The marginal and joint distributions of the total tree lengths across loci in populations with variable size

Alexey Miroshnikov[1,2] and Matthias Steinrücken[1]

[1]University of Massachusetts Amherst, Department of Biostatistics and Epidemiology
[2]University of California, Los Angeles, Department of Mathematics

Abstract

In recent years, a number of methods have been developed to infer complex demographic histories, especially historical population size changes, from genomic sequence data. Coalescent Hidden Markov Models have proven to be particularly useful for this type of inference. Due to the Markovian structure of these models, an essential building block is the joint distribution of local genealogical trees, or statistics of these genealogies, at two neighboring loci in populations of variable size. Here, we present a novel method to compute the marginal and the joint distribution of the total length of the genealogical trees at two loci separated by a given recombination distance for samples of arbitrary size. To our knowledge, no method to compute these distributions has been presented in the literature to date. We show that they can be obtained from the solution of certain hyperbolic systems of partial differential equations. We present a numerical algorithm, based on the method of characteristics, that can be used to efficiently and accurately solve these systems and compute the marginal and the joint distributions. We demonstrate its utility to study properties of the joint distribution. Our flexible method can be straightforwardly extended to include the distributions of other statistics of the genealogies as well, and can also be applied in structured populations.

Keywords: coalescent theory, variable population size, hyperbolic systems of PDEs

AMS subject classification: 92D, 60J27, 60J28, 35L40

1 Introduction

In recent years, a number of methods have been developed to infer complex demographic histories, especially historical population size changes, from genomic sequence data. Besides advancing our understanding of the genetic processes that shape contemporary genomic variation, unraveling the demographic history underlying human evolution is also an important step towards understanding disease related genetic variation. Recent rapid population growth, for instance, severely affected the distribution of rare genetic variants (Keinan and Clark, 2012), which have been linked to complex genetic diseases.

Approaches that have proven to be particularly successful for demographic inference, but also other population genetic applications, are based on Coalescent Hidden Markov Models (HMM) (Li and Durbin, 2011; Sheehan et al., 2013; Schiffels and Durbin, 2014; Steinrücken et al., 2016; Rasmussen et al., 2014; Cheng and Mailund, 2015). In a population-sample of genomic sequences, the genealogical relationships vary along the genome, due to intra-chromosomal recombination. The Coalescent-HMMs approximate the intricate correlation structure between these local genealogies under variable population size and in other complex demographic scenarios by a Markov chain, the Sequentially Markovian Coalescent (Wiuf and Hein, 1999; McVean and Cardin, 2005). Due to the Markovian structure of this approximation, an essential building block is the joint distribution of the local genealogical trees, or statistics of these genealogies, at two neighboring loci separated by a given recombination distance. The correlation between local genealogies

is also important for studying linkage disequilibrium (McVean, 2002), that is, allelic association across loci, which is a severe confounding factor in genome-wide association studies.

In this work, we present a novel efficient and accurate method to compute the joint distribution of the total branch length of the genealogical trees at two neighboring loci for a sample of arbitrary size n in populations of varying size, as well as the single-locus marginal distribution. To our knowledge, no method to compute these distributions has been presented in the literature to date that can be applied to arbitrary sample sizes. Moreover, even computing the marginal distribution of the total tree length at a single locus has only received limited attention (Pfaffelhuber et al., 2011).

The inter-coalescent times $T_k^{(n)}$, that is the time period during which k lineages persist in the genealogical tree for a sample of size n can be used to compute the total branch length at a single locus as

$$\mathcal{L} = \sum_{k=2}^{n} k T_k^{(n)}, \tag{1.1}$$

since in the period $T_k^{(n)}$, k lineages contribute towards the total length. In the case of a panmictic population of constant size, formulas for the first two moments of the total tree length can be readily obtained using standard arguments for sums of the independently exponentially distributed random variables $T_k^{(n)}$. However, non-constant population size histories introduce intricate dependencies among the inter-coalescent times, and thus it is not straightforward to generalize this approach. Polanski et al. (2003) introduced a method to compute the expected inter-coalescence times under variable population size. However, the coalescence rates of ancestral lineages in the genealogical process depend on past population sizes, whereas the rate for ancestral recombination is constant along each ancestral lineage. The approach of Polanski et al. (2003) depends on the fact that all rates of the process are rescaled uniformly with the same factor, and thus it cannot be extended to the case when ancestral recombination between two linked loci is taken into account.

Ferretti et al. (2013) used another approach to investigate the correlation between the times to the most recent common ancestor at two neighboring loci. The authors approached the problem using coalescent arguments to quantify the changes recombination induces on the local trees, but it is unclear how to generalize their approach efficiently to the total length of the genealogical trees. Lastly, Li and Durbin (2011) presented analytic formulas for the joint distribution of the local genealogies for a sample of size two under variable population size. However, due to the increase in complexity of the local genealogy with increasing sample size, their approach cannot be generalized efficiently for larger sample sizes.

In this article, we derive a novel efficient and accurate approach to compute the marginal and joint distributions of the total length of the genealogical trees at two neighboring loci in a population of variable size. In Section 2, we introduce the requisite notation and the stochastic processes that are involved in computing the marginal and joint distributions. We further introduce a hyperbolic system of partial differential equations (PDEs) in Section 3 that can be solved to compute the distributions of interest. We provide a proof of the main proposition used to derive these equations in Appendix A. In Section 3, we also provide details of our novel numerical algorithm based on the method of characteristics that can be used to efficiently compute the solutions to these PDEs. We demonstrate the accuracy of the method, and study properties of the joint distribution function in Section 4. Finally, we discuss future applications and extensions of this method in Section 5.

2 Background and Notation

In this section, we will introduce the necessary background and notation for the stochastic processes that we employ to compute the marginal and joint distribution of the length of the genealogical trees. We will also provide some details about computing the distribution of these processes, since our main result extends upon the underlying ideas.

2.1 Ancestral Process at a Single Locus

The genealogical relationship of a sample of n haploid individuals in a panmictic population of constant size is commonly modeled using Kingman's coalescent (Kingman, 1982; Wakeley, 2008), and this process and its extensions have found widespread applications. It is a Markov process that describes the dynamics of the ancestral lineages of the sample backwards in time. Here we focus on the ancestral process $A(t)$ (Tavaré and Zeitouni, 2004, Chapter 4.1). This coarser process records only the number of ancestral lineages in the coalescent process at time t before present, which is sufficient to compute the total branch length of the coalescent tree. The initial number of lineages is equal to the sample size n. Furthermore, at time t, each pair of lineages coalesces at rate one, thus if there are $A(t) = k$ lineages at time t, then coalescence of any two lineages happens at rate $\binom{k}{2}$. This dynamics is followed until all lineages coalesced into a single lineage, and this time is denoted by T_{MRCA}, the time to the most recent common ancestor.

Variable population size is commonly modeled by a positive, real-valued function $\lambda(t)$, which provides the coalescent rate for each pair of ancestral lineages at time t in the past (Tavaré and Zeitouni, 2004, Chapter 4.1). If the size of the population changes at different points in the past, the rate of coalescence at a given time is inversely proportional to the relative population size at that time. Intuitively, for two lineages to coalesce, they have to find a common ancestor. If the population consists of a large number of individuals, this happens at a lower rate, whereas in small populations, the ancestral lineages coalesce more quickly. In the remainder of this paper, we assume that $\lambda(t)$ is continuous. If $\lambda(t)$ is piece-wise continuous, we can obtain the same results by considering each continuous piece separately. For convenience, we further introduce the cumulative coalescent rate at time t as

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

These considerations yield the following definition.

Definition 2.1 (Ancestral Process with variable population size)**.** *The* ancestral process with variable population size $\{A(t)\}_{t\in\mathbb{R}_+}$ *is a time-inhomogeneous Markov chain on* $\{1,\dots,n\}$ *with initial state* $A(0) = n$*, and the transition rates at time* t *are given by the infinitesimal generator matrix*

$$Q(t) = \lambda(t)Q,$$

with

$$Q_{k,j} := \begin{cases} -\binom{k}{2}, & \text{if } j = k, \\ \binom{k}{2}, & \text{if } j = k-1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Remark 2.2. *Note that we do require* $A(0) = n$*, and thus this definition of the ancestral process does depend on the sample size* n*. However, for different sample sizes* n'*, the rates of the process are given by equation* (2.1) *as well, only the initial state changes. The dynamics of the process is essentially the same, independent of the sample size, and we therefore do not include the dependence on the sample size explicitly in the notation for the remainder of this article.*

The ancestral process can be used to formally define the time to the most recent common ancestor as

$$T_{\text{MRCA}} := \inf\big\{t \in \mathbb{R}_+ : A(t) \le 1\big\},$$

the time when the number of lineages reaches one. Furthermore, with

$$p_k(t) := \mathbb{P}\big\{A(t) = k\big\},$$

for $k \in \{1,\dots,n\}$, the distribution of the ancestral process can be obtained by solving the Kolmogorov-forward-equation (Stroock, 2008, Chapter 5), a system of ordinary differential equations (ODEs) given by

$$\frac{d}{dt}\big(p_1(t),\dots,p_n(t)\big) = \big(p_1(t),\dots,p_n(t)\big)Q(t). \tag{2.2}$$

Equivalently, perhaps more familiar to the reader, this system can be expressed as

$$\frac{d}{dt}p_k(t) = \lambda(t)\binom{k+1}{2}p_{k+1}(t) - \lambda(t)\binom{k}{2}p_{k+1}, \tag{2.3}$$

for all $k \in \{1, \dots, n\}$. The latter version is more explicit about the influence of the number of ancestral lineages and the coalescent-speed function on the dynamics of the ODEs. The relevant solution is given by

$$\big(p_1(0), \dots, p_n(0)\big) = (0, \dots, 0, 1)$$

and

$$\big(p_1(t), \dots, p_n(t)\big) = \big[e^{\Lambda(t)\cdot Q}\big]_{n,\cdot} \tag{2.4}$$

for $t \in \mathbb{R}_+$. In Tavaré and Zeitouni (2004), the authors provides an analytic expression for these probabilities using the spectral decomposition of the rate matrix $Q(t)$. However, the resulting formulas are numerically unstable, so for practical purposes it can be more efficient to solve the system of ODEs numerically using step-wise solution schemes. Furthermore, note that

$$\mathbb{P}\{T_{\mathrm{MRCA}} \leq t^*\} = \big[e^{\Lambda(t^*)\cdot Q}\big]_{n,1} \tag{2.5}$$

holds for $t^* \in \mathbb{R}_+$, thus equation (2.4) can also be used to compute the cumulative distribution function of the time to the most recent common ancestor.

