_{(z_i - \frac{\delta_z}{2})d}^{z_i d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C^-_{z_i} f^2.$$

Therefore, taking $C_{z_i} := C^+_{z_i} + C^-_{z_i}$, it follows that

$$\int_{(z_i - \frac{\delta_z}{2})d}^{(z_i + \frac{\delta_z}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{z_i} f^2.$$

Summing over $i = 1, \ldots, \ell$ implies the statement of the claim with $C_I := C_{z_1} + \ldots + C_{z_\ell}$. □ Claim

Next, define

$$\begin{aligned} J &:= \left[(z_0 + \frac{\delta_z}{2})d, (z_\ell + \frac{\delta_z}{2})d\right] \setminus I \\ &= \bigcup_{i=0}^{\ell-1} \left((z_i + \frac{\delta_z}{2})d, (z_{i+1} - \frac{\delta_z}{2})d\right) \end{aligned} \tag{4.117}$$

In the next two claims, we will utilize Lemma 36 to bound

$$\int_J \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt. \tag{4.118}$$

For each $i \in \{0, 1, \ldots, \ell - 1\}$, define

$$y_i := z_i + \frac{\delta_z}{2}$$

and

$$\delta_i := z_{i+1} - z_i - \frac{3}{4}\delta_z.$$

We claim that $\delta_i > 0$ for all i. In order to prove this, it is sufficient to demonstate that $z_{i+1} - z_i > \delta_z$ for all $i \in \{0, 1, \ldots, \ell - 1\}$. Indeed, suppose $z_{i+1} - z_i < \delta_z$ for some $i \in \{1, \ldots, \ell - 1\}$. Since $g(z_i) = 0$, **A.4** (ii) implies that $g(z_{i+1}) \geq \tilde{A} > 0$, a contradiction since $g(z_{i+1}) = 0$. By similar reasoning, **A.2** implies $z_1 - z_0 \geq \delta_a$. Moreover since $\delta_a \geq \delta_z$, it follows that $z_1 - z_0 \geq \delta_z$.

The next claim demonstrates that for each $i = 0, \ldots, \ell - 1$, the function g is uniformly bounded and bounded away from zero on the interval $[y_i, y_i + \delta_i]$.

Claim 15. *For each $i \in \{0, \ldots, \ell - 1\}$, there exist constants $A_i, B_i > 0$ such that*

$$A_i \leq g(x) \leq B_i$$

whenever $x \in [y_i, y_i + \delta_i]$.

Proof of Claim 15. There are two cases, $i = 0$ and $i \geq 1$, which are essentially similar. We start by proving the case with $i = 0$. By **A.4** (ii),

$$\tilde{A} \leq g(\xi) \leq \tilde{B}$$

for all $\xi \in [a + \delta_a, z_1 - \frac{\delta_z}{4}]$. In addition, Eq. (4.103) implies that if $\xi \in [a + \frac{\delta_z}{2}, a + \delta_a]$ then

$$\min\left\{A\left(\frac{\delta_z}{2}\right)^{k_a}, A\delta_a^{k_a}\right\} \leq g(\xi) \leq \max\left\{B\left(\frac{\delta_z}{2}\right)^{k'_a}, B\delta_a^{k'_a}\right\}.$$

Taking $A_0 = \min\left\{\tilde{A}, A\left(\frac{\delta_z}{2}\right)^{k_a}, A\delta_a^{k_a}\right\}$ and $B_0 = \max\left\{\tilde{B}, B\left(\frac{\delta_z}{2}\right)^{k'_a}, B\delta_a^{k'_a}\right\}$, it follows that

$$A_0 \leq g(\xi) \leq B_0$$

for all $\xi \in [a + \frac{\delta_z}{2}, z_1 - \frac{\delta_z}{4}] = [y_0 + y_0 + \delta_0]$. This completes the case with $i = 0$.

Next assume $i \in \{1, \ldots, \ell - 1\}$. Since

$$[y_i, y_i + \delta_i] = [y_i, z_i + \delta_z] \cup [z_i + \delta_z, z_{i+1} - \delta_z] \cup [z_{i+1} - \delta_z, y_i + \delta_i]$$

it will be sufficient to obtain bounds for $g(\xi)$ on each of the tree sets on the right-hand side.

- By **A.4** (ii),

$$\tilde{A} \leq g(\xi) \leq \tilde{B} \tag{4.119}$$

  whenever $\xi \in [z_i + \delta_z, z_{i+1} - \delta_z]$

- By **A.4** (i), g has almost-polynomial growth on $[z_i, z_i + \delta_z]$ with constants $\tilde{A}, \tilde{B}$ and exponents k, k' with $2k' - k \geq 0$; that is,

$$\tilde{A}(\xi - z_i)^k \leq g(\xi) \leq \tilde{B}(\xi - z_i)^{k'}$$

  whenever $\xi \in [z_i, z_i + \delta_z]$. Therefore since $k, k' \geq 0$,

$$\tilde{A}\left(\frac{\delta_z}{2}\right)^k \leq g(\xi) \leq \tilde{B}\delta_z^{k'} \tag{4.120}$$

  whenever $\xi \in [z_i + \frac{\delta_z}{2}, z_i + \delta_z] = [y_i, z_i + \delta_z]$

- By **A.4** (i), g has almost-polynomial decay on $[z_{i+1} - \delta_z, z_{i+1}]$ with constants $\tilde{A}, \tilde{B}$ and exponents $\hat{k}, \hat{k}'$ with $2\hat{k}' - \hat{k} \geq 0$. Therefore since $\hat{k}, \hat{k}' \geq 0$,

$$\tilde{A}(z_{i+1} - \xi)^{\hat{k}} \leq g(\xi) \leq \tilde{B}(z_{i+1} - \xi)^{\hat{k}'}$$

  whenever $\xi \in [z_{i+1} - \delta_z, z_{i+1}]$. Therefore,

$$\tilde{A}\left(\frac{\delta_z}{4}\right)^{\hat{k}} \leq g(\xi) \leq \tilde{B}\delta_z^{\hat{k}'} \tag{4.121}$$

  whenever $x \in [z_{i+1} - \delta_z, z_{i+1} - \frac{\delta_z}{4}] = [z_{i+1} - \delta_z, y_i + \delta_i]$

By Eqs. (4.119) to (4.121), for all $i \in \{1, \ldots, \ell - 1\}$, the statement of the claim holds with $A_i := \tilde{A} \cdot \min\left\{1, \left(\frac{\delta_z}{2}\right)^k, \left(\frac{\delta_z}{4}\right)^{\hat{k}}\right\}$ and $B_i := \tilde{B} \cdot \max\left\{1, \delta_z^{k'}, \delta_z^{\hat{k}'}\right\}$.

□ Claim

In addition to Eqs. (4.104) and (4.116), we now further assume that

$$\frac{f}{d} < \min_{i \in \{0,\ldots,\ell-1\}} \left\{ \frac{(1-\lambda_i)\lambda_i \delta_i}{2\lambda_i \delta_i + y_i} \right\} \tag{4.122}$$

The next claim utilizes Claim 15 and Lemma 36 to bound Eq. (4.118).

Claim 16. *There exists a constant $C_J > 0$ not depending on f or d such that*

$$\int_J \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_J f^2$$

for all $f < \frac{1}{2}$ and $d \geq d_{\min}$ such that $\frac{f}{d}$ is sufficiently small that both Eqs. (4.104) *and* (4.116) *hold.*

Proof of Claim 16. Let $i \in \{0, \ldots, \ell - 1\}$. Since

$$\lim_{\lambda \to 1} y_i + \lambda \delta_i = z_{i+1} - \frac{\delta_z}{4} > z_{i+1} - \frac{\delta_z}{2},$$

it follows that there exists $\lambda_i \in [\frac{1}{2}, 1)$ sufficiently close to 1 that

$$y_i + \lambda_i \delta_i \geq z_{i+1} - \frac{\delta_z}{2}.$$

Therefore

$$\int_{(z_i + \frac{\delta_z}{2})d}^{(z_{i+1} - \frac{\delta_z}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq \int_{y_i d}^{(y_i + \lambda_i \delta_i)d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt. \tag{4.123}$$

To bound the integral on the right-hand side, we will use Lemma 36 (ii.). In particular, by Eq. (4.122) and Claim 15 the assumptions of Lemma 36 are satisfied (with $y = y_i$, $\delta = \delta_i$, $k = k' = 0$, $A = A_i$, $B = B_i$, and $\lambda = \lambda_i$ in the statement of the lemma); therefore by

there exists a constant $C_{18} > 0$ not depending on f or d such that

$$\int_{y_i d}^{(y_i+\lambda_i\delta_i)d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{18} f^2. \tag{4.124}$$

Therefore by Eqs. (4.123) and (4.124),

$$\int_{(z_i+\frac{\delta_z}{2})d}^{(z_{i+1}-\frac{\delta_z}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_{18} f^2.$$

By Eq. (4.117), summing over $i = 0, \ldots, \ell - 1$ implies the statement of the claim. □ Claim

Since $\delta_z \leq \delta_a$, it holds that

$$\left[(a + \frac{\delta_a}{2})d, (z_\ell + \frac{\delta_z}{2})d\right] \subseteq \left[(a + \frac{\delta_z}{2})d, (z_\ell + \frac{\delta_z}{2})d\right] = I \cup J.$$

Therefore by monotonicity of the integral along with the bounds from Claims 14 and 16, we obtain the following claim:

Claim 17 (Intermediate t)**.** *There exists a constant $C_3^* := C_I + C_J > 0$ not depending on f or d such that*

$$\int_{(a+\frac{\delta_a}{2})d}^{(z_\ell+\frac{\delta_z}{2})d} \left(\sqrt{p(t)} - \sqrt{q(t)}\right)^2 dt \leq C_3^* f^2$$

for all $f < \frac{1}{2}$ and $d \geq d_{\min}$ such that $\frac{f}{d}$ is sufficiently small that the inequalities in Eqs. (4.104)*,* (4.116) *and* (4.122) *hold.*

Claims 11, 13 and 17 imply Eq. (4.52), which completes the proof. □

Chapter 5

Maximum Likelihood Inference and Long-Branch Attraction

Abstract:

Maximum likelihood estimation (MLE) is among the most widely-used methods for inferring phylogenetic trees from sequence data, however it is thought to perform poorly when the underlying tree parameter contains two or more distantly-related taxa. This phenomenon, known as long-branch attraction (LBA), is of considerable interest in evolutionary phylogenetics but is not fully understood.

With the aim of better understanding LBA, this chapter explores the problem of computing solutions to the maximum likelihood problem for 3- and 4-leaf trees under the 2-state symmetric mutation model (CFN model). We prove two main results. First, using Hadamard conjugation we compute a closed-form solution to the maximum likelihood problem for unrooted 3-leaf trees, given generic data. That is, we provide a formula for the numerical edge-length parameters, as a function of the data, which maximizes the log-likelihood function. For the 4-leaf case, by contrast, it is likely that no such closed-form solution exists [156]. Instead, we utilize the results for the 3-leaf

case, along with techniques from numerical algebraic geometry and phylogenetic invariants, to implement a fast algorithm to compute an *exact solution* to the 4-leaf MLE problem. This algorithm computes the maximum likelihood tree to arbitrary precision without the need to execute a heuristic (e.g. hill-climbing) search. Our second main result is a proof of the correctness of this algorithm. In both cases, particular attention is paid to account for submodels in which one or more branch of the underlying tree parameter has infinite length, as it is believed that such cases may provide special insight into the problem of LBA.

5.1 Introduction and Preliminaries

This chapter is concerned with inferring the evolutionary history of a set of a species or other taxa from sequence data using maximum likelihood. Our chapter is structured as follows. In Section 5.1 we introduce an evolutionary model, the maximum likelihood estimation for phylogenetic trees, provide background on phylogenetic invariants, and discuss what is known about long branch attraction, which motivates the work here. In Section 5.1.4 we recall an important parameterization of the Cavendar-Farris-Neyman (CFN) model and introduce results pertaining to the CFN model, including the essential technique of Hadamard conjugation. In Section 5.2, we present our first main result, a closed-form "analytic" solution to the maximum-likelihood problem for 3-leaf trees under the CFN model. In Section 5.3 we present our second main result, an algorithm for computing an exact solution the maximum-likelihood problem for four-leaf trees, and a proof of the correctness of the algorithm.

5.1.1 Long-branch attraction

A motivating goal for this chapter is to better understand a phenomenon usually referred to as **long branch attraction (LBA)**, which broadly construed, is understood as a type of estimation bias in which long branches are incorrectly inferred to be more closely

related to each other than they really are, though there is no widely-accepted and precise definition [3]. Long-branch attraction is thought to occur under biologically plausible parameter regimes, is frequently cited as a source of estimation error, and is thought to have the potential to cause major errors in inference [157, 158]. While LBA is most well understood in the context of parsimony, it has been shown to affect maximum likelihood estimation as well, however in the latter case it is less well understood, in large part because of the paucity of analytic solutions to the maximum likelihood problem in phylogenetics [3].

The estimation of trees with 3 and 4 leaves are of particular interest for two reasons. First, the inference of 3- and 4-leaf subtrees trees forms the foundation of a number of methods for reconstructing larger trees. Some examples of such methods include [159, 57, 77, 78, 31]. Second, these are the simplest cases in which the bias introduced by long branches under maximum likelihood estimation becomes evident [3]. Yet even in these simple cases, long branch attraction is mathematically not fully understood. In 2000, [160] computed a closed form solutions to the problem of inferring the *topology* of 3-leaf trees via maximum likelihood, but not the branch lengths. Analytic solutions for certain 4-leaf trees under molecular clock assumptions were obtained in [161, 156, 162], though in [156] it was shown that for ultrametric 4-leaf caterpillar trees, the critical points of the likelihood function as a closed-form expression using radicals. In particular, in [156, 161], the authors compute an analytic solution using rational expressions for the ML problem for ultrametric 4-leaf tree with balanced rooted topology under the CFN model. No closed form solution has been presented for the maximum likelihood estimate of 3-leaf trees for any mode of site substitution.

Two recent chapters have made significant progress in understanding long branch attraction. In [3], authors Parks and Goldman considered 3- and 4-leaf trees with the Jukes-Cantor model of site substitution. Using simulation and analytic solutions for boundary cases, the authors tested two hypotheses about the nature of LBA, which they call **long branch joining (LBJ)**, which refers to a propensity for maximum likelihood analysis

to return an incorrect tree topology in which the two longest branches are sister, and **long branch closeness (LBC)**, in which the maximum likelihood tree has the correct tree topology but branch lengths which erroneously place the two long branches closer to each other in the estimate than in the true tree. The authors provide strong evidence via simulation to show that while LBC does not occur, LBJ not only occurs, but also becomes worse as the length of the long branches increases relative to the size of the data. In addition, through a combination of simulation and analytic calculations, the authors demonstrated a close connection between distance-matrix estimates and maximum likelihood estimates, showing that properties of the former could reliably predict features of the latter.

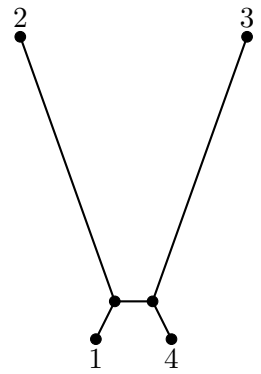


Figure 5.1: When the path between leaves 1 and 4 is very small relative to the branch lengths of leaves 2 and 3, the maximum likelihood topology is more likely to be 23|14 rather than the (true) topology 12|34; this phenomenon is colloquially known as long-branch attraction (LBA), or more accurately, long-branch joining (LBJ).

Another important recent contribution was [163], in which authors Susko and Roger provided the first mathematical proof that for any fixed data of finite length, there exists parameter regimes for four-leaf trees in which maximum likelihood is more likely to return the wrong topology, thus proving the existence of LBJ (see Fig. 5.1). Their proof relies on a limiting argument in which the length of the path between the two "short" leaves (i.e., the leaves 1 and 4 in Fig. 5.1) is taken to tend to zero, so that with high probability their bases do not differ. As a result, the maximum likelihood tree is more likely to have topology 23|14.

But this is not the end of the story; simulated regimes in [3] showed that the LBJ effect got worse as the two long branches got longer, *even keeping the length of the path between the short edges fixed*. In other words, there appears to be something about branches being long that *ipso facto* biases maximum likelihood analyses, which means the proof presented in [163] provides only a partial picture of the ways that long branches can bias maximum likelihood estimation. The aim of this chapter is to provide analytic or exact solutions to the maximum likelihood problem for 3- and 4-leaf trees with the aim of providing tools to help test and characterize the nature of that effect; the application of these results to long-branch attraction is left for future work.

5.1.2 Data and model of evolution

Tree parameter

Let $\mathcal{X}$ be a finite set. A **semi-labelled tree** (on $\mathcal{X}$) T is an ordered pair $(T;\theta)$ where T is a tree with vertex set $V(T)$ and the **labelling map** $\phi : \mathcal{X} \to V$ is a map such that $v \in \phi(\mathcal{X})$ whenever $v \in V(T)$ and $\deg(v) \leq 2$. A semi-labelled tree on $\mathcal{X}$ is also called an **$\mathcal{X}$-tree**. If ϕ is a bijection into the leaves of T, then T is called a **phylogenetic $\mathcal{X}$-tree**.

Associated to each $\mathcal{X}$-tree is a vector of nonnegative **branch lengths** $(d_e)_{e\in E(T)}$, where $E(T)$ is the edge set of T. We regard $\mathcal{X}$ as a set of taxa, with T representing a hypothesis about their evolutionary or genealogical history of the taxa; the branch lengths are regarded as representing a measure of evolutionary distance measured in *expected number of mutations per site*. Typically we take $\mathcal{X} = [n]$ and in this chapter will consider the cases with $n = 3$ or $n = 4$.

Rather than using the evolutionary distances d as edge parameters of T, for our analyses it will be more convenient to use an alternative parameterization of the branch lengths as a vector $\theta \in [0,1]^{2n-3}$, where

$$\theta_e := e^{-2d_e} \tag{5.1}$$

for all $e \in E(T)$. The numerical edge parameters $(\theta_e)_{e\in E(T)}$ have been referred to as "path-set variables" [156]; we refer to them here as the **Hadamard parameters**.

For a binary phylogenetic $[n]$-tree T, a **split** is a bipartition of the leaf set $[n]$ into two nonempty subsets A and B, and is denoted $A|B$. An edge $e \in E(T)$ is said to **induce** the split $A_e|B_e$ if removing e from T results in two disconnected components whose respective labelled nodes are A_e and B_e. We say that split $A|B$ is **compatible** with T if and only if there exists an edge $e \in E(T)$ such that $A|B = A_e|B_e$. For any $\mathcal{X}$-tree $\mathcal{T}$, denote the set of compatible splits (for T) by

$$\Sigma(\mathcal{T}) := \{A_e|B_e : e \in E(\mathcal{T})\}$$

The **topology** of T is determined by the set of compatible splits. There is only one unrooted topology for $n = 3$ (shown in Fig. 5.2); for 4-leaf trees, there are three unrooted topologies which correspond to the split obtained by the internal branch, and we denote these **quartet topologies** by 12|34, 13|24, and 23|14.

Site substitution model

We assume that the data X is generated according to the fully symmetric **Cavendar-Farris-Neyman (CFN)** model of site substitution (also known as the N_2 model in [79]), which takes as input a tree parameter with branch lengths. The CFN model, time-reversible Markov chain on a tree, is the simplest model of site substitution, possessing only two nucleotide states, which we denote by $+1$ (pyrimidine) and -1 (purine). Under this model, the probability of a nucleotide in state i transitioning to state j over an edge of length t can be shown to be

$$p_{ij}(t) = \begin{cases} \frac{1}{2}\left(1 + e^{-2t}\right) & \text{if} \quad i = j \\ \frac{1}{2}\left(1 - e^{-2t}\right) & \text{if} \quad i \neq j \end{cases} \tag{5.2}$$

for $i, j \in \{-1, +1\}$ ([160], and for a more general reference, see [79, p.197]). Moreover, we assume a uniform root distribution, from which it follows that

$$\mathbb{P}[X = \sigma] = \mathbb{P}[X = -\sigma] \tag{5.3}$$

for all $\sigma \in \{-1,1\}^n$ (c.f. Lemma 8.6.1.(ii) in [79], see also [164, p.221]).

The distribution of X under the CFN model depends on the tree parameter, including branch lengths. For an $[n]$-tree T with Hadamard branch parameters $(\theta_e)_{e\in E(T)}$ and topology τ, let

$$p_\sigma(\theta,\tau) := \mathbb{P}\left[X=\sigma\right],$$

where $\mathbb{P}$ is the distribution of X under the CFN model on T.

For any nonnegative integer r, let $\Delta_{r-1} \subset \mathbb{R}^r$ denote the probability simplex of dimension $r-1$; that is, $\Delta_{r-1} := \{x \in \mathbb{R}^r : x_1,\ldots,x_r \geq 0 \text{ and } x_1+\ldots+x_r = 1\}$. As a parameterized statistical model for an unrooted, n-leaf tree with topology $\tau \in \{12|34, 13|24, 23|14\}$, the CFN model is the image of the map

$$\begin{aligned} \Psi_\tau : \Theta_\tau &\to \Delta_{2^n-1} \subseteq \mathbb{R}^{2^n} \\ \theta &\mapsto p(\theta,\tau) \end{aligned} \tag{5.4}$$

where

$$p(\theta,\tau) := \left(p_\sigma(\theta,\tau)\right)_{\sigma\in\{-1,+1\}^n}.$$

Identifiability of Model Parameters

The usual assumption prescribed for the CFN model (which is *not* made in this chapter) is that $\Theta_\tau = (0,1)^{|E(T)|}$ (so that $0 < \theta_e < 1$ for all $e \in E(T)$, or equivalently, that $0 < d_e < \infty$ for all $e \in E(T)$). Under that assumption, Ψ_τ is injective [165], and hence the edge parameters $\theta = (\theta_e)_{e\in E(T)}$ are **identifiable**. This means that if T and T' are two n-leaf trees with the same topology τ and with edge parameters θ and θ' respectively such that $\theta \neq \theta'$, then the distribution of X will be different under T and T').

On the other hand in this chapter, we consider an extension taking $\Theta_\tau = [0,1]^{|E(T)|}$, thereby allowing for branch lengths which are infinite or zero (when measured in expected number of mutations per site), in order to better understand the behavior of maximum likelihood estimation in the limit as one or more branch lengths tend to zero or to infinity. This seemingly slight extension of the model substantially adds to the complexity of the

analysis. In particular, as a consequence of this extension, it is no longer the case that the numerical parameters θ are identifiable, which presents certain complications, described in detail later in the chapter. On the other hand, it also has the effect of guaranteeing the existence of the maximum likelihood estimate.

Under the CFN model, the discrete parameter τ, representing the unrooted tree topology, is **generically identifiable**, which means that for distinct unrooted n-leaf topologies τ, τ', the intersection

$$\mathrm{im}(\Psi_\tau) \cap \mathrm{im}(\Psi_{\tau'}) \subseteq \Delta_{2^n-1}$$

is of strictly lower dimension than both $\mathrm{im}(\Psi_{\tau'})$ and $\mathrm{im}(\Psi_\tau)$, and hence is a set of Lebesgue measure zero [166, Chapter 16.1].

Data

In practice, DNA sequence data is typically arranged as a **multiple sequence alignment**, an $n \times N$ matrix, with each row corresponding to a leaf of T and each column representing an aligned site position. It is standard (albeit unrealistic) to assume that sites evolved independently, and as such our data consists of N random column vectors

$$X^{(1)}, \ldots, X^{(N)} \overset{iid}{\sim} X$$

where X is a random variable taking values in $\{+1, -1\}^n$ whose distribution will be described below, and which is regarded as a vector of nucleotides observed at the leaves of T such that X_i is the nucleotide observed at the vertex with label i for each $i \in [n]$. Under the CFN model, the distribution of X depends on the tree parameter T (both the topology τ of T and its branch lengths). Due to the exchangeability of $X^{(1)}, \ldots, X^{(N)}$, the data can be summarized by a **site frequency vector**

$$\mathbf{s} := (s_\sigma)_{\sigma \in \{-1,1\}^n} \tag{5.5}$$

where

$$s_\sigma := \#\left\{i \in [N] : X^{(i)} = \sigma\right\}.$$

Throughout this chapter, we make the assumption that the site frequency vector satisfies

$$s_\sigma + s_{-\sigma} > 0 \tag{5.6}$$

for all $\sigma \in \{-1, +1\}^n$. In other words, this says that each pattern

$$aaaa, aaab, aaba, aabb, abaa, abab, abba, abbb,$$

(where a and b represent different nucleotides) is observed at least once. This assumption considerably simplifies the number of cases which must be considered. Since the number of site patterns is 2^{n-1}, this assumption is reasonable here (with $n = 3$ or $n = 4$), however it would be unrealistic for a tree with many more leaves (e.g., $n > 30$) given the size of genomic datasets [162].

α-Site Patterns

In light of Eq. (5.3), it is possible using a change of coordinates to represent the distribution of X using a vector of 2^{n-1} entries rather than 2^n entries. To do so we introduce the following notation, first given by [167], in which we augment the subsets of $[n-1]$ with a lexicographic ordering, assumed throughout this chapter. Specifically, the ordering assumed on the subsets of $[n-1]$ is $(\alpha_1, \ldots, \alpha_{2^{n-1}})$, where

$$\alpha_k = \{i \in [n-1] : \text{the } i\text{-th digit of the binary representation of } k,$$
$$\text{counting from the right is } 1\}.$$

For the cases we will consider ($n = 3$ and $n = 4$), this yields the orderings

$$(\emptyset, \{1\}, \{2\}, \{1, 2\})$$

and

$$(\emptyset, \{1\}, \{2\}, \{1,2\}, \{3\}, \{1,3\}, \{2,3\}, \{1,2,3\}).$$

For any vector $\sigma \in \{-1,+1\}^n$ and any $\alpha \in [n-1]$, we say that σ has **α-split pattern** if there exists $k \in \{-1,+1\}$ such that $\sigma_i = k$ if and only if $i \in \alpha$. For example, if $n = 4$ then $(+1,-1,-1,+1)$ and $(-1,+1,+1,-1)$ both have $\{2,3\}$-split pattern; these are the only vectors in $\{-1,+1\}^4$ with this split pattern, since $\sigma, \tilde{\sigma} \in \{-1,+1\}^n$ share the same split pattern if and only if $\sigma = \pm\tilde{\sigma}$.

Define

$$\bar{p} := (\bar{p}_\alpha)_{\alpha \subseteq [n-1]},$$

where, for each $\alpha \subseteq [n-1]$,

$$\bar{p}_\alpha := \mathbb{P}[X \text{ has } \alpha\text{-split pattern}].$$

In other words, $\bar{p}_\alpha$ is the probability that the entries of X correspond to the split $\alpha|([n]\backslash\alpha)$. By Eq. (5.3) (see also [168]), the distribution $\mathbb{P}$ is summarized without loss of information by $\bar{p}$, and hence the CFN model for a fixed tree topology τ may be regarded as the image of the map

$$\begin{aligned} \bar{\Psi}_\tau : \Theta_\tau &\to \Delta_{2^{n-1}-1} \subseteq \mathbb{R}^{2^{n-1}} \\ \theta &\mapsto \bar{p}(\tau, \theta) \end{aligned} \tag{5.7}$$

and in our setting with $n = 4$ and $\Theta_\tau = [0,1]^{|E(T)|}$, this obtains

$$\overline{\mathcal{M}}_\tau := \left\{\bar{p}(\tau,\theta) : \theta \in [0,1]^5\right\} \subseteq \Delta_7. \tag{5.8}$$

Similarly, the data may be summarized by the following sufficient statistic

$$\bar{\mathbf{s}} := (\bar{s}_\alpha)_{\alpha \in [n-1]}, \tag{5.9}$$

where

$$\bar{s}_\alpha := \#\left\{i \in [N] : X^{(i)} \text{ has } \alpha\text{-split pattern}\right\}.$$

5.1.3 The phylogenetic maximum likelihood problem

Given a tree T with unrooted tree topology τ and associated branch lengths $\theta = (\theta_e)_{e \in E(T)}$, denote the distribution of X under the CFN model parameterized by T by $p(\tau, \theta) = (p_\sigma)_{\sigma \in \{-1,1\}^n}$ where $p_\sigma = \mathbb{P}[X = \sigma]$. Given data taking the form Eq. (5.5), or equivalently Eq. (5.9), the **log-likelihood function** is the function

$$\begin{aligned} \ell(\theta) &:= \sum_{\sigma \in \{-1,1\}^n} s_\sigma \log p_\sigma(\tau, \theta) \\ &= -N \log 2 + \sum_{\alpha \in [n-1]} \bar{s}_\alpha \log \bar{p}_\alpha(\tau, \theta) \end{aligned} \tag{5.10}$$

(Both of these forms of the log-likelihood will be used in this chapter, as the two notations s_σ, p_σ and $\bar{s}_\alpha, \bar{p}_\alpha$ are each uniquely well-suited to particular contexts.)

The **maximum likelihood problem** is to find the parameters τ and $\theta \in [0,1]^{|E(T)|}$ which maximize Eq. (5.10). Since the log-likelihood is an upper semicontinuous function and the parameter space is compact, such a maximum is guaranteed to exist.

For $n = 3$, there is only one unrooted tree topology, and so this problem reduces to that of finding the numerical parameters $\theta \in [0,1]^3$ which maximize Eq. (5.10).

For $n = 4$, the problem is to find the parameters $\tau \in \{12|34, 13|24, 23|14\}$ and $\theta \in [0,1]^5$ which maximize Eq. (5.10). That is, in the four-leaf case it suffices to select a model from three models with maximal dimension and boundary cases (i.e., when one or more numerical parameters equals zero or one). Identifying these boundary cases is the main content of the proof of Theorem 10, since the proliferation of boundary cases means that finding the exact MLE necessitates a framework to allow for parsing a large number of overlapping boundary cases.

In model selection for phylogenetics, the main objective is to recover the tree parameter from data. The number of trees to consider is exponential in the number of leaves.

Practical methods use quartets (subtrees with four leaves) to perform the model selection. In a quartet method, one does model selection by working with many four-leaf models to identify the n-leaf tree parameter.

In practice, maximum likelihood inference is among the most commonly-used in phylogenetic analyses, and in contrast to the simple (but more analytically tractable) model considered in this chapter, maximum likelihood estimation is typically undertaken with sophisticated models of site evolution, utilizing heuristic (e.g. hill-climbing) methods and multiple start points to explore tree space in order to obtain parameters which maximizes the likelihood. A variety of excellent and widely-used implementations (e.g., [169, 170, 171, 172]) all of which have been used in thousands (or tens of thousands) of studies.

Nonetheless, there remains interest in computing *analytic* (i.e., closed-form) solutions in simpler cases [161, 156, 162, 160] as well as *exact* solutions via algebraic methods [173] with the goal of providing more rigorous understanding of the ways that maximum likelihood can fail. For example, it is well-known that the maximum likelihood tree need not be unique [174], that for certain data there exists a continuum of trees which maximize the likelihood [162], and that there exist data for which the maximum likelihood estimate does not exist (or at least, is not a tree with finite branch lengths) [173].

5.1.4 Hadamard conjugation

In this section we introduce an important reparametrizaton of the CFN model, as well as a central tool in our analyses: Hadamard conjugation.