We can employ the ancestral process to compute the total tree length as follows. If at a given time t there are k ancestral lineages or branches in the coalescent tree, each branch extends further into the past. Thus, we can say that the total sum of branch lengths in the coalescent tree grows at a rate of k. Once all lineages have coalesced into a single common ancestral lineage, the most recent common ancestor is reached, and the coalescent tree stops growing. This motivates the following definition.

Definition 2.3. *The* accumulated tree length $L(t) \in \mathbb{R}_+$ *by time* $t \in \mathbb{R}_+$ *is given by*

$$L(t) = \int_0^t \mathbb{1}_{\{A(s)>1\}} A(s) ds.$$

With this definition, the *total tree length* or the *total sum of the branch lengths* at a single locus is given by

$$\mathcal{L} := L\big(T_{\mathrm{MRCA}}\big).$$

Note that

$$\mathcal{L} = \sum_{k=2}^{n} k T_k^{(n)}$$

holds, which is equal to equation (1.1). Here $T_k^{(n)}$ is the period of time for which k lineages persist in the ancestral process, the inter-coalescent time. The main goal of this paper is to study the distribution of $\mathcal{L}$ for populations with arbitrary coalescent-rate function $\lambda(t)$ marginally at a single locus and jointly at two loci, which can be computed using a system of hyperbolic PDEs that is closely related to the ODE (2.3). For the two-locus case, we will now introduce the joint ancestral process at two linked loci.

2.2 Ancestral Process with Recombination

The joint genealogy of the ancestral lineages for two loci, locus a and b, separated by a recombination distance ρ is commonly modeled by the coalescent with recombination (Hudson, 1990). The initial state in the coalescent with recombination for a sample of size n is comprised of n lineages, each ancestral to both loci of one sampled haplotype. As in the single-locus coalescent with variable population size, at time t, each pair of lineages can coalesce at rate $\lambda(t)$. In addition, ancestral recombination events happen at rate $\rho/2$ along

each active lineage. At a recombination event, the lineage splits into two new lineages, each ancestral to the respective haplotype of the original lineage at only one of the two loci. Note that recombination happens along each lineage at a constant rate and, unlike the coalescent rate, is not affected by the population size, and thus it does not scale with $\lambda(t)$.

Again, we do not focus on the exact genealogical relationships, but only on the number of lineages at time t that are ancestral to a certain locus, given by the *ancestral process with recombination* $A^\rho(t)$. The process A^ρ for a sample of size two under constant population size is described in detail by Simonsen and Churchill (1997). Here we use an extension of this process to samples of arbitrary size n and variable population size.

Definition 2.4 (Ancestral Process with Recombination)**.** *For a sample of size $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$, the* ancestral process with recombination *in a population of variable size*

$$A^\rho(t) = \big(K_{ab}(t), K_a(t), K_b(t)\big)$$

is a time-inhomogeneous Markov chain with state space

$$\mathcal{S}^\rho := \big\{s \in \mathbb{N}_0^3 \big| s_1 + \max\{s_2, s_3\} \leq n\big\} \big\backslash \big\{(0,0,0),(0,1,0),(0,0,1)\big\}.$$

The component $K_{ab}(t)$ gives the number of lineages that are ancestral to both loci, $K_a(t)$ is the number ancestral to locus a only, and $K_b(t)$ is the number ancestral to locus b only. The initial state is

$$A^\rho(0) = (n, 0, 0),$$

all n lineages ancestral to both loci. The transition rates are given by the infinitesimal generator matrix

$$\tilde{Q}(t) = \lambda(t) Q^c + Q^\rho,$$

where all off-diagonal entries of Q^c (coalescence) are zero, except

$$\begin{aligned}
Q^c_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a,k_b)} &= \binom{k_{ab}}{2}, \\
Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a-1,k_b)} &= \binom{k_a}{2} + k_{ab}k_a, \\
Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a,k_b-1)} &= \binom{k_b}{2} + k_{ab}k_b, \\
&\textit{and} \\
Q^c_{(k_{ab},k_a,k_b),(k_{ab}+1,k_a-1,k_b-1)} &= k_a k_b,
\end{aligned} \tag{2.6}$$

and all off-diagonal entries of Q^ρ (recombination) are zero, except

$$Q^\rho_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a+1,k_b+1)} = \frac{\rho}{2} k_{ab}.$$

The state $(1,0,0)$ is defined to be the absorbing state, so all rates leaving this state are set to zero. Furthermore, the diagonal entries of both matrices are set to minus the sum of the off-diagonal entries in the corresponding row.

Remark 2.5. *i) Two versions of the coalescent with recombination are commonly used in the literature, one version for the infinitely-many-sites (IMS) model (Hudson, 1990; Griffiths and Marjoram, 1997), and another version for the finitely-many-sites (FMS) model (Paul et al., 2011; Steinrücken et al., 2015). In the IMS version, the chromosome is modeled as the interval $[0,1]$, and whenever recombination occurs, it occurs at a uniformly chosen point in this interval. As a result, recombination always occurs at a novel site, and two neighboring local genealogies are separated by at most one recombination event. In the FMS version, multiple recombination events can occur between two loci. It can be obtained from the IMS version by considering*

the local genealogies at two fixed loci along the continuous chromosome that are separated by a certain fixed recombination distance. Our definition of the ancestral process with recombination is in line with the FMS version for two loci.
ii) The ancestral process with recombination can be defined for an arbitrary number of loci. However, in the remainder of the paper, we will only use the process for two loci.
iii) In the literature, some authors use the 'full' coalescent with recombination and others the 'reduced' coalescent with recombination. The difference between the two is that the 'full' version always keeps track of both ancestral lineages that branch off at a recombination event, whereas in the 'reduced' version, lineages that don't leave any descendant ancestral material in the contemporary sample are not traced. Our definition of the ancestral process with recombination is compatible with the 'reduced' version. Thus, the number of ancestral lineages is bounded, which is not the case in the 'full' version.
iv) Following the ideas of Wiuf and Hein (1999), the correlation structure between all local genealogies along a chromosome can be approximated using the Sequentially Markovian Coalescent (SMC) (McVean and Cardin, 2005), or the modified version SMC' (Marjoram and Wall, 2006). In the SMC, if a lineage has been hit by a recombination event and branches into two, subsequently, the two resulting branches are not allowed to coalesce with each other, whereas such events are permitted under the SMC'. Thus, under the SMC', the rates for coalescence of lineages with no overlapping ancestral material (equation (2.6)*) are as given in Definition 2.4, whereas under the SMC, these rates have to be set to zero.*

Again, the Kolmogorov-forward-equation can be used to compute the distribution of the ancestral process $A^\rho(t)$ as the solution of

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\tilde{Q}(t), \tag{2.7}$$

where the row-vector $\mathbf{p}(t)$ is defined by

$$\mathbf{p}(t) := \Big(\mathbb{P}\big\{A^\rho(t) = s\big\}\Big)_{s\in\mathcal{S}^\rho}.$$

Note that the rate matrix $Q(t)$ in the ODE (2.2) for the ancestral process at a single locus is triangular for all t. This simplifies approaches to compute solutions substantially, as the solutions can be obtained sequentially for each state of the corresponding Markov chain. In the ancestral process with recombination for two loci on the other hand, with a positive probability, the underlying Markov chain can transition back to a state it already visited before. Consequently, the rate matrix $\tilde{Q}(t)$ in the ODE (2.7) is not triangular, and it is also not possible to transform it into a triangular matrix by permuting the rows and columns.

Since a triangular rate matrix simplifies analytical and numerical approaches significantly, we introduce an approximation to the full ancestral process with recombination that exhibits this property and compute the distributions of the tree lengths under this approximation. To achieve this, we explicitly account for the number of recombination events that have occurred up to a certain time t. For ease of exposition, we further limit the maximal number of recombination events to one. Since in most organism the per generation recombination probability is very small between loci that are physically close, this approximation is justified. Furthermore, numerical experiments supporting this approximation are provided in Section 4.

Definition 2.6 (Ancestral Process with Limited Recombination)**.** *For a sample of size $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$, the* ancestral process with limited recombination

$$\bar{A}^\rho(t) = \big(\bar{K}_{ab}(t), \bar{K}_a(t), \bar{K}_b(t), \bar{R}(t)\big)$$

is a time-inhomogeneous Markov chain with state space

$$\begin{aligned}\bar{\mathcal{S}}^\rho :=& \Big(\{1,\dots,n\}\times\{(0,0,0)\}\\ &\cup\{1,\dots,n\}\times\{0,1\}\times\{0,1\}\times\{1\}\Big)\\ &\qquad\setminus\big\{(n,1,1,1),(n,1,0,1),(n,0,1,1)\big\}.\end{aligned}$$

The components $\bar{K}_{ab}(t)$, $\bar{K}_a(t)$, $\bar{K}_b(t)$ have the same interpretation as before, and $\bar{R}(t)$ is the number of recombination events that have happened by time t. The initial state is

$$\bar{A}^\rho(0) = (n, 0, 0, 0),$$

and the transition rates are given by the infinitesimal generator matrix

$$\bar{Q}(t) = \lambda(t)\bar{Q}^c + \bar{Q}^\rho, \tag{2.8}$$

where the entries of $\bar{Q}^c$ (coalescence) are given by

$$\bar{Q}^c_{(k_{ab},k_a,k_b,r),(k_{ab},k_a,k_b,r)} = Q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a,k_b)},$$

and all off-diagonal entries of Q^ρ (recombination) are zero, except

$$\bar{Q}^\rho_{(k_{ab},k_a,k_b,0),(k_{ab}-1,k_a+1,k_b+1,1)} = \frac{\rho}{2}k_{ab},$$

allowing at most one recombination event. The diagonal entries are set to minus the sum of the off-diagonal entries in the corresponding row. The states $(1,0,0,0)$ and $(1,0,0,1)$ are absorbing states, so all rates leaving these states are set to zero.

For later convenience, define the relation $\prec$ on $\bar{\mathcal{S}}^\rho$ as

$$s \prec s' :\Leftrightarrow \bar{Q}_{s',s}(t) > 0, \tag{2.9}$$

that is, $s \prec s'$ holds if s can be reached from s' in one step. Note that embedded into the ancestral process with recombination (limited or not) is a single-locus ancestral process for locus a and for locus b. Thus, we can define the the branch length of the genealogical tree at locus a and b similar to the one-locus case as follows, and study their joint distribution.

Definition 2.7. *For a given time $t \in \mathbb{R}_+$, the* accumulated tree lengths *$L^a(t) \in \mathbb{R}^+$ at locus a and $L^b(t) \in \mathbb{R}_+$ at locus b are given by*

$$L^a(t) = \int_0^t \mathbb{1}_{\{\bar{K}_{ab}(s)+\bar{K}_a(s)>1\}}\big(\bar{K}_{ab}(s) + \bar{K}_a(s)\big)ds,$$

and

$$L^b(t) = \int_0^t \mathbb{1}_{\{\bar{K}_{ab}(s)+\bar{K}_b(s)>1\}}\big(\bar{K}_{ab}(s) + \bar{K}_b(s)\big)ds.$$

Remark 2.8. *This definition of the accumulated tree length can be applied to $\bar{A}^\rho$, as well as A^ρ. We will not distinguish these cases in our notation, since in the remainder of the paper, we will use $\bar{A}^\rho$.*

The *total tree length* at locus a is thus given by

$$\mathcal{L}^a := L^a\big(T^a_{\mathrm{MRCA}}\big), \tag{2.10}$$

and at locus b by

$$\mathcal{L}^b := L^b\big(T^b_{\mathrm{MRCA}}\big). \tag{2.11}$$

Here, T^a_{MRCA} is the time to the most recent common ancestor at locus a

$$T^a_{\mathrm{MRCA}} := \inf\big\{t \in \mathbb{R}_+ : \bar{K}_{ab}(t) + \bar{K}_a(t) \le 1\big\},$$

and thus its distribution is given by

$$\mathbb{P}\{T^a_{\mathrm{MRCA}} \le t^*\} = \mathbb{P}\big\{\bar{K}_{ab}(t^*) + \bar{K}_a(t^*) \le 1\big\}$$

for $t^* \in \mathbb{R}_+$. Similar relations hold for locus b. We will now study the joint distribution of $\mathcal{L}^a$ and $\mathcal{L}^b$, and also the marginal $\mathcal{L}$. Note that these quantities are computed under the ancestral process with limited recombination, but we will demonstrate in Section 4 that they give an accurate approximation to the respective quantities under the true ancestral process.

3 Marginal and Joint Distribution of the Total Tree Length

The main goal of this paper is to present a method to compute the marginal and joint cumulative distribution function (CDF) of the total tree length at two linked loci. Thus, we aim at computing

$$\mathbb{P}\{\mathcal{L} \leq x\} \tag{3.1}$$

and

$$\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} \tag{3.2}$$

for $x, y \in \mathbb{R}_+$. Equation (2.5) can be used to compute the marginal distribution of T_{MRCA}, which can be obtained as the sum of the inter-coalescence times. The total branch length, however, is a more general linear combination of these times. It can be shown that because of this and the time-inhomogeneity of the ancestral processes, the approach underlying equation (2.5) can not be applied to compute the distributions of $\mathcal{L}$, $\mathcal{L}^a$, and $\mathcal{L}^b$, and we need to devise a different approach.

To this end, with $t \in \mathbb{R}_+$, we introduce the time-dependent cumulative distribution functions

$$F_k(t, x) := \mathbb{P}\big\{A(t) = k, L(t) \leq x\big\} \tag{3.3}$$

for $k \in \{1, \ldots, n\}$ and

$$F_s(t, x, y) := \mathbb{P}\big\{\bar{A}^\rho(t) = s, L^a(t) \leq x, L^b(t) \leq y\big\} \tag{3.4}$$

for $s \in \bar{\mathcal{S}}^\rho$.

We will show that the CDFs (3.1) and (3.2) can be computed from the time-dependent CDFs (3.3) and (3.4). Furthermore, we will present numerical schemes, to efficiently and accurately compute the time-dependent CDFs (3.3) and (3.4).

3.1 Distribution of the Total Tree length at a Single Locus

The following lemma shows that the CDF (3.1) can be computed from the time-dependent CDF (3.3).

Lemma 3.1. *With definition* (3.3)*, the relation*

$$\mathbb{P}\{\mathcal{L} \leq x\} = \mathbb{P}\big\{A(\bar{t}) = 1, L(\bar{t}) \leq x\big\} = F_1(\bar{t}, x)$$

holds for $x \in \mathbb{R}_+$ *and* $\bar{t} \geq x/2$.

Proof. First, observe that

$$2T_{\mathrm{MRCA}} \leq \int_0^{T_{\mathrm{MRCA}}} \mathbb{1}_{\{A(s)>1\}} A(s)\, ds = \mathcal{L},$$

since $A(s) \geq 2$ holds for $s < T_{\mathrm{MRCA}}$. Thus, on the event $\big\{\mathcal{L} \leq x\big\}$, the relation $T_{\mathrm{MRCA}} \leq x/2 \leq \bar{t}$ holds, which implies $A(\bar{t}) = 1$, and therefore

$$\big\{\mathcal{L} \leq x\big\} = \big\{A(\bar{t}) = 1, \mathcal{L} \leq x\big\}.$$

On the event $\big\{A(\bar{t}) = 1\big\}$, $\bar{t} \geq T_{\mathrm{MRCA}}$ and $L(\bar{t}) = \mathcal{L}$ hold, and thus

$$\big\{A(\bar{t}) = 1, \mathcal{L} \leq x\big\} = \big\{A(\bar{t}) = 1, L(\bar{t}) \leq x\big\},$$

which proves the statement of the lemma. □

Lemma 3.1 shows that the CDF of $\mathcal{L}$ can be computed from the time-dependent CDF $F_1(t, x)$. Due to the structure of the underlying Markov chain, it is necessary to compute the time-dependent CDFs for all states in order to compute it for the absorbing state. Thus, in the remainder of this section, we focus on computing the time-dependent CDFs for all $k \in \{1, \ldots, n\}$. Proposition A.14 derived in Appendix A can be applied to show that the time-dependent CDFs solve a certain system of linear hyperbolic PDEs. This yields the following corollary.