For any even subset $Y \subseteq [n]$, define the **path set** $P(\mathcal{T}, Y)$ induced by Y on $\mathcal{T}$ to be the set of $\frac{1}{2}|Y|$ edge-disjoint paths in $\mathcal{T}$, each of which connects a pair of leaves labelled by elements from Y, taking $P(\mathcal{T}, \emptyset) = \emptyset$. This set is unique if $\mathcal{T}$ is a binary tree [79].

By Eq. (5.2), the transition probability from state i to j along a given edge e, denoted $M_e(i, j)$, is given by

$$M_e(i, j) = \frac{1}{2}(1 + ij\theta_e) \tag{5.11}$$

for $i, j \in \{+1, -1\}$.

The **edge spectrum** is the vector

$$\gamma := (\gamma_\alpha)_{\alpha \subseteq [n-1]},$$

where

$$\gamma_\alpha := \begin{cases} -\sum_{e \in E(T)} d_e & : \quad \alpha = \emptyset \\ d_e & : \quad e \text{ induces the split } \alpha | ([n] \backslash \alpha) \\ 0 & : \quad \text{else} \end{cases}$$

Inductively define $H_0 := [1]$, and for $k \geq 0$ define

$$H_{k+1} := \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix} \tag{5.12}$$

Let $H := H_{n-1}$. Then H is a $2^{n-1} \times 2^{n-1}$ matrix, and by our choice of ordering for $[n-1]$, we have $H = (h_{\alpha,\beta})_{\alpha,\beta \subseteq [n-1]}$ where

$$h_{\alpha,\beta} = (-1)^{|\alpha \cap \beta|}.$$

In particular, H is a symmetric Hadamard matrix with $H^{-1} = \frac{1}{2^{n-1}} H$.

Letting $x \mapsto \exp(x)$ denote the exponential function applied component-wise, the map $\mathbb{R}^{2^{n-1}} \to \mathbb{R}^{2^{n-1}}$ given by

$$\gamma \mapsto H^{-1} \exp(H\gamma)$$

is referred to as **Hadamard conjugation**. As in many previous results (e.g., [156, 162], and not solely for the CFN model), Hadamard conjugation turns out to be an essential tool and is closely related to the discrete Fourier transform [175, 166, 176], as in fact in here multiplication by H can be regarded as a discrete Fourier transformation. Hadamard conjugation allows us to translate between the edge spectrum and the expected site pattern spectrum, as shown in the following theorem, discovered independently by [167, 177]:

Theorem 8 (Hadamard conjugation theorem)**.** *Let γ be the edge spectrum of a phyloge-*

netic $[n]$-tree T, and let $H := H_{n-1}$. Then

$$p = H^{-1} \exp(H\gamma).$$

For proof and a detailed discussion, we refer the reader to [79]. In particular, we will utilize the following proposition, itself a consequence of Theorem 8.

Proposition 3 (Corollary 8.6.6 in [79])**.** *Let $\theta_e \in [0,1]$ for all $e \in E(T)$. Then for all subsets α of $[n-1]$,*

$$\bar{p}_\alpha = \frac{1}{2^{n-1}} \sum_{\substack{Y \subseteq X: \\ Y \text{ even}}} \left[(-1)^{|Y \cap \alpha|} \prod_{e \in P(\mathcal{T}, Y)} \theta_e \right]. \tag{5.13}$$

Note that Proposition 3 holds even if the root distribution is not taken to be uniform, however we do not consider that case in this chapter. The key takeaway is that the probability of any site pattern can be computed as a *polynomial* function of the Hadamard parameters. This fact, and the formula it provides, will be critical in the work that follows.

Example 2 (4-leaf tree)**.** Suppose $n = 4$, and T has topology $12|34$. Here, the subsets of $[3]$ are ordered as follows:

$$(\emptyset, \{1\}, \{2\}, \{1,2\}, \{3\}, \{1,3\}, \{2,3\}, \{1,2,3\})$$

For $i \in [4]$, let d_i be the branch length of leaf i, and let d_5 be the internal branch length

(corresponding to the edge with split $\{1,2\}|\{3,4\}$). Then

$$H_3 = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \end{bmatrix} \quad \text{and} \quad \gamma = \begin{bmatrix} -(d_1+d_2+d_3+d_4+d_5) \\ d_1 \\ d_2 \\ d_5 \\ d_3 \\ 0 \\ 0 \\ d_4 \end{bmatrix},$$

where γ is indexed by the ordered subsets of [3]. Therefore

$$\exp(H_3\gamma) = \exp\left(\begin{bmatrix} 0 \\ -2(d_1+d_4+d_5) \\ -2(d_2+d_4+d_5) \\ -2(d_1+d_2) \\ -2(d_3+d_4) \\ -2(d_1+d_3+d_5) \\ -2(d_2+d_3+d_5) \\ -2(d_1+d_2+d_3+d_4) \end{bmatrix}\right) = \begin{bmatrix} 1 \\ \theta_1\theta_4\theta_5 \\ \theta_2\theta_4\theta_5 \\ \theta_1\theta_2 \\ \theta_3\theta_4 \\ \theta_1\theta_3\theta_5 \\ \theta_2\theta_3\theta_5 \\ \theta_1\theta_2\theta_3\theta_4 \end{bmatrix}. \tag{5.14}$$

The entries of this vector constitute a monomial parameterization of the CFN model on T, a fact which is widely-used in algebraic statistics in the study of more general group-based models [166], where they are called **Fourier coordinates**. (To see why these are in fact the Fourier coordinates, understood as the Fourier transform of the vector of probability coordinates $\bar{p} = (\bar{p}_\alpha)_{\alpha\subseteq[3]}$, it is enough to note that H_3 may be regarded as a discrete Fourier transformation and $\exp(H_3\gamma) = H_3 p$ by Theorem 8). Because the Fourier coordinates factorize into Hadamard parameters, as shown in Eq. (5.14), the Fourier coordinates thus have a simple biological interpretation. Consider the entry $\theta_1\theta_4\theta_5$

for example. By Eq. (5.1), $\theta_1\theta_4\theta_5 = e^{-2(d_1+d_4+d_5)}$. Since $d_1 + d_4 + d_5$ is the length of the path on T from leaf 1 to leaf 4, we we might expect $\theta_1\theta_4\theta_5$ to measure, in some sense, the correlation between X_1 and X_4. And indeed this turns out to be the case. Since $E[X_1] = E[X_4] = 0$, we have $\mathrm{cov}(X_1, X_4) = E[X_1X_4] = \theta_1\theta_4\theta_5$ by Eq. (5.11). Moreover, since the standard deviations of X_1 and X_4 are both equal to one, it follows that $\theta_1\theta_5\theta_4$ is prescisely the Pearson correlation coefficient between X_1 and X_4.

This example also illustrates how Proposition 3 can be obtained. Multiplying both sides of Eq. (5.14) by $H_3^{-1} = \frac{1}{8}H_3$ (i.e., the inverse discrete Fourier transform), gives the formula for $\bar{p}_\alpha$ in Proposition 3, with $n = 4$. By a change of notation, and using Eq. (5.3), we obtain the formula

$$\begin{aligned}\mathbb{P}[X = \sigma] = \frac{1}{16}\,(1 &+ \sigma_1\sigma_2\theta_1\theta_2 + \sigma_3\sigma_4\theta_3\theta_4 + \sigma_1\sigma_3\theta_1\theta_3\theta_5 + \sigma_1\sigma_4\theta_1\theta_4\theta_5 \\ &+ \sigma_2\sigma_3\theta_2\theta_3\theta_5 + \sigma_2\sigma_4\theta_2\theta_4\theta_5 + \sigma_1\sigma_2\sigma_3\sigma_4\theta_1\theta_2\theta_3\theta_4)\,.\end{aligned}$$

By nothing more than a relabelling of leaves, if we suppose that T has leaves $i, j, k, l \in [4]$ with $\{i, j, k, l\} = [4]$ and topology $ij|kl$, Proposition 3 implies that for all $\sigma \in \{-1, 1\}^4$,

$$\begin{aligned}\mathbb{P}[X = \sigma] = \frac{1}{16}\,(1 &+ \sigma_i\sigma_j\theta_i\theta_j + \sigma_k\sigma_l\theta_k\theta_l + \sigma_i\sigma_k\theta_i\theta_k\theta_5 + \sigma_i\sigma_l\theta_i\theta_l\theta_5 \\ &+ \sigma_j\sigma_k\theta_j\theta_k\theta_5 + \sigma_j\sigma_l\theta_j\theta_l\theta_5 + \sigma_i\sigma_j\sigma_k\sigma_l\theta_i\theta_j\theta_k\theta_l)\,.\end{aligned} \tag{5.15}$$

We will use this formula when computing likelihoods for 4-leaf trees.

Interpretation of Hadamard parameters under the CFN model

We regard a vector of Hadamard parameters $\theta = (\theta_e)_{e\in E(T)}$ as **biologically plausible** if $\theta_e \in (0, 1)$ for all $e \in E(T)$. Since $-\frac{1}{2}\log\theta_e$ is the expected number of mutations on edge e, it follows that $\theta_e \in (0, 1)$ if and only if $d_e \in (0, \infty)$. In other words, biologically plausible Hadamard parameters correspond to trees with branch lengths having positive and finite expected number of mutations per site.

Unlike evolutionary distances, which are additive, Hadamard parameters are multi-

plicative, in the sense that the "length" of a path $P \subseteq E(T)$ is $\prod_{e \in P} \theta_e$.

In this work, we allow for $\theta_e \in [0, 1]$ in order to better study the behavior of maximum likelihood inference in the setting of extremely short or long branches, since this is the setting where long-branch attraction is hypothesized to occur. As we saw in Example 2, any $\theta_e \in [0, 1]$ can be regarded as a measure of correlation between the state of the Markov process at the endpoints of the edge e. Suppose $e = (u, v) \in E(T)$ and let X_u and X_v denote the state of the Markov process at nodes u and v respectively. Eq. (5.11) implies that if $\theta_e = 1$ then $X_u = X_v$ with probability 1. On the other hand, if $\theta_e = 0$ then X_u and X_v are independent; to see why this is the case, observe that using Eq. (5.11), it holds for all $i, j \in \{-1, 1\}$ that

$$\begin{aligned}
\mathbb{P}[X_u = i, X_v = j] &= \mathbb{P}[X_u = i]\,\mathbb{P}[X_v = j \mid X_u = i] \\
&= \frac{1}{2} \cdot \frac{1}{2} \\
&= \mathbb{P}[X_u = i]\,\mathbb{P}[X_v = j]\,.
\end{aligned}$$

The observation has an important consequence which is summarized in the next lemma.

Lemma 40 (Independence caused by "infinitely long" branches)**.** *Suppose T is an unrooted n-leaf tree. Let $e \in E(T)$ and let $A_e|B_e$ denote the split induced by e on the leaf set $[n]$. If $\theta_e = 0$ then the random vectors $(X_i : i \in A_e)$ and $(X_i : i \in B_e)$ are independent.*

Proof. Let $A := (X_i : i \in A_e)$ and $B := (X_i : i \in A_e)$. Write $e = (e_A, e_B)$, and without loss of generality assume that e_A and e_B are labeled such that any path from a leaf in A_e to e_A does not contain e_B. Let Z_A, Z_B denote the nucleotide states of vertices e_A and e_B respectively. It follows by Eq. (5.11) and symmetry of the process that Z_A and Z_B are independent.

Let $a \in \{1,-1\}^{|A_e|}$ and $b \in \{1,-1\}^{|B_e|}$. Then using the Markov property,

$$\begin{aligned}
\mathbb{P}[A=a,B=b] &= \sum_{i,j\in\{-1,1\}} \mathbb{P}[A=a,B=b \mid Z_A=i,Z_B=j]\,\mathbb{P}[Z_A=i,Z_B=j] \\
&= \sum_{i,j\in\{-1,1\}} \mathbb{P}[A=a \mid Z_A=i]\,\mathbb{P}[B=b \mid Z_B=j]\,\mathbb{P}[Z_A=i,Z_B=j] \\
&= \sum_{i,j\in\{-1,1\}} \mathbb{P}[A=a \mid Z_A=i]\,\mathbb{P}[B=b \mid Z_B=j]\,\mathbb{P}[Z_A=i]\,\mathbb{P}[Z_B=j] \\
&= \left(\sum_{i\in\{-1,1\}} \mathbb{P}[A=a,Z_A=i]\right)\left(\sum_{j\in\{-1,1\}} \mathbb{P}[B=b,Z_B=j]\right) \\
&= \mathbb{P}[A=a]\,\mathbb{P}[B=b]\,.
\end{aligned}$$

□

5.1.5 Phylogenetic invariants

The study of algebraic invariants, understood broadly as polynomial ideals which characterize statistical models, has grown substantially over the last two decades, with substantial progress made in understanding, computing, and characterizing the algebraic invariants for a wide range of models, including a broad class of site substitution models on both trees and networks (see, e.g., [178, 166, 168, 179, 180]. While much of the focus is on resolving questions of parameter identifiability (i.e. whether the model parameters can be recovered from the distribution), another key area of focus has been toward the application of invariants to maximum likelihood estimation [181, 168, 182].

The use of algebraic invariants to solve maximum likelihood problems in phylogenetics appears to have been first used in [162], and a general framework has been developed [181, 182] (for good introductions to this topic, see [168, p.132–135], and [166, chapters 7 and 15]). The key insight in the application of invariants to maximum likelihood estimation is that it allows one to reformulate the problem of maximizing the likelihood from an *unconstrained* optimization problem into a *constrained* optimization problem (with algebraic invariants as constraints) using the method of Lagrange multipliers, thus obviating the

need to search around the parameter space to maximize the likelihood function [181].

In particular, the use of Lagrange multipliers allows the reduction of the problem to that of finding the solutions to a system of polynomials, which can be computed precisely using techniques from numerical algebraic geometry, such as homotopy continuation. These techniques allow for the "exact" computation of maximum likelihood estimate, in the sense that they return theoretically correct solutions for a.e. data, and that the solutions can be computed to an arbitrary level accuracy[173, 183]. More recently, [173] presented an algorithm incorporating semialgebraic constraints (i.e. polynomial inequalities) and demonstrated an application of this approach on 3-leaf phylogenetic trees. In this chapter, we utilize a similar approach to compute exact solutions to the maximum likelihood problem for four-leaf trees. We sketch the general approach and the particular invariants used here.

Informally, for a fixed topology $\tau \in \{12|34, 13|24, 23|14\}$, this approach seeks to maximize the log-likelihood function ℓ in Eq. (5.10) over all $\bar{p} = (\bar{p}_\alpha)_{\alpha \subseteq [n-1]}$ subject to certain polynomial constraints

$$g_i(\bar{p}) = 0, i = 1, \ldots, c$$

which are common to all probability distributions which may arise under a given site substitution model (e.g. the CFN model, Jukes-Cantor, etc) on a tree with topology τ. This problem can be solved using the method of Lagrange multipliers, in which the gradient of the Lagrangian is a system of polynomials in $2^{n-1} + c$ variables the solutions of which correspond to critical points of Eq. (5.10). Such solutions can be computed using numerical algebraic techniques, such as homotopy continuation, to arbitrary degree of precision. Moreover, since for generic data there are at most finitely many critical points, this approach enables one to compute the maxima of Eq. (5.10) to arbitrary level of precision, without the need for heuristic search. We rely on this approach in our analyses of 4-leaf trees, where closed form solutions are not possible.

For the remainder of this section, it will be convenient to write $p_1, \ldots, p_8$ in place of

$(\bar{p}_\alpha)_{\alpha\in[3]}$.

We note that $\bar{\Psi}_\tau$, defined in Eq. (5.7), is a polynomial map by Proposition 3, and may be extended to a polynomial map $\mathbb{C}^5 \to \mathbb{C}^8$. Therefore the Zariski closure of its image, $\mathcal{V}_\tau := \overline{\bar{\Psi}_\tau(\mathbb{C}^5)}$, is an algebraic variety (see, e.g., [184]), called a **phylogenetic variety**. The ideal generated by $\mathcal{V}_\tau$ is the **phylogenetic ideal** $\mathcal{I}_\tau := I(\mathcal{V}_\tau)$, consisting of all polynomials of the form $f \in \mathbb{C}[p_1, \ldots, p_8]$ such that

$$f(p_1, \ldots, p_8) = 0$$

for all $p \in \overline{\mathcal{M}}_\tau$, and can be thought to implicitly define the statistical model. The generators of the phylogenetic ideal are referred to as **phylogenetic invariants**[1]. Since the set $\overline{\mathcal{M}}_\tau$ is a proper subset of $\mathcal{V}_\tau$, the phylogenetic invariants always vanish on $\overline{\mathcal{M}}_\tau$ and may thus be regarded as *polynomial relations* which constrain the possible distributions which may be obtained by the CFN model on a 4-leaf tree with topology τ. Both phylogenetic varieties and invariants are well-studied objects (for good introductions, see [166, 168]).

For unrooted 3-leaf tree under the CFN model, there are no phylogenetic invariant [178]. On the other hand, the generators of $\mathcal{I}_{12|34}, \mathcal{I}_{13|23}$ and $\mathcal{I}_{23|14}$ are well-known and usually expressed in terms of the Fourier coordinates, which are obtained via an invertible linear transformation of the probability coordinates $p_1, \ldots, p_8$ [164, p.221] (see also [166, 175]). To keep this exposition self contained, we compute the minimal generators of $\mathcal{I}_{12|34}, \mathcal{I}_{13|23}, \mathcal{I}_{23|14}$ and $\mathcal{I}_{\text{star}}$ directly in the probability coordinates $p_1, \ldots, p_8$ using `Macaulay2`; the code we used can be found in Fig. A.4 in the appendix. In particular,

[1] Some authors use the term "phylogenetic invariants" to refer to any elements of a phylogenetic ideal.

letting

$$
\begin{aligned}
g_0 &:= p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 - 1 \\
g_1 &:= p_3p_5 - p_4p_6 + p_2p_7 + p_3p_7 + p_4p_7 + p_5p_7 + p_6p_7 + p_7^2 + p_2p_8 + p_7p_8 - p_7 \\
g_2 &:= p_2p_5 + p_2p_6 + p_3p_6 + p_4p_6 + p_5p_6 + p_6^2 - p_4p_7 + p_6p_7 + p_3p_8 + p_6p_8 - p_6 \\
g_3 &:= p_2p_3 + p_2p_4 + p_3p_4 + p_4^2 + p_4p_5 + p_4p_6 + p_4p_7 - p_6p_7 + p_4p_8 + p_5p_8 - p_4,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\mathcal{I}_{12|34} &= \langle g_0, g_1, g_2 \rangle \\
\mathcal{I}_{13|24} &= \langle g_0, g_1, g_3 \rangle \\
\mathcal{I}_{23|14} &= \langle g_0, g_2, g_3 \rangle \\
\mathcal{I}_{\text{star}} &= \langle g_0, g_1, g_2, g_3 \rangle ,
\end{aligned}
\tag{5.16}
$$

where the angle bracket notation is used to indicate the generators of the ideal; e.g., so that

$$
\langle g_0, g_1, g_2 \rangle = \{ h_0 g_0 + h_1 g_1 + h_2 g_2 : h_0, h_1, h_2 \in \mathbb{C}[p_1, \ldots, p_n] \} .
$$

5.2 Analytic Solution to The 3-Leaf MLE Problem

In this section we present an analytic solution to the problem of finding all numerical parameters $\theta_1, \theta_2, \theta_3 \in [0, 1]$ which maximize the log-likelihood function ℓ, given generic data from a 3-leaf tree with CFN constraints. Despite longstanding interest in this and related problems (see, e.g., [160, 173, 163, 3]), the existence of such a solution does not appear to have been known or published prior to this work.

5.2.1 Definitions and Assumptions

The data takes the form of N independent, identically distributed random variables $X^{(1)}, \ldots, X^{(N)} \sim X$, where X is drawn according to the CFN process on the unrooted 3-leaf $\mathcal{T}$ shown in Fig. 5.2 with unknown but fixed numerical edge parameters $\theta^{\mathcal{T}} =$

$(\theta_1^{\mathcal{T}}, \theta_2^{\mathcal{T}}, \theta_3^{\mathcal{T}}) \in (0,1)^3$.

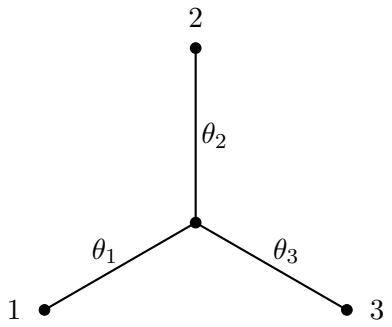


Figure 5.2: Three-leaf tree with Hadamard edge parameters $\theta_1, \theta_2, \theta_3$.

Since there is only one possible topology for an unrooted 3-leaf tree, the maximum likelihood problem reduces to the problem of determining the numerical parameters $\theta \in [0,1]^3$ maximize Eq. (5.10). The main result of this section is a theorem which presents an analytic solution to the maximum likelihood problem for generic data generated in an i.i.d. manner according to the CFN site substitution process on an unrooted 3-leaf tree.

By Eq. (5.13), for each $\alpha \subseteq [2]$, the α-th component of p is

$$\bar{p}_\alpha(\theta) = \frac{1}{4}\left(1 + \sum_{1 \le i < j \le 3} (-1)^{|\{i,j\} \cap \alpha|} \theta_i \theta_j\right). \tag{5.17}$$

By a change of notation, this can be rewritten as

$$\mathbb{P}\left[X \in \{-\sigma, \sigma\}\right] = \frac{1}{4}(1 + \sigma_1\sigma_2\theta_1\theta_2 + \sigma_1\sigma_3\theta_1\theta_3 + \sigma_2\sigma_3\theta_2\theta_3)$$

for all $\sigma \in \{-1, +1\}^3$, and therefore by Eq. (5.3), it follows that

$$\mathbb{P}\left[X = \sigma\right] = \frac{1}{8}(1 + \sigma_1\sigma_2\theta_1\theta_2 + \sigma_1\sigma_3\theta_1\theta_3 + \sigma_2\sigma_3\theta_2\theta_3) \tag{5.18}$$

for all $\sigma \in \{+1, -1\}^3$.

The key statistics used are the following:

Definition 3 (The statistics $M_{ij}^+, M_{ij}^-, B_{ij}, \mathbf{B}$)**.** *For for all $i, j \in [n]$ such that $i \neq j$, define*

$$M_{ij}^+ := \sum_{\substack{\sigma \in \{+1,-1\}^n \\ \sigma_i \sigma_j = 1}} s_\sigma \quad \textit{and} \quad M_{ij}^- := \sum_{\substack{\sigma \in \{+1,-1\}^n \\ \sigma_i \sigma_j = -1}} s_\sigma, \tag{5.19}$$

as well as

$$B_{ij} := \frac{M_{ij}^+ - M_{ij}^-}{N}$$

and

$$\mathbf{B} := (B_{ij} : 1 \leq i < j \leq n).$$

In words, M_{ij}^+ is the number of samples for which leaves i and j share the same nucleotide state and M_{ij}^- is the number for which the nucleotides observed at leaves i and j differ. It follows by definition that $M_{ij}^+ + M_{ij}^- = N$ and that $M_{ij}^+ = M_{ji}^+$ and $M_{ij}^- = M_{ji}^-$ for all distinct $i, j \in [3]$.

The statistic B_{ij} measures the observed correlation of the observations at leaves i and j of the tree. To see this, observe that by the law of large numbers, $B_{ij} \to \mathbb{P}[X_i = X_j] - \mathbb{P}[X_i \neq X_j]$ as $N \to \infty$. Moreover, since $E[X_i] = E[X_j] = 0$,

$$\begin{aligned}
\operatorname{Cov}(X_i, X_j) &= \mathbb{E}[X_i X_j] \\
&= \mathbb{P}[X_i = X_j] - \mathbb{P}[X_i \neq X_j] \\
&= \frac{1}{2}(1 + \theta_i \theta_j) - \frac{1}{2}(1 - \theta_i \theta_j) \qquad \text{using Eq. (5.11)} \\
&= \theta_i \theta_j
\end{aligned}$$

This calculation shows that B_{ij} is an estimate of the covariance of X_i and X_j, which is the quantity $\theta_i \theta_j$.

Remark 4 (Permutation Notation)**.** *For the case when $n = 3$, it will often be useful to index the statistics in Definition 3 using permutations. Let A_3 denote the alternating*

group of degree 3, which can be expressed in cycle notation as

$$A_3 = \{(1), (123), (132)\}.$$

For each $\pi \in A_3$, we write

$$M_\pi^+ := M^+_{\pi(1),\pi(2)}, \quad M_\pi^- := M^-_{\pi(1),\pi(2)}, \quad \textit{and} \quad B_\pi := B_{\pi(1),\pi(2)}.$$

Example 3 ($n = 3$)**.** *In the case of $n = 3$ considered here it is easy to check that*

$$\mathbf{B} = (B_{12}, B_{13}, B_{23})$$

where

$$\begin{aligned} B_{12} &= \frac{1}{N}\left(\bar{s}_\emptyset - \bar{s}_{\{1\}} - \bar{s}_{\{2\}} + \bar{s}_{\{1,2\}}\right) \\ B_{13} &= \frac{1}{N}\left(\bar{s}_\emptyset - \bar{s}_{\{1\}} + \bar{s}_{\{2\}} - \bar{s}_{\{1,2\}}\right) \\ B_{23} &= \frac{1}{N}\left(\bar{s}_\emptyset + \bar{s}_{\{1\}} - \bar{s}_{\{2\}} - \bar{s}_{\{1,2\}}\right). \end{aligned} \tag{5.20}$$

The use of the statistic $\mathbf{B}$ will permit us to obtain a simple criterion for when a maximum likelihood estimate corresponding to a 3-leaf tree with finite branch lengths exists. This criteria involves the following set:

Definition 4 (The set $\mathcal{D}$)**.** *Define*

$$\mathcal{D} := \left\{x \in (0,1)^3 : x_i x_j < x_k \textit{ for all distinct } i, j, k \in [3]\right\}$$

As we will show in the next theorem, it turns out that given some fixed data $\mathbf{s}$, a maximum likelihood estimate with finite branch lengths exists precisely if and only if $\mathbf{B} \in \mathcal{D}$. Before stating the theorem, it is worthwhile to examine the semialgebraic constraints (i.e. the polynomial inequalities) that define $\mathcal{D}$, to understand where they are coming from in the context of this problem. Observe that if $\mathbf{B} \in \mathcal{D}$, then (to consider just one of the inequalities) we have $B_{12}B_{13} < B_{23}$. When N is large, $B_{ij} \approx \theta_i\theta_j$ and hence

this inequality is approximately

$$\theta_1\theta_2\theta_1\theta_3 < \theta_2\theta_3,$$

which by Eq. (5.1) is

$$e^{-2(d_1+d_2+d_1+d_3)} < e^{-2(d_2+d_3)},$$

or equivalently

$$d_2 + d_3 < (d_2 + d_1) + (d_1 + d_3).$$

In words, the evolutionary distance from leaf 2 to leaf 3 is less than or equal to the distance from 2 to 1 plus the distance from 1 to 3. The other inequalities $B_{13}B_{23} < B_{12}$ and $B_{12}B_{23} < B_{13}$ are similar, and we conclude that, taken together, the semialgebraic constraints defining $\mathcal{D}$ are nothing but the triangle inequality in disguise.

Finally, we make the following two assumptions about the data $\mathbf{s}$:

A.1 $\bar{s}_\alpha > 0$ for all $\alpha \subseteq [2]$.

A.2 B_{12}, B_{13} and B_{23} are nonzero and distinct.

In words, assumption **A.1** states that each site pattern aaa, aab, aba, abb is observed at least once in the data. One useful consequence of this is that $M^+_{ij}, M^-_{ij} \in (0, N)$ whenever $i \neq j$. Assumption **A.2** is an assumption about the genericity of the data, in the sense that it is equivalent to assuming that

$$B_{12}B_{13}B_{23}(B_{12} - B_{13})(B_{13} - B_{23})(B_{12} - B_{23}) \neq 0.$$

These two assumptions considerably simplify the problem with very little loss of generality, as both assumptions are likely to be satisfied when N is large.

5.2.2 Main Result

The main result of this section is the following theorem.

Theorem 9 (Global MLE for the 3-leaf tree)**.** *Assume that* ***A.1*** *and* ***A.2*** *hold. Then* ℓ *has a maximizer on the set* $[0,1]^3$*. Denote the set of all such maximizers as*

$$E_{\text{MLE}} := \left\{ \widehat{\theta} \in [0,1]^3 : \ell(\widehat{\theta}) = \max_{\theta \in [0,1]^3} \ell(\theta) \right\},$$

and let $\pi_1, \pi_2, \pi_3 \in A_3$ *such that*

$$B_{\pi_1} < B_{\pi_2} < B_{\pi_3}.$$

If $\mathbf{B} \in \mathcal{D}$*, then*

$$E_{\text{MLE}} = \left\{ \left(\sqrt{\frac{B_{12}B_{13}}{B_{23}}}, \sqrt{\frac{B_{12}B_{23}}{B_{13}}}, \sqrt{\frac{B_{13}B_{23}}{B_{12}}} \right) \right\}. \tag{5.21}$$

On the other hand, if $\mathbf{B} \notin \mathcal{D}$*, then the following conclusions hold:*

(i) If $B_{\pi_2}, B_{\pi_3} > 0$ *then*

$$E_{\text{MLE}} = \left\{ \theta \in [0,1]^3 : \theta_{\pi_1(1)} = B_{\pi_1(1),\pi_1(3)},\ \theta_{\pi_1(2)} = B_{\pi_1(2),\pi_1(3)},\ \textit{and}\ \theta_{\pi_1(3)} = 1 \right\}.$$

(ii) If $B_{\pi_3} > 0$ *and* $B_{\pi_2} < 0$ *then*

$$E_{\text{MLE}} = \left\{ \theta \in [0,1]^3 : \theta_{\pi_3(1)}\theta_{\pi_3(2)} = B_{\pi_3},\ \ \textit{and}\ \ \theta_{\pi_3(3)} = 0 \right\}.$$

(iii) If $B_{\pi_3} < 0$ *then*

$$E_{\text{MLE}} = \bigcup_{\substack{A \subseteq [3]: \\ |A| \geq 2}} \left\{ \theta \in [0,1]^3 : \theta_i = 0 \textit{ for all } i \in A \right\}.$$

In addition to providing necessary and sufficient conditions for the MLE to exist as a tree with finite branch lengths, Theorem 9 also characterizes the ways that this can fail to occur. As a result, this theorem provides insight into problems considered in [173, 3]. Before sketching the proof of Theorem 9, we give two examples to illustrate its use and significance. In first example, we consider an application of Theorem 9 to a data point

shown in [173] not to have an MLE with biologically plausible branch lengths.