Corollary 3.2. *The row-vector*

$$\mathbf{F}(t,x) := \big(F_1(t,x), \dots, F_n(t,x)\big)$$

can be obtained for all points in $\mathcal{U} = \Big\{(x,t) : 0 < x < nt,\, t > 0\Big\}$ *as the strong solution of*

$$\partial_t \mathbf{F}(t,x) + V \partial_x \mathbf{F}(t,x) = \mathbf{F}(t,x) Q(t), \tag{3.5}$$

with

$$V = \texttt{diag}(0, 2, 3, \dots, n),$$

boundary conditions

$$\begin{aligned} \mathbf{F}(t,x) &= \big(P\big\{A(t)=1\big\}, \dots, \mathbb{P}\big\{A(t)=n-1\big\}, 0\big), \quad x = nt \\ \mathbf{F}(t,0) &= \big(0,\, 0,\, \dots,\, 0\big), \quad t > 0\,, \end{aligned} \tag{3.6}$$

and matrix $Q(t)$ *as defined in equation* (2.1)*.*

Proof. Define the function

$$v(k) := k \cdot \mathbb{1}_{\{k>1\}}$$

on the state space $\mathcal{S} = \{1, \dots, n\}$ of the ancestral process. This function and the generator $Q(t)$ satisfy the requirements of Proposition A.14, and thus, the statement of the corollary follows from this proposition. □

Remark 3.3. *The n-th component of the boundary condition* (3.6) *is equal to 0 and not* $\mathbb{P}\big\{A(t)=n\big\}$*. This holds for technical reasons that will be detailed in the proof of Proposition A.14.*

Note that the right-hand side of equation (3.5) is essentially equal to the right-hand side of equation (2.3), because the only stochastic element in the underlying dynamics is the ancestral process $A(t)$. Given a certain number of lineages $\big\{A(t)=k\big\}$, the accumulation towards the total tree length happens deterministically at rate k, and is captured by the term $V\partial_x \mathbf{F}(t,x)$. We will now present a numerical scheme to efficiently and accurately compute the time-dependent CDF $\mathbf{F}(t,x)$.

3.1.1 Applying the Method of Characteristics

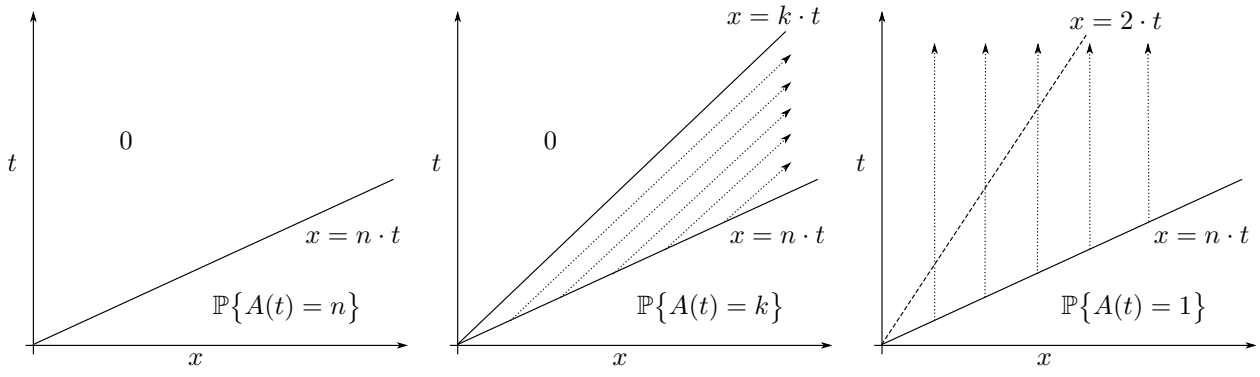
The system of PDEs introduced in Corollary 3.2 can be solved using the method of characteristics (Renardy and Rogers, 2004, Chapter 3). To this end, note that due to the triangular structure of the matrix $Q(t)$, for a given component with $k \in \{1, \dots, n\}$, the right-side of equation (3.5) does only depend on F_ℓ whit $\ell \geq k$. Thus, the system of PDEs (3.1) can be solved separately for each k, starting at $k = n$, and decreasing it step-by-step.

Furthermore, note that

$$F_k(t,x) = \mathbb{P}\big\{A(t)=k, L(t) \leq x\big\} = \begin{cases} 0, & \text{if } x < v(k)t, \\ \text{solution to (3.5)}, & \text{if } v(k)t \leq x < n \cdot t, \\ \mathbb{P}\big\{A(t)=k\big\}, & \text{if } n \cdot t \leq x \end{cases} \tag{3.7}$$

holds for $k \in \{1, \dots, n\}$, since if the ancestral process has k lineages at time t, it must have accumulated at least $v(k)t$ and at most nt towards the total tree length. Note that $v(1) = 0$. Moreover, for $k = n$, the region $v(k)t \leq x < n \cdot t$ is empty, and thus $F_n(t,x)$ has a discontinuity along the line $n \cdot t$. See Figure 1 for a visualization of the different regions for different values of k. To devise an accurate and efficient numerical scheme for computing the time-dependent CDFs in the interior region, we use the method of characteristics to solve the respective PDE

$$\partial_t F_k(t,x) + v(k) \partial_x F_k(t,x) = F_k(t,x) Q_{k,k}(t) + F_{k+1}(t,x) Q_{k+1,k}(t). \tag{3.8}$$



(a) The two only regions for the initial state $k = n$. The function has a discontinuity at $x = nt$.

(b) The three regions and the characteristics in the interior for an intermediate state with $1 < k < n$.

(c) Regions and characteristics for the absorbing state $k = 1$. The characteristics are parallel to the t-axis.

Figure 1: The different regions and characteristics of $F_k(t, x)$ for different values of k. In (c), according to Lemma 3.1, $F_k(t, x)$ does not depend on t beyond the dashed line $x = 2t$.

Since for $k = n$, the interior region is empty, we consider $k \neq n$ and introduce the family of characteristics

$$\tau \to \Big(t_0 + \tau, x_0 + v(k)\tau\Big)^\top \quad \text{with} \quad t_0 = \frac{x_0}{n}$$

Taking the derivative of $F_k(t, x)$ along such a characteristic yields

$$\begin{aligned}
&\frac{d}{d\tau} F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) \\
&\quad = \left(\frac{d}{d\tau}\Big[t_0 + \tau\Big] \cdot \partial_t F_k(t, x) + \frac{d}{d\tau}\Big[x_0 + v(k)\tau\Big] \cdot \partial_x F_k(t, x) \right)\Bigg|_{(t,x)=(\frac{x_0}{n}+\tau, x_0+v(k)\tau)} \\
&\quad = \Big(\partial_t F_k(t, x) + v(k)\partial_x F_k(t, x) \Big)\Big|_{(t,x)=(\frac{x_0}{n}+\tau, x_0+v(k)\tau)} \\
&\quad = F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) Q_{k,k}(t_0 + \tau) + F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big) Q_{k+1,k}(t_0 + \tau).
\end{aligned} \tag{3.9}$$

Here we used the chain rule and the fact that the third line is equal to the left-hand side of equation (3.8). Formally, the derivations (3.9) do not hold for all τ. It can be shown, however, that the equality holds for almost all τ; we omit the technical details here for readability. Thus, for given x_0, as a function of τ, the function $\tau \to F_k(t_0 + \tau, x_0 + v(k)\tau)$ solves the equation

$$\frac{d}{d\tau} F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) = -q_k^{(1)}(\tau) F_k\Big(t_0 + \tau, x_0 + v(k)\tau\Big) + g_k^{(1)}(\tau),$$

with

$$q_k^{(1)}(\tau) := -Q_{k,k}(t_0 + \tau) = \frac{k(k-1)}{2}\lambda(t_0 + \tau)$$

and

$$\begin{aligned}
g_k^{(1)}(\tau) :=& F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big) Q_{k+1,k}(t_0 + \tau) \\
=& F_{k+1}\Big(t_0 + \tau, x_0 + v(k)\tau\Big) \frac{(k+1)k}{2}\lambda(t_0 + \tau),
\end{aligned}$$

Since this is a non-homogeneous linear first-order ODE, the solution can be readily obtained as

$$F_k\Big(t_0+\tau, x_0+v(k)\tau\Big) = e^{-H_k^{(1)}(\tau)}\left(\int_0^\tau g_k^{(1)}(\alpha)e^{H_k^{(1)}(\alpha)}d\alpha + F_k\big(t_0,x_0\big)\right), \tag{3.10}$$

with

$$H_k^{(1)}(\tau) := \int_0^\tau q_k^{(1)}(\alpha)d\alpha = \frac{k(k-1)}{2}\big(\Lambda(u)-\Lambda(t_0)\big). \tag{3.11}$$

The initial conditions for $\tau = 0$ are given by the boundary values of the associated PDE as

$$F_k(t_0,x_0) = \mathbb{P}\big\{A(t_0)=k\big\}.$$

Now, to obtain the value of the function $F_k(t,x)$, for given t and x, one just needs to identify the right characteristic and the parameters x_0 and τ such that $(t_0+\tau, x_0+v(k)\tau)^\top = (t,x)^\top$. Since the characteristics are parallel, it can be uniquely identified. Using these values of x_0 and τ in the solution (3.10) yields $F_k(t,x)$. However, we will not pursue this strategy to compute the requisite values of $F_k(t,x)$. Instead, we present a numerical upstream scheme in Appendix B.1 that can be used to compute $F_k(t,x)$ efficiently on a suitable grid to ultimately obtain values for the CDF $\mathbb{P}\{\mathcal{L}\le x\}$.

3.2 Joint Distribution of the Total Tree Length

In this section we present a method to compute the joint CDF of the total tree length

$$\mathbb{P}\{\mathcal{L}^a \le x, \mathcal{L}^b \le y\}$$

at two loci a and b separated by a given recombination distance ρ. Again, we approach this problem by first computing the time-dependent joint CDF

$$F_s(t,x,y) = \mathbb{P}\big\{\bar{A}^\rho(t)=s, L^a(t)\le x, L^b(t)\le y\big\}.$$

We will follow closely along the lines of the method presented in Section 3.1, where we replace the ancestral process A by the ancestral process with limited recombination $\bar{A}^\rho$, and compute the integrals (2.10) and (2.11), to ultimately compute the joint CDF.