Example 4. Suppose we are given data of the form

$$\mathbf{s} = (s_{(-1,-1,-1)}, s_{(-1,-1,+1)}, s_{(-1,+1,-1)} \ldots, s_{(+1,+1,+1)}) = (17, 5, 27, 5, 16, 5, 19, 6).$$

In [173], it was shown using algebraic methods that for this choice of data no maximum likelihood estimate exists with branch lengths $\theta = (\theta_1, \theta_2, \theta_3) \in (0, 1)^3$, but instead that the likelihood is maximized when one the branch length of leaf 2 goes to infinity (i.e., when $\theta_2 = 0$). We verify this result by observing that

$$\begin{aligned} B_{12} &= \frac{(17 + 5 + 19 + 6) - (27 + 5 + 16 + 5)}{100} = -0.06 \\ B_{13} &= \frac{(17 + 27 + 5 + 6) - (5 + 5 + 16 + 19)}{100} = 0.10 \\ B_{23} &= \frac{(17 + 5 + 16 + 6) - (5 + 27 + 5 + 19)}{100} = -0.12. \end{aligned}$$

Since $B_{12}, B_{23} < 0$ and $B_{13} > 0$, therefore $\mathbf{B} \notin \mathcal{D}$ and we conclude this data corresponds to case (ii) of Theorem 9, which implies that the log-likelihood is maximized when $\theta_1\theta_3 = 0.10$ and $\theta_2 = 0$, which agrees with the result in [173]. Indeed, up to our simplifying assumptions **A.1** and **A.2**, Theorem 9 characterizes all the ways that the likelihood estimate might be maximized when one or more branch lengths tend to infinity.

In [3], an important connection was made between maximum likelihood and distance matrix estimates on a 3-leaf tree under the Jukes-Cantor model of site substitution; in particular it was demonstrated through simulation that the failure of distance-based branch-length estimates to satisfy the triangle inequality as well as nonnegativity constraints was a good predictor of maximum likelihood failing to return a tree with biologically plausible branch lengths. The distances considered in [3] for the Jukes-Cantor model are

$$D_{ij} := -\frac{3}{4} \log \left(1 - \frac{4}{3} M_{ij}^{-}\right).$$

This is standard Jukes-Cantor correction, which gives an estimate of distance, measured

in expected number of mutations per site. In the setting of the CFN model considered here, the analogous distance estimate is

$$\tilde{D}_{ij} := -\frac{1}{2}\log(1-2M_{ij}^{-}) = -\frac{1}{2}\log(B_{ij}).$$

Modulo these changes to the distance formulas, required to account for the difference in substitution model, the results in Theorem 9 coincide exactly with the predictions made in that chapter [3, Table 1], providing them with a theoretical underpinning. An illustration of this is given in the next example.

Example 5. Suppose

$$\mathbf{s} = (21, 12, 9, 8, 7, 11, 17, 15)$$

Then

$$\begin{aligned} B_{12} &= \frac{(21+12+17+15)-(9+8+7+11)}{100} = 0.3 \\ B_{13} &= \frac{(21+9+11+15)-(12+8+7+17)}{100} = 0.12 \\ B_{23} &= \frac{(21+8+7+15)-(12+9+11+17)}{100} = 0.02. \end{aligned}$$

According to the results in [3], the appropriate prediction for this data is that since the distance estimates do not satisfy the triangle inequality, and specifically because $\tilde{D}_{23} \geq \tilde{D}_{12} + \tilde{D}_{13}$, it should follow that leaf 1 in the MLE will have a branch length of zero expected mutations per site (or equivalently, $\theta_1 = 1$ in our notation).

Theorem 9 confirms this prediction. In this case $B_{ij} > 0$ for all $i, j \in [3]$. Since $B_{12}B_{13} = 0.036 > 0.02 = B_{23}$, it follows that $\mathbf{B} \notin \mathcal{D}$, and hence the data falls into case (i) of Theorem 9. In the notation of Theorem 9, we have

$$B_{\pi_1} := \min\{B_{12}, B_{13}, B_{23}\} = B_{23}$$

which implies that $\pi_1 = (123)$. Theorem 9 case (i) then implies that the likelihood is

maximized when $\theta = (\theta_1, \theta_2, \theta_3) = (1, B_{12}, B_{13}) = (1, 0.3, 0.12)$.

One takeaway from Theorem 9, illustrated in the previous example, is that it shows how the maximum likelihood estimate obeys certain semialgebraic constraints (e.g., those which define the set $\mathcal{D}$), returning a tree with biologically plausible branch lengths only if they are satisfied. This suggests that understanding the semialgebraic constraints satisfied by phylogenetic models may prove to be an essential tool to understanding the sometimes complex behavior of maximum likelihood estimation observed in phylogenetics.

5.2.3 Overview of the proof of Theorem 9

Our proof of Theorem 9 considers separately two cases:

1. the *interior case*, i.e., the problem of maximizing ℓ over all $\theta \in (0,1)^3$, and
2. the *boundary cases*, corresponding to when $\theta \in \partial(0,1)^3$.

We present only the interior case here (in Section 5.2.4), as this analysis is new and will be required when we turn our focus to trees with four leaves. A complete proof of Theorem 9, including consideration of all boundary cases, can be found in Appendix A.1. In particular, for the boundary cases (found in the appendix), we follow the approach taken by the authors of [3], who analyzed the maximum likelihood problem in the context of the Jukes-Cantor model and obtained analytic solutions for all boundary cases for that model by decomposing the boundary $\partial(0,1)^3$ into 26 components, consisting of 8 vertices, 12 edges, and 6 faces, and maximizing ℓ on each of those individually. Our approach is similar, though we group certain edges and faces together in those cases in which the analysis is similar.

5.2.4 Maximizing the log-likelihood function on $(0,1)^3$

In this subsection we consider the problem of maximizing Eq. (5.10) on the set $(0,1)^3$. This set corresponds to those trees whose branches are of finite and nonzero length, when measured in expected number of mutations per site. Since $(0,1)^3$ is open, the existence of

a local maximum is not guaranteed. The main result of this subsection gives, for generic positive data, necessary and sufficient conditions for ℓ to have a local maximum in $(0,1)^3$, and a formula if it exists; it also shows ℓ has at most one local maximum on $(0,1)^3$.

We begin with an important definition and a technical lemma.

Let $\widehat{\phi} : \mathbb{R}^3 \backslash \{x : x_1x_2x_3 = 0\} \to \mathbb{C}^3$ be defined by

$$\widehat{\phi}(x_1, x_2, x_3) := \left(\sqrt{\frac{x_1x_2}{x_3}}, \sqrt{\frac{x_1x_3}{x_2}}, \sqrt{\frac{x_2x_3}{x_1}} \right). \tag{5.22}$$

Further, let ϕ be the restriction of $\widehat{\phi}$ to $\mathcal{D}$:

$$\phi := \widehat{\phi}|_{\mathcal{D}}.$$

The next lemma summarizes some useful properties of ϕ

Lemma 41. *The function $\phi : \mathcal{D} \to (0,1)^3$ is a continous bijection, with inverse function $\phi^{-1} : (0,1)^3 \to \mathcal{D}$ given by*

$$\phi^{-1}(y) = (y_1y_2, y_1y_3, y_2y_3). \tag{5.23}$$

Proof of Lemma 41. First, note that if $x = (x_1, x_2, x_3) \in \mathcal{D}$ then $\phi(x) \in (0,1)^3$ by definition of $\widehat{\phi}$ and $\mathcal{D}$. Moreover, since $x_1, x_2, x_3 \neq 0$, whenever $x = (x_1, x_2, x_3) \in \mathcal{D}$, it follows that ϕ is continuous on $\mathcal{D}$.

Next, to show that ϕ is injective, let $x, \tilde{x} \in \mathcal{D}$ such that $\phi(x) = \phi(\tilde{x})$. Let ϕ_1 and ϕ_2 denote the first and second components of ϕ respectively. Then $\phi_1(x)\phi_2(x) = \phi_1(\tilde{x})\phi_2(\tilde{x})$ or equivalently, $x_1^2 = \tilde{x}_1^2$. Since $x, \tilde{x} \in (0,1)^3$, this implies $x_1 = \tilde{x}_2$. Similar arguments show that $x_2 = \tilde{x}_2$ and $x_3 = \tilde{x}_3$, and hence $x = \tilde{x}$. Therefore ϕ is injective.

Next we show that ϕ is surjective. Let $y \in (0,1)^3$ be arbitrary. Then it is easy to see that the point $y' = (y_1', y_2', y_3') := (y_1y_2, y_1y_3, y_2y_3)$ satisfies

$$y_i'y_j' < y_k'$$

for all choices of distinct $i, j, k \in [3]$, and hence $y' \in \mathcal{D}$. Moreover, $\phi(y') = y$ by definition

of $\widehat{\phi}$. Therefore ϕ is surjective and has inverse given by the formula $\phi^{-1}(y) = y'$, which is precisely the formula in Eq. (5.23). □

We are now ready to state the main lemma of this subsection, which solves the problem of maximizing ℓ on the set $(0,1)^3$, or in other words, solves the maximum likelihood problem for biologically plausible parameters.

Lemma 42 (MLE for 3 Leaf Tree – Interior Case)**.** *Assume that* ***A.1*** *and* ***A.2*** *hold. Let*

$$x^* := (B_{12}, B_{13}, B_{23})$$

and let

$$\theta^* := \widehat{\phi}(x^*) = \left(\sqrt{\frac{B_{12}B_{13}}{B_{23}}}, \sqrt{\frac{B_{12}B_{23}}{B_{13}}}, \sqrt{\frac{B_{13}B_{23}}{B_{12}}}\right)$$

If $x^* \in \mathcal{D}$ then θ^* is the unique local maximum of ℓ in $(0,1)^3$ and has log-likelihood

$$\ell(\theta^*) = \sum_{\alpha \subseteq [2]} \bar{s}_\alpha \log\left(\frac{\bar{s}_\alpha}{N}\right) - N \log 2 \tag{5.24}$$

Conversely, if $x^* \notin \mathcal{D}$ then ℓ has no local maximum of on $(0,1)^3$.

Proof of Lemma 42. For ease of notation, we will write

$$\bar{s}_1 := \bar{s}_\emptyset, \quad \bar{s}_2 := \bar{s}_{\{1\}}, \quad \bar{s}_3 := \bar{s}_{\{2\}}, \text{ and } \bar{s}_4 := \bar{s}_{\{1,2\}}. \tag{5.25}$$

It follows by Eq. (5.10) and Eq. (5.17) that

$$\begin{aligned}\ell(\theta \mid \mathbf{s}) = {} & \bar{s}_1 \log(1 + \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3) + \bar{s}_2 \log(1 - \theta_1\theta_2 - \theta_1\theta_3 + \theta_2\theta_3) \\ & + \bar{s}_3 \log(1 - \theta_1\theta_2 + \theta_1\theta_3 - \theta_2\theta_3) + \bar{s}_4 \log(1 + \theta_1\theta_2 - \theta_1\theta_3 - \theta_2\theta_3) \\ & - N \log 8.\end{aligned} \tag{5.26}$$

An initial attempt to compute the critical points of $\ell(\theta)$ directly by taking the gradient of ℓ yields a polynomial system which is difficult to solve analytically, so instead we modify

this approach by first considering a different function whose extrema are closely related to those of ℓ.

To define this function, first let $\mathcal{D}_F \subseteq \mathbb{R}^3$ be the intersection of half-spaces defined by the following inequalities

$$\begin{aligned} 1 + x + y + z &> 0 \\ 1 - x - y + z &> 0 \\ 1 - x + y - z &> 0 \\ 1 + x - y - z &> 0 \end{aligned} \tag{5.27}$$

and let $F : \mathcal{D}_F \to \mathbb{R}$ be defined by

$$\begin{aligned} F(x, y, z) := \bar{s}_1 \log(1 + x + y + z) + \bar{s}_2 \log(1 - x - y + z) \\ + \bar{s}_3 \log(1 - x + y - z) + \bar{s}_4 \log(1 + x - y - z) - N \log 8 \end{aligned} \tag{5.28}$$

The significance of F is owes to the observation that $\ell = F \circ \phi^{-1}$, which is proved in the next claim.

Claim 1: For all $\theta \in (0, 1)^3$,

$$\ell(\theta) = F \circ \phi^{-1}(\theta). \tag{5.29}$$

Proof of Claim 1. Since domain of ℓ is $(0, 1)^3$, it follows from Lemma 41 and Eqs. (5.26) and (5.28) that Eq. (5.29) holds provided that F is defined on the image of ϕ^{-1}. Therefore, since $\text{im}(\phi^{-1}) = \mathcal{D}$ and $\text{dom}(F) = \mathcal{D}_F$ it will suffice to show that $\mathcal{D} \subseteq \mathcal{D}_F$.

Let $u \in \mathcal{D}$. Then by Lemma 41 there exists $w_1, w_2, w_3 \in (0, 1)$ such that

$$u = (w_1 w_2, w_1 w_3, w_2 w_3).$$

To show that $u \in \mathcal{D}_F$, it suffices by Eq. (5.27) to show that

$$\begin{aligned} 1 + w_1w_2 + w_1w_3 + w_2w_3 &> 0 \\ 1 - w_1w_2 - w_1w_3 + w_2w_3 &> 0 \\ 1 - w_1w_2 + w_1w_3 - w_2w_3 &> 0 \\ 1 + w_1w_2 - w_1w_3 - w_2w_3 &> 0. \end{aligned}$$

The first of these four equations holds trivially. As for the other three, we will only prove $1 + w_1w_2 - w_1w_3 - w_2w_3 > 0$, as the other two inequalities can be proved in the same manner. Let $h(w_1, w_2) := 1 + w_1w_1 - w_1 - w_2$. Since $w_3 < 1$, it is sufficient to show that $h(w_1, w_2) \geq 0$ on for all $w_1, w_2 \in [0, 1]$. Indeed, using calculus it easy to see that h is minimized when at least one of the arguments w_1, w_2 equals zero, and that the minimum is zero. Therefore $1 + w_1w_2 - w_1w_3 - w_2w_3 > 0$. We conclude that $\mathcal{D} \subseteq \mathcal{D}_F$, as required to prove the claim. □

The next two claims serve to characterize the extrema of F.

Claim 2. The point $x^* = (B_{12}, B_{13}, B_{23})$ is the unique critical point of F.

Proof of Claim 2. It is first necessary to verify that x^* is in the domain of F. To do this, it will suffice to show that x^* satisties the inequalities in Eq. (5.27).

Using Eq. (5.20) and the observation $\bar{s}_1 + \bar{s}_2 + \bar{s}_3 + \bar{s}_4 = N$, it is easy to check that

$$\begin{aligned} 1 + B_{12} + B_{13} + B_{23} &= \frac{4}{N}\bar{s}_1 \\ 1 - B_{12} - B_{13} + B_{23} &= \frac{4}{N}\bar{s}_2 \\ 1 - B_{12} + B_{13} - B_{23} &= \frac{4}{N}\bar{s}_3 \\ 1 + B_{12} - B_{13} - B_{23} &= \frac{4}{N}\bar{s}_4 \end{aligned} \tag{5.30}$$

Therefore by **A.1**, it follows that $x^* = (B_{12}, B_{13}, B_{23})$ satisfies the inequalities in Eq. (5.27), as required. Therefore x^* is in the domain of F.

We now proceed with a standard critical point calculation. Letting $v_1 = (1, 1, 1)^\top$,

$v_2 = (-1,-1,1)^\top$, $v_3 = (-1,1,-1)^\top$, $v_4 = (1,-1,-1)^\top$, and taking partial derivatives of F in Eq. (5.28) with respect to the variables x, y and z, it follows that for all $u = (x,y,z)^\top \in \mathcal{D}_F$,

$$\nabla F(u) = \begin{bmatrix} A_1 + A_2 - A_3 - A_4 \\ A_1 - A_2 + A_3 - A_4 \\ A_1 - A_2 - A_3 + A_4 \end{bmatrix}, \tag{5.31}$$

where

$$A_i := \frac{s_i}{1 + v_i^\top u}, \quad \text{for each } i = 1,2,3,4. \tag{5.32}$$

Setting $\nabla F = 0$ and using Eq. (5.31), we deduce that the following system of equations holds:

$$\begin{aligned} A_1 - A_2 &= 0 \\ A_2 - A_3 &= 0 \\ A_3 - A_4 &= 0. \end{aligned}$$

Substituting the formulas for the A_i's from Eq. (5.32) and rearranging terms, we obtain

$$\begin{aligned} \left(v_2^\top \bar{s}_1 - v_1^\top \bar{s}_2\right) u &= \bar{s}_2 - \bar{s}_1 \\ \left(v_3^\top \bar{s}_2 - v_2^\top \bar{s}_3\right) u &= \bar{s}_3 - \bar{s}_2 \\ \left(v_4^\top \bar{s}_3 - v_3^\top \bar{s}_4\right) u &= \bar{s}_4 - \bar{s}_3. \end{aligned}$$

Writing this out, this is the matrix equation

$$\begin{bmatrix} \bar{s}_1 - \bar{s}_4 & -(\bar{s}_1 + \bar{s}_4) & -(\bar{s}_1 + \bar{s}_4) \\ -(\bar{s}_4 + \bar{s}_3) & \bar{s}_4 + \bar{s}_3 & \bar{s}_3 - \bar{s}_4 \\ \bar{s}_2 - \bar{s}_3 & -(\bar{s}_3 + \bar{s}_2) & \bar{s}_3 + \bar{s}_2 \end{bmatrix} u = \begin{bmatrix} \bar{s}_2 - \bar{s}_1 \\ \bar{s}_3 - \bar{s}_2 \\ \bar{s}_4 - \bar{s}_3 \end{bmatrix}.$$

Using the fact that $\bar{s}_1 + \bar{s}_2 + \bar{s}_3 + \bar{s}_4 = N$, one can check that the 3×3 matrix in the above

equation has inverse

$$\frac{1}{2N}\begin{bmatrix} -\frac{\bar{s}_3+\bar{s}_2}{\bar{s}_4} & -\frac{(\bar{s}_1+\bar{s}_4)(\bar{s}_3+\bar{s}_2)}{\bar{s}_4\bar{s}_3} & -\frac{\bar{s}_1+\bar{s}_4}{\bar{s}_3} \\ -\frac{\bar{s}_4+\bar{s}_2}{\bar{s}_4} & \frac{\bar{s}_4\bar{s}_3-\bar{s}_1\bar{s}_2}{\bar{s}_4\bar{s}_3} & -\frac{\bar{s}_1+\bar{s}_3}{\bar{s}_3} \\ -\frac{\bar{s}_4+\bar{s}_3}{\bar{s}_4} & \frac{\bar{s}_4\bar{s}_2-\bar{s}_1\bar{s}_3}{\bar{s}_4\bar{s}_3} & \frac{\bar{s}_4+\bar{s}_3}{\bar{s}_3} \end{bmatrix}.$$

Therefore

$$\begin{aligned} u &= \frac{1}{2N}\begin{bmatrix} -\frac{\bar{s}_3+\bar{s}_2}{\bar{s}_4} & -\frac{(\bar{s}_1+\bar{s}_4)(\bar{s}_3+\bar{s}_2)}{\bar{s}_4\bar{s}_3} & -\frac{\bar{s}_1+\bar{s}_4}{\bar{s}_3} \\ -\frac{\bar{s}_4+\bar{s}_2}{\bar{s}_4} & \frac{\bar{s}_4\bar{s}_3-\bar{s}_1\bar{s}_2}{\bar{s}_4\bar{s}_3} & -\frac{\bar{s}_1+\bar{s}_3}{\bar{s}_3} \\ -\frac{\bar{s}_4+\bar{s}_3}{\bar{s}_4} & \frac{\bar{s}_4\bar{s}_2-\bar{s}_1\bar{s}_3}{\bar{s}_4\bar{s}_3} & \frac{\bar{s}_4+\bar{s}_3}{\bar{s}_3} \end{bmatrix}\begin{bmatrix} \bar{s}_4-\bar{s}_1 \\ \bar{s}_3-\bar{s}_4 \\ \bar{s}_2-\bar{s}_3 \end{bmatrix} \\ &= \frac{1}{N}\begin{bmatrix} \bar{s}_1-\bar{s}_2-\bar{s}_3+\bar{s}_4 \\ \bar{s}_1-\bar{s}_2+\bar{s}_3-\bar{s}_4 \\ \bar{s}_1+\bar{s}_2-\bar{s}_3-\bar{s}_4 \end{bmatrix}. \end{aligned}$$

By Eq. (5.20), the right-hand side is precisely the vector $(B_{12}, B_{13}, B_{23})^\top$, and therefore we conclude that $x^* = (B_{12}, B_{13}, B_{23})$ is the unique critical point of F on its domain $\mathcal{P}$. This completes the proof of the claim. □

Let H_F denote the Hessian matrix of F; that is,

$$H_F := \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} & \frac{\partial^2 F}{\partial x \partial z} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} & \frac{\partial^2 F}{\partial y \partial z} \\ \frac{\partial^2 F}{\partial z \partial x} & \frac{\partial^2 F}{\partial z \partial y} & \frac{\partial^2 F}{\partial z^2} \end{bmatrix}.$$

It is clear that F is twice-differentiable on its domain, so $H_F(x)$ is defined for all $x \in \mathcal{D}_F$.

Claim 3. $H_F(x^*)$ is negative definite.

Proof of Claim 3. Since $H_F(x^*)$ is a real symmetric matrix, it will suffice to show that its eigenvalues are all negative, as this will imply that $H_F(x^*)$ is negative definite.

We first compute the characteristic polynomial of H_F using the following Julia (v1.8.5) code:

```
using HomotopyContinuation, LinearAlgebra
@var x y z λ c[1:4] t[1:4]
v = [1+x+y+z,1-x-y+z,1-x+y-z,1+x-y-z] # dummy variable
logL = c'log.(v) # the log-likelihood (up to a constant)
H = differentiate(differentiate(logL,[x,y,z]),[x,y,z])'
P_char = det(λ*I-H)
P_simplified = expand(subs(P_char,v=>t))
```

The result of this code gives

$$\begin{aligned} P_{\text{char}}(\lambda) &:= \det(\lambda I - A) \\ &= \lambda^3 + 3\lambda^2 \left(\sum_{i=1}^{4} \frac{\bar{s}_i}{t_i^2} \right) + 8\lambda \left(\sum_{1 \le i < j \le 4} \frac{\bar{s}_i \bar{s}_j}{t_i^2 t_j^2} \right) + 16 \sum_{\substack{\alpha \subseteq [4]: \\ |\alpha| = 3}} \left(\prod_{i \in \alpha} \frac{\bar{s}_i}{t_i^2} \right), \end{aligned}$$

where $t_1 = 1 + x + y + z$, $t_2 = 1 - x - y + z$, $t_3 = 1 - x + y - z$, $t_4 = 1 + x - y - z$, and $\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4$ are defined in Eq. (5.25).

We need to show that P_{char} has only negative roots, as this will imply that all three eigenvalues of H_F are negative, and hence that $H_F(x^*)$ is negative definite. Since H_F is a real symmetric matrix, the roots of P_{char} are all real numbers, and therefore it will be enough to show that P_{char} has no nonnegative roots. Indeed, since all the coefficients of P_{char} are positive, Decartes' rule of signs implies that P_{char} has no positive roots. Moreover since $(\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4) \neq (0, 0, 0, 0)$ by **A.1**, the constant term in P_{char} is nonzero, and hence $P_{\text{char}}(0) \neq 0$ as well. We conclude that P_{char} has no nonnegative roots, as required to prove the claim. □

Using the results of the previous three claims, the next two claims will together characterize the local maxima of ℓ on $(0, 1)^3$.

Claim 4: If $x^* \in \mathcal{D}$ then $\theta^* \in (0, 1)^3$ and θ^* is a local maximum of ℓ.

Proof of Claim 4. Observe that F and ϕ^{-1} are both differentiable on their respective domains. Therefore if $\theta \in (0, 1)^3$ and $x = \phi^{-1}(\theta)$, then using the chain rule to differentiate

Eq. (5.29) gives

$$\nabla \ell(\theta) = \nabla F(x) \cdot J_{\phi^{-1}}(\theta). \tag{5.33}$$

Suppose $x^* \in \mathcal{D}$. Then by Lemma 41, $\theta^* = \phi(x^*) \in (0,1)^3$ and $x^* = \phi^{-1}(\theta^*)$. Therefore Eq. (5.33) implies

$$\nabla \ell(\theta^*) = \nabla F(x^*) \cdot J_{\phi^{-1}}(\theta^*).$$

By Claim 2, x^* is a critical point of F, i.e., $\nabla F(x^*) = 0$. Therefore

$$\nabla \ell(\theta^*) = 0.$$

This shows that θ^* is a critical point of ℓ.

Next we show that θ^* is a local maximum of ℓ. Since $\ell = F \circ \phi^{-1}$ by Eq. (5.29), and since F and ϕ^{-1} are both twice differentiable on their respective domains, therefore it follows by the chain rule for Hessian matrices (see, e.g., [185, p.125-126]), that ℓ is also twice differentiable and has Hessian matrix

$$H_\ell(\theta) = \left(J_{\phi^{-1}}(\theta)\right)^\top \left(H_F(x)\right) J_{\phi^{-1}}(\theta) + \sum_{i=1}^{3} \left(\frac{\partial F}{\partial x_i}(x)\right) \left(H_{(\phi^{-1})_i}(\theta)\right)$$

where

$$(\phi^{-1})_1(\theta) := \theta_1\theta_2, \quad (\phi^{-1})_2(\theta) := \theta_1\theta_3, \quad \text{and} \quad (\phi^{-1})_3(\theta) := \theta_2\theta_3$$

for all $\theta \in (0,1)$. Since x^* is a critical point of F due to Claim 2, we have $\frac{\partial F}{\partial x_i}(x^*) = 0$ for each $i = 1, 2, 3$. Therefore

$$H_\ell(\theta^*) = \left(J_{\phi^{-1}}(\theta^*)\right)^\top \left(H_F(x^*)\right) J_{\phi^{-1}}(\theta^*).$$

Since $H_F(x^*)$ is a negative definite matrix by Claim 3, and since $\det J_{\phi^{-1}}(\theta^*) \neq 0$ by Eq. (5.34), we conclude that $H_\ell(\theta^*)$ and $H_F(x^*)$ are similar matrices, and hence $H_\ell(\theta^*)$ is negative definite as well. Therefore by the second derivative test (see, e.g., [185, Theorem 4, p.140]), the point θ^* is a local maximum of ℓ. This completes the proof of the claim. $\square$

Claim 5: If $\theta \in (0,1)^3$ is a local maximum of ℓ then $x^* \in \mathcal{D}$ and $\theta = \theta^*$.

Proof of Claim 5. Let $\theta \in (0,1)^3$ be a local maximum of ℓ and let $x = \phi^{-1}(\theta)$. Since θ a critical point, Eq. (5.33) implies

$$0 = \nabla F(x) \cdot J_{\phi^{-1}}(\theta).$$

Since

$$\det J_{\phi^{-1}}(\theta) = \begin{vmatrix} \theta_2 & \theta_3 & 0 \\ \theta_1 & 0 & \theta_3 \\ 0 & \theta_1 & \theta_2 \end{vmatrix} = -2\theta_1\theta_2\theta_3 \neq 0, \tag{5.34}$$

it follows that $J_{\phi^{-1}}(\theta)$ is nonsingular, and hence

$$\nabla F(x) = 0.$$

Therefore x is a critical point of F. Since the only critical point of F is x^* by Claim 3, it follows that $x = x^*$. Since $x \in D$ by Lemma 41, this implies that $x^* \in \mathcal{D}$. Therefore

$$\begin{aligned} \theta &= \phi(x) && \text{by definition of } x \\ &= \phi(x^*) && \text{since } x^* = x \in \mathcal{D} \\ &= \theta^* && \text{by definition of } \theta^*. \end{aligned}$$

This completes the proof of the claim. □

We can now use Claims 4 and 5 to prove the theorem. If $x^* \in \mathcal{D}$ then Claims 4 and 5 imply that θ^* is the unique local maxium in $(0,1)^3$. On other other hand, if $x^* \notin \mathcal{D}$, then the contraposition of Claim 5 implies ℓ has no local maximum in $(0,1)^3$. This proves the first part of the theorem; it remains only to prove Eq. (5.24).

Indeed, if $x^* \in \mathcal{D}$ then since $\theta^* = \phi(B_{12}, B_{13}, B_{23})$, it follows by definition of ϕ that

$$\theta_i^* \theta_j^* = B_{ij} \tag{5.35}$$

for all $i, j \in [3]$ such that $i < j$. Plugging Eq. (5.35) into Eq. (5.30),

$$\begin{aligned}
1 + \theta_1^*\theta_2^* + \theta_1^*\theta_3^* + \theta_2^*\theta_3^* &= \frac{4}{N}\bar{s}_1 \\
1 - \theta_1^*\theta_2^* - \theta_1^*\theta_3^* + \theta_2^*\theta_3^* &= \frac{4}{N}\bar{s}_2 \\
1 - \theta_1^*\theta_2^* + \theta_1^*\theta_3^* - \theta_2^*\theta_3^* &= \frac{4}{N}\bar{s}_3 \\
1 + \theta_1^*\theta_2^* - \theta_1^*\theta_3^* - \theta_2^*\theta_3^* &= \frac{4}{N}\bar{s}_4
\end{aligned}$$

Therefore by Eq. (5.26),

$$\ell(\theta^*) = \sum_{i=1}^{4} \bar{s}_i \log\left(\frac{\bar{s}_i}{N}\right) - N \log 2,$$

which is precisely Eq. (5.24). This completes the proof of the lemma. □

5.3 An Exact Algorithm for the 4-Leaf MLE Problem

We now turn to the case of 4-leaf phylogenetic trees. There has been substantial interest in direct computation of ML points for site substitution models on 4-leaf trees [162, 156, 161, 3]. In [156], the authors use Hadamard conjugation and techniques similar to those utilized in the proof of Lemma 42 to compute an analytic formula for the ML point of a four-leaf tree with balanced topology under the molecular clock assumption. In that paper the authors also demonstrated that, even under the simplifying assumption of the molecular clock, the solutions to the critical equations admit no general closed form analytic formula (i.e. expressed using radicals of the data).

In this section, we do the next best thing, which is to present an algorithm to compute the MLE for a 4-leaf phylogenetic tree under the CFN model with arbitrary substitution probabilities, and a proof of the correctness of this algorithm.

5.3.1 Statement of the 4-Leaf Maximum Likelihood Problem

We make two assumptions about the data, which are analogous to the assumptions **A.1** and **A.2** made for the 3-leaf case. These assumptions are:

A.3 $\bar{s}_\alpha > 0$ for all $\alpha \subseteq [3]$.

A.4 $B_{12}, B_{13}, B_{14}, B_{23}, B_{24}$ and B_{34} are nonzero and distinct.

The problem we seek to solve is the following.

Problem 1 (Global 4-leaf maximum likelihood problem)**.** *Given data*

$$\mathbf{s} = (s_\sigma)_{\sigma \in \{-1,+1\}^4} \in \mathbb{R}^{16}_{\geq 0}$$

with $\sum_{\sigma \in \{-1,+1\}^4} s_\sigma = N \in \mathbb{Z}_{>0}$ *satisfying assumptions* ***A.3*** *and* ***A.4****, maximize the function*

$$\ell(\tau, \theta) := \sum_{\alpha \subseteq [3]} \bar{s}_\alpha \log \bar{p}_\alpha(\tau, \theta) - N \log 2 \tag{5.36}$$

over all $\tau \in \{12|34, 13|23, 23|14\}$ *and all* $\theta \in [0,1]^5$.