The analog to Lemma 3.1 is as follows.

Lemma 3.4. *With definition* (3.4)*, the relation*

$$\begin{aligned}\mathbb{P}\big\{\mathcal{L}^a\le x, \mathcal{L}^b\le y\big\} &= \mathbb{P}\big\{\bar{A}^\rho(\bar{t})\in\Delta, L^a(\bar{t})\le x, L^b(\bar{t})\le y\big\}\\ &= F_{(1,0,0,0)}(\bar{t},x,y) + F_{(1,0,0,1)}(\bar{t},x,y)\end{aligned}$$

holds for $x,y\in\mathbb{R}_+$, $\bar{t}\ge\max\{x,y\}/2$, and $\Delta = \big\{(1,0,0,0),(1,0,0,1)\big\}$, the absorbing states of $\bar{A}^\rho$.

Proof. The proof is similar to the proof of lemma 3.1. With

$$\bar{A}^\rho(t) = \big(\bar{K}_{ab}(t), \bar{K}_a(t), \bar{K}_b(t), \bar{R}(t)\big),$$

note that

$$2T^a_{\mathrm{MRCA}} \le \int_0^{T^a_{\mathrm{MRCA}}} \mathbb{1}_{\{\bar{K}_{ab}(s)+\bar{K}_a(s)>1\}}\big(\bar{K}_{ab}(s)+\bar{K}_a(s)\big)ds = \mathcal{L}^a,$$

and similarly $2T^b_{\mathrm{MRCA}}\le\mathcal{L}^b$. Thus, on the event $\big\{\mathcal{L}^a\le x,\mathcal{L}^b\le y\big\}$, the relations $T^a_{\mathrm{MRCA}}\le\max\{x,y\}/2\le\bar{t}$ and $T^b_{\mathrm{MRCA}}\le\bar{t}$ hold. This implies $\bar{K}_{ab}(\bar{t})+\bar{K}_a(\bar{t})=1$ and $\bar{K}_{ab}(\bar{t})+\bar{K}_a(\bar{t})=1$, which in turn implies $\bar{A}^\rho(\bar{t})\in\Delta=\big\{(1,0,0,0),(1,0,0,1)\big\}$, since these two states are the only admissible states that can satisfy these conditions. Incidentally, these are also the absorbing states of $\bar{A}^\rho$. Thus,

$$\big\{\mathcal{L}^a\le x,\mathcal{L}^b\le y\big\} = \big\{\bar{A}^\rho(\bar{t})\in\Delta, \mathcal{L}^a\le x,\mathcal{L}^b\le y\big\}$$

holds. Furthermore, on the event $\{\bar{A}^\rho(\bar{t}) \in \Delta\}$, $T^a_{\mathrm{MRCA}} \leq \bar{t}$ and $T^b_{\mathrm{MRCA}} \leq \bar{t}$ hold, which imply $L^a(\bar{t}) = \mathcal{L}^a$ and $L^b(\bar{t}) = \mathcal{L}^b$. This in turn implies

$$\{\bar{A}^\rho(\bar{t}) \in \Delta, \mathcal{L}^a \leq x, \mathcal{L}^b \leq y\} = \{\bar{A}^\rho(\bar{t}) \in \Delta, L^a(\bar{t}) \leq x, L^b(\bar{t}) \leq y\}.$$

Finally, note that

$$\{\bar{A}^\rho(\bar{t}) = (1,0,0,1)\} \cap \{\bar{A}^\rho(\bar{t}) = (1,0,0,0)\} = \emptyset,$$

which proves the statement of the lemma. □

Again, Lemma 3.4 shows that the joint CDF of $\mathcal{L}^a$ and $\mathcal{L}^b$ can be computed from the time-dependent CDFs $F_{(1,0,0,0)}(t,x,y)$, and $F_{(1,0,0,1)}(t,x,y)$. In order to derive a system of PDEs like (3.5) for the time-dependent joint CDF of the tree length at two loci, we would require a version of Proposition A.14 that handles two dimensions x, y. We will provide a detailed proof of such an extended theorem in a separate paper. Here, we include the statement as a conjecture and demonstrate the correctness empirically in Section 4.

To this end, define the functions

$$v^a(k_{ab}, k_a, k_b, r) := \mathbb{1}_{\{k_{ab}+k_a>1\}}(k_{ab} + k_a)$$

and

$$v^b(k_{ab}, k_a, k_b, r) := \mathbb{1}_{\{k_{ab}+k_b>1\}}(k_{ab} + k_b)$$

that yield for $(k_{ab}, k_a, k_b, r) \in \bar{\mathcal{S}}^\rho$ the number of lineages ancestral to locus a and b, respectively, and define

$$V^a := \mathtt{diag}\Big(\big(v^a(s)\big)_{s\in\bar{\mathcal{S}}^\rho}\Big) \qquad \text{and} \qquad V^b := \mathtt{diag}\Big(\big(v^b(s)\big)_{s\in\bar{\mathcal{S}}^\rho}\Big).$$

We then have the following conjecture.

Conjecture 3.5. *The time-dependent joint CDF of the tree lengths*

$$\mathbf{F}(t,x,y) = \big(\mathbf{F}_s(t,x,y)\big)_{s\in\bar{\mathcal{S}}^\rho}$$

can be obtained for all points in $U = \Big\{(t,x,y) : 0 < x < nt, 0 < y < nt, t > 0\Big\}$ *as the strong solution of*

$$\partial_t \mathbf{F}(t,x,y) + V^a \partial_x \mathbf{F}(t,x,y) + V^b \partial_y \mathbf{F}(t,x,y) = \mathbf{F}(t,x,y)\bar{Q}(t), \tag{3.12}$$

with boundary conditions

$$\mathbf{F}(t,x,y) = \begin{cases} \Big(\mathbb{P}\{\bar{A}^\rho(t) = s, L^b(t) \leq y\}\Big)_{s\in\bar{\mathcal{S}}^\rho} \cdot \mathbb{1}_{\{v^a(s)\neq n\}}, & \text{if } x = nt, \\ \Big(\mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \leq x\}\Big)_{s\in\bar{\mathcal{S}}^\rho} \cdot \mathbb{1}_{\{v^b(s)\neq n\}}, & \text{if } y = nt, \\ 0, & \text{if } x = 0 \text{ or } y = 0, \end{cases}$$

for $(x,y,t) \in \partial U$ *and* $\bar{Q}(t)$ *as defined in* (2.8).

Remark 3.6. *Note that due to symmetry of* $\bar{A}^\rho$*,*

$$\mathbb{P}\{\bar{A}^\rho(t) = s, L^a(t) \leq x\} = \mathbb{P}\{\bar{A}^\rho(t) = s, L^b(t) \leq x\}$$

holds.

Again, we now provide a numerical scheme to efficiently and accurately compute the time-dependent joint CDF.

3.2.1 Applying the Methods of Characteristics

The numerical scheme to compute the time-dependent joint CDF is again an upstream scheme based on the method of characteristics and follows essentially along the lines of the scheme presented for the marginal case. The relation $\prec$ defined in (2.9) implies a partial ordering on the state space $\bar{\mathcal{S}}^\rho$, and the matrix $\bar{Q}(t)$ is triangular with respect to this ordering. Thus, again, the values of F_s only depends on $F_{s'}$ with $s \prec s'$, and they can be computed for each s separately. For given $s \in \bar{\mathcal{S}}^\rho$,

$$
\begin{aligned}
F_s(t,x,y) &= \mathbb{P}\big\{\bar{A}^\rho(t) = s, L^a(t) \le x, L^b(t) \le y\big\} \\
&= \begin{cases}
0, & \text{if } x < v^a(s)\cdot t \text{ or } y < v^b(s)\cdot t, \\
\text{solution to (3.12)}, & \text{if } v^a(s)\cdot t \le x < n\cdot t \text{ and } v^b(s)\cdot t \le y < n\cdot t, \\
\mathbb{P}\big\{\bar{A}^\rho(t) = s, L^a(t) \le x\big\}, & \text{if } v^a(s)\cdot t \le x < n\cdot t \text{ and } n\cdot t \le y, \\
\mathbb{P}\big\{\bar{A}^\rho(t) = s, L^b(t) \le y\big\}, & \text{if } n\cdot t \le x \text{ and } v^b(s)\cdot t \le y < n\cdot t, \\
\mathbb{P}\big\{\bar{A}^\rho(t) = s\big\}, & \text{if } n\cdot t \le x \text{ and } n\cdot t \le y
\end{cases}
\end{aligned}
\tag{3.13}
$$

holds. Figure 2 shows the different regions of $F_s(t,x,y)$ for a fixed t. Moreover, for each $s \in \bar{\mathcal{S}}^\rho$, the PDE that has to be satisfied in the region $v^a(s)\cdot t \le x < n\cdot t$ and $v^b(s)\cdot t \le y < n\cdot t$ can be re-written as

$$
\partial_t F_s(t,x,y) + \big(v^a(s), v^b(s)\big)\nabla F_s(t,x,y) = F_s(t,x,y)\bar{Q}_{s,s}(t) + \sum_{s \prec s'} F_{s'}(t,x,y)\bar{Q}_{s',s}(t), \tag{3.14}
$$

where $\nabla f = (\partial_x f, \partial_y f)^\top$. Again, taking the derivative of $F_s(t,x,y)$ along the characteristics

$$
\tau \to \Big(t_0 + \tau, \mathbf{x}_0 + \tau \mathbf{v}(s)\Big)^\top,
$$

with $t_0 := \frac{1}{n}\max\{x_0, y_0\}$, $\mathbf{x}_0 := (x_0, y_0)$, and $\mathbf{v}(s) := \big(v^a(s), v^b(s)\big)$, yields the right-hand side of equation (3.14). Thus, $F_s(\cdot,\cdot,\cdot)$ satisfies the ODE

$$
\frac{d}{d\tau} F_s\Big(t_0 + \tau, \mathbf{x}_0 + \tau\mathbf{v}(s)\Big) = -q_s^{(2)}(\tau) F_s\Big(t_0 + \tau, \mathbf{x}_0 + \tau\mathbf{v}(s)\Big) + g_s^{(2)}(\tau),
$$

with

$$
q_s^{(2)}(\tau) = -\bar{Q}_{s,s}(t_0 + \tau)
$$

and

$$
g_s^{(2)}(\tau) = \sum_{s \prec s'} F_{s'}\Big(t_0 + \tau, \mathbf{x}_0 + \tau\mathbf{v}(s)\Big)\bar{Q}_{s',s}(t_0 + \tau).
$$