There are important complications which arise when moving from the 3-leaf to the 4-leaf MLE problem. An unrooted phylogenetic [3]-tree has only three edge parameters $\theta_1, \theta_2, \theta_3 \in [0,1]$ and only one possible unrooted topology (shown in Fig. 5.2). Therefore the space of phylogenetic [3]-trees may be identified with the unit cube $[0,1]^3$. In the proof of Theorem 9, the consideration of **boundary cases** (i.e. those trees in which $\theta_i \in \{0,1\}$ for one or more $i \in [3]$) corresponded neatly to the boundary of $[0,1]^3$. With the goal of maximizing $\ell(\theta)$ over $[0,1]^3$, the approach that was used in the proof of Theorem 9 was to maximize ℓ on the interior of the cube, and then on the interiors of its faces and edges, and finally evaluating ℓ on the vertices of the cube. This was essentially what was done in the proof of Theorem 9, albeit with certain boundary cases grouped together when the analysis was similar.

In the 4-leaf case, where the goal is to maximize ℓ on all phylogenetic [4]-trees, we seek to apply a similar approach, however the geometry of the boundary cases is significantly

more cumbersome than the boundary of the unit cube. Specifically, a phylogenetic [4]-tree has 5 edge parameters $\theta_1, \ldots, \theta_5 \in [0, 1]$ (with $\theta_1, \ldots, \theta_4$ corresponding to leaves $1, \ldots, 4$, and θ_5 to the internal edge) and three possible quartet topologies 12|34, 13|23, and 23|14. Therefore the space of phylogenetic [4]-trees can be regarded as three disjoint copies of the 5-dimensional unit cube $[0, 1]^5$.

The boundary cases for the four-leaf case thus correspond to union of the boundaries of three disjoint 5-dimensional hypercubes. As a result, the boundary cases are far more numerous (e.g., for each fixed quartet topology, there are $3^5 - 1 = 242$ ways that one or more elements of $\{\theta_1, \ldots, \theta_5\}$ equals zero or one). Because of this, we will introduce a general framework which allows us to group together those boundary cases which require similar analyses (i.e. the notion of "reduced topologies" described below).

Before introducing that framework, we highlight another issue, which is that in many of the boundary cases which must be considered, the goal of recovering the full vector $\theta = (\theta_1, \ldots, \theta_5)$ of Hadamard parameters and even the quartet topology of T is not possible, as such parameters are in general not identifiable. This issue arose in the consideration of boundary cases in the proof of Theorem 9 (see Remark 7), however was limited to nonidentifiability of the edge parameters. In the 4-leaf case, not only the numerical parameters, but also the quartet topology may not be identifiable, as illustrated in the following example.

Example 6 (Non-identifiability in 4-leaf case)**.** Consider the boundary case consisting of phylogenetic [4]-trees with topology 12|34 and Hadamard edge parameters $\theta = (\theta_1, \ldots, \theta_5)$ such that $\theta_4 = 0$ and $\theta_1, \theta_2, \theta_3\theta_5 \in (0, 1)$. This "tree" is shown in Fig. 5.3 (on the left).

The issues that arise in this example are that, given $\theta_4 = 0$ and $\tau = 12|34$, the parameters θ_3 and θ_5 are not independently identifiable, and in addition, the fact that $\theta_4 = 0$ implies that the topology 12|34 cannot be recovered from the data either, as we now explain. By Lemma 40, $(X_1, X_2, X_3) \perp\!\!\!\perp X_4$, and consequently (e.g., using Proposition 3) one can show that (a) the distribution of X does not depend on the the numerical parameters θ_3 and θ_5 individually but only on their product $\theta_3\theta_5$, so θ_3 and θ_5 cannot be

recovered from the distribution of X, and (b) the topology 12|34 cannot be recovered from the distribution of X; this is because the same distribution for X can be attained by any phylogenetic [4]-tree T with edge parameters $\theta = (\theta_1, \ldots, \theta_5) \in [0,1]^5$ satisfying any one of three following conditions:

(i.) T has topology 12|34 and $\tilde{\theta}_1 = \theta_1$, $\tilde{\theta}_2 = \theta_2$, $\tilde{\theta}_3 = \theta_3\theta_5$, and $\theta_4 = 0$

(ii.) T has topology 13|24 and $\tilde{\theta}_1 = \theta_1$, $\tilde{\theta}_2 = \theta_2\theta_5$, $\tilde{\theta}_3 = \theta_3$, and $\theta_4 = 0$

(iii.) T has topology 23|14, and $\tilde{\theta}_1 = \theta_1\theta_5$, $\tilde{\theta}_2 = \theta_2$, $\tilde{\theta}_3 = \theta_3$, and $\theta_4 = 0$.

To resolve this situation, we reduce the problem by suppressing the degree 2 internal node, and then combining the parameters θ_3 and θ_5 into a single parameter $\tilde{\theta}_3$, which results in the the second graphical model shown in Fig. 5.3 (right). We then content ourselves with obtaining the parameters $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3) \in (0,1)^3$ which maximize the likelihood, with the understanding that this corresponds to a continuum of phylogenetic [4]-trees (specifically all phylogenetic [4]-trees satisfying conditions (i.) (ii.) or (iii.)). The distribution of X is the same under all of these trees; it is the distribution obtained by running the CFN substitution model independently on the disconnected components of the graph shown in the right of Fig. 5.3. The procedure we used to obtain the graph on the right was to "reduce" the tree by suppressing the internal node and combining the parameters θ_3 and θ_5 into one $\tilde{\theta}_3$. The reduction step shown here is an example of a more general reduction framework, which we which we introduce next.

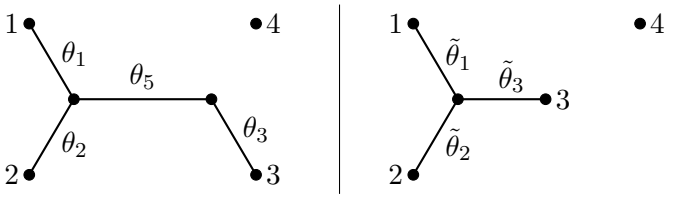


Figure 5.3: Example of a tree with non-identifiable parameters when $\theta_4 = 0$ (left). A reduced version, in which the internal node of degree two is suppressed, is shown on right.

5.3.2 Reduction Framework

Here we introduce a notion of topological reduction to filter out parameters which cannot be recovered in cases like that shown in Fig. 5.3, and which importantly will allow us to simplify the analysis of 4-leaf boundary cases to manageable number of distinct cases. Let T be an unrooted phylogenetic $[n]$-tree. Note that this implies that all leaf nodes of T are labeled and all internal nodes of T are unlabeled. Let $G(T)$ be the graphical model, with associated edge parameters, obtained by Algorithm 1.

One important purpose of Algorithm 1 is that it re-parameterizes the branch lengths in terms **identifiable combinations**, or identifiable functions of the numerical parameters [186]. Example 6 provides an example of this, as in that example the numerical parameters θ_3 and θ_5 cannot be recovered from the distribution of X, but their product $\tilde{\theta}_3 = \theta_3\theta_5$ can be.

Algorithm 1: Reduction Algorithm

Input: A phylogenetic $[n]$-tree with unrooted topology τ and Hadamard branch lengths $\theta \in [0,1]^{2n-3}$.

Output: A graph $G(T)$ consisting of n_c disjoint semi-labelled trees $T_1, \ldots, T_{n_c}$ each having Hadamard branch lengths in $(0,1)$.

1 Delete all $e \in E(T)$ such that $\theta_e = 0$.
2 Delete all components not containing a labelled vertex.
3 If $e = \{u, v\} \in E(T)$ such that $\theta_e = 1$, then contract e and assign to the new, single vertex the labels of both u and v.
4 Recursively delete each unlabeled degree 1 vertex and its incident edge.
5 For each unlabelled vertex v of degree 2 with incident edges $\{v, u\}$ and $\{v, \tilde{u}\}$, suppress v and assign to the new edge $\{u, \tilde{u}\}$ the numerical parameter $\theta_{\{v,u\}}\theta_{\{v,\tilde{u}\}}$

The next lemma shows that the output of Algorithm 1 is a well-defined forest of edge-weighted semi-labelled trees, and that the distribution of X is invariant under reduction.

Lemma 43 (Reduction properties)**.** *For every unrooted phylogenetic $[n]$-tree T, the output $G(T)$ of Algorithm 1 is well-defined and consists of a collection of n_c disconnected components $T_1, \ldots, T_{n_c}$ satisfying the property that for all $i \in [n_c]$ there exists $A_i \subseteq [n]$ such that*

(i.) the collection $\{A_1, \ldots, A_{n_c}\}$ is a partition of $[n]$, and

(ii.) for all $i \in [n_c]$, T_i is a semi-labelled tree on A_i.

In addition, under the CFN substitution model, the distribution at the labeled nodes of $G(T)$ is the same as the distribution at the leaves of T.

Proof. The sets $A_1, \ldots, A_{n_c}$ are determined by Step 1 of the algorithm, in which the deletion of edges breaks T into c uniquely determined disjoint components. Moreover, in the remaining steps of the algorithm labels are never deleted but only (possibly) reassigned within each component. This proves (i.).

The components obtained in Step 1 corresponding to $A_1, \ldots, A_{n_c}$ are obtained by deletion of edges, and therefore are trees. In the remaining steps these trees are subject to contraction of edges, deletion of pendant edges, and suppression of degree 2 vertices; these operations preserve treelikeness. Moreover, Steps 2, 4 and 5 result in the deletion or suppression of all vertices of degree less than 2. This proves (ii.).

It remains to prove that the distribution is unchanged by Algorithm 1. Let $(\theta_e)_{e \in E(T)}$ denote the Hadamard edge parameters of T and denote the edge parameters of $G(T)$ by $(\tilde{\theta}_e)_{e \in E(G(T))}$. As usual, let $X = (X_1, \ldots, X_n)$ be the nucleotide states observed at the labeled vertices $1, \ldots, n$, such that X_i is the state at vertex with label i, and let $\mathbb{P}$ be the distribution of X under the CFN process on T. Denote by $\mathbb{P}_{G(T)}$ the distribution of X under the CFN process run independently on the components of $G(T)$. We need to show that for all $\sigma \in \{-1, +1\}^n$,

$$\mathbb{P}_{G(T)}\left[X = \sigma\right] = \mathbb{P}\left[X = \sigma\right].$$

The key is to observe that if $j \in [n_c]$, and $B \subseteq A_j$ such that $|B| \equiv 0 \pmod 2$, then

$$\prod_{e \in \mathcal{P}(G(T), B)} \tilde{\theta}_e = \prod_{e \in \mathcal{P}(T, B)} \theta_e. \tag{5.37}$$

Therefore by Proposition 3,

$$\mathbb{P}_{G(T)}\left[X_i = \sigma_i \text{ for all } i \in A_j\right] = \mathbb{P}\left[X_i = \sigma_i \text{ for all } i \in A_j\right]$$

Therefore by independence of the components of $G(T)$,

$$\mathbb{P}_{G(T)}\left[X = \sigma\right] = \prod_{j=1}^{n_c} \mathbb{P}_{G(T)}\left[X_i = \sigma_i : i \in A_j\right]$$

Therefore by Eq. (5.37) and by (i.)

$$\begin{aligned}\mathbb{P}_{G(T)}\left[X = \sigma\right] &= \prod_{j=1}^{n_c} \mathbb{P}\left[X_i = \sigma_i \text{ for all } i \in A_j\right] \\ &= \mathbb{P}\left[X = \sigma\right]\end{aligned}$$

where the second equality is justified by Lemma 40. Since $\sigma \in \{-1 + 1\}^n$ was arbitrary, it follows that $\mathbb{P}_{G(T)} \overset{d}{=} \mathbb{P}$. □

Definition 5 (Reduced Topology)**.** *The* ***reduced topology*** *of T is the multiset*

$$\Sigma_{\mathrm{R}}(T) := \left[\Sigma(T_i) : i \in n_c\right]$$

where $\Sigma(T_i)$ is the set of edge splits of T_i; if T_i consists of a single node, then $\Sigma(T_i) = \emptyset$.

Definition 5 generalizes the notion of a split system to a collection of disjoint semi-labeled trees. A reduced topology r is regarded as a graph consisting of $|r|$ disconnected components whose edges are determined by their respective split systems. The edges in r are denoted $E(r)$. For lack of a better name, the pair (r, θ), where $\theta = (\theta_e)_{e \in E(r)}$ is referred to a as a **reduced tree** (though it need not be a tree). The output of Algorithm 1 is a reduced tree.

The necessity of using a multiset in Definition 5 is to allow for repeated instances of the empty set, which correspond to disconnected labeled singleton vertices; this is illustrated in the following example.

Example 7 (3-leaf reduced topologies)**.** If T is a [3]-tree with $\theta_1, \theta_2 \in (0, 1)$, and $\theta_3 = 1$,

then the reduced topology of T is

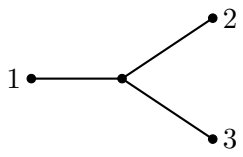
$$\Sigma_{\mathrm{R}}(T) = [\{1|23, 2|13\}]$$

On the other hand, if $\theta_1 = 0$, $\theta_2\theta_3 \in (0,1)$, then the reduced topology is

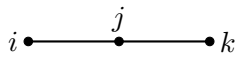
$$\Sigma_{\mathrm{R}}(T) = [\{2|3\}, \emptyset].$$

More generally, it is not too difficult to enumerate all of the reduced topologies which can be obtained from a phylogenetic [3]-tree. If T is a phylogenetic [3]-tree with Hadamard edge parameters $\theta_1, \theta_2, \theta_3 \in [0,1]$, then $\Sigma_{\mathrm{R}}(T)$ is an element of one of the following seven equivalence classes, with the corresponding graphical model shown. Note that in this example, when writing a split $A|B$, we do not require $A \subseteq [2]$.

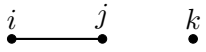
1. $\Sigma_{\mathrm{R}}(T) = [\{1|23, 2|13, 12|3\}]$



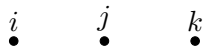
2. $\Sigma_{\mathrm{R}}(T) \in \{[\{i|jk, k|ij\}] : i, j, k \in \{1,2,3\} \text{ pairwise distinct}\}$



3. $\Sigma_{\mathrm{R}}(T) \in \{[\{i|j\}, \emptyset] : i, j \in [3], i \neq j\}$



4. $\Sigma_{\mathrm{R}}(T) = [\emptyset, \emptyset, \emptyset]$



5. $\Sigma_{\mathrm{R}}(T) \in \{[\{ij|k\}] : i, j, k \in [3] \text{ pairwise distinct}\}$

6. $\Sigma_{\mathrm{R}}(T) = [\emptyset, \emptyset]$

7. $\Sigma_{\mathrm{R}}(T) = [\emptyset]$

In Theorem 9 we showed that the MLE for the 3-leaf case always has reduced topology of the form (1), (2), (3), or (4). Actually, this fact is utterly trivial under assumption **A.1**. If $\Sigma_R(T)$ contains a node with more than one distinct label, say i and j, then then $\mathbb{P}[X_i = X_j] = 1$, and as such the likelihood of observing data with a site pattern σ such that $\sigma_i \neq \sigma_j$ is zero. It follows that T is a phylogenetic [3]-tree with $\Sigma_R(T)$ in one of the equivalence classes shown in (5), (6) or (7), then the likelihood of T is zero by the assumption given in **A.1**.

Finite vs Infinite Branch Models

Another useful distinction is between models corresponding to trees with finite branch lengths and models corresponding to trees with one or more infinitely long branches.

To make this precise, let $\bar{p}(\theta, \tau) = (\bar{p}_\alpha(\theta, \tau)_{\alpha \subseteq [3]} \in \Delta_7$ be the distribution of X under the CFN model on with phylogenetic [4]-tree parameter $T = (\theta, \tau)$, where $\theta \in [0, 5]$ is a vector of Hadamard branch lengths and τ is the unrooted topology of the tree. We regard the 4-leaf CFN model as the set

$$\mathcal{M}_{\text{CFN}} = \bar{\mathcal{M}}_{12|34} \cup \bar{\mathcal{M}}_{13|24} \cup \bar{\mathcal{M}}_{23|14}$$

where $\bar{\mathcal{M}}_{12|34}$, $\bar{\mathcal{M}}_{13|24}$, and $\bar{\mathcal{M}}_{23|14}$ are defined in Eq. (5.8). Define

$$\mathcal{M}_{\text{IBM}} := \bigcup_{B \subseteq [5]} \bigcap_{i \in B} \mathcal{B}_{i,0} \cap \mathcal{M}_{\text{CFN}}.$$

where

$$\mathcal{B}_{i,0} := \{p(T) : \theta_i = 0\}.$$

We refer to elements of $\mathcal{M}_{\text{IBM}}$ as **infinite branch models** as they correspond to phylogenetic trees possessing one or more branches which are "infinitely long" when measured in evolutionary distance (i.e. expected number of mutations per site). As demonstrated in Lemma 40, under an infinite branch model, the entries of the vector $X = (X_1, \ldots, X_4)$ are partitioned into two or more independent sets. Under the reduction framework presented here, such trees are (more properly) regarded as a union of disconnected trees, since

$$\mathcal{M}_{\text{IBM}} = \{p(T) : |\Sigma(T)| > 0\}.$$

By contrast, we regard the **finite branch models** as the elements of the set

$$\mathcal{M}_{\text{FBM}} = \mathcal{M}_{\text{CFN}} \cap (\mathcal{M}_{\text{IBM}})^c = \left\{p(\theta, \tau) : \theta \in (0, 1]^5\right\} = \{p(T) : |\Sigma_{\text{R}}(T)| = 1\}.$$

While consideration of infinite branch models adds complexity to this problem (e.g. the issue of nonidentifiability discussed in Example 6), we hope that understanding the behavior of the maximum likelihood estimate on the infinite branch models may help to

provide insights into the ways that maximum likelihood can fail on trees with one or more very long branch lengths.

The next lemma, a consequence of Theorem 8, shows that this problem of nonidentifiable branch lengths does not arise when considering only distributions in $\mathcal{M}_{\text{FBM}}$, a fact proved in [167]. The adaptation stated here gives an explicit formula for recovering the branch lengths $\theta_1, \ldots, \theta_5$ from a probability distribution $\bar{p}$ on these models.

Lemma 44 (Identifiability of Edge Parameters)**.** *For each* $\tau \in \{12|34, 13|24, 23|14\}$ *, the function* $\bar{\Psi}_\tau : (0,1]^5 \to \Delta_3 \subseteq \mathbb{C}^4$ *given by* $\theta \mapsto \bar{p}(\theta, \tau)$ *has inverse*

$$\bar{\Psi}_\tau^{-1} = \psi_\tau \circ H_3,$$

where H_3 *is the* 8×8 *Hadamard matrix defined by Eq.* (5.12) *and*

$$\psi_{12|34}(x) := \left(\sqrt{\frac{x_4 x_6}{x_7}}, \sqrt{\frac{x_3 x_4}{x_2}}, \sqrt{\frac{x_5 x_7}{x_3}}, \sqrt{\frac{x_3 x_5}{x_7}}, \sqrt{\frac{x_3 x_6}{x_8}}\right)$$

$$\psi_{13|23}(x) := \left(\sqrt{\frac{x_2 x_6}{x_5}}, \sqrt{\frac{x_3 x_4}{x_2}}, \sqrt{\frac{x_6 x_7}{x_4}}, \sqrt{\frac{x_3 x_5}{x_7}}, \sqrt{\frac{x_4 x_5}{x_8}}\right)$$

$$\psi_{23|14}(x) := \left(\sqrt{\frac{x_2 x_6}{x_5}}, \sqrt{\frac{x_3 x_7}{x_5}}, \sqrt{\frac{x_5 x_7}{x_3}}, \sqrt{\frac{x_2 x_5}{x_6}}, \sqrt{\frac{x_4 x_5}{x_8}}\right).$$

Moreover, if τ^* *is the star toplogy, then*

$$\bar{\Psi}_{\tau^*}^{-1} = \psi_{\tau^*} \circ H_3$$

where ψ_{τ^*} *may be taken to be* $\psi_{12|34}$*,* $\psi_{13|24}$*, or* $\psi_{23|14}$*.*

Proof. It will suffice to consider the case in which $\tau = 12|34$, as the other cases are similar. Let $\theta \in (0,1]^5$ and let $\gamma := (\gamma_\alpha)_{\alpha \subseteq [3]}$, where

$$\gamma_\alpha = \begin{cases} \frac{1}{2}\sum_{e \in E(T)} \log \theta_e & : \quad \alpha = \emptyset \\ -\frac{1}{2} \log \theta_e & : \quad e \text{ induces the split } \alpha | [n] \backslash \alpha \\ 0 & : \quad \text{else} \end{cases}$$

Therefore by Theorem 8

$$\bar{p}(\theta) = H_3^{-1} \exp . (H_3 \gamma) ,$$

and hence by Eq. (5.14)

$$H_3 \bar{p}(\theta) = \exp . (H_3 \gamma) = \begin{bmatrix} 1 \\ \theta_1\theta_4\theta_5 \\ \theta_2\theta_4\theta_5 \\ \theta_1\theta_2 \\ \theta_3\theta_4 \\ \theta_1\theta_3\theta_5 \\ \theta_2\theta_3\theta_5 \\ \theta_1\theta_2\theta_3\theta_4 \end{bmatrix}$$

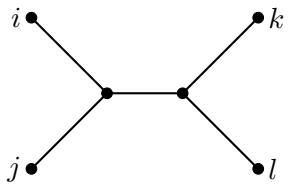
In particular, this equation implies that $(H_3 p(\theta))_i > 0$ for all i. It is easy to check that $\theta = \psi_{12|34}(H_3 \bar{p})$. This proves the lemma. □

Four-leaf reduced topologies

Next we apply the reduction framework introduced in Section 5.3.2 to the study of 4-leaf trees. Consideration of reduced topologies considerably simplifies the number of boundary cases to consider when computing the maximum likelihood estimate. There are 10 classes of reduced topologies which correspond to phylogenetic [4]-trees with nonzero likelihoods, as we discuss below. We list them here, in descending order of the number of edges on the reduced topology.

(R_1) Binary quartets:

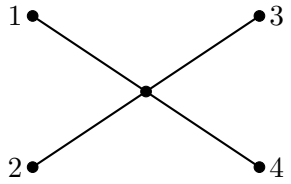
$$R_1 := \{[\{ij|kl, i|jkl, k|ijl, j|ikl, ijk|l\}] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



There are three distinct elements of R_1, corresponding to the three unrooted 4-leaft tree topologies. Observe that that $\Sigma_{\mathrm{R}}(T) \in R_1$ if and only if $\theta_e \in (0,1)^5$ for all $e \in E(T)$.

(R_2) Star tree:

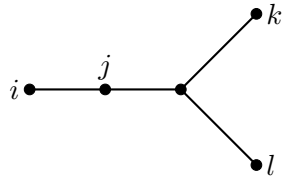
$$R_2 := \{[\{1|234, 2|134, 3|124, 123|4\}]\}$$



R_2 has only one element. Note that $\Sigma_{\mathrm{R}}(T) \in R_2$ if and only if $\theta_5 = 1$ and $\theta_1, \theta_2, \theta_3, \theta_4 \in (0,1)$

(R_3) Y-shaped tree:

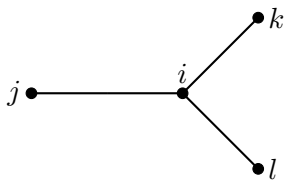
$$R_3 := \{[\{ij|kl, i|jkl, k|ijl, l|ijk\}] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



Note that $\Sigma_{\mathrm{R}}(T) \in R_3$ exactly when $\theta_i = 1, \theta_{\{i,j\}}, \theta_j, \theta_k, \theta_l \in (0,1)$ for some choice of pairwise distinct $i, j, k, l \in [4]$. There are 12 distinct elements of R_3, as swapping the labels k and l does not change $\Sigma_{\mathrm{R}}(T)$.

(R_4) Claw:

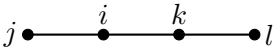
$$R_4 := \{[\{j|ikl, k|ijl, l|ijk\}] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



There are 4 elements of R_4. Note that $\Sigma_{\mathrm{R}}(T) \in R_4$ if and only if there is some choice of pairwise distinct $i, j, k, l \in [4]$ such that $\theta_i = \theta_5 = 1$ and $\theta_j \in (0, 1)$ for all $j \in [4] \backslash \{i, 5\}$.

(R_5) Line:

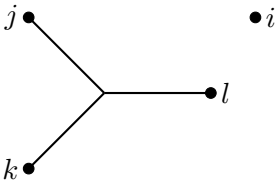
$$R_5 := \{[\{ij|kl, j|ikl, l|ijk\}] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



Note that R_5 has 12 elements. Moreover, $\Sigma_{\mathrm{R}}(T) \in R_5$ if and only if for some choice of pairwise distinct $i, j, k, l \in [4]$, T has topology $ij|kl$ with $\theta_i, \theta_k = 1$ and $\theta_5, \theta_j, \theta_k \in (0, 1)$.

(R_6) Infinite $\{k\}$-branch model:

$$R_6 := \{[\{j|kl, k|jl, l|jk\}, \emptyset] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



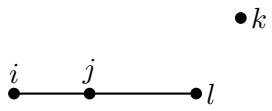
R_6 consists of four elements, which we abbreviate IBM(k), $k \in [4]$. Note that $\Sigma_{\mathrm{R}}(T) \in R_6$ if and only if there exists a $k \in [4]$ such that $\theta_k = 0$ and

$$\prod_{e \in \mathcal{P}(T,A)} \theta_e \in (0, 1)$$

for all $A \subseteq [4] \backslash \{k\}$.

(R_7) Degenerate IBM(k)

$$R_7 := \{[\{ij|l, i|jl\}, \emptyset] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$



R_7 contains 12 distinct elements. For a 4-leaf tree T with topology $ij|kl$, it holds that $\Sigma_{\mathrm{R}}(T) \in R_7$ if and only if there exists a choice of pairwise distinct $i, j, k, l \in [4]$ such that T has topology $ij|kl$, that $\theta_k = 0$ and that one of the following two conditions hold:

- $\theta_j = 1$ and $\theta_i, \theta_l\theta_5 \in (0, 1)$,
- $\theta_l = \theta_5 = 1$ and $\theta_i, \theta_j \in (0, 1)$.

(R_8) Infinite $\{i, j\}$-branch model

$$R_8 := \{[\{k|l\}, \{i|j\}] : i, j, k, l \in [4] \text{ pairwise distinct}\}$$

R_8 has three distinct elements: $[1|2, 3|4]$, $[1|3, 2|4]$, and $[1|4, 2|3]$. Note $\Sigma_{\mathrm{R}}(T) \in R_8$ if and only if for some choice of pairwise distinct $i, j, k, l \in [4]$, T has topology $ij|kl$ with $\theta_5 = 0$ and $\theta_i\theta_j, \theta_k\theta_l \in (0, 1)$.

(R_9) Infinite $\{i\}, \{j\}$-branch model (IBM(i, j))

$$R_9 := \{[\{k|l\}, \emptyset, \emptyset] : k, l \in [4], k \neq l\}$$

R_9 consists of 6 elements. $\Sigma_{\mathrm{R}}(T)$ if and only if there exists a pair i, j such that for all $A \subseteq [4]$ with $|A| = 2$, the following condition holds:

$$\prod_{e \in \mathcal{P}(T,A)} \theta_e = 0 \text{ if and only if } \{i, j\} \cap A \neq \emptyset.$$

(R_{10}) Full independence model:

$$R_{10} := \{[\emptyset, \emptyset, \emptyset, \emptyset]\}$$

Note that $\Sigma_{\mathrm{R}}(T) = [\emptyset, \emptyset, \emptyset, \emptyset]$ if and only if $\prod_{e \in \mathcal{P}(T,A)} \theta_e = 0$ for all $A \subseteq [4]$. In this case, Lemma 40 implies that the random variables $X_1, \ldots, X_4 \overset{iid}{\sim} \mathrm{unif}\{-1, +1\}$. Hence the log-likelihood of any data s on a tree with $\Sigma_{\mathrm{R}}(T) = [\emptyset, \emptyset, \emptyset, \emptyset]$ is

$$\ell(p(\Sigma_{\mathrm{R}}(T), \theta)) = -N \log 16.$$

for all θ.

5.3.3 Restatement of the 4-leaf MLE problem

In addition to the reduced topology classes listed in the previous section, there are a number of others (listed in Appendix A.3), all of which correspond to graphs containing a vertex with more than one label; as noted in Example 7, such graphs correspond to trees with likelihood zero due to assumption **A.1**. Therefore to maximize $\ell(\theta)$ over $\theta \in [0, 1]^4$,

it suffices to maximize ℓ over the set of reduced trees with reduced topologies in

$$\mathcal{R} := \{R_1, \ldots, R_{10}\}.$$

In light of this observation, we can restate Problem 1 in a manner which is much easier to work with. For each $r \in R_i$ with $i \in [10]$, define

$$\tilde{p}(r, \theta) := (\tilde{p}_\alpha(r, \theta))_{\alpha \subseteq [3]}$$

where $\tilde{p}_\sigma(r, \theta)$ is the probability that $X = \sigma$ under the CFN process on a graph with topology r and Hadamard branch parameters $(\theta_e)_{e \in E(r)}$.

Remark 5 (Notation)**.** *If $r \in R_1 \cup R_2$, then r is a binary or star tree, and hence $\tilde{p}(r, \theta) = \bar{p}(r, \theta)$; in that case, we will use $\bar{p}$ and $\bar{p}_\alpha$ rather than $\tilde{p}$ and $\tilde{p}_\alpha$ in order to stay consistent with the notation introduced Section 5.1.2*

With this definition in hand, the global 4-leaf maximum likelihood problem can be reformulated as follows.

Problem 2 (Global 4-leaf maximum likelihood problem: reduced version)**.** *Given data*

$$\mathbf{s} = (s_\sigma)_{\sigma \in \{-1,+1\}^4} \in \mathbb{R}^{16}_{\geq 0}$$

*with $\sum_{\sigma \in \{-1,+1\}^4} \bar{s}_\sigma = N \in \mathbb{Z}_{>0}$ satisfying assumptions **A.3** and **A.4**, maximize the function*

$$\ell(r, \theta) := \sum_{\alpha \subseteq [3]} \bar{s}_\alpha \log \tilde{p}_\alpha(r, \theta) - N \log 2 \tag{5.38}$$

over all $r \in R_1 \cup \cdots \cup R_{10}$ and all $\theta = (\theta_e)_{e \in E(r)} \in (0, 1)^{|E(r)|}$.

Solving this problem yields a solution to Problem 1, since if $(\widehat{r}, \widehat{\theta})$ is a maximizer of Eq. (5.38), one can use the discussion in Section 5.3.2 to recover the set of all phylogenetic [4]-trees which are mapped by the reduction algorithm to $(\widehat{r}, \widehat{\theta})$. Moreover if T is a phylogenetic [4]-tree with topology $\tau \in \{12|34, 13|23, 23|14\}$ and Hadamard branch parameters

$\theta \in [0,1]^5$ such that $(\widehat{r}, \widehat{\theta})$ is the reduced tree of T, then since likelihoods are invariant under reduction by Lemma 43, it follows that $T = (\tau, \theta)$ is a maximizer of Eq. (5.36).

To solve Problem 2, we break it into a number of smaller problems, maximizing Eq. (5.38) on each of the elements of $R_1 \cup \ldots \cup R_{10}$ individually:

Problem 3 (Maximize the log-likelihood for fixed $r \in R_i$.)**.** *Fix $r \in R_i$ for some $i \in [10]$. Find all local maxima of Eq.* (5.38) *over all $\theta = (\theta_e)_{e \in E(r)} \in (0,1)^{|E(r)|}$.*

5.3.4 Solutions to relevant subproblems

In this section, we solve Problem 3 for $r \in R_1 \cup \ldots \cup R_{10}$.

Closed-form solutions for $R_3, \ldots, R_{10}$

The next proposition shows that when $r \in R_3 \cup \ldots \cup R_{10}$, a closed-form solution exists to Problem 3.

Proposition 4 (Existence of closed-form solutions: 4-Leaf Boundary Cases)**.** *Fix $r \in R_i$ for some $i \in \{3, \ldots, 10\}$, and assume that the data $\mathbf{s}$ satisfies assumptions **A.3** and **A.4** hold. Then Problem 3 has a closed-form algebraic solution. Specifically, the function $\theta \mapsto \ell(r, \theta)$ has at most one local maximum; if this maximum exists, it has a unique maximizer $\widehat{\theta} = (\widehat{\theta}_e)_{e \in E(r)} \in (0,1)^{|E(r)|}$ such that for each $e \in E(r)$, the Hadamard branch parameter $\widehat{\theta}_e$ can be written as a closed-form expression (possibly involving radicals) of the data.*

Proof. The proof consists of considering each case $R_3, \ldots, R_{10}$ separately, and computing the numerical parameters which maximimze $\theta \mapsto \ell(r, \theta)$ using calculus. This is the content of Lemmas 55 to 61, found in the Appendix A.4, which provide formulas for the maximizers; in each case, the maximizer is an algebraic expression of the data. □

Solutions for R_1, R_2 using numerical algebraic geometry

In this section we describe the techniques used to maximize $\theta \mapsto \ell(r, \theta)$ when $r \in R_1 \cup R_2$. This section is based on algorithms developed in [183, 182, 173]. Here we use the notation

$$(s_1, \ldots, s_8) := \bar{\mathbf{s}} = (\bar{s}_\alpha)_{\alpha \subseteq [3]}$$

and

$$(p_1, \ldots, p_8) := \bar{p} = (\bar{p}_\alpha)_{\alpha \subseteq [3]}\,.$$

Using this notation, as a function of the probability coordinates $p_1, \ldots, p_8$, log-likelihood can be expressed as

$$\ell(\bar{p}) = \sum_{i=1}^{8} s_i \log p_i - N \log 2. \tag{5.39}$$

The problem is to maximize ℓ subject to constraints which are the phylogenetic invariants depending on the topology of T, i.e., those shown in Eq. (5.16), as these generate the phylogenetic ideal. We solve this problem by using the method of Lagrange multipliers to obtain a polynomial system which can be solved numerically with homotopy continuation.

Case 1. (Binary trees). Here we consider the case where $r \in R_1$, so that r is a binary quartet tree with topology $\tau \in \{12|34, 13|24, 23|14\}$. As stated in Problem 3, the goal is to find all vectors of Hadamard parameters $\theta \in (0, 1)^5$ which maximize Eq. (5.38). This is can be accomplished by first finding all maximizers of Eq. (5.39) in $\mathcal{I}_\tau$ using the method of Lagrange multipliers, and then recovering the corresponding θ with Lemma 44. This procedured is described as follows.

Let g_{i_1}, g_{i_2}, and g_{i_3} be the three generators of $\mathcal{I}_\tau$ shown in Eq. (5.16). Let

$$\mathbf{g} = (g_{i_1}, g_{i_2}, g_{i_3})^\top$$

and let $\lambda = (\lambda_1, \lambda_2, \lambda_3)^\top$ where the λ_i's are the Lagrange multipliers. Then the Lagrangian

takes the form

$$\mathcal{L} = \sum_{i=1}^{8} s_i \log(p_i) + \mathbf{g}(p_1, \ldots, p_8)^\top \lambda. \tag{5.40}$$

Let $J_{\mathbf{g}} = J_{\mathbf{g}}(p_1, \ldots, p_8)$ denote the Jacobian of $\mathbf{g}$ with respect to $p_1, \ldots, p_8$. Taking the partial derivatives of Eq. (5.40) with respect to the variables $p_1, \ldots, p_8, \lambda_1, \lambda_2, \lambda_3$, we obtain the system of (rational) equations

$$\nabla_{p,\lambda} \mathcal{L} = 0,$$

which written out is

$$\begin{aligned} \operatorname{diag}\left(\frac{s_1}{p_1}, \ldots, \frac{s_8}{p_8}\right) + J_{\mathbf{g}}^\top \lambda &= 0 \\ g(p_1, \ldots, p_8) &= 0. \end{aligned}$$

Multiplying by the matrix $\Lambda_p := \operatorname{diag}(p_1, \ldots, p_8)$ to clear the p_i's from the denominators, gives the polynomial system

$$\begin{aligned} \bar{\mathbf{s}} + \Lambda_p J_{\mathbf{g}}^\top \lambda &= 0 \\ \mathbf{g}(p_1, \ldots, p_8) &= 0. \end{aligned} \tag{5.41}$$

The solutions to Eq. (5.41) correspond to the critical points of ℓ on $\mathcal{V}_\tau$, so if a local maximum exists it must satisfy Eq. (5.41).

For generic data, the system Eq. (5.41) has finitely many complex solutions; this number is is the **maximum likelihood degree**, and for the CFN model considered here is known to be 14 for the binary quartet topologies and 92 for the star tree [187]. Since Eq. (5.41) is a polynomial system, we can use numerical homotopy continuation to compute its solutions; for this purpose we use parameter homotopy implemented in the Julia package `HomotopyContinuation`, which also allows for certification of the solutions and refinement of the solution set to arbitrary accuracy using Newton's method [188]. This method provides theoretically correct solutions to polynomial systems given generic

data [173]. Thus, for generic data, we obtain finite solution set

$$\mathcal{C}_{p,\tau} := \{\bar{p} \in \Delta_7 : \bar{p} \text{ is a critical point of Eq. (5.41)}\}$$

which contains the any maximizer(s) of $\bar{p} \mapsto \ell(\bar{p})$ on $\mathcal{V}_\tau \cap \Delta_7$. By Lemma 44, each element in $\mathcal{C}_{\max}$ corresponds to a unique assignment of edge parameters $\theta_1, \ldots, \theta_5 \in \mathbb{R}$ on a 4-leaf tree with topology τ, and these can be obtained by applying $\bar{\Phi}_\tau^{-1}$ to the elements of $\mathcal{C}_\theta$. While it is possible to test the remaining critical points to determine if they are local maxima (e.g., using the bordered determinental criterion [185, p.155]), doing so is not necessary. Rather, it is sufficient to note that if there exists an assignment of Hadamard branch parameters $\theta_1, \ldots, \theta_5 \in (0,1)$ to the edges of g which maximizes $\theta \mapsto \ell(g, \theta)$, then it is contained in the set

$$\mathcal{C}_{\theta,\tau} := \operatorname{im} \bar{\Phi}_\tau^{-1}[\mathcal{C}_{p,\tau}] \cap (0,1)^3 \tag{5.42}$$

Case 2. (Star Tree). If $\Sigma_{\mathrm{R}}(T) \in R_2$, then by Eq. (5.16), $\mathcal{I}_{\text{star}} = \langle g_0, g_1, g_2, g_3 \rangle$, so we perform the same procedure as in Case 1, but with constraints g_0, g_1, g_2 and g_3.

Remark 6. *Since Eq.* (5.15) *is a polynomial in* $\theta_1, \ldots, \theta_5$*, it is possible to use homotopy continuation to solve a polynomial system in the branch lengths* $\theta_1, \ldots, \theta_5$ *directly; however the method detailed here, of first finding the probability coordinates* $p_1, \ldots, p_8$ *which maximize* ℓ *and then recovering the branch lengths using the formulas from Lemma 44 is substantially faster.*

5.3.5 Main Result: Statement of Algorithm and Proof of Correctness

We are now ready to state the main result of this section, an algorithm for solving Problem 2 and a proof of its correctness. For generic (i.e. almost all) data, Algorithm 2 returns the theoretically correct solution to Problem 2. By the discussion in Section 5.3.3, this also yields a solution to Problem 1.

Most of the work for proving the correctness of Algorithm 2 has already been done, and the result is summarized here in the following theorem:

Theorem 10 (Correctness of Algorithm 2)**.** *There exists a generic subset $S \subseteq \mathbb{R}^{16}_{>0}$ such that for all $\mathbf{s} \in S$, Algorithm 2 returns all reduced trees maximizing Eq.* (5.38)*.*

Proof. Let (r^*, θ^*) be a global maximizer of Eq. (5.38).

If $r^* \in R_1 \cup R_2$, then $\bar{p}^* := (\bar{p}_\alpha(r^*, \theta^*))_{\alpha \subseteq [3]}$ is a global maximizer of Eq. (5.39), and therefore by the work in Section 5.3.4, there is a generic set $\tilde{S} \subseteq R^{16}_{>0}$ such that if $\mathbf{s} \in \tilde{S}$, then the set defined in Eq. (5.42) obtained via numerical algebraic geometry contains (r^*, θ^*). Therefore $(r^*, \theta^*) \in \mathcal{C}$ by Step 2.

On the other hand, suppose $r^* \in R_3 \cup \cdots \cup R_{10}$. Let

$$S := \tilde{S} \cap \left\{ \mathbf{s} \in \mathbb{R}^{16}_{>0} : \prod_{i,j,k,l \in [4] \text{ pairwise distinct } : (i,j) \neq (k,l)} (B_{ij} - B_{kl}) \neq 0 \right\}.$$

If $\mathbf{s} \in S$ then both assumptions **A.3** and **A.4** are satisfied, and therefore Step 3 returns the correct local maxima for all cases where $r \in R_3 \cup \cdots \cup R_{10}$ by Proposition 4. Therefore $(r^*, \theta^*) \in \mathcal{C}$ by Step 3.

Since $\mathcal{C}$ is a finite set for $\mathbf{s} \in \tilde{S}$, and since (r^*, θ^*) is a global maximizer, it will be returned by Step 4. □

Algorithm 2: Algorithm for computing an global solution to Problem 2

Input: A point $\mathbf{s} \in \mathbb{R}^{16}_{\geq 0}$ satisfying assumptions **A.3** and **A.4**.
Output: The set of reduced trees which maximize Eq. (5.38).

1 Let $\mathcal{C}$ be an empty list.
2 For each $r \in R_1 \cup R_2$, use numerical algebraic geometry to calculate $\mathcal{C}_{\theta,\tau}$, where $\tau \in \{12|34, 13|24, 23|14, \text{star}\}$ is the quartet topology of r, and collect the elements of $\mathcal{C}_{\theta,\tau}$ into $\mathcal{C}$.
3 For each $r \in R_3 \cup \ldots \cup R_{10}$, compute the solution to Problem 3 using the formulas from Lemmas 55 to 61. Collect the reduced trees which are maximizers into $\mathcal{C}$.
4 Output the element(s) of $\mathcal{C}$ with the greatest log-likelihood.

Appendix A

Supplementary Material

A.1 Proof of 3-Leaf Theorem

This section presents the remainder of the proof of Theorem 9. In this section, we use the notation from Section 5.2.

A.1.1 Maximizing ℓ on $\partial(0,1)^3$

In this subsection we consider the problem of maximizing ℓ on the boundary $\partial(0,1)^3$. As discussed in Section 5.2, this corresponds to the boundary of the unit cube, consisting of 6 faces, 12 edges, and 8 vertices. The lemmas in this section consider the problem of maximizing ℓ on various groupings of these components.

The eight vertices of the unit cube are simply the elements of the set $\{0,1\}^3$. The twelve edges are the sets

$$E_{(\cdot,j,k)} := \big\{(\theta_1, j, k) \in \mathbb{R}^3 : \theta_1 \in (0,1)\big\}$$
$$E_{(j,\cdot,k)} := \big\{(j, \theta_2, k) \in \mathbb{R}^3 : \theta_2 \in (0,1)\big\}$$
$$E_{(j,k,\cdot)} := \big\{(j, k, \theta_3) \in \mathbb{R}^3 : \theta_3 \in (0,1)\big\}$$

defined for $j, k \in \{0, 1\}$. The 6 faces are the sets of the form

$$F_{\pi,i} := \left\{(\theta_1, \theta_2, \theta_3) : \theta_{\pi(1)}, \theta_{\pi(2)} \in (0, 1), \theta_{\pi(3)} = i\right\},$$

where $\pi \in A_3$ and $i \in \{0, 1\}$.

The task at hand is to maximize ℓ on each of these boundary sets. We begin with the next lemma, which utilizes assumption **A.1** to dispatch the edge and vertex boundary cases which are "trivial" in the sense of never containing the maximum.

Lemma 45 (Trivial cases: $\mathcal{E}_{\text{triv}}$)**.** *Assume that **A.1** holds. Let*

$$\mathcal{E}_{\text{triv}} := E_{(1,1,\cdot)} \cup E_{(1,\cdot,1)} \cup E_{(\cdot,1,1)} \cup \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

If $\theta \in \mathcal{E}_{\text{triv}}$ then $\ell(\theta \mid s) = -\infty$.

Proof. If $\theta \in E_{(1,1,\cdot)} \cup E_{(1,\cdot,1)} \cup E_{(\cdot,1,1)} \cup \{(0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$ then there exist two leaves $i, j \in [3]$ with $i \neq j$ such that $\theta_i = \theta_j = 1$. By **A.1**, there exists a $\sigma \in \{1, -1\}^3$ with $\sigma_i \neq \sigma_j$ and $s_\sigma > 0$. Moreover, by Eq. (5.11), the probability of a transition occurring along the path between leaves i and j is zero. In particular, since $\mathbb{P}[X = \sigma] \leq \mathbb{P}[X_i \neq X_j]$, this implies that $\mathbb{P}[X = \sigma] = 0$. Therefore

$$s_\sigma \log \mathbb{P}[X = \sigma] = -\infty.$$

Therefore $\ell(\theta) = -\infty$ by Eq. (5.10). □

The next lemma considers the remaining 4 vertices of the unit cube $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and $(0, 0, 0)$, as well as the edges $E_{(0,0,\cdot)}$, $E_{(0,\cdot,0)}$, and $E_{(\cdot,0,0)}$. The union of these boundaries is the following set:

$$\begin{aligned}\mathcal{E}_{\text{ind}} &:= \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0)\} \cup E_{(0,0,\cdot)} \cup E_{(0,\cdot,0)} \cup E_{(\cdot,0,0)}. \\ &= \bigcup_{\substack{A \subseteq [3]: \\ |A| \geq 2}} \left\{\theta \in [0, 1]^3 : \theta_i = 0 \text{ for all } i \in A\right\}.\end{aligned}$$

The next lemma shows that $\mathcal{E}_{\text{ind}}$ consists of all choices of numerical parameter values under which the X_1, X_2, and X_3 are independent.

Lemma 46 (Log-likelihood of ℓ on $\mathcal{E}_{\text{ind}}$)**.** *If $\theta \in \mathcal{E}_{\text{ind}}$ then*

$$\ell(\theta) = -N \log 8.$$

Proof. Suppose $\theta \in \mathcal{E}_{\text{ind}}$, and let $i, j \in [3]$ such that $i \neq j$. Then the path between leaves i and j contains an edge e such that $\theta_e = 0$. Therefore by Lemma 40, the random variables X_1, X_2 and X_3, are mutually independent. Therefore for all $\sigma \in \{-1, 1\}^3$,

$$\begin{aligned} \mathbb{P}[X = \sigma] &= \mathbb{P}[X_1 = \sigma_1]\, \mathbb{P}[X_2 = \sigma_2]\, \mathbb{P}[X_3 = \sigma_3] \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{8}. \end{aligned} \tag{A.1}$$

Therefore by Eq. (5.10),

$$\ell(\theta) = \sum_{\sigma \in \{1,-1\}^3} s_\sigma \log \mathbb{P}[X = \sigma] = -N \log 8.$$

□

The next lemma considers the problem of maximizing ℓ when exactly one paramter in $\{\theta_1, \theta_2, \theta_3\}$ is assumed to be zero. This pertains to each of the 3 faces $F_{\pi,0}$, $\pi \in A_3$, as well as the remaining six edges $E_{(\cdot,0,1)}$, $E_{(\cdot,1,0)}$, $E_{(0,\cdot,1)}$, $E_{(1,\cdot,0)}$, $E_{(0,1,\cdot)}$, and $E_{(1,0,\cdot)}$. Such boundaries represent one of two graphical models, like those shown in Fig. A.1. We group the boundary cases up into the following three sets, defined for each $\pi \in A_3$:

$$G_\pi := \left\{(\theta_1, \theta_2, \theta_3) : (\theta_{\pi(1)}, \theta_{\pi(2)}) \in (0,1]^2 \backslash \{(1,1)\} \text{ and } \theta_{\pi(3)} = 0\right\}$$

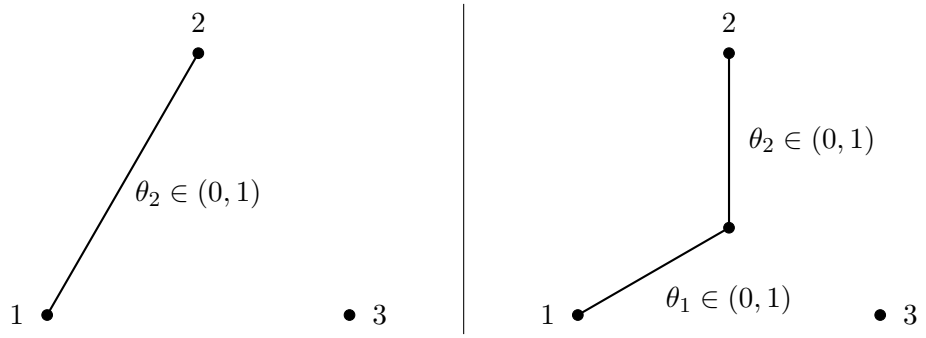


Figure A.1: The graphical model of corresponding to a 3-leaf "tree" with edge parameters $\theta \in E_{(1,\cdot,0)}$ (left) or $\theta \in F_{(\cdot,\cdot,0)}$ (right). In both cases, the biological meaning of $\theta_3 = 0$ is that species 3 is "infinitely far away" from species 1 and 2, when distances are measured in expected number of mutations per site. For this reason $(X_1, X_2) \perp\!\!\!\perp X_3$ (this is proved in Lemma 40), and hence vertex 3 is regarded as a disconnected vertex. By symmetry of the CFN process, $X_3 \sim \text{unif}\{-1, 1\}$.

In the figure on the right, a degree two root is depicted, but it is important to note that its position is not identifiable, as the distribution of X depends only on the product $\theta_1\theta_2$ (as shown in the proof of Lemma 47). In the figure on the left, the root is not depicted; this is because $\theta_1 = 1$, which means that with probability one no site substitutions occur between leaf 1 and the root, and hence that the root coincides with leaf 1; as such, the degree of correlation between observations at leaves 1 and 2 is a function solely of the distance from 2 to the root, which is determined by θ_2.

Interpreted geometically, G_π consists of the union of one face and two of its adjacent edges:

$$G_{(1)} = F_{(1),0} \cup E_{(1,\cdot,0)} \cup E_{(\cdot,1,0)}$$
$$G_{(123)} = F_{(123),0} \cup E_{(0,1,\cdot)} \cup E_{(0,\cdot,1)}$$
$$G_{(132)} = F_{(132),0} \cup E_{(1,0,\cdot)} \cup E_{(\cdot,0,1)}.$$

Lemma 47 (Maximizers of ℓ on G_π)**.** *Let $\pi \in A_3$. The local maxima of ℓ on G_π are the points in the set*

$$\mathcal{G}_\pi = \left\{(\theta_1, \theta_2, \theta_3) : \theta_{\pi(1)}\theta_{\pi(2)} = B_\pi\right\} \cap G_\pi,$$

all of which have log-likelihood

$$\ell(\theta) = M_{ij}^+ \log(1 + B_{ij}) + M_{ij}^- \log(1 - B_{ij}) - N \log 8 \tag{A.2}$$

or equivalently

$$\ell(\theta) = M_\pi^+ \log\left(\frac{M_\pi^+}{4N}\right) + M_\pi^- \log\left(\frac{M_\pi^-}{4N}\right). \tag{A.3}$$

In particular, this implies that if $B_\pi \notin (0,1)$ then ℓ has no local maxima on F.

Proof. Let $\theta \in \mathcal{G}_\pi$. For simplicity, write $i = \pi(1)$, $j = \pi(2)$, and $k = \pi(3)$. Since $\theta_k = 0$, Eq. (5.18) implies that

$$\mathbb{P}[X = \sigma] = \frac{1}{8}\left(1 + \sigma_i\sigma_j\theta_i\theta_j\right). \tag{A.4}$$

Since the distribution depends only on the product $\theta_i\theta_j$, and not on θ_i or θ_j independently, the log-likelihood restricted to the set G_π may thus be regarded as function of the single variable $x := \theta_i\theta_j \in (0,1)$. Plugging Eq. (A.4) into Eq. (5.10), we obtain

$$\begin{aligned}\ell(\theta) &= \sum_{\sigma\in\{1,-1\}^3} s_\sigma \log\left(\frac{1}{8}(1+\sigma_i\sigma_j x)\right) \\ &= \sum_{\sigma\in\{1,-1\}^3} s_\sigma \log\left(1+\sigma_i\sigma_j x\right) - N\log 8\end{aligned}$$

where the second equality follows from $\sum_{\sigma\in\{1,-1\}} s_\sigma = N$. Regrouping terms,

$$\ell(x) = -N\log 8 + M_{ij}^+ \log(1+x) + M_{ij}^- \log(1-x). \tag{A.5}$$

Differentiating gives

$$\ell'(x) = \frac{M_{ij}^+}{1+x} - \frac{M_{ij}^-}{1-x}$$

Solving the equation $\ell'(x) = 0$, we find that f has at most one critical point on $(0,1)$, which is the point

$$x = B_{ij},$$

provided that $B_{ij} \in (0,1)$. If this is the case, then x is a local maximum by the second derivative test, since

$$\ell''(x) = -\left(\frac{M_{ij}^+}{(1+x)^2} + \frac{M_{ij}^-}{(1-x)^2}\right) < 0.$$

Plugging $x = B_{ij}$ into Eq. (A.5) implies Eq. (A.2), since $M_{ij}^{\pm} = B_{\pi}^{\pm}$ and $B_{ij} = B_{\pi}$.

Eq. (A.3) then follows from Eq. (A.2) along with the observations that $M_{\pi}^{+} + M_{\pi}^{-} = N$, $1 + B_{\pi} = \frac{2M_{\pi}^{+}}{N}$, and $1 - B_{\pi} = \frac{2M_{\pi}^{-}}{N}$. □

Remark 7 (Parameter Identifiability)**.** *Observe that one consequence of the proofs of Lemma 46 and Lemma 47 is that if one or more of the numerical parameters* $\{\theta_1^{\mathcal{T}}, \theta_2^{\mathcal{T}}, \theta_3^{\mathcal{T}}\}$ *are equal to zero, then the full vector* $\theta^{\mathcal{T}}$ *cannot be recovered. For example, if two of the numerical parameters in* $\{\theta_1^{\mathcal{T}}, \theta_2^{\mathcal{T}}, \theta_3^{\mathcal{T}}\}$ *are zero, then the other, nonzero parameter* $\theta_i^{\mathcal{T}}$ *is not uniquely determined; specifically this follows immediately by Eq.* (A.1) *which tells us that the distribution of* X *does not depend on the value of* $\theta_i^{\mathcal{T}}$*. Similarly in the proof of Lemma 47, we saw that if exactly one of the numerical parameters equals zero, then (by Eq.* (A.1)*) implies that the distribution* $p(\theta)$ *depends only on the product of the other two, meaning that they cannot be independently identified.*

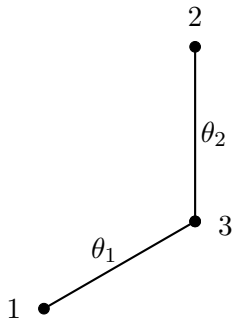


Figure A.2: An example of a tree with numerical parameters $\theta \in F_{\pi,1} = \{\theta : \theta_1, \theta_2 \in (0,1), \theta_3 = 1\}$ with $\pi = (1)$ the identity permutation. Since $\theta_3 = 1$, Eq. (5.11) implies that no transitions can occur on leaf 3, and hence vertex 3 is regarded to lie on the path between leaves 1 and 2.

The final lemma of this section considers the problem of maximzing ℓ on the three remaining faces $F_{\pi,1}$, $\pi \in A_3$. An example of the graphical model correponding to these cases is shown in Fig. A.2.

Lemma 48 (Maximizers of ℓ on $F_{\pi,1}$)**.** *Let* $\pi \in A_3$*, and let* $\theta \in F_{\pi,1}$*. Then the set of local*

maxima of ℓ on $F_{\pi,1}$ is

$$\mathcal{F}_\pi := \left\{(\theta_1,\theta_2,\theta_3) : \theta_{\pi(1)} = B_{\pi(1),\pi(3)},\, \theta_{\pi(2)} = B_{\pi(2),\pi(3)}\right\} \cap F_{\pi,1}.$$

Moreover, if $\theta \in \mathcal{F}_\pi$ then

$$\ell(\theta) = -N\log 8 + \sum_{\tilde{\pi}\in A_3\setminus\{\pi\}} M^+_{\tilde{\pi}} \log(1+B_{\tilde{\pi}}) + M^-_{\tilde{\pi}} \log(1-B_{\tilde{\pi}}) \tag{A.6}$$

or equivalently

$$\ell(\theta) = \sum_{\tilde{\pi}\in A_3\setminus\{\pi\}} M^+_{\tilde{\pi}} \log\left(\frac{M^+_{\tilde{\pi}}}{\sqrt{2}N}\right) + M^-_{\tilde{\pi}} \log\left(\frac{M^-_{\tilde{\pi}}}{\sqrt{2}N}\right). \tag{A.7}$$

Proof. Let $\pi \in A_3$, and let $i = \pi(1)$, $j = \pi(2)$, and $k = \pi(3)$. Let $\theta_i, \theta_j \in (0,1)$ be arbitrary. Since $\theta_k = 1$, it follows by Eq. (5.18)

$$\begin{aligned}\mathbb{P}\left[X=\sigma\right] &= \frac{1}{8}\left(1+\sigma_i\sigma_j\theta_i\theta_j + \sigma_i\sigma_k\theta_i + \sigma_j\sigma_k\theta_j\right)\\ &= \frac{1}{8}(1+\sigma_i\sigma_k\theta_i)(1+\sigma_j\sigma_k\theta_j).\end{aligned}$$

Hence Eq. (5.10) can be written as

$$\ell(\theta) = -N\log 8 + \sum_{\sigma\in\{-1,1\}^3} s_\sigma \log\left(1+\sigma_i\sigma_k\theta_i\right) + \sum_{\sigma\in\{-1,1\}^3} s_\sigma \log\left(1+\sigma_j\sigma_k\theta_j\right).$$

Therefore

$$\begin{aligned}\ell(\theta) = &-N\log 8 + M^+_{ik}\log\left(1+\theta_i\right) + M^-_{ik}\log\left(1-\theta_i\right) + M^+_{jk}\log\left(1+\theta_j\right)\\ &+ M^-_{jk}\log\left(1-\theta_j\right).\end{aligned} \tag{A.8}$$

Differentiating with respect to θ_i and θ_j, it follows that for each $u \in \{i,j\}$,

$$\frac{\partial\ell}{\partial\theta_u} = \frac{M^+_{uk}}{1+\theta_u} - \frac{M^-_{uk}}{1-\theta_u}.$$

Solving the system $\nabla\ell(\theta) = 0$, we obtain the solution satisfying

$$(\theta_i, \theta_j) = (B_{ik}, B_{jk}) \tag{A.9}$$

and if $B_{ik}, B_{jk} \in (0,1)$ then Eq. (A.9) determines the unique critical point of ℓ on the set $F_{\pi,1}$; on the other hand, if $B_{ik} \notin (0,1)$ or $B_{jk} \notin (0,1)$, then ℓ has no critical points on $F_{\pi,1}$.

Moreover, this critical point on $F_{\pi,1}$ is a maximum by the second derivative test since for all $\theta \in F_{\pi,1}$, the Hessian matrix

$$H_\ell(\theta) = \begin{bmatrix} -\left(\frac{M_{ik}^+}{(1+\theta_i)^2} + \frac{M_{ij}^-}{(1-\theta_i)^2}\right) & 0 \\ 0 & -\left(\frac{M_{jk}^+}{(1+\theta_j)^2} + \frac{M_{jk}^-}{(1-\theta_j)^2}\right) \end{bmatrix}$$

is negative definite.

Finally, if Eq. (A.9) holds, plugging $\theta_i = B_{ik}$ and $\theta_j = B_{jk}$ into Eq. (A.8) gives

$$\begin{aligned} \ell(\theta) = -N\log 8 + M_{ik}^+ \log(1+B_{ik}) + M_{ik}^- \log(1-B_{ik}) \\ + M_{jk}^+ \log(1+B_{jk}) + M_{jk}^- \log(1-B_{jk}), \end{aligned} \tag{A.10}$$

which is nothing but Eq. (A.6) written in a different notation. It remains to prove Eq. (A.7). Observe that $1+B_\pi = \frac{2M_\pi^+}{N}$ and $1-B_\pi = \frac{2M_\pi^-}{N}$ for all $\pi \in A_3$. Eq. (A.7) can be obtained by making these substitutions in Eq. (A.6) and then simplifying using logarithm properties and $M_\pi^+ + M_\pi^- = N$. □

The results of Appendix A.1.1 and Section 5.2.4 are summarized in Fig. A.3.

A.1.2 Comparison of Likelihoods

To prove Theorem 9 will require some comparisons of the log-likelihoods of elements of $\mathcal{E}_{\text{int}}, \mathcal{E}_{\text{ind}}, \mathcal{G}_\pi$, and $\mathcal{F}_\pi$, ($\pi \in A_3$). In this subsection, we prove several lemmas toward this end.

Set S	**Criteria for** $S \neq \emptyset$	$\ell(\theta), \theta \in S$
$\mathcal{E}_{\text{int}}$	$(B_{12}, B_{13}, B_{13}) \in \mathcal{D}$	Eq. (5.24)
$\mathcal{E}_{\text{triv}}$	always nonempty	$-\infty$
$\mathcal{E}_{\text{ind}}$	always nonempty	$-N \log 8$
$\mathcal{G}_\pi$ $(\pi \in A_3)$	$B_\pi \in (0,1)$	Eq. (A.3)
$\mathcal{F}_\pi$ $(\pi \in A_3)$	$B_{\pi'}, B_{\pi''} \in (0,1)$, where $\{\pi', \pi''\} = A_3 \backslash \{\pi\}$	Eq. (A.7)

Figure A.3: Summary of the results of Lemma 42, 45, 46, 47, and 48, which compute the maximizer(s) of ℓ on the interior and boundary of the unit cube (namely, the sets $E_{\text{int}}, \mathcal{E}_{\text{triv}}, \mathcal{E}_{\text{ind}}, G_\pi$, and $F_{\pi,1}, \pi \in A_3$, which together partition the closed unit cube). The corresponding sets of maximizers on each of these sets ($\mathcal{E}_{\text{int}}, \mathcal{E}_{\text{triv}}, \mathcal{E}_{\text{ind}}, \mathcal{G}_\pi$, and $\mathcal{F}_\pi$, $\pi \in A_3$, respectively) are level sets of ℓ, although some of them may be emtpy depending on the data; the necessary and sufficient conditions for each of them to be nonempty, as well as the value that the log-likelihood function takes on each of them, are shown in the second and third columns of the table..