The characteristics for $F_s(t,x,y)$ are depicted in Figure 2. Like in the marginal case, this is a non-homogeneous linear first-order ODE and can be readily solved. The solution involves integrating $q_s^{(2)}(\tau)$, which leads to

$$
F_s\Big(t_0 + \tau, x_0 + \mathbf{v}(s)\tau\Big) = e^{-H_k^{(2)}(\tau)}\left(\int_0^\tau g_s^{(2)}(\alpha) e^{H_k^{(2)}(\alpha)} d\alpha + F_s\big(t_0, x_0\big)\right), \tag{3.15}
$$

with

$$
H_s^{(2)}(\tau) = \int_0^\tau q_s^{(2)}(\alpha) d\alpha = -\bar{Q}_{s,s}^\rho(u - t_0) - \bar{Q}_{s,s}^c\big(\Lambda(u) - \Lambda(t_0)\big). \tag{3.16}
$$

We provide the details of our numerical upstream scheme to efficiently and accurately compute solutions to equation (3.15) in Appendix B.2.

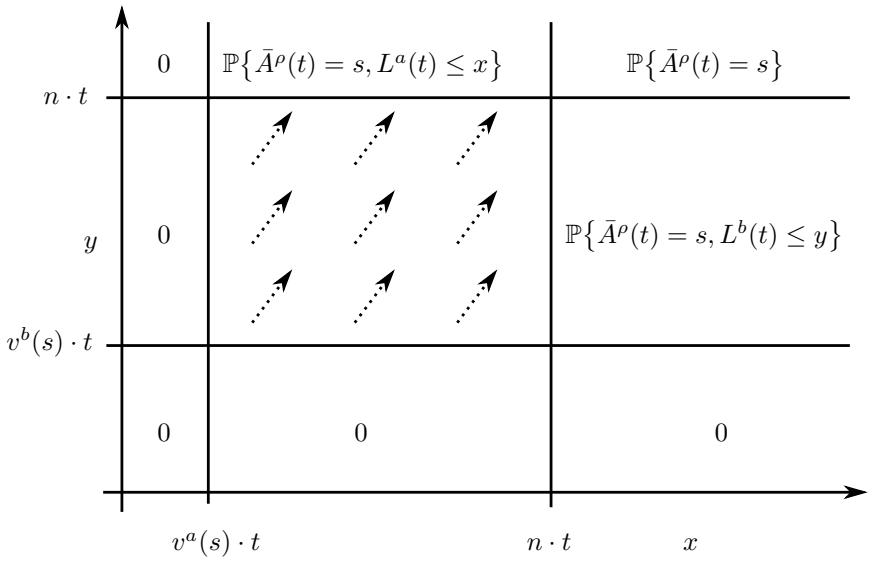


Figure 2: The different regions and (projected) characteristics of $F_s(t, x, y)$ for an intermediate state $s \in \bar{\mathcal{S}}^\rho$ at a given time t. The characteristics also extend in the t-direction at unit speed. Note that for the states s with $v^a(s) = n$ or $v^b(s) = n$ the interior region is empty.

4 Empirical evaluation

In this section, we demonstrate that the numerical algorithms presented in Section B.1 and B.2 can be used to accurately and efficiently compute the time-dependent marginal CDF (3.3) and joint CDF (3.4), as well as the regular marginal CDF (3.1) and joint CDF (3.2), for different population size histories and different recombination rates. Furthermore, we show how our method can be used to study properties of the marginal and joint distributions, and compute their moments. We implemented the numerical algorithms in MATLAB, and the code is available upon request.

For ease of exposition, we use a sample size of $n = 10$ in the remainder of this paper, except in one example, were we explicitly mention that we use $n = 5$. Furthermore, we consider a wide range of relevant recombination rates when studying the joint distribution. Moreover, we mainly focus on three population size histories, depicted in Figure 3. The first one is a history with an ancient bottleneck, followed by exponential growth up to the present. Specifically, for $t > 0.15$, the relative population size is set to 2, and for $0.025 < t < 0.15$, it is set to 0.25. Then, the population grows exponentially from size 0.25 at $t = 0.025$ up to $t = 0$ (the present), at an exponential rate of g. We refer to the coalescent-rate function under this population size history by λ_1, and if not mentioned otherwise, the growth rate is set to $g = 200$. This size history is a rough sketch of the human population size history, with an out-of-Africa bottleneck, followed by recent exponential growth at a rate of 1% per generation. In addition, we consider a pure bottleneck, where the relative ancestral size is 2 until time $t = 0.05$, and N_B from $t = 0.05$ until the present. We refer to the rate function under this population size history by λ_2, and if not otherwise mentioned, we set $N_\mathrm{B} = 0.2$. Finally, we refer to the rate function in a population that is of constant size 1 by λ_3.

4.1 Accuracy

In this section we demonstrate that the numerical algorithms presented in this paper can be used to compute the requisite CDFs accurately. Naturally, the accuracy will depend on the exact choice of the grid for the numerical algorithm. We will present results for a particular grid here, and discuss the issues for choosing

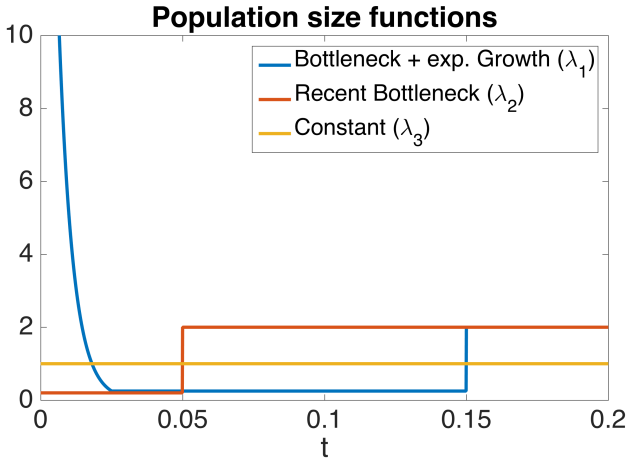
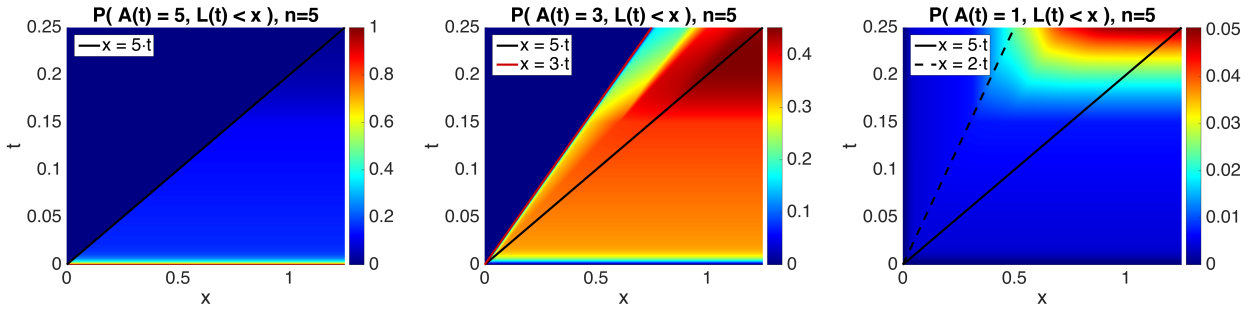


Figure 3: The three population size histories we will mainly consider in this paper: An ancient bottleneck followed by exponential growth (λ_1), a recent bottleneck (λ_2), and a constant population size (λ_3).



(a) $\mathbb{P}\{A(t) = 5, L(t) \leq x\}$ (b) $\mathbb{P}\{A(t) = 3, L(t) \leq x\}$ (c) $\mathbb{P}\{A(t) = 1, L(t) \leq x\}$

Figure 4: Heatmaps of $\mathbb{P}\{A(t) = k, L(t) \leq x\}$ as a function of t and x, for different k, computed using our numerical algorithm.

an adequate grid in Section 5. We set $n = 5$ and compute the time-dependent marginal CDF

$$\mathbb{P}\{A(t) = k, L(t) \leq x\}$$

for $k = 5$, 3, and the absorbing state 1, and show the respective surfaces as functions of t and x in Figure 4. Here we used the population size history with exponential growth λ_1. These surfaces exhibit the properties sketched in Figure 1, and the different regions can be observed. Below the line $x = nt$, the functions are independent of x. Furthermore, the functions are zero above the line $t = kt$, except for $k = 1$, where the function is independent of t above the line $x = 2t$. For different k, the time it takes to reach these states and the associated tree lengths show the expected distributions.

To verify that the values of the time-dependent CDFs $\mathbb{P}\{A(t) = k, L(t) \leq x\}$ and

$$\mathbb{P}\{\bar{A}^\rho(t) \in \Delta, L^a(t) \leq x, L^b(t) \leq y\}$$

are computed accurately, we estimated the respective values for different t, x, and y from simulations under the respective ancestral processes A and $\bar{A}^\rho$. To this end, we simulated a certain number of trajectories N, counted how many of the trajectories had accumulated a certain tree length by the given time, and divided

t	x	p	$\hat{p}$ ($N = 10^5$)	$\hat{p}$ ($N = 10^7$)
0.6	1.0	0.040947	0.040450 ($\pm$ 0.002)	0.040940 ($\pm$ 0.0002)
0.6	5.9	0.103601	0.103890 ($\pm$ 0.002)	0.103671 ($\pm$ 0.0002)
2.2	3.4	0.259234	0.257360 ($\pm$ 0.003)	0.259307 ($\pm$ 0.0003)
2.2	5.9	0.473988	0.473760 ($\pm$ 0.004)	0.474183 ($\pm$ 0.0004)
3.7	8.3	0.698705	0.698750 ($\pm$ 0.003)	0.698755 ($\pm$ 0.0003)

Table 1: Time-dependent CDF $\mathbb{P}\big\{A(t) = 1, L(t) \leq x\big\}$ for different values of t and x. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations using different numbers of trajectories N. The confidence bounds are indicated in parentheses.

t	x	y	p	$\hat{p}$ ($N = 10^5$)	$\hat{p}$ ($N = 10^7$)
0.6	1.2	2.4	0.053702	0.052500 ($\pm$ 0.002)	0.053600 ($\pm$ 0.0002)
1.2	2.4	2.4	0.153384	0.152730 ($\pm$ 0.003)	0.153199 ($\pm$ 0.0003)
1.2	3.6	4.8	0.239071	0.239300 ($\pm$ 0.003)	0.238984 ($\pm$ 0.0003)
2.2	4.4	8.8	0.362669	0.362290 ($\pm$ 0.003)	0.362342 ($\pm$ 0.0003)
3.7	7.4	11.1	0.640609	0.639080 ($\pm$ 0.003)	0.640461 ($\pm$ 0.0003)
3.7	14.8	14.8	0.755157	0.755310 ($\pm$ 0.003)	0.755093 ($\pm$ 0.0003)

Table 2: Time-dependent joint CDF $\mathbb{P}\big\{\bar{A}^\rho(t) \in \Delta, L^a(t) \leq x, L^b(t) \leq y\big\}$ for different values of t, x, and y. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations with different numbers of trajectories N. The confidence bounds are indicated in parentheses.

this by the total number of trajectories. By the central limit theorem, this converges to the true probability, and it also allows us to provide confidence bounds for the estimates obtained using a certain number of trajectories. Some values computed using our algorithm and the respective estimates from simulations are shown in Table 1 for the marginal time-dependent CDF, and in Table 2 for the joint time-dependent CDF. We present these values for the absorbing states $k = 1$ and $s \in \Delta$, respectively. Here we used $n = 10$, the exponential growth model for the population size history λ_1, and recombination rate $\rho = 0.001$. The tables show that the values computed using our algorithm always fall into the confidence bounds of the estimates from the simulations using different numbers of trajectories, thus demonstrating that our algorithm computes the respective values accurately.

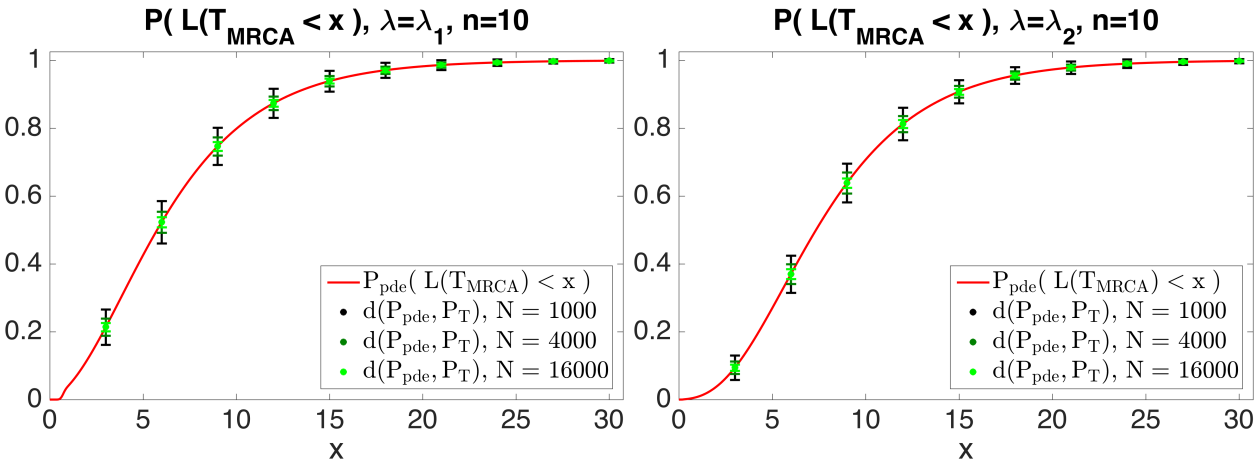
The marginal and the joint CDF of the total tree length

$$\mathbb{P}\{\mathcal{L} \leq x\}$$

and

$$\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$$

can be computed from the respective time-dependent CDFs. To demonstrate the accuracy of our numerical algorithms, we again compared the numerical values to values obtained from simulations. Like before, we simulated a certain number of trajectories, and estimated the respective probabilities. Figure 5 shows the marginal CDFs for $n = 10$ under exponential growth (λ_1) and the bottleneck scenario (λ_2). Furthermore, the simulations can be used to bound the difference $d(P_{\text{pde}}, P_{\text{T}})$ between the value computed using the numerical scheme P_{pde} and the true value P_{T}, since the values estimated from simulations converge to P_{T} as $N \to \infty$. This bound is also indicated in Figure 5 for different values of N and decreases as N gets larger, as expected. For the joint CDF, we present values computed using our numerical procedure for different x and y, and compare them to simulated values, including confidence bounds. We set $n = 10$, and used $\rho = 0.001$. The values for the model with exponential growth (λ_1) are shown in Table 3, and for the bottleneck scenario (λ_2) in Table 4. Again, the values computed using the numeric algorithm always fall into the confidence



(a) $\mathbb{P}\{\mathcal{L} \leq x\}$ under exponential population growth (λ_1). (b) $\mathbb{P}\{\mathcal{L} \leq x\}$ in the bottleneck scenario (λ_2).

Figure 5: The CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ as a function of x is depicted by the red line. Additionally, the green bars indicate the bound on the distance between the numerical value P_{pde} and the true value P_{T} for different N, thus the true value is guaranteed to fall within these bounds.

x	y	p	$\hat{p}$ $(N = 256,000)$	$\hat{p}$ $(N = 16,384,000)$
1.5	3.0	0.075326	0.074914 ($\pm$ 0.002)	0.075030 ($\pm$ 0.0002)
3.0	6.0	0.213703	0.213324 ($\pm$ 0.002)	0.213565 ($\pm$ 0.0002)
6.0	6.0	0.522821	0.521578 ($\pm$ 0.002)	0.522707 ($\pm$ 0.0003)
12.0	18.0	0.873357	0.872840 ($\pm$ 0.002)	0.873319 ($\pm$ 0.0002)
30.0	30.0	0.998499	0.998504 ($\pm$ 0.0002)	0.998516 ($\pm$ 0.00002)

Table 3: The CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ for different values of x and y under λ_1, with $n = 10$ and $\rho = 0.001$. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations for different N. The confidence bounds are indicated in parentheses.

bounds. In these tables, it becomes particularly apparent that in order to guarantee a high accuracy using simulations, a very large number of trajectories needs to be simulated, which is very time-consuming. Our numerical scheme yields a high accuracy, and does not suffer from these issues.