We will make use of the information inequality, which we state next (for proof, see, e.g., Theorem 2.6.3 in [189]).

Theorem 11 (Information Inequality)**.** *Let $k \geq 1$. If $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_k)$ and $\tilde{q} = (\tilde{q}_1, \ldots, \tilde{q}_k)$ satisfy $\tilde{p}, \tilde{q} \in \Delta_{k-1}$ then*

$$\sum_{i=1}^{k} \tilde{p}_i \log \tilde{q}_i \leq \sum_{i=1}^{k} \tilde{p}_i \log \tilde{p}_i,$$

with equality if and only if $\tilde{p} = \tilde{q}$.

The next two lemmas utilize the information inequality to show that elements of $\mathcal{E}_{\text{int}}$ have greater log-likelihood than elements of $\mathcal{F}_\pi$ and $\mathcal{G}_\pi$, for all $\pi \in A_3$.

Lemma 49 ($\mathcal{E}_{\text{int}}$ vs $\mathcal{F}_\pi$)**.** *If $\theta^* \in \mathcal{E}_{\text{int}}$ then*

$$\ell(\theta^*) > \ell(\theta)$$

for all $\theta \in \mathcal{F}_\pi$ and all $\pi \in A_3$.

Proof. We prove the case with $\pi = (1)$ as the proofs for the cases with $\pi \in \{(123), (132)\}$ are similar.

Let $\theta \in \mathcal{F}_\pi$. By Eq. (A.7),

$$\begin{aligned}\ell(\theta) = M_{13}^+ \log\left(\frac{M_{13}^+}{\sqrt{2}N}\right) + M_{13}^- \log\left(\frac{M_{13}^-}{\sqrt{2}N}\right) \\ + M_{23}^+ \log\left(\frac{M_{23}^+}{\sqrt{2}N}\right) + M_{23}^- \log\left(\frac{M_{23}^-}{\sqrt{2}N}\right)\end{aligned} \tag{A.11}$$

Observe that

$$\begin{aligned}M_{13}^+ &= \bar{s}_\emptyset + \bar{s}_{\{2\}} \\ M_{13}^- &= \bar{s}_{\{1\}} + \bar{s}_{\{1,2\}} \\ M_{23}^+ &= \bar{s}_\emptyset + \bar{s}_{\{1\}} \\ M_{23}^- &= \bar{s}_{\{2\}} + \bar{s}_{\{1,2\}}.\end{aligned}$$

Therefore we can rewrite Eq. (A.11) as

$$\begin{aligned}\ell(\theta) = \left(\bar{s}_\emptyset + \bar{s}_{\{2\}}\right) \log\left(\frac{M_{13}^+}{\sqrt{2}N}\right) + \left(\bar{s}_{\{1\}} + \bar{s}_{\{1,2\}}\right) \log\left(\frac{M_{13}^-}{\sqrt{2}N}\right) \\ + \left(\bar{s}_\emptyset + \bar{s}_{\{1\}}\right) \log\left(\frac{M_{23}^+}{\sqrt{2}N}\right) + \left(\bar{s}_{\{2\}} + \bar{s}_{\{1,2\}}\right) \log\left(\frac{M_{23}^-}{\sqrt{2}N}\right).\end{aligned}$$

Regrouping terms gives

$$\begin{aligned}\ell(\theta) = \bar{s}_\emptyset \log\left(\frac{M_{13}^+ M_{23}^+}{N^2}\right) + \bar{s}_{\{1\}} \log\left(\frac{M_{13}^- M_{23}^+}{N^2}\right) + \bar{s}_{\{2\}} \log\left(\frac{M_{13}^+ M_{23}^-}{N^2}\right) \\ + \bar{s}_{\{1,2\}} \log\left(\frac{M_{13}^- M_{23}^-}{N^2}\right) - N \log 2.\end{aligned} \tag{A.12}$$

To apply Theorem 11, it is first necessary to verify that

$$\left(\frac{M_{13}^+ M_{23}^+}{N^2}, \frac{M_{13}^- M_{23}^+}{N^2}, \frac{M_{13}^+ M_{23}^-}{N^2}, \frac{M_{13}^- M_{23}^-}{N^2}\right) \in \Delta_3.$$

The entries of this vector are clearly nonnegative, so it suffices to show that they sum to

1. Indeed, using $M^+_{13} + M^-_{13} = N$ and $M^+_{23} + M^-_{23} = N$, we have

$$\begin{aligned}\frac{M^+_{13}M^+_{23}}{N^2} + \frac{M^-_{13}M^+_{23}}{N^2} + \frac{M^+_{13}M^-_{23}}{N^2} + \frac{M^-_{13}M^-_{23}}{N^2} &= \frac{M^+_{23}\left(M^+_{13} + M^-_{13}\right)}{N^2} + \frac{M^-_{23}\left(M^+_{13} + M^-_{13}\right)}{N^2} \\ &= \frac{\left(M^+_{13} + M^-_{13}\right)\left(M^+_{23} + M^-_{23}\right)}{N^2} \\ &= 1.\end{aligned}$$

Therefore by applying Theorem 11 to the right hand side of Eq. (A.12),

$$\begin{aligned}\ell(\theta) &\leq \sum_{\alpha \subseteq [2]} \bar{s}_\alpha \log\left(\frac{\bar{s}_\alpha}{N}\right) - N \log 2 \\ &= \ell(\theta^*)\end{aligned}$$

where the last equality follow from Eq. (5.24). □

Lemma 50 ($\mathcal{E}_{\text{int}}$ vs $\mathcal{G}_\pi$)**.** *Assume that* ***A.1*** *holds. If* $\theta^* \in \mathcal{E}_{\text{int}}$ *then*

$$\ell(\theta^*) > \ell(\theta)$$

for all $\theta \in \mathcal{G}_\pi$ *and all* $\pi \in A_3$.

Proof. We prove the case with $\pi = (1)$, as the proofs for the cases with $\pi \in \{(123), (132)\}$ are similar. If $\mathcal{E}_{\text{int}} = \emptyset$ then there is nothing to show, so henceforth assume $\mathcal{E}_{\text{int}} \neq \emptyset$. By Lemma 42, $\mathcal{E}_{\text{int}} = \{\theta^*\}$ and

$$\ell(\theta^*) = \sum_{\alpha \subseteq [2]} \bar{s}_\alpha \log\left(\frac{\bar{s}_\alpha}{N}\right) - N \log 2. \tag{A.13}$$

Also by Lemma 42 it must be the case that

$$B_{12}, B_{13}, B_{23} > 0. \tag{A.14}$$

Moreover since $B_{12} > 0$, it follows that $\mathcal{G}_{(1)} \neq \emptyset$ by Lemma 47.

Let $\theta \in \mathcal{G}_\pi$. By Eq. (A.3),

$$\begin{aligned}\ell(\theta) &= M_{12}^+ \log\left(\frac{M_{12}^+}{4N}\right) + M_{12}^- \log\left(\frac{M_{12}^-}{4N}\right) \\ &= \bar{s}_\emptyset \log\left(\frac{M_{12}^+}{2N}\right) + \bar{s}_{\{1,2\}} \log\left(\frac{M_{12}^+}{2N}\right) + \bar{s}_{\{1\}} \log\left(\frac{M_{12}^-}{2N}\right) + \bar{s}_{\{2\}} \log\left(\frac{M_{12}^-}{2N}\right) \\ &\quad - N \log 2.\end{aligned}$$

Therefore by Theorem 11 and by Eq. (A.13),

$$\begin{aligned}\ell(\theta) &\leq \sum_{\alpha \subseteq [3]} \bar{s}_\alpha \log\left(\frac{\bar{s}_\alpha}{N}\right) - N \log 2 \\ &= \ell(\theta^*).\end{aligned}$$

Suppose that equality holds in the above, i.e., that $\ell(\theta) = \ell(\theta^*)$. Then by Theorem 11,

$$\frac{M_{12}^+}{2} = \bar{s}_\emptyset = \bar{s}_{\{1,2\}} \quad \text{and} \quad \frac{M_{12}^-}{2} = \bar{s}_{\{1\}} = \bar{s}_{\{2\}}.$$

Therefore,

$$\begin{aligned}B_{23} &= \frac{M_{23}^+ - M_{23}^-}{N} \\ &= \frac{1}{N}\left(\bar{s}_\emptyset + \bar{s}_{\{1\}} - \bar{s}_{\{1,2\}} - \bar{s}_{\{2\}}\right) \\ &= 0,\end{aligned}$$

but this contradicts Eq. (A.14). We conclude that $\ell(\theta) < \ell(\theta^*)$. □

The next lemma compares the log-likelihods of elements in $\mathcal{G}_\pi$ and $\mathcal{G}_{\tilde{\pi}}$, for distinct elements $\pi, \tilde{\pi} \in A_3$.

Lemma 51 ($\mathcal{G}_\pi$ vs $\mathcal{G}_{\tilde{\pi}}$ for $\pi \neq \tilde{\pi}$)**.** *Let $\pi, \tilde{\pi} \in A_3$ such that $\pi \neq \tilde{\pi}$, and let $\theta \in \mathcal{G}_\pi$ and $\tilde{\theta} \in \mathcal{G}_{\tilde{\pi}}$. Assume that **A.1** holds. If*

$$0 < B_\pi < B_{\tilde{\pi}} \tag{A.15}$$

then

$$\ell(\theta) < \ell(\tilde{\theta}).$$

Proof. Since $B_\pi = 2M_\pi^+ - N$ and $B_{\tilde{\pi}} = 2M_{\tilde{\pi}}^+ - N$, therefore Eq. (A.15) can be rewritten

$$0 < 2M_\pi^+ - N \leq 2M_{\tilde{\pi}}^+ - N$$

and therefore it holds that

$$\frac{N}{2} < M_\pi^+ \leq M_{\tilde{\pi}}^+ < N, \tag{A.16}$$

where the last inequality holds by **A.1**.

If $\hat{\pi} \in \{\pi, \tilde{\pi}\}$ and $\hat{\theta} \in \mathcal{G}_{\hat{\pi}}$, then Eq. (A.3) implies

$$\begin{aligned}\ell(\hat{\theta}) &= M_{\hat{\pi}}^+ \log\left(\frac{M_{\hat{\pi}}^+}{4N}\right) + M_{\hat{\pi}}^- \log\left(\frac{M_{\hat{\pi}}^-}{4N}\right) \\ &= M_{\hat{\pi}}^+ \log\left(\frac{M_{\hat{\pi}}^+}{4N}\right) + \left(N - M_{\hat{\pi}}^+\right) \log\left(\frac{N - M_{\hat{\pi}}^+}{4N}\right).\end{aligned} \tag{A.17}$$

The right-hand side has the form $f(x) := x \log\left(\frac{x}{4N}\right) + (N - x) \log\left(\frac{N-x}{4N}\right)$. Since

$$f'(x) = \log\left(\frac{x}{N-x}\right),$$

which is positive for all $x \in (N/2, N)$, it follows that f is strictly increasing on $(\frac{N}{2}, N)$. Combining this fact with Eqs. (A.16) and (A.17) implies $\ell(\tilde{\theta}) \geq \ell(\theta)$, with equality only if $M_\pi^+ = M_{\tilde{\pi}}^+$, or equivalently $B_\pi = B_{\tilde{\pi}}$. □

Another comparison can be made between the likelihoods of elements in $\mathcal{F}_\pi$ and $\mathcal{G}_{\tilde{\pi}}$ when $\pi, \tilde{\pi} \in A_3$ are distinct, shown in the next lemma.

Lemma 52 ($\mathcal{F}_\pi$ vs $\mathcal{G}_{\tilde{\pi}}$ for $\pi \neq \tilde{\pi}$)**.** *Assume that **A.1** holds. Let π_1, π_2 and π_3 denote the three distinct elements of A_3. If $\theta \in \mathcal{F}_{\pi_1}$ then*

$$\ell(\tilde{\theta}) < \ell(\theta)$$

for all $\tilde{\theta} \in \mathcal{G}_{\pi_2} \cup \mathcal{G}_{\pi_3}$.

Proof. Suppose $\theta \in \mathcal{F}_{\pi_1}$, and suppose that $\tilde{\theta} \in \mathcal{G}_{\tilde{\pi}}$ for some $\tilde{\pi} \in A_3 \backslash \{\pi_1\}$. Without loss of generality, assume $\tilde{\pi} = \pi_3$. Then by Eqs. (A.2) and (A.6)

$$\begin{aligned} \ell(\theta) - \ell(\tilde{\theta}) &= \sum_{i=2}^{3} M_{\pi_i}^{+} \log\left(1 + B_{\pi_i}\right) + M_{\pi_i}^{-} \log\left(1 - B_{\pi_i}\right) \\ &\quad - M_{\pi_3}^{+} \log(1 + B_{\pi_3}) - M_{\pi_3}^{-} \log(1 - B_{\pi_3}) \\ &= M_{\pi_2}^{+} \log\left(1 + B_{\pi_2}\right) + M_{\pi_2}^{-} \log\left(1 - B_{\pi_2}\right). \end{aligned} \tag{A.18}$$

Since $M_{\pi_2}^{+} = \frac{NB_{\pi_2}+N}{2}$ and $M_{\pi_2}^{-} = \frac{NB_{\pi_2}-N}{2}$, it follows that

$$\ell(\theta) - \ell(\tilde{\theta}) = \frac{N}{2}\left[(1 + B_{\pi_2}) \log\left(1 + B_{\pi_2}\right) + (1 - B_{\pi_2}) \log\left(1 - B_{\pi_2}\right)\right]. \tag{A.19}$$

As a function of B_{π_2}, the right-hand side is strictly increasing on $(0, 1)$, which can be seen by differentiating and observing that the derivative is positive on this interval. Moreover, we note that $B_{\pi_2} \in (0, 1)$, a fact which follows from the hypothesis that $\mathcal{F}_{\pi_1} \neq \emptyset$ (see Fig. A.3). Therefore, since the right-hand side of Eq. (A.19) is strictly increasing on $(0, 1)$, and since $B_{\pi_2} \in (0, 1)$, it follows that

$$\ell(\theta) - \ell(\tilde{\theta}) > 0.$$

This completes the proof of the lemma.

□

Lemma 53 ($\mathcal{F}_{\pi}$ vs $\mathcal{F}_{\tilde{\pi}}$, $\pi \neq \tilde{\pi}$)**.** *Let* $\pi, \tilde{\pi} \in A_3$ *be distinct, and suppose that* $\theta \in \mathcal{F}_{\pi}$ *and* $\tilde{\theta} \in \mathcal{F}_{\tilde{\pi}}$. *Then*

$$\ell(\tilde{\theta}) < \ell(\theta)$$

if and only if

$$B_{\pi} < B_{\tilde{\pi}}$$

Proof. By Eq. (A.6)

$$\begin{aligned}\ell(\theta) - \ell(\tilde{\theta}) &= M_\pi^+ \log(1+B_\pi) + M_\pi^- \log(1-B_\pi) - M_\pi^+ \log(1+B_\pi) \\ &\quad - M_\pi^- \log(1-B_\pi)\,. \\ &= \frac{N}{2}\left[f(B_\pi) - f(B_{\tilde{\pi}})\right] \end{aligned} \tag{A.20}$$

where $f(x) := (1+x)\log(1+x) + (1-x)\log(1-x)$. Note that

$$f'(x) = \log\left(\frac{1+x}{1-x}\right) = \log\left(1 + \frac{2x}{1-x}\right)$$

which is positive for all $x \in (0,1)$. Therefore f is increasing on $(0,1)$. Moreover, $B_\pi, B_{\tilde{\pi}} \in (0,1)$ since $\mathcal{F}_\pi, \mathcal{F}_{\tilde{\pi}} \neq \emptyset$ (see Fig. A.3). Taken together, these facts along with Eq. (A.20) imply that $\ell(\theta) - \ell(\tilde{\theta}) > 0$ if and only if $B_\pi > B_{\tilde{\pi}}$. □

The next lemma shows that elements of $\mathcal{G}_\pi, \pi \in A_3$ have greater log-likelihood than elements in $\mathcal{E}_{\text{ind}}$.

Lemma 54 ($\mathcal{G}_\pi$ vs $\mathcal{E}_{\text{ind}}$)**.** *If $\theta \in \mathcal{G}_\pi$ for some $\pi \in A_3$ then*

$$\ell(\theta) > -N\log 8.$$

Proof. If $\theta \in \mathcal{G}_{\pi'}$ then Lemma 47 implies $B_\pi > 0$. Therefore since $B_\pi = \frac{M_\pi^+ - M_\pi^{\neq}}{N}$ and $M_\pi^+ + M_\pi^- = N$, it follows that

$$M_\pi^+ > \frac{N}{2}. \tag{A.21}$$

Let $x := M_\pi^+ = N - M_\pi^-$. By Eq. (A.3),

$$\begin{aligned}\ell(\theta) &= M_\pi^+ \log\left(\frac{M_\pi^+}{4N}\right) + M_\pi^- \log\left(\frac{M_\pi^-}{4N}\right) \\ &= x\log(x) + (N-x)\log(N-x) - N\log(4N)\,. \end{aligned} \tag{A.22}$$

The function $x \mapsto x\log(x) + (N-x)\log(N-x)$ is strictly increasing on the interval

$(\frac{N}{2}, N)$, which can be seen by observing that its derivative is $x \mapsto \log\left(\frac{x}{N-x}\right)$, which is positive on this interval. Since $x \in (\frac{N}{2}, N)$ by Eq. (A.21), it follows that

$$\begin{aligned}\ell(\theta) &> \frac{N}{2}\log\left(\frac{N}{2}\right) + \frac{N}{2}\log\left(\frac{N}{2}\right) - N\log(4N) \\ &= N\log(N) - N\log(2) - N\log(4) - N\log(N) \\ &= -N\log(8)\,.\end{aligned}$$

□

A.1.3 Proof of Theorem 9

Using the lemmas from Section 5.2.4 and Appendices A.1.1 and A.1.3, we are now ready to prove Theorem 9.

Proof of Theorem 9. We claim that $\theta \mapsto \ell(\theta)$ is upper semicontinuous on its domain $[0,1]^3$. To see this, first observe that the function $L : [0,1]^3 \to [0,\infty)$ defined by

$$L(\theta) := \prod_{\alpha \subseteq [2]} \left(\frac{\bar{p}_\alpha(\theta)}{2}\right)^{\bar{s}_\alpha}$$

is continuous since for all $\alpha \subseteq [2]$, $\theta \mapsto \bar{p}_\alpha(\theta)$ is a polynomial in the variables $\theta_1, \theta_2, \theta_3 \in [0,1]$ by Eq. (5.18). Therefore, since $\ell = \log \circ L$ and since $\log(\cdot)$ is increasing and upper semicontinuous on $[0,\infty)$, it follows that ℓ is upper semicontinuous on $[0,1]^3$. This proves the claim.

Since ℓ is upper semicontinuous, it has at least one maximizer on $[0,1]^3$. In order to find the maximizer(s), observe that since

$$[0,1]^3 = (0,1)^3 \sqcup \mathcal{E}_{\text{triv}} \sqcup \mathcal{E}_{\text{ind}} \sqcup \left(\bigsqcup_{\pi \in A_3} G_\pi\right) \sqcup \left(\bigsqcup_{\pi \in A_3} F_\pi\right),$$

it suffices to consider the maximizers of ℓ on each of these sets. By Lemma 45,

$$\ell(\theta) = -\infty$$

whenever $\theta \in \mathcal{E}_{\text{triv}}$. Therefore if $\widehat{\theta} \in [0,1]^3$ is a global maximum of ℓ, then

$$\widehat{\theta} \in \mathcal{E}_{\text{int}} \sqcup \mathcal{E}_{\text{ind}} \sqcup \left(\bigsqcup_{\pi \in A_3} \mathcal{G}_\pi \right) \sqcup \left(\bigsqcup_{\pi \in A_3} \mathcal{F}_\pi \right). \tag{A.23}$$

In Lemmas 42, 46, 47, and 48, we computed the log-likelihood of the points in each of the eight sets in this disjoint union (see Fig. A.3 for a summary of these results), so the rest of the proof will simply be a comparison of the likelihoods of elements of these sets.

We start by proving that if $\mathbf{B} \in \mathcal{D}$, then the maximum is given by Eq. (5.21). Suppose $\mathbf{B} \in \mathcal{D}$. Then by Lemma 42, $\mathcal{E}_{\text{int}}$ is nonempty and consists of a single element

$$\theta^* = \left(\sqrt{\frac{B_{12}B_{13}}{B_{23}}}, \sqrt{\frac{B_{12}B_{23}}{B_{13}}}, \sqrt{\frac{B_{13}B_{23}}{B_{12}}} \right).$$

Moreover, Lemmas Lemma 49, 50, 52, and 54 together imply that θ^* is the global maximizer of ℓ. This proves Eq. (5.21).

Henceforth assume $\mathbf{B} \notin \mathcal{D}$, so that $\mathcal{E}_{\text{int}} = \emptyset$ by Lemma 42. Therefore

$$\widehat{\theta} \in \mathcal{E}_{\text{ind}} \sqcup \left(\bigsqcup_{\pi \in A_3} \mathcal{G}_\pi \right) \sqcup \left(\bigsqcup_{\pi \in A_3} \mathcal{F}_\pi \right). \tag{A.24}$$

Next we will prove part (i) in the statement of the lemma. Suppose that $B_{\pi_3}, B_{\pi_2} > 0$. It will suffice to show that $\widehat{\theta} \in \mathcal{F}_{\pi_1}$.

By the criteria shown in Fig. A.3, it holds that $\mathcal{G}_{\pi_3} \neq \emptyset$ and $\mathcal{F}_{\pi_1} \neq \emptyset$. Let $\theta' \in \mathcal{G}_{\pi_3}$ and $\theta'' \in \mathcal{F}_{\pi_1}$. By Lemmas 51 and 54,

$$\ell(\theta) < \ell(\theta') \tag{A.25}$$

whenever $\theta \in \mathcal{E}_{\text{ind}} \cup \mathcal{G}_{\pi_1} \cup \mathcal{G}_{\pi_2}$. In addition, since $\pi_1 \neq \pi_3$, Lemma 52 implies

$$\ell(\theta') < \ell(\theta''). \tag{A.26}$$

Therefore by Eqs. (A.24) to (A.26),

$$\widehat{\theta} \in \mathcal{F}_{\pi_1} \sqcup \mathcal{F}_{\pi_2} \sqcup \mathcal{F}_{\pi_3}. \tag{A.27}$$

Finally, observe that by Lemma 53,

$$\ell(\theta) < \ell(\theta'') \tag{A.28}$$

whenever $\theta \in \mathcal{F}_{\pi_1} \sqcup \mathcal{F}_{\pi_2}$. By Eqs. (A.27) and (A.28), we conclude that $\widehat{\theta} \in \mathcal{F}_{\pi_1}$. This proves part (i) of the lemma.

Next, suppose that $B_{\pi_3} > 0$ and $B_{\pi_2} < 0$. In order to prove part (ii) in the statement of the lemma, it will suffice to show that $\widehat{\theta} \in \mathcal{G}_{\pi_3}$. By the criteria in Fig. A.3, $\mathcal{F}_\pi = \emptyset$ for all $\pi \in A_3$, and $\mathcal{G}_\pi = \emptyset$ for all $\pi \in A_3 \backslash \{\pi_1, \pi_2\}$. Therefore by Eq. (A.24),

$$\widehat{\theta} \in \mathcal{E}_{\text{ind}} \sqcup \mathcal{G}_{\pi_3}.$$

By Lemma 50, the elements of $\mathcal{G}_{\pi_3}$ have strictly larger log-likelihood than the elements of $\mathcal{E}_{\text{ind}}$. Therefore $\widehat{\theta} \in \mathcal{G}_{\pi_3}$. This proves part (ii) of the lemma.

It remains to prove part (iii) in the statement of the lemma. Suppose $B_{\pi_3} < 0$. By the criteria in Fig. A.3, $\mathcal{F}_\pi = \emptyset$ and $\mathcal{G}_\pi = \emptyset$ for all $\pi \in A_3$. Therefore by Eq. (A.24), $\widehat{\theta} \in \mathcal{E}_{\text{ind}}$. This prove part (iii), which completes the proof of the lemma.

□

A.2 Macaulay2 Code for Computing Phylogenetic Invariants

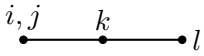
```
--- First define our rings. Here, we take  (p1,...,p8) = (p_j:j in J)
--- = (p_xxxx,p_xyyy,p_xyxx,p_xxyy,p_xxyx,p_xyxy,p_xyyx,p_xxxy).
R1 = QQ[p1,p2,p3,p4,p5,p6,p7,p8]
R2 = QQ[θ1,θ2,θ3,θ4,θ5]
--- Let S be an ordered list consisting a representative of each
--- unique site pattern, with the same order as p:
S = {(1,1,1,1), (-1,1,1,1), (1,-1,1,1), (1,1,-1,-1),
    (1,1,-1,1), (1,-1,1,-1), (1,-1,-1,1),  (1,1,1,-1)}
--- For a 4-leaf tree with topology 12|34, the Hadamard conjugation
--- theorem gives the following formula for p:
f12=(s1,s2,s3,s4)->(1+s1*s2*θ1*θ2 + s3*s4*θ3*θ4+s1*s3*θ1*θ3*θ5 +
    s1*s4*θ1*θ4*θ5 + s2*s3*θ2*θ3*θ5 + s2*s4*θ2*θ4*θ5 +
    s1*s2*s3*s4*θ1*θ2*θ3*θ4)/8
p = apply(S,f12)
--- Define this as a map from R1 to R2 and compute the minimal generators
--- of its kernel:
M = map(R2,R1,p)
I = trim kernel M
--- This returns three generators, which are the phylogenetic invariants:
I_0 == p1+p2+p3+p4+p5+p6+p7+p8-1
I_1 == p3*p5-p4*p6+p2*p7+p3*p7+p4*p7+p5*p7+p6*p7+p7^2+p2*p8+p7*p8-p7
I_2 == p2*p5+p2*p6+p3*p6+p4*p6+p5*p6+p6^2-p4*p7+p6*p7+p3*p8+p6*p8-p6
--- For the other two binary quartet topologies 13|24 and 23|14, as well
--- as the star tree topology, replace f12 in the code above with f13,
--- f23, or f_star (respectively) below:
f13=(s1,s2,s3,s4)->(1+s1*s3*θ1*θ3 + s2*s4*θ2*θ4+s1*s2*θ1*θ2*θ5
    + s1*s4*θ1*θ4*θ5 + s3*s2*θ3*θ2*θ5 + s3*s4*θ3*θ4*θ5
    + s1*s3*s2*s4*θ1*θ3*θ2*θ4)/8
f23=(s1,s2,s3,s4)-> (1+s1*s4*θ1*θ4 + s3*s2*θ3*θ2+s1*s3*θ1*θ3*θ5
    + s1*s2*θ1*θ2*θ5 + s4*s3*θ4*θ3*θ5 + s4*s2*θ4*θ2*θ5
    + s1*s4*s3*s2*θ1*θ4*θ3*θ2)/8
f_star=(s1,s2,s3,s4)-> (1+s1*s4*θ1*θ4 + s3*s2*θ3*θ2+s1*s3*θ1*θ3
     + s1*s2*θ1*θ2 + s4*s3*θ4*θ3 + s4*s2*θ4*θ2
     + s1*s4*s3*s2*θ1*θ4*θ3*θ2)/8
-- --  e.g., using f13 for topology 13|24 gives
-- I_1 == p3*p5-p4*p6+p2*p7+p3*p7+p4*p7+p5*p7+p6*p7+p7^2+p2*p8+p7*p8-p7
-- I_2 == p2*p3+p2*p4+p3*p4+p4^2+p4*p5+p4*p6+p4*p7-p6*p7+p4*p8+p5*p8-p4
```

Figure A.4: Code for computing the CFN invariants of the 4-leaf tree.

A.3 Trivial 4-leaf reduced topologies

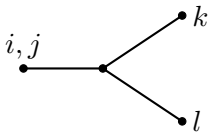
1. $\theta_i = \theta_j = \theta_k = 1$ and $\theta_l, \theta_{\{i,j\}} \in (0,1)$

$$\Sigma_{\mathrm{R}}(T) = \{\{ij|kl, l|ijk\}\}$$



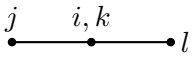
2. $\theta_i = \theta_j = 1$ and $\theta_k, \theta_l \in (0,1)$

$$\Sigma_{\mathrm{R}}(T) = \{\{ij|kl, k|ijl, l|ijk\}\}$$



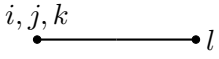
3. $\theta_j, \theta_l \in (0,1)$ for some distinct pair $j, l \in [4]$ and $\theta_e = 1$ for all $e \neq j, l$

$$\Sigma_{\mathrm{R}}(T) = \{\{j|ikl, l|ijk\}\}$$



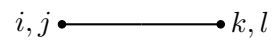
4. $\theta_i = \theta_j = \theta_{\{1,2\}} = \theta_k = 1$ and $\theta_l \in (0,1)$

$$\Sigma_{\mathrm{R}}(T) = \{\{ijk|l\}\}$$



5. $\theta_i = 1$ for all $i \in [4]$, $\theta_{\{i,j\}} \in (0,1)$.

$$\Sigma_{\mathrm{R}}(T) = \{\{ij|kl\}\}$$



6. Point: $\theta_e = 1$ for all $e \in E(T)$

$$\Sigma_{\mathrm{R}}(T) = \{\emptyset\}$$

A.4 Exact Solutions for 4-Leaf Boundary Cases

Lemma 55 (MLE for R_6)**.** *Suppose* $\Sigma_{\mathrm{R}}(T) \in R_6$*. Then*

$$\Sigma_{\mathrm{R}}(T) = [\{j|kl, k|jl, l|jk\}, \emptyset]$$

for some choice of pairwise disjoint $j, k, l \in [4]$*. In this setting, the log-likelihood is a function of the three edge parameters* θ_j, θ_k*, and* θ_l*, corresponding to the branches with leaves* j, k *and* l *respectively. Assume the data satisfies* ***A.4****, so that* $B_{jk}, B_{jl}, B_{kl} \neq 0$*. If* $(B_{ji}, B_{jl}, B_{kl}) \in \mathcal{D}$*, then values of* $\theta_i, \theta_j, \theta_j$ *which maximize* ℓ *on* $(0,1)^3$ *are*

$$(\theta_j, \theta_k, \theta_l) = \left(\sqrt{\frac{B_{jk}B_{jl}}{B_{kl}}}, \sqrt{\frac{B_{jk}B_{kl}}{B_{jl}}}, \sqrt{\frac{B_{jl}B_{kl}}{B_{jk}}}\right). \tag{A.29}$$

On the other hand, if $(B_{ji}, B_{jl}, B_{kl}) \notin \mathcal{D}$ *the* ℓ *has no maximizer in* $(0,1)^3$*.*

Proof. For each $\tilde{\sigma} \in \{-1, +1\}^3$, define

$$J_{\tilde{\sigma}} := \left\{\sigma \in \{-1, +1\}^4 : \sigma_\lambda = \tilde{\sigma}_\lambda \text{ for all } \lambda \in \{j, k, l\}\right\}$$

and let

$$s_{\tilde{\sigma}} := \sum_{\sigma \in J_{\tilde{\sigma}}} s_\sigma.$$

Then

$$\begin{aligned}\ell(\theta_j, \theta_k, \theta_l) &= \sum_{\sigma \in \{-1,+1\}^4} s_\sigma \log \mathbb{P}\left[(X_j, X_k, X_l) = (\sigma_j, \sigma_k, \sigma_l)\right] \\ &= \sum_{\tilde{\sigma} \in \{-1,+1\}^3} s_{\tilde{\sigma}} \log \mathbb{P}\left[(X_j, X_k, X_l) = (\tilde{\sigma}_j, \tilde{\sigma}_k, \tilde{\sigma}_l)\right]\end{aligned} \tag{A.30}$$

This shows that, up to a relabeling of the leaves, ℓ has the same form as the log-likelihood in the unrooted 3-leaf case (c.f., Eq. (5.10)). Therefore maximizing $\ell(\theta_j, \theta_k, \theta_l)$ over all $\theta_j, \theta_k, \theta_l \in (0, 1)$ is the same problem as that considered in Lemma 42 (maximizing the log-likelihood over all 3-leaf unrooted trees with branch lengths in the interior of the unit cube). Relabeling the leaves in the statement of Lemma 42 from $1, 2$ and 3 to i, j and k, it follows that if $(B_{ji}, B_{jl}, B_{kl}) \in \mathcal{D}$, then $\ell(\theta_j, \theta_k, \theta_l)$ is maximized on $(0, 1)^3$ at the point

$$(\theta_j, \theta_k, \theta_l) = \left(\sqrt{\frac{B_{jk}B_{jl}}{B_{kl}}}, \sqrt{\frac{B_{jk}B_{kl}}{B_{jl}}}, \sqrt{\frac{B_{jl}B_{kl}}{B_{jk}}}\right),$$

and, on the other hand, if $B_{jk}, B_{jl}, B_{kl} \notin \mathcal{D}$ then no local maximum exists. □

Lemma 56 (MLE for R_3)**.** *Suppose* $\Sigma_{\mathrm{R}}(T) \in R_3$*. Then*

$$\Sigma_{\mathrm{R}}(T) = [\{ij|kl, i|jkl, k|ijl, l|ijk\}]$$

for some choice of pairwise disjoint $i, j, k, l \in [4]$*. Let* $\theta_i, \theta_k, \theta_l$ *denote the edge parameters of leaves* i, k, l*, and let* θ_5 *denote the internal branch parameter.*

Assume $B_{jk}, B_{jl}, B_{kl} \neq 0$*. If* $B_{ij} > 0$*, and* $(B_{jk}, B_{jl}, B_{kl}) \in \mathcal{D}$*, then* ℓ *has a unique maximizer on* $(0, 1)^4$ *given by*

$$(\theta_i, \theta_j, \theta_k, \theta_l) = \left(\frac{B_{ij}}{N}, \sqrt{\frac{B_{jk}B_{jl}}{B_{kl}}}, \sqrt{\frac{B_{jk}B_{kl}}{B_{jl}}}, \sqrt{\frac{B_{jl}B_{kl}}{B_{jk}}}\right). \tag{A.31}$$

If $B_{ij} \leq 0$ *or* $(B_{ji}, B_{jl}, B_{kl}) \notin \mathcal{D}$ *the* ℓ *has no maximizer on* $(0, 1)^4$*.*

Proof. Let $\sigma \in \{-1,+1\}^4$. Then since $X_i \perp\!\!\!\perp (X_k, X_l)$ given X_j,

$$\begin{aligned}
\mathbb{P}[X=\sigma] &= \mathbb{P}[X_i=\sigma_i \mid X_j=\sigma_j]\,\mathbb{P}[(X_k,X_l)=(\sigma_k,\sigma_l) \mid X_j=\sigma_j]\,\mathbb{P}[X_j=\sigma_j] \\
&= \mathbb{P}[X_i=\sigma_i \mid X_j=\sigma_j]\,\mathbb{P}[(X_j,X_k,X_l)=(\sigma_j,\sigma_k,\sigma_l)] \\
&= \frac{1}{2}(1+\sigma_i\sigma_j\theta_i)\,\mathbb{P}[(X_j,X_k,X_l)=(\sigma_j,\sigma_k,\sigma_l)],
\end{aligned}$$

where the last step is justified by Eq. (5.11). Therefore,

$$\begin{aligned}
\ell(\theta) &= \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log \mathbb{P}[X=\sigma] \\
&= \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log\left(\frac{1}{2}(1+\sigma_i\sigma_j\theta_i)\right) \\
&\quad + \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log \mathbb{P}[(X_j,X_k,X_l)=(\sigma_j,\sigma_k,\sigma_l)].
\end{aligned}$$

The two sums on the right-hand side, which we will term ℓ_1 and ℓ_2, can be maximized individually because the first sum ℓ_1 depends only on the parameter θ_i, while second sum ℓ_2 depends only on θ_j, θ_k and θ_l.

To maximize ℓ_1, first observe that

$$\begin{aligned}
\ell_1(\theta_i) &= \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log\left(\frac{1}{2}(1+\sigma_i\sigma_j\theta_i)\right) \\
&= M_{ij}^{+}\log(1+\theta_i) + M_{ij}^{-}\log(1-\theta_i) - N\log 2
\end{aligned}$$

Therefore

$$\ell_1'(\theta_i) = \frac{M_{ij}^{+}}{1+\theta_i} - \frac{M_{ij}^{-}}{1-\theta_i}.$$

Solving the equation $\ell_1'(\theta_i)=0$, we obtain at most one critical point, which is

$$\theta_i = B_{ij} \tag{A.32}$$

provided that this point lies in the open unit interval. Since

$$\ell_1''(\theta_i) = -\left(\frac{M_{ij}^+}{(1+\theta_i)^2} + \frac{M_{ij}^-}{(1-\theta_i)^2}\right) < 0$$

for all $\theta_i \neq \pm 1$, it follows that Eq. (A.32) gives the unique maximizer of ℓ_1 on $(0,1)$.

Next, since ℓ_2 has the form Eq. (A.30), we can apply Lemma 56 to maximize ℓ_2 on $(0,1)^3$. Combining the results of Lemma 55 this with the above result for maximizing ℓ_1 implies the statement of the lemma. □

Lemma 57 (MLE for R_4)**.** *Suppose* $\Sigma_{\mathrm{R}}(T) \in R_4$*. The numerical parameters which maximize the likelihood are*

$$\theta_j = B_{ij}, \quad \theta_k = B_{ik}, \quad \textit{and} \quad \theta_l = B_{il}$$

provided that all of these quantities lie in $(0,1)$*.*

Proof. Using Eq. (5.11) and the conditional independence of X_j, X_k and X_l given X_i,

$$\begin{aligned}\mathbb{P}[X=\sigma] &= \mathbb{P}[X_i=\sigma_i] \prod_{\lambda\in[4]\setminus\{i\}} \mathbb{P}[X_\lambda = \sigma_\lambda \mid X_i = \sigma_i] \\ &= \frac{1}{16} \prod_{\lambda\in[4]\setminus\{i\}} (1+\sigma_\lambda\sigma_i\theta_\lambda)\end{aligned}$$

(Alternatively: take $\theta_5 = \theta_i = 1$ in Eq. (5.15) and factor the result.) Therefore

$$\begin{aligned}\ell(\theta) &= \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log \mathbb{P}[X=\sigma] \\ &= \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log\left(\frac{1}{16}\prod_{\lambda\in[4]\setminus\{i\}} (1+\sigma_\lambda\sigma_i\theta_\lambda)\right).\end{aligned}$$

Therefore since $\sum_\sigma u_\sigma = N$, interchanging summations gives

$$\begin{aligned}\ell(\theta) &= -N\log(16) + \sum_{\lambda\in[4]\setminus\{i\}} \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log\left(1+\sigma_\lambda\sigma_i\theta_\lambda\right) \\ &= -N\log(16) + \sum_{\lambda\in[4]\setminus\{i\}} \left[M^+_{\lambda,i}\log\left(1+\theta_\lambda\right) + M^-_{\lambda,i}\log\left(1-\theta_\lambda\right)\right]\end{aligned}$$

Therefore

$$\frac{\partial\ell}{\partial\theta_\lambda} = \frac{M^+_{\lambda,i}}{1+\theta_\lambda} - \frac{M^-_{\lambda,i}}{1-\theta_\lambda}$$

Solving $\frac{\partial\ell}{\partial\theta_\lambda}(\theta) = 0$, we find that ℓ has at most one critical point in $(0,1)^3$, which is at at

$$\theta_\lambda = \frac{M^+_{\lambda,i} - M^-_{\lambda,i}}{N} = B_{\lambda i},$$

for each $\lambda \in \{j,k,l\}$, provided that each of these is in $(0,1)$. By checking the second derivatives, we see this point is indeed a local maximum.

□

Lemma 58 (MLE for R_5)**.** *Suppose* $\Sigma_{\mathrm{R}}(T) \in R_5$*. Then*

$$\Sigma_{\mathrm{R}}(T) = [\{ij|kl, j|ikl, l|ijk\}]$$

for some choice of pairwise distinct $i,j,k,l \in [4]$*. If*

$$(B_{12}, B_{13}, B_{23}) \in (0,1)^3 \tag{A.33}$$

then

$$(\theta_i, \theta_l, \theta_5) = (B_{ij}, B_{kl}, B_{ik})$$

is the unique maximizer of $\ell(\theta_i,\theta_j,\theta_5)$ *on* $(0,1)^3$*. If Eq.* (A.33) *does not hold, then* $\ell(\theta_i,\theta_j,\theta_5)$ *has no maxima on* $(0,1)^3$*.*

Proof. Using conditional independence properties and Eq. (5.11),

$$\begin{aligned}\mathbb{P}[X=\sigma] &= \mathbb{P}[X_l=\sigma_l \mid X_k=\sigma_k]\,\mathbb{P}[X_k=\sigma_k \mid X_i=\sigma_i]\,\mathbb{P}[X_i=\sigma_i \mid X_j=\sigma_j]\,\mathbb{P}[X_j=\sigma_j] \\ &= \frac{1}{16}\left(1+\sigma_k\sigma_l\theta_l\right)\left(1+\sigma_i\sigma_k\theta_5\right)\left(1+\sigma_i\sigma_j\theta_j\right).\end{aligned}$$

Therefore

$$\begin{aligned}\ell(\theta_j,\theta_l,\theta_5) &= \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log \mathbb{P}[X=\sigma] \\ &= -N\log(16) + \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log\left(1+\sigma_k\sigma_l\theta_l\right) \\ &\quad + \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log\left(1+\sigma_i\sigma_k\theta_5\right) + \sum_{\sigma\in\{-1,+1\}^4} u_\sigma \log\left(1+\sigma_i\sigma_j\theta_j\right) \\ &= -N\log(16) + M_{kl}^{+}\log(1+\theta_l) + M_{kl}^{-}\log(1-\theta_l) + M_{ik}^{+}\log(1+\theta_5) \\ &\quad + M_{ik}^{-}\log(1-\theta_5) + M_{ij}^{+}\log(1+\theta_j) + M_{ij}^{-}\log(1-\theta_j)\end{aligned}$$

Therefore

$$\frac{\partial\ell}{\partial\theta_j} = \frac{M_{ij}^{+}}{1+\theta_j} - \frac{M_{ij}^{-}}{1-\theta_j}$$

$$\frac{\partial\ell}{\partial\theta_l} = \frac{M_{kl}^{+}}{1+\theta_l} - \frac{M_{kl}^{-}}{1-\theta_l}$$

and

$$\frac{\partial\ell}{\partial\theta_5} = \frac{M_{ik}^{+}}{1+\theta_5} - \frac{M_{ik}^{-}}{1-\theta_5}$$

This has solution

$$(\theta_i,\theta_l,\theta_5) = (B_{ij},B_{kl},B_{ik})$$

and this critical point, if it is in $(0,1)^3$, is the unique maximum of ℓ; this is because the Hessian of ℓ, being a diagonal matrix with negative entries on the diagonal, is negative definite. □

Lemma 59 (MLE for R_7)**.** *Suppose* $\Sigma_{\mathrm{R}}(T) \in R_7$*. Then*

$$\Sigma_{\mathrm{R}}(T) = [\{ij|l, i|jl\}, \emptyset]$$

for some choice of pairwise distinct $i, j, k, l \in [4]$*. Let* θ_i *and* θ_l *denote the Hadamard edge parameters of leaves* i *and* l *respectively. If*

$$B_{j\lambda} \in (0,1), \quad \lambda \in \{i, j\} \tag{A.34}$$

Then $(\theta_i, \theta_j) = (B_{ji}, B_{jl})$ *is the unique global maximum of* $\ell(\theta_i, \theta_l)$ *on* $(0,1)^2$*. If Eq.* (A.34) *does not hold, then* $\ell(\theta_j, \theta_l)$ *has no maximum on* $(0,1)^2$*.*

Proof. Since $X_k \perp\!\!\!\perp (X_i, X_j, X_l)$, and since $X_i \perp\!\!\!\perp X_l$ given X_j,

$$\begin{aligned}
\mathbb{P}[X = \sigma] &= \mathbb{P}[X_k = \sigma_k]\, \mathbb{P}[(X_i, X_j, X_l) = (\sigma_i, \sigma_j, \sigma_l)] \\
&= \frac{1}{2}\mathbb{P}[(X_i = \sigma_i \mid X_j = \sigma_j]\, \mathbb{P}[(X_l = \sigma_l \mid X_j = \sigma_j]\, \mathbb{P}[X_j = \sigma_j] \\
&= \frac{1}{16}(1 + \sigma_j\sigma_i\theta_i)(1 + \sigma_j\sigma_l\theta_l)
\end{aligned}$$

where the final step is justified by Eq. (5.11).

Therefore

$$\begin{aligned}
\ell(\theta_i, \theta_j) &= -N\log(16) + \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log(1 + \sigma_j\sigma_i\theta_i) + \sum_{\sigma\in\{-1,+1\}^4} s_\sigma \log(1 + \sigma_j\sigma_l\theta_l) \\
&= -N\log(16) + M^+_{ji}\log(1+\theta_i) + M^-_{ji}\log(1-\theta_i) \\
&\quad + M^+_{jl}\log(1+\theta_l) + M^-_{jl}\log(1-\theta_l)
\end{aligned}$$

Therefore

$$\frac{\partial \ell}{\partial \theta_\lambda} = \frac{M^+_{j\lambda}}{1+\theta_\lambda} - \frac{M^-_{j\lambda}}{1-\theta_\lambda}, \quad \lambda \in \{i, l\}$$

This system has a unique solution

$$\theta_\lambda = \frac{M^+_{j\lambda} - M^-_{j\lambda}}{N} = B_{j\lambda}, \quad \lambda \in \{i, j\}\,. \tag{A.35}$$

Since the Hessian of ℓ is negative definite for all $\theta_i, \theta_j \in (0,1)$, ℓ is concave on $(0,1)^2$ and hence if the point defined in Eq. (A.35) is in $(0,1)^2$, then it is the unique global maximum of ℓ on $(0,1)^2$; otherwise ℓ has no maximum on $(0,1)^2$. □

Lemma 60 (MLE for R_8)**.** *Suppose $\Sigma_{\mathrm{R}}(T) \in R_9$. Then there exists a choice of pairwise distinct $i, j, k, l \in [4]$ such that*

$$\Sigma_{\mathrm{R}}(T) = [\{i|j, l|k\}]$$

Let $\theta_{ij}, \theta_{kl} \in (0,1)$ be the Hadamard branch lengths of $G(T)$. Here ℓ is a function of θ_{ij} and θ_{kl}. If

$$B_{ij}, B_{kl} \in (0,1) \tag{A.36}$$

Then $(\theta_{ij}, \theta_{kl}) = (B_{ij}, B_{kl})$ is the unique maximum of ℓ on $(0,1)^2$. If Eq. (A.36) do not both hold, then ℓ has no maximum on $(0,1)^2$.

Proof. Let $\sigma \in \{-1,+1\}^4$. Since $(X_i, X_j) \perp\!\!\!\perp (X_j, X_l)$,

$$\begin{aligned}\mathbb{P}[X = \sigma] &= \mathbb{P}[X_i = \sigma_i, X_j = \sigma_j]\,\mathbb{P}[X_k = \sigma_k, X_l = \sigma_l] \\ &= \frac{1}{4}\mathbb{P}[X_i = \sigma_i \mid X_j = \sigma_j]\,\mathbb{P}[X_k = \sigma_k \mid X_l = \sigma_l]\end{aligned}$$

Therefore

$$\begin{aligned}\ell(\theta_{ij}, \theta_{kl}) &= \sum_{\sigma \in \{-1,+1\}^4} s_\sigma \log \mathbb{P}[X = \sigma] \\ &= -N \log(16) + \sum_{\sigma \in \{-1,+1\}^4} s_\sigma \log(1 + \sigma_i\sigma_j\theta_{ij}) + \sum_{\sigma \in \{-1,+1\}^4} s_\sigma \log(1 + \sigma_k\theta_l\theta_{kl}) \\ &- N \log(16) + M^+_{ij} \log(1 + \theta_{ij}) + M^-_{ij} \log(1 - \theta_{ij}) + M^+_{kl} \log(1 + \theta_{kl}) \\ &\quad + M^-_{kl} \log(1 - \theta_{kl})\,.\end{aligned}$$

Therefore

$$\frac{\partial \ell}{\partial \theta_{ij}} = \frac{M_{ij}^{+}}{1+\theta_{ij}} - \frac{M_{ij}^{-}}{1-\theta_{ij}}$$

and

$$\frac{\partial \ell}{\partial \theta_{kl}} = \frac{M_{kl}^{+}}{1+\theta_{kl}} - \frac{M_{kl}^{-}}{1-\theta_{kl}}$$

Setting this system equal to zero and solving for $(\theta_{ij}, \theta_{kl})$, it follows that ℓ as at most one critical point in $(0,1)^2$, when $(\theta_{ij}, \theta_{kl}) = (B_{ij}, B_{kl}) \in (0,1)^2$. The Hessian of ℓ is

$$H_\ell = \begin{bmatrix} -\left(\frac{M_{ij}^{+}}{(1+\theta_{ij})^2} + \frac{M_{ij}^{-}}{(1-\theta_{ij})^2}\right) & 0 \\ 0 & -\left(\frac{M_{kl}^{+}}{(1+\theta_{kl})^2} + \frac{M_{kl}^{-}}{(1-\theta_{kl})^2}\right) \end{bmatrix}$$

which is negative definite, and hence this critical point is a maximum. This prove the statement of the lemma. □

Lemma 61 (MLE for R_9)**.** *Suppose $\Sigma_{\mathrm{R}}(T) \in R_9$. Then there exists a choice of pairwise distinct $i, j, k, l \in [4]$ such that*

$$\Sigma_{\mathrm{R}}(T) = [\{k|l\}, \emptyset, \emptyset].$$

Let $\theta_{kl} \in (0,1)$ be the Hadamard parameter of the edge with endpoints k and l. If

$$B_{kl} \in (0,1) \tag{A.37}$$

Then $\ell = \ell(\theta_{kl})$ is maximized on $(0,1)$ when

$$\theta_{kl} = B_{kl}$$

If Eq. (A.37) *does not hold, then $\ell(\theta_{kl})$ has no maximum on $(0,1)$.*

Proof. Similar to the proof of Lemma 60. □

Bibliography

[1] Joseph Felsenstein. "Cases in which parsimony or compatibility methods will be positively misleading". In: *Systematic zoology* 27.4 (1978), pp. 401–410.

[2] Junhyong Kim. "General Inconsistency Conditions for Maximum Parsimony: Effects of Branch Lengths and Increasing Numbers of Taxa". In: *Systematic Biology* 45.3 (Sept. 1996), pp. 363–374. ISSN: 1063-5157. DOI: 10.1093/sysbio/45.3.363. eprint: https://academic.oup.com/sysbio/article-pdf/45/3/363/19501755/45-3-363.pdf. URL: https://doi.org/10.1093/sysbio/45.3.363.

[3] Sarah L Parks and Nick Goldman. "Maximum likelihood inference of small trees in the presence of long branches". In: *Systematic Biology* 63.5 (2014), pp. 798–811.

[4] Sebastien Roch and Mike Steel. "Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent". In: *Theoretical Population Biology* 100 (2015), pp. 56–62. ISSN: 0040-5809. DOI: http://dx.doi.org/10.1016/j.tpb.2014.12.005. URL: http://www.sciencedirect.com/science/article/pii/S0040580914001075.

[5] Laura Salter Kubatko and James H Degnan. "Inconsistency of phylogenetic estimates from concatenated data under coalescence". In: *Systematic biology* 56.1 (2007), pp. 17–24. DOI: 10.1080/10635150601146041.

[6] James H Degnan and Noah A Rosenberg. "Discordance of species trees with their most likely gene trees". In: *PLoS genet* 2.5 (2006), e68. DOI: 10.1371/journal.pgen.0020068.

[7] James H Degnan. "Anomalous unrooted gene trees". In: *Systematic biology* 62.4 (2013), pp. 574–590. DOI: 10.1093/sysbio/syt023.

[8] James H. Degnan and Noah A. Rosenberg. "Gene tree discordance, phylogenetic inference and the multispecies coalescent". In: *Trends in Ecology & Evolution* 24.6 (2009), pp. 332–340. ISSN: 0169-5347. DOI: 10.1016/j.tree.2009.01.009.

[9] Claudia Solís-Lemus, Mengyao Yang, and Cécile Ané. "Inconsistency of species tree methods under gene flow". In: *Systematic biology* 65.5 (2016), pp. 843–851. DOI: 10.1093/sysbio/syw030.

[10] Colby Long and Laura Kubatko. "The effect of gene flow on coalescent-based species-tree inference". In: *Systematic biology* 67.5 (2018), pp. 770–785.

[11] James H Degnan. "Modeling Hybridization Under the Network Multispecies Coalescent". In: *Systematic Biology* 67.5 (May 2018), pp. 786–799. ISSN: 1063-5157. DOI: 10.1093/sysbio/syy040. eprint: https://academic.oup.com/sysbio/article-pdf/67/5/786/25517429/syy040.pdf. URL: https://doi.org/10.1093/sysbio/syy040.

[12] Xiyun Jiao et al. "The impact of cross-species gene flow on species tree estimation". In: *Systematic Biology* 69.5 (2020), pp. 830–847.

[13] Conrad L Schoch et al. "NCBI Taxonomy: a comprehensive update on curation, resources and tools". In: *Database* 2020 (2020), baaa062. URL: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=10320&lvl=3&keep=1&srchmode=1&unlock&mod=1&log_op=modifier_toggle.

[14] Tae-Kun Seo, Benjamin D Redelings, and Jeffrey L Thorne. "Correlations between alignment gaps and nucleotide substitution or amino acid replacement". In: *Proceedings of the National Academy of Sciences* 119.34 (2022), e2204435119.

[15] Tanya Golubchik et al. "Mind the gaps: evidence of bias in estimates of multiple sequence alignments". In: *Molecular biology and evolution* 24.11 (2007), pp. 2433–2442.

[16] Lam-Tung Nguyen et al. "IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies". In: *Molecular biology and evolution* 32.1 (2015), pp. 268–274.

[17] Silvina Soledad Maidana et al. "Evidence of natural interspecific recombinant viruses between bovine alphaherpesviruses 1 and 5". In: *Virus research* 242 (2017), pp. 122–130.

[18] Claudia Solís-Lemus, Paul Bastide, and Cécile Ané. "PhyloNetworks: A Package for Phylogenetic Networks". In: *Molecular Biology and Evolution* 34.12 (Sept. 2017), pp. 3292–3298. ISSN: 0737-4038. DOI: 10.1093/molbev/msx235. eprint: https://academic.oup.com/mbe/article-pdf/34/12/3292/21946966/msx235.pdf. URL: https://doi.org/10.1093/molbev/msx235.

[19] David A Baum and StaceyD Smith. "Tree thinking". In: *An Introduction to Phylogenetic Biology. Roberts and Company Publishers* (2013).

[20] Marc HV Van Regenmortel. "Virus species, a much overlooked but essential concept in virus classification". In: *Intervirology* 31.5 (1990), pp. 241–254.

[21] Matthew William Hahn. *Molecular population genetics*. Oxford University Press, 2018.

[22] Magnus Nordborg. *Coalescent Theory*. Ed. by David J Balding, Ida Moltke, and John Marioni. John Wiley & Sons, 2019.

[23] John Frank Charles Kingman. "The coalescent". In: *Stochastic processes and their applications* 13.3 (1982), pp. 235–248. DOI: 10.1016/0304-4149(82)90011-4.

[24] Bruce Rannala et al. "The Multi-species Coalescent Model and Species Tree Inference". In: *Phylogenetics in the Genomic Era*. Ed. by C. Scornavacca, F. Delsuc, and N. Galtier. No commercial publisher — Authors open access book. This book is freely available online at https://hal.inria.fr/PGE, 2020. Chap. 3.3, 3.3:1–3.3:21.

[25] Bruce Rannala and Ziheng Yang. "Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci". In: *Genetics* 164.4 (2003), pp. 1645–1656. DOI: 10.1093/genetics/164.4.1645.

[26] Wayne Maddison. "Gene Trees in Species Trees". In: *Systematic Biology* 46.3 (1997), pp. 523–536. DOI: 10.1093/sysbio/46.3.523.

[27] Iker Rivas-González et al. "Pervasive incomplete lineage sorting illuminates speciation and selection in primates". In: *Science* 380.6648 (2023), eabn4409. DOI: 10.1126/science.abn4409. eprint: https://www.science.org/doi/pdf/10.1126/science.abn4409. URL: https://www.science.org/doi/abs/10.1126/science.abn4409.

[28] Cecile Ané. "Reconstructing concordance trees and testing the coalescent model from genome-wide data sets". In: *In Estimating species trees: Practical and theoretical aspects*. 2010.

[29] David Bryant and Matthew W Hahn. "The concatenation question". In: *Phylogenetics in the Genomic Era*. Ed. by C. Scornavacca, F. Delsuc, and N. Galtier. No commercial publisher— Authors open access book. This book is freely available online at https://hal.inria.fr/PGE, 2020. Chap. 3.4, 3.4:1–3.4:23.

[30] Paschalia Kapli, Ziheng Yang, and Maximilian J Telford. "Phylogenetic tree building in the genomic age". In: *Nature Reviews Genetics* 21.7 (2020), pp. 428–444. DOI: 10.1038/s41576-020-0233-0.

[31] Siavash Mirarab et al. "ASTRAL: genome-scale coalescent-based species tree estimation". In: *Bioinformatics* 30.17 (2014), pp. i541–i548. DOI: 10.1093/bioinformatics/btu462.

[32] Scott V. Edwards et al. "Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics". In: *Molecular Phylogenetics and Evolution* 94 (2016), pp. 447–462. ISSN: 1055-7903. DOI: 10.1016/j.ympev.2015.10.027.

[33] Hayley C Lanier and L Lacey Knowles. "Is recombination a problem for species-tree analyses?" In: *Systematic Biology* 61.4 (2012), pp. 691–701. DOI: 10.1093/sysbio/syr128.

[34] Mark S Springer and John Gatesy. "Delimiting coalescence genes (c-genes) in phylogenomic data sets". In: *Genes* 9.3 (2018), p. 123. DOI: 10.3390/genes9030123.

[35] Mark S. Springer and John Gatesy. "The gene tree delusion". In: *Molecular Phylogenetics and Evolution* 94 (2016), pp. 1–33. ISSN: 1055-7903. DOI: 10.1016/j.ympev.2015.07.018.

[36] Tianqi Zhu and Ziheng Yang. "Complexity of the simplest species tree problem". In: *Molecular Biology and Evolution* 38.9 (Jan. 2021), pp. 3993–4009. ISSN: 0737-4038. DOI: 10.1093/molbev/msab009. eprint: https://academic.oup.com/mbe/article-pdf/38/9/3993/39882887/msab009.pdf.

[37] C. Semple and M. Steel. *Phylogenetics*. Vol. 22. Mathematics and its Applications series. Oxford University Press, 2003, p. 250.

[38] Tal Pupko and Itay Mayrose. *A gentle introduction to probabilistic evolutionary models*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. 2020. URL: https://hal.archives-ouvertes.fr/hal-02535622.

[39] James A Cavender. "Taxonomy with confidence". In: *Mathematical biosciences* 40.3-4 (1978), pp. 271–280.

[40] James S Farris. "A probability model for inferring evolutionary trees". In: *Systematic Biology* 22.3 (1973), pp. 250–256.

[41] Jerzy Neyman. "Molecular studies of evolution: a source of novel statistical problems". In: *Statistical decision theory and related topics*. Elsevier, 1971, pp. 1–27.

[42] Thomas H Jukes and Charles R Cantor. "Evolution of protein molecules". In: *Mammalian protein metabolism* 3 (1969), pp. 21–132. DOI: 10.1016/B978-1-4832-3211-9.50009-7.

[43] Miguel Arenas. "Trends in substitution models of molecular evolution". In: *Frontiers in genetics* 6 (2015), p. 319.

[44] Leo Breiman. *Probability*. SIAM, 1992.

[45] Seth Sullivant. *Algebraic statistics*. Vol. 194. American Mathematical Soc., 2018.

[46] Elizabeth S Allman and John A Rhodes. "Phylogenetic invariants for the general Markov model of sequence mutation". In: *Mathematical biosciences* 186.2 (2003), pp. 113–144.

[47] Bernd Sturmfels and Seth Sullivant. "Toric ideals of phylogenetic invariants". In: *Journal of Computational Biology* 12.4 (2005), pp. 457–481.

[48] Elizabeth S Allman et al. "Identifiability of two-tree mixtures for group-based models". In: *IEEE/ACM transactions on computational biology and bioinformatics* 8.3 (2010), pp. 710–722. URL: %5Curl%7Bhttps://dl.acm.org/doi/pdf/10.1109/TCBB.2010.79%7D.

[49] Elizabeth S Allman and John A Rhodes. "The identifiability of tree topology for phylogenetic models, including covarion and mixture models". In: *Journal of Computational Biology* 13.5 (2006), pp. 1101–1113. URL: %5Curl%7Bhttps://pubmed.ncbi.nlm.nih.gov/16796553/%7D.

[50] Colby Long and Seth Sullivant. "Identifiability of 3-class Jukes–Cantor mixtures". In: *Advances in Applied Mathematics* 64 (2015), pp. 89–110. URL: %5Curl%7Bhttps://arxiv.org/pdf/1406.7256v2.pdf%7D.

[51] John A Rhodes and Seth Sullivant. "Identifiability of large phylogenetic mixture models". In: *Bulletin of mathematical biology* 74.1 (2012), pp. 212–231. URL: %5Curl%7Bhttps://jarhodesuaf.github.io/papers/LargeMixtures_BMB2012.pdf%7D.

[52] Elizabeth Gross and Colby Long. "Distinguishing phylogenetic networks". In: *SIAM Journal on Applied Algebra and Geometry* 2.1 (2018), pp. 72–93.

[53] Joseph Cummings, Benjamin Hollering, and Christopher Manon. *Invariants for level-1 phylogenetic networks under the Cavendar-Farris-Neyman Model*. 2021. DOI: 10.48550/ARXIV.2102.03431. URL: https://arxiv.org/abs/2102.03431.

[54] Elizabeth S Allman, Hector Baños, and John A Rhodes. "Identifiability of species network topologies from genomic sequences using the logDet distance". In: *Journal of Mathematical Biology* 84.5 (2022), pp. 1–38.

[55] Robert C Griffiths and Paul Marjoram. "An ancestral recombination graph". In: *Progress in Population Genetics and Human Evolution*. Ed. by P. Donnelly and S Tavaré. Vol. 87. 1997, p. 257.

[56] Tandy Warnow. *Computational phylogenetics: an introduction to designing methods for phylogeny estimation*. Cambridge University Press, 2017.

[57] James H Degnan et al. "Properties of consensus methods for inferring species trees from gene trees". In: *Systematic Biology* 58.1 (2009), pp. 35–54. DOI: 10.1093/sysbio/syp008.

[58] Jianzhi Zhang. "Evolution by gene duplication: an update". In: *Trends in ecology & evolution* 18.6 (2003), pp. 292–298.

[59] Peng Du and Luay Nakhleh. "Species tree and reconciliation estimation under a duplication-loss-coalescence model". In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2018, pp. 376–385.

[60] M. D. Rasmussen and M. Kellis. "Unified modeling of gene duplication, loss, and coalescence using a locus tree". In: *Genome Research* 22.4 (2012), pp. 755–765. DOI: 10.1101/gr.123901.111.

[61] Ricard Albalat and Cristian Cañestro. "Evolution by gene loss". In: *Nature Reviews Genetics* 17.7 (2016), pp. 379–391.

[62] Eugene V Koonin. "Orthologs, paralogs, and evolutionary genomics". In: *Annu. Rev. Genet.* 39 (2005), pp. 309–338.

[63] Zhi Yan et al. "Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs". In: *Systematic Biology* 71.2 (2022), pp. 367–381.

[64] Dominik Schrempf and Gergely Szöllösi. "The Sources of Phylogenetic Conflicts". In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher — Authors open access book, 2020, 3.1:1–3.1:23. URL: https://hal.archives-ouvertes.fr/hal-02535482.

[65] Peng Du, Matthew W Hahn, and Luay Nakhleh. "Species Tree Inference under the Multispecies Coalescent on Data with Paralogs is Accurate". In: *bioRxiv* (2019). DOI: 10.1101/498378.

[66] Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. "Multi-allele species reconstruction using ASTRAL". In: *Molecular Phylogenetics and Evolution* 130 (2019), pp. 286–296. DOI: 10.1016/j.ympev.2018.10.033.

[67] Brandon Legried et al. "Polynomial-Time Statistical Estimation of Species Trees under Gene Duplication and Loss". In: *bioRxiv* (2019). DOI: 10.1101/821439. eprint: https://www.biorxiv.org/content/early/2019/10/29/821439.full.pdf. URL: https://www.biorxiv.org/content/early/2019/10/29/821439.

[68] Ziheng Yang. "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods". In: *Journal of Molecular evolution* 39 (1994), pp. 306–314.

[69] Joseph Felsenstein. "Inferring phylogenies". In: *Inferring phylogenies*. 2004, pp. 664–664.

[70] Elizabeth S Allman, Cécile Ané, and John A Rhodes. "Identifiability of a Markovian model of molecular evolution with gamma-distributed rates". In: *Advances in Applied Probability* 40.1 (2008), pp. 229–249.

[71] John Wakeley. "Substitution-rate variation among sites and the estimation of transition bias." In: *Molecular Biology and Evolution* 11.3 (1994), pp. 436–442.

[72] Tal Pupko and Itay Mayrose. "Probabalistic Methods and Rate Heterogeneity". In: *Elements of Computational Systems Biology*. Wiley, 2010.

[73] Elchanan Mossel and Sebastien Roch. "Distance-based species tree estimation under the coalescent: information-theoretic trade-off between number of loci and sequence length". In: *The Annals of Applied Probability* 27.5 (2017), pp. 2926–2955.

[74] Johannes Bergsten. "A review of long-branch attraction". In: *Cladistics* 21.2 (2005), pp. 163–193.

[75] Sebastien Roch, Michael Nute, and Tandy Warnow. "Long-Branch Attraction in Species Tree Estimation: Inconsistency of Partitioned Likelihood and Topology-Based Summary Methods". In: *Systematic Biology* 68.2 (Sept. 2018), pp. 281–297. ISSN: 1063-5157. DOI: 10.1093/sysbio/syy061. eprint: https://academic.oup.com/sysbio/article-pdf/68/2/281/27739135/syy061.pdf. URL: https://doi.org/10.1093/sysbio/syy061.

[76] Dominik Schrempf and Gergely Szöllősi. "The sources of phylogenetic conflicts". In: *Phylogenetics in the Genomic Era*. Ed. by C. Scornavacca, F. Delsuc, and N. Galtier. No commercial publisher— Authors open access book. This book is freely available online at https://hal.inria.fr/PGE, 2020. Chap. 3.1, 3.1:1–3.1:23.

[77] Liang Liu, Lili Yu, and Scott V Edwards. "A maximum pseudo-likelihood approach for estimating species trees under the coalescent model". In: *BMC evolutionary biology* 10.1 (2010), pp. 1–18. DOI: 10.1186/1471-2148-10-302.

[78] Bret R. Larget et al. "BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis". In: *Bioinformatics* 26.22 (Sept. 2010), pp. 2910–2911. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq539.

[79] Charles Semple, Mike Steel, et al. *Phylogenetics*. Vol. 24. Oxford University Press on Demand, 2003.

[80] Mike Steel. *Phylogeny: discrete and random processes in evolution*. SIAM, 2016.

[81] Fábio K Mendes, Andrew P Livera, and Matthew W Hahn. "The perils of intralocus recombination for inferences of molecular convergence". In: *Philosophical Transactions of the Royal Society B* 374.1777 (2019), p. 20180244. DOI: 10.1098/rstb.2018.0244.

[82] Miguel Arenas. "The importance and application of the ancestral recombination graph". In: *Frontiers in genetics* 4 (2013), p. 206. DOI: 10.3389/fgene.2013.00206.

[83] Gautam Dasarathy et al. "Coalescent-based species tree estimation: a stochastic farris transform". In: *arXiv preprint arXiv:1707.04300* (2017).

[84] Kun-Chieh Wang. *Phylogenetic reconstruction accuracy in the face of heterogeneity, recombination, and reticulate evolution*. The University of Wisconsin-Madison, 2017.

[85] Fabio Pardi and Olivier Gascuel. *Distance-based methods in phylogenetics*. 2016.

[86] Michael Lynch. "Evolution of the mutation rate". In: *TRENDS in Genetics* 26.8 (2010), pp. 345–352. DOI: 10.1016/j.tig.2010.05.003.

[87] Michael Lynch and Georgi K Marinov. "The bioenergetic costs of a gene". In: *Proceedings of the National Academy of Sciences* 112.51 (2015), pp. 15690–15695. DOI: 10.1073/pnas.1514974112.

[88] Ziheng Yang. "Complexity of the simplest phylogenetic estimation problem". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1439 (2000), pp. 109–116. DOI: 10.1098/rspb.2000.0974.

[89] Zhiwei Wang and Kevin J Liu. "A performance study of the impact of recombination on species tree analysis". In: *BMC genomics* 17.10 (2016), pp. 165–174. DOI: 10.1186/s12864-016-3104-5.

[90] Michael Conry. "Determining the Impact of Recombination on Phylogenetic Inference". PhD thesis. The Florida State University, 2020.

[91] J. Felsenstein. *Inferring Phylogenies*. Sinauer, 2003. ISBN: 9780878931774. URL: https://books.google.com/books?id=GI6PQgAACAAJ.

[92] O. Gascuel. *Mathematics of Evolution and Phylogeny*. OUP Oxford, 2005. ISBN: 9780198566106. URL: https://books.google.com/books?id=VjA8ThtLs7IC.

[93] Z. Yang. *Molecular Evolution: A Statistical Approach*. OUP Oxford, 2014. ISBN: 9780191023309. URL: https://books.google.com/books?id=T-LoAwAAQBAJ.

[94] Mike Steel. *Phylogeny: Discrete and Random Processes in Evolution*. Philadelphia, PA, USA: SIAM-Society for Industrial and Applied Mathematics, 2016. ISBN: 161197447X.

[95] Tandy Warnow. *Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation*. Cambridge University Press, 2017. DOI: 10.1017/9781316882313.

[96] Sébastien Roch. "Hands-on Introduction to Sequence-Length Requirements in Phylogenetics". In: *Bioinformatics and Phylogenetics: Seminal Contributions of Bernard Moret*. Ed. by Tandy Warnow. Cham: Springer International Publishing, 2019, pp. 47–86. ISBN: 978-3-030-10837-3. DOI: 10.1007/978-3-030-10837-3_4. URL: https://doi.org/10.1007/978-3-030-10837-3_4.

[97] E. Mossel. "On the impossibility of reconstructing ancestral data and phylogenies". In: *J. Comput. Biol.* 10.5 (2003), pp. 669–678. URL: http://www.stat.berkeley.edu/~mossel/publications/jcb_impossibility.pdf.

[98] Radu Mihaescu, Cameron Hill, and Satish Rao. "Fast phylogeny reconstruction through learning of ancestral sequences". In: *Algorithmica* 66 (2013), pp. 419–449.

[99] Cécile Ané, Lam Si Tung Ho, and Sebastien Roch. "Phase transition on the convergence rate of parameter estimation under an Ornstein-Uhlenbeck diffusion on a tree". In: *J. Math. Biol.* 74.1-2 (2017), pp. 355–385. ISSN: 0303-6812. DOI: 10.1007/s00285-016-1029-x. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s00285-016-1029-x.

[100] Sebastien Roch and Allan Sly. "Phase transition in the sample complexity of likelihood-based phylogeny inference". In: *Probab. Theory Related Fields* 169.1-2 (2017), pp. 3–62. ISSN: 0178-8051. DOI: 10.1007/s00440-017-0793-x. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s00440-017-0793-x.

[101] Wai-Tong Fan and Sebastien Roch. "Necessary and sufficient conditions for consistent root reconstruction in Markov models on trees". In: *Electron. J. Probab.* 23 (2018), Paper No. 47, 24. DOI: 10.1214/18-ejp165. URL: https://doi-org.ezproxy.library.wisc.edu/10.1214/18-ejp165.

[102] Arun Ganesh and Qiuyi (Richard) Zhang. "Optimal Sequence Length Requirements for Phylogenetic Tree Reconstruction with Indels". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. Phoenix, AZ, USA: Association for Computing Machinery, 2019, pp. 721–732. ISBN: 9781450367059. DOI: 10.1145/3313276.3316345. URL: https://doi.org/10.1145/3313276.3316345.

[103] Elchanan Mossel and Yuval Peres. "Information flow on trees". In: *The Annals of Applied Probability* 13.3 (2003), pp. 817–844.

[104] E. Mossel. "Survey: Information flow on trees". In: *Graphs, morphisms and statistical physics*. Ed. by J. Nestril and P. Winkler. Amer. Math. Soc., 2004, pp. 155–170. URL: http://www.stat.berkeley.edu/~mossel/publications/inf_survey.ps.

[105] E. Mossel. "Phase transitions in Phylogeny". In: *Trans. Amer. Math. Soc.* 356.6 (2004), pp. 2379–2404. ISSN: 0002-9947. URL: http://www.stat.berkeley.edu/~mossel/publications/phylogeny_ising.ps.

[106] Elchanan Mossel and Mike Steel. "A phase transition for a random cluster model on phylogenetic trees". In: *Mathematical Biosciences* 187.2 (2004), pp. 189–203.

[107] Christian Borgs et al. "The Kesten-Stigum Reconstruction Bound Is Tight for Roughly Symmetric Binary Channels." In: *FOCS*. 2006, pp. 518–530.

[108] Sebastien Roch. "Toward extracting all phylogenetic information from matrices of evolutionary distances". In: *Science* 327.5971 (2010), pp. 1376–1379.

[109] Constantinos Daskalakis, Elchanan Mossel, and Sébastien Roch. "Evolutionary trees and the Ising model on the Bethe lattice: a proof of Steel's conjecture". In: *Probab. Theory Related Fields* 149.1-2 (2011), pp. 149–189. ISSN: 0178-8051. DOI: 10.1007/s00440-009-0246-2. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s00440-009-0246-2.

[110] Elchanan Mossel, Sébastien Roch, and Allan Sly. "On the inference of large phylogenies with long branches: how long is too long?" In: *Bull. Math. Biol.* 73.7 (2011), pp. 1627–1644. ISSN: 0092-8240. DOI: 10.1007/s11538-010-9584-6. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s11538-010-9584-6.

[111] James H. Degnan and Noah A. Rosenberg. "Gene tree discordance, phylogenetic inference and the multispecies coalescent". In: *Trends in Ecology and Evolution* 24.6 (2009), pp. 332–340. ISSN: 0169-5347.

[112] Luay Nakhleh. "Computational approaches to species phylogeny inference and gene tree reconciliation". In: *Trends in ecology & evolution* 28.12 (2013), pp. 719–728.

[113] Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher — Authors open access book, 2020, p.p. 1–568. URL: https://hal.archives-ouvertes.fr/hal-02535070.

[114] Bruce Rannala and Ziheng Yang. "Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci". In: *Genetics* 164.4 (2003), pp. 1645–1656.

[115] Lars Arvestad, Jens Lagergren, and Bengt Sennblad. "The gene evolution model and computing its associated probabilities". In: *Journal of the ACM* 56.2 (2009), p. 7. DOI: 10.1145/1502793.1502796.

[116] Nicolas Galtier. "A Model of Horizontal Gene Transfer and the Bacterial Phylogeny Problem". In: *Systematic Biology* 56.4 (2007), pp. 633–642. DOI: 10.1080/10635150701546231. eprint: http://sysbio.oxfordjournals.org/content/56/4/633.full.pdf+html. URL: http://sysbio.oxfordjournals.org/content/56/4/633.abstract.

[117] Simone Linz, Achim Radtke, and Arndt von Haeseler. "A Likelihood Framework to Measure Horizontal Gene Transfer". In: *Molecular Biology and Evolution* 24.6 (Mar. 2007), pp. 1312–1319. ISSN: 0737-4038. DOI: 10.1093/molbev/msm052. eprint: https://academic.oup.com/mbe/article-pdf/24/6/1312/13637523/msm052.pdf. URL: https://doi.org/10.1093/molbev/msm052.

[118] Sebastien Roch and Sagi Snir. "Recovering the treelike trend of evolution despite extensive lateral genetic transfer: a probabilistic analysis". In: *Journal of Computational Biology* 20.2 (2013), pp. 93–112.

[119] Frederick A. Matsen and Mike Steel. "Phylogenetic Mixtures on a Single Tree Can Mimic a Tree of Another Topology". In: *Systematic Biology* 56.5 (2007), pp. 767–775.

[120] Elchanan Mossel and Sebastien Roch. "Phylogenetic mixtures: concentration of measure in the large-tree limit". In: *Ann. Appl. Probab.* 22.6 (2012), pp. 2429–2459. ISSN: 1050-5164. DOI: 10.1214/11-AAP837. URL: https://doi-org.ezproxy.library.wisc.edu/10.1214/11-AAP837.

[121] Sebastien Roch and Tandy Warnow. "On the Robustness to Gene Tree Estimation Error (or lack thereof) of Coalescent-Based Species Tree Methods". In: *Systematic Biology* 64.4 (2015), pp. 663–676. DOI: 10.1093/sysbio/syv016. eprint: http://sysbio.oxfordjournals.org/content/64/4/663.full.pdf+html. URL: http://sysbio.oxfordjournals.org/content/64/4/663.abstract.

[122] Paul Simion, Frédéric Delsuc, and Herve Philippe. "To What Extent Current Limits of Phylogenomics Can Be Overcome?" In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher — Authors open access book, 2020, 2.1:1–2.1:34. URL: https://hal.archives-ouvertes.fr/hal-02535366.

[123] Chen Meng and Laura Salter Kubatko. "Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model". In: *Theoretical Population Biology* 75.1 (2009), pp. 35–45. ISSN: 0040-5809. DOI: https://doi.org/10.1016/j.tpb.2008.10.004. URL: http://www.sciencedirect.com/science/article/pii/S0040580908001111.

[124] Elizabeth S Allman, Hector Baños, and John A Rhodes. "NANUQ: a method for inferring species networks from gene trees under the coalescent model". In: *Algorithms for Molecular Biology* 14 (2019), pp. 1–25.

[125] Qiuyi Li et al. "The Multilocus Multispecies Coalescent: A Flexible New Model of Gene Family Evolution". In: *bioRxiv* (2020). DOI: 10.1101/2020.05.07.081836. eprint: https://www.biorxiv.org/content/early/2020/05/10/2020.05.07.081836.full.pdf. URL: https://www.biorxiv.org/content/early/2020/05/10/2020.05.07.081836.

[126] Alexey Markin and Oliver Eulenstein. "Quartet-Based Inference Methods are Statistically Consistent Under the Unified Duplication-Loss-Coalescence Model". Preprint available at https://arxiv.org/abs/2004.04299. 2020.

[127] Claudia Solís-Lemus and Cécile Ané. "Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting". In: *PLOS Genetics* 12.3 (Mar. 2016), pp. 1–21. DOI: 10.1371/journal.pgen.1005896. URL: https://doi.org/10.1371/journal.pgen.1005896.

[128] Gautam Dasarathy, Robert Nowak, and Sebastien Roch. "Data requirement for phylogenetic inference from multiple loci: a new distance method". In: *IEEE/ACM transactions on computational biology and bioinformatics* 12.2 (2014), pp. 422–432.

[129] Elchanan Mossel and Sebastien Roch. "Distance-based species tree estimation under the coalescent: information-theoretic trade-off between number of loci and sequence length". In: *Ann. Appl. Probab.* 27.5 (2017), pp. 2926–2955. ISSN: 1050-5164. DOI: 10.1214/16-AAP1273. URL: https://doi-org.ezproxy.library.wisc.edu/10.1214/16-AAP1273.

[130] Sebastien Roch, Michael Nute, and Tandy Warnow. "Long-Branch Attraction in Species Tree Estimation: Inconsistency of Partitioned Likelihood and Topology-Based Summary Methods". In: *Systematic Biology* 68.2 (Mar. 2019), pp. 281–297. ISSN: 1063-5157. DOI: 10.1093/sysbio/syy061. eprint: http://oup.prod.sis.lan/sysbio/article-pdf/68/2/281/27739135/syy061.pdf. URL: https://doi.org/10.1093/sysbio/syy061.

[131] E. Allman, C. Long, and J. Rhodes. "Species Tree Inference from Genomic Sequences Using the Log-Det Distance". In: *SIAM Journal on Applied Algebra and Geometry* 3.1 (2019), pp. 107–127. DOI: 10.1137/18M1194134. eprint: https://doi.org/10.1137/18M1194134. URL: https://doi.org/10.1137/18M1194134.

[132] Bruce Rannala et al. "The Multi-species Coalescent Model and Species Tree Inference". In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher — Authors open access book, 2020, 3.3:1–3.3:21. URL: https://hal.archives-ouvertes.fr/hal-02535622.

[133] Alexei J Drummond and Andrew Rambaut. "BEAST: Bayesian evolutionary analysis by sampling trees". In: *BMC evolutionary biology* 7.1 (2007), pp. 1–8. DOI: 10.1186/1471-2148-7-214.

[134] J. H. Degnan and N. A. Rosenberg. "Discordance of species trees with their most likely gene trees". In: *PLoS Genet.* 2.5 (May 2006), e68.

[135] Elizabeth S Allman, James H Degnan, and John A Rhodes. "Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent". In: *Journal of Mathematical Biology* 62.6 (2011), pp. 833–862. DOI: 10.1007/s00285-010-0355-7.

[136] Bret R Larget et al. "BUCKy: Gene Tree/Species Tree Reconciliation with Bayesian Concordance Analysis". In: *Bioinformatics* 26.22 (2010), pp. 2910–2911. DOI: 10.1093/bioinformatics/btq539.

[137] S. Mirarab et al. "ASTRAL: genome-scale coalescent-based species tree estimation". In: *Bioinformatics* 30.17 (2014), pp. i541–i548. DOI: 10.1093/bioinformatics/btu462.

[138] Shubhanshu Shekhar, Sebastien Roch, and Siavash Mirarab. "Species Tree Estimation Using ASTRAL: How Many Genes Are Enough?" In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 15.5 (Sept. 2018), pp. 1738–1747. ISSN: 2374-0043. DOI: 10.1109/tcbb.2017.2757930. URL: http://dx.doi.org/10.1109/TCBB.2017.2757930.

[139] K. B. Athreya and P. E. Ney. *Branching processes*. Die Grundlehren der mathematischen Wissenschaften, Band 196. New York: Springer-Verlag, 1972, pp. xi+287.

[140] Roman Vershynin. *High-dimensional probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. An introduction with applications in data science, With a foreword by Sara van de Geer. Cambridge University Press, Cambridge, 2018, pp. xiv+284. ISBN: 978-1-108-41519-4. DOI: 10.1017/9781108231596. URL: https://doi-org.ezproxy.library.wisc.edu/10.1017/9781108231596.

[141] Mike Steel. "A basic limitation on inferring phylogenies by pairwise sequence comparisons". In: *Journal of theoretical biology* 256.3 (2009), pp. 467–472.

[142] Juanjuan Chai and Elizabeth A Housworth. "On Rogers' proof of identifiability for the GTR+ Γ+ I model". In: *Systematic biology* 60.5 (2011), pp. 713–718.

[143] Elchanan Mossel and Sebastien Roch. "Identifiability and inference of non-parametric rates-across-sites models on large-scale phylogenies". In: *Journal of mathematical biology* 67.4 (2013), pp. 767–797.

[144] Gautam Dasarathy, Robert Nowak, and Sebastien Roch. "Data requirement for phylogenetic inference from multiple loci: a new distance method". In: *IEEE/ACM transactions on computational biology and bioinformatics* 12.2 (2014), pp. 422–432.

[145] Michael A Steel and László A Székely. "Inverting random functions II: Explicit bounds for discrete maximum likelihood estimation, with applications". In: *SIAM Journal on Discrete Mathematics* 15.4 (2002), pp. 562–575.

[146] Shubhanshu Shekhar, Sebastien Roch, and Siavash Mirarab. "Species tree estimation using ASTRAL: how many genes are enough?" In: *IEEE/ACM transactions on computational biology and bioinformatics* 15.5 (2017), pp. 1738–1747.

[147] Max Hill, Brandon Legried, and Sebastien Roch. "Species tree estimation under joint modeling of coalescence and duplication: sample complexity of quartet methods". In: *arXiv preprint arXiv:2007.06697* (2020).

[148] Yao-ban Chan, Qiuyi Li, and Celine Scornavacca. "The large-sample asymptotic behaviour of quartet-based summary methods for species tree inference". In: *Journal of Mathematical Biology* 85.3 (2022), p. 22.

[149] Michael DeGiorgio and James H Degnan. "Robustness to divergence time underestimation when inferring species trees from estimated gene trees". In: *Systematic biology* 63.1 (2014), pp. 66–82.

[150] Gautam Dasarathy et al. "A stochastic Farris transform for genetic data under the multispecies coalescent with applications to data requirements". In: *Journal of mathematical biology* 84.5 (2022), p. 36.

[151] GA Watterson. "On the number of segregating sites in genetical models without recombination". In: *Theoretical population biology* 7.2 (1975), pp. 256–276.

[152] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

[153] Alexandre B Tsybakov. *Introduction to nonparametric estimation*. Trans. by Vladimir Zaiats. Springer Series in Statistics, 2009. DOI: 10.1007/978-0-387-79052-7.

[154] David Pollard. *A user's guide to measure theoretic probability*. 8. Cambridge University Press, 2002.

[155] Aad W Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[156] Benny Chor, Amit Khetan, and Sagi Snir. "Maximum likelihood on four taxa phylogenetic trees: analytic solutions". In: *Proceedings of the seventh annual international conference on Research in computational molecular biology*. 2003, pp. 76–83.

[157] Frank E. Anderson and David L. Swofford. "Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA". In: *Molecular Phylogenetics and Evolution* 33.2 (2004), pp. 440–451. ISSN: 1055-7903. DOI: https://doi.org/10.1016/j.ympev.2004.06.015. URL: https://www.sciencedirect.com/science/article/pii/S1055790304002052.

[158] Johannes Bergsten. "A review of long-branch attraction". In: *Cladistics* 21.2 (2005), pp. 163–193.

[159] Claudia Solís-Lemus and Cécile Ané. "Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting". In: *PLOS Genetics* 12.3 (Mar. 2016), pp. 1–21. DOI: 10.1371/journal.pgen.1005896. URL: https://doi.org/10.1371/journal.pgen.1005896.

[160] Ziheng Yang. "Complexity of the simplest phylogenetic estimation problem". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267.1439 (2000), pp. 109–116.

[161] Benny Chor and Sagi Snir. "Molecular clock fork phylogenies: Closed form analytic maximum likelihood solutions". In: *Systematic biology* 53.6 (2004), pp. 963–967.

[162] Benny Chor et al. "Multiple Maxima of Likelihood in Phylogenetic Trees: An Analytic Approach". In: *Molecular Biology and Evolution* 17.10 (Oct. 2000), pp. 1529–1541. ISSN: 0737-4038. DOI: 10.1093/oxfordjournals.molbev.a026252. eprint: https://academic.oup.com/mbe/article-pdf/17/10/1529/23444883/mbev_17_10_1529.pdf. URL: https://doi.org/10.1093/oxfordjournals.molbev.a026252.

[163] Edward Susko and Andrew J Roger. "Long Branch Attraction Biases in Phylogenetics". In: *Systematic Biology* 70.4 (Feb. 2021), pp. 838–843. ISSN: 1063-5157. DOI: 10.1093/sysbio/syab001. eprint: https://academic.oup.com/sysbio/article-pdf/70/4/838/38663996/syab001.pdf. URL: https://doi.org/10.1093/sysbio/syab001.

[164] Bernd Sturmfels and Seth Sullivant. "Toric Ideals of Phylogenetic Invariants". In: *Journal of computational biology: a journal of computational molecular cell biology* 12 2 (2004), pp. 204–28.

[165] Michael D. Hendy. "A combinatorial description of the closest tree algorithm for finding evolutionary trees". In: *Discrete Mathematics* 96.1 (1991), pp. 51–58. ISSN: 0012-365X. DOI: https://doi.org/10.1016/0012-365X(91)90469-I. URL: https://www.sciencedirect.com/science/article/pii/0012365X9190469I.

[166] Seth Sullivant. *Algebraic statistics*. Vol. 194. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2018, pp. xiii+490. ISBN: 978-1-4704-3517-2. DOI: 10.1090/gsm/194. URL: https://doi-org.ezproxy.library.wisc.edu/10.1090/gsm/194.

[167] Michael D Hendy and David Penny. "Spectral analysis of phylogenetic data". In: *Journal of classification* 10 (1993), pp. 5–24.

[168] Elizabeth S. Allman and John A. Rhodes. "Phylogenetic invariants". In: *Reconstructing evolution*. Oxford Univ. Press, Oxford, 2007, pp. 108–146.

[169] Lam-Tung Nguyen et al. "IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies". In: *Molecular biology and evolution* 32.1 (2015), pp. 268–274.

[170] Alexandros Stamatakis. "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies". In: *Bioinformatics* 30.9 (2014), pp. 1312–1313.

[171] Ziheng Yang. "PAML 4: Phylogenetic Analysis by Maximum Likelihood". In: *Molecular Biology and Evolution* 24.8 (May 2007), pp. 1586–1591. ISSN: 0737-4038. DOI: 10.1093/molbev/msm088. eprint: https://academic.oup.com/mbe/article-pdf/24/8/1586/3853532/msm088.pdf. URL: https://doi.org/10.1093/molbev/msm088.

[172] Morgan N Price, Paramvir S Dehal, and Adam P Arkin. "FastTree 2–approximately maximum-likelihood trees for large alignments". In: *PloS one* 5.3 (2010), e9490.

[173] Dimitra Kosta and Kaie Kubjas. "Maximum likelihood estimation of symmetric group-based models via numerical algebraic geometry". In: *Bull. Math. Biol.* 81.2 (2019), pp. 337–360. ISSN: 0092-8240. DOI: 10.1007/s11538-018-0523-2. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s11538-018-0523-2.

[174] Mike Steel. "The Maximum Likelihood Point for a Phylogenetic Tree is not Unique". In: *Systematic Biology* 43.4 (1994), pp. 560–564. ISSN: 10635157, 1076836X. URL: http://www.jstor.org/stable/2413552 (visited on 08/21/2023).

[175] Jane Ivy Coons and Seth Sullivant. "Toric geometry of the Cavender-Farris-Neyman model with a molecular clock". In: *Advances in Applied Mathematics* 123 (2021), p. 102119.

[176] Michael D Hendy, David Penny, and Mike A Steel. "A discrete Fourier analysis for evolutionary trees." In: *Proceedings of the National Academy of Sciences* 91.8 (1994), pp. 3339–3343.

[177] Steven N Evans and Terrence P Speed. "Invariants of some probability models used in phylogenetic inference". In: *The Annals of Statistics* (1993), pp. 355–377.

[178] Luis David Garcia-Puente and Jacob Porter. *Small Phylogenetic Trees https://www.coloradocollege.edu/aapps/ldg/small-trees/small-trees_0.html*. 2007. (Visited on 10/03/2023).

[179] Elizabeth Gross et al. "Distinguishing level-1 phylogenetic networks on the basis of data generated by Markov processes". In: *Journal of Mathematical Biology* 83 (2021), pp. 1–24.

[180] Hector Banos et al. "Phylogenetic trees". In: *Journal of Software for Algebra and Geometry* 11.1 (2022), pp. 1–7.

[181] Elizabeth S. Allman and John A. Rhodes. "Phylogenetic invariants for the general Markov model of sequence mutation". In: *Mathematical Biosciences* 186.2 (2003), pp. 113–144. ISSN: 0025-5564. DOI: https://doi.org/10.1016/j.mbs.2003.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0025556403001329.

[182] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. "Solving the likelihood equations". In: *Found. Comput. Math.* 5.4 (2005), pp. 389–407. ISSN: 1615-3375. DOI: 10.1007/s10208-004-0156-8. URL: https://doi-org.ezproxy.library.wisc.edu/10.1007/s10208-004-0156-8.

[183] Elizabeth Gross et al. "Numerical algebraic geometry for model selection and its application to the life sciences". In: *Journal of The Royal Society Interface* 13.123 (2016), p. 20160256.

[184] David Cox et al. "Ideals, varieties, and algorithms". In: *American Mathematical Monthly* 101.6 (1994), pp. 582–586.

[185] Jan R Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. 3rd. John Wiley & Sons, 2007.

[186] Nicolette Meshkat, Marisa Eisenberg, and Joseph J DiStefano III. "An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases". In: *Mathematical biosciences* 222.2 (2009), pp. 61–72.

[187] Luis David Garcia Puente, Garrote-Lopez. Marina, and Elima Shehu. "Computing algebraic degrees of phylogenetic varieties". In: *arXiv:2210.02116* (2022).

[188] Paul Breiding and Sascha Timme. "HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia". In: *International Congress on Mathematical Software*. Springer. 2018, pp. 458–465.

[189] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 2006.