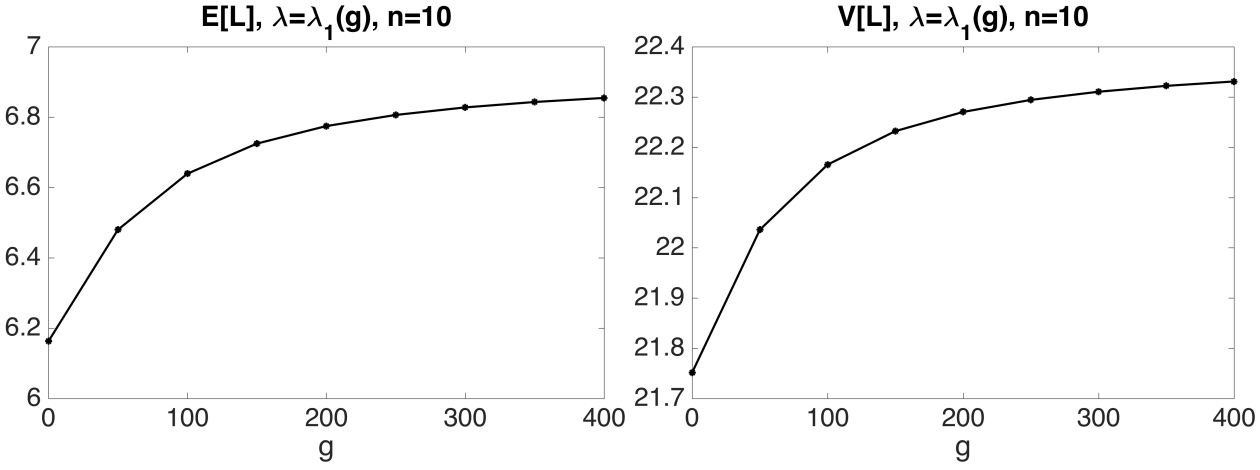
4.2 Properties of the Distributions

The results provided in the previous section show that our numerical algorithm can be used to accurately and efficiently compute the marginal and joint CDF of the total tree length in populations with variable size. We will now demonstrate the utility of our numerical method to study properties of the respective distributions.

The numerical values of the marginal CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ can be readily applied to compute approximations of the expected value and the variance of the total tree length $\mathcal{L}$. Figure 6 shows the different values of the expectation and the variance under exponential growth (λ_1), with varying growth-rates g. A rate of $g = 0$ corresponds to no growth, and exhibits the smallest expected value. As we increase the growth-rate, the expected value increases as well. This is to be expected, as with increasing growth-rate, the recent population size increases, and coalescent happens at smaller rates. The lineages stay separated for longer and accumulate towards the total length at a higher total rate. However, as the growth-rate increases further,

x	y	p	$\hat{p}$ $(N = 256,000)$	$\hat{p}$ $(N = 16,384,000)$
1.5	3.0	0.019794	0.019238 ($\pm$ 0.0006)	0.019579 ($\pm$ 0.00007)
3.0	6.0	0.094393	0.094414 ($\pm$ 0.002)	0.094172 ($\pm$ 0.0002)
6.0	6.0	0.369581	0.369059 ($\pm$ 0.002)	0.369544 ($\pm$ 0.0003)
12.0	18.0	0.812236	0.812328 ($\pm$ 0.002)	0.812109 ($\pm$ 0.0002)
30.0	30.0	0.997696	0.997922 ($\pm$ 0.0002)	0.997721 ($\pm$ 0.00003)

Table 4: The CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ for different values of x and y under λ_2, with $n = 10$ and $\rho = 0.001$. p is computed using the numeric algorithm, and $\hat{p}$ is estimated from simulations for different N. The confidence bounds are indicated in parentheses.



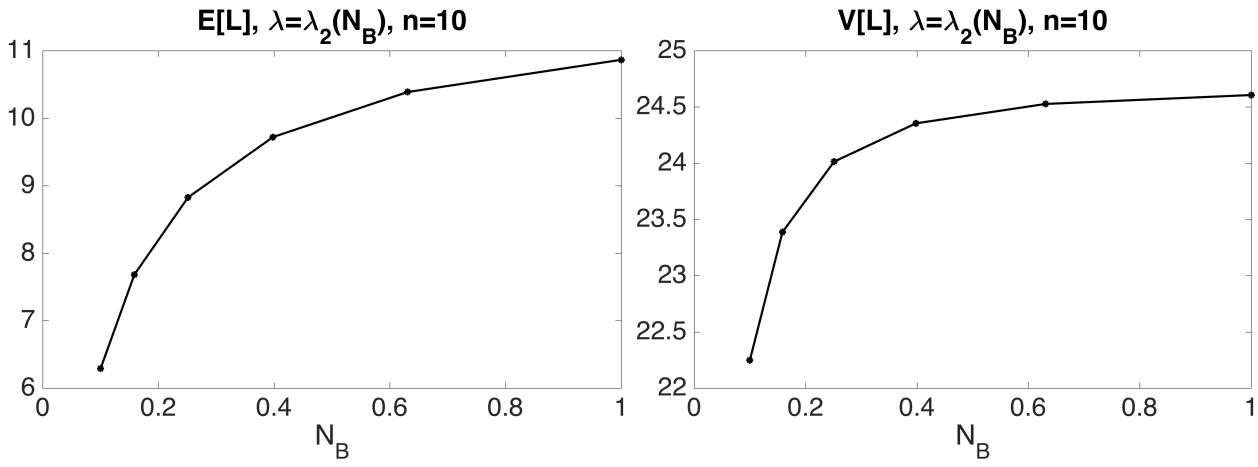
(a) The expected value of $\mathcal{L}$.

(b) The variance of $\mathcal{L}$.

Figure 6: Approximations to the expected value and the variance of $\mathcal{L}$ computed using our numerical procedure. Here we used the model of exponential population growth (λ_1), with different values for the growth-rate g.

the expected value seems to level off. Again, this is reasonable, since once the population is large enough for the lineages to stay separated the entire duration of the growth period, increasing the population size further will not affect the total tree length significantly anymore. The functional form of the variance looks similar to the expected value, with a slightly sharper increase in the beginning. Figure 7 shows the expected value and the variance under the bottleneck model (λ_2) for different values of the bottleneck size N_B. The plotted values follow a similar trend as in the exponential growth case, since the recent population size increases as well with the bottleneck size. However, the absolute value of these moments increases to substantially larger values before it levels off. This is due to the fact that in the exponential-growth model, the population size during the ancient bottleneck is set to 0.25, whereas in the bottleneck scenario, the size during the complete bottleneck increases with N_B. Thus, in the later case, the whole tree is not subject to a severe bottleneck if N_B is large, and the lineages stay separated longer.

Figure 8 shows the joint CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ as a function of x and y for different population size scenarios and different recombination rates ρ, computed on a suitable grid using our numerical algorithm. Naturally, the CDF converges towards 1 as x and y increase, and due to the symmetry of the ancestral process $\bar{A}^\rho$ the CDF is symmetric when interchanging x and y. Furthermore, note that the isolines in the plots for $\rho = 0.0001$ show pronounced right angles along the line $x = y$. This is due to the fact, that for small ρ the trees at the two loci are highly correlated. When $\mathcal{L}^a$ and $\mathcal{L}^b$ are highly correlated, then their joint CDF



(a) The expected value of $\mathcal{L}$.

(b) The variance of $\mathcal{L}$.

Figure 7: Approximations to the expected value and the variance of $\mathcal{L}$, under the bottleneck model (λ_2), with different values for the bottleneck size N_B.

is essentially a function of $\min(x, y)$, and this leads to the pronounced right angle along $x = y$ and the fact that the isolines are nearly parallel to the x- and y-axis. As the recombination rate increases, the two tree lengths become increasingly uncorrelated, and the pattern disappears. For different population size histories, the CDFs exhibit different patterns, however, explaining these patterns by the population size history does not seem straightforward. In all plots, the isoline for 0.2 is around $x = y = 5$, for the case λ_1 even lower. Thus, under λ_1, there is an elevated probability for very short trees. This seems to be due to the fact that under λ_1, there is a long bottleneck with a very small population size, which favors short trees. However, under the constant population size model λ_3, the CDF increases rapidly as x and y increase, whereas the function is less steep for λ_1 and λ_2. This behavior seems to be dominated by the ancient population sizes. Under λ_3, the ancient population size is 1, whereas it is 2 under λ_1 and λ_2, thus allowing for very long trees with a higher probability in the latter case.

Finally, we employ our numerical values of the joint CDF on the specified grid obtained using our algorithm to compute an approximation to the correlation coefficient between the tree lengths

$$\operatorname{corr}(\mathcal{L}^a, \mathcal{L}^b) := \frac{\operatorname{cov}(\mathcal{L}^a, \mathcal{L}^b)}{\sqrt{\mathbb{V}\mathcal{L}^a}\sqrt{\mathbb{V}\mathcal{L}^b}},$$

where $\operatorname{cov}(\cdot, \cdot)$ denotes the covariance. This coefficient is 1 for completely correlated random variables and 0, if they are independent. Figure 9 shows this correlation coefficient under the population size history λ_1 and λ_2 for different values of ρ, for a sample of size $n = 10$. Recall that our numerical procedure was built on the approximate ancestral process $\bar{A}^\rho$ for computational efficiency, where we limited the number of recombination events to 1. To compare the correlation under the process $\bar{A}^\rho$ with the correlation under the regular ancestral process with recombination A^ρ, we also plotted estimated values for the latter. We obtained these estimates from repeated simulations using the widely applied coalescent-simulation tool `ms` (Hudson, 2002) that is based on the regular coalescent with recombination (using $N = 10^5$ repetitions). Naturally, the correlation is close to 1 for small recombination rates. As the recombination rate increases, the correlation decreases, both under the approximate process $\bar{A}^\rho$, as well as the regular process A^ρ. The lines are basically indistinguishable until they start separating around $\rho = 0.05$. This is to be expected, since the approximation we introduced limits the number of recombination events to 1, and thus as the recombination rate increases, the approximation error also increases. However, Figure 9 shows that the approximate process can be used without loss in accuracy for a large range of recombination rates relevant for human genetics, where average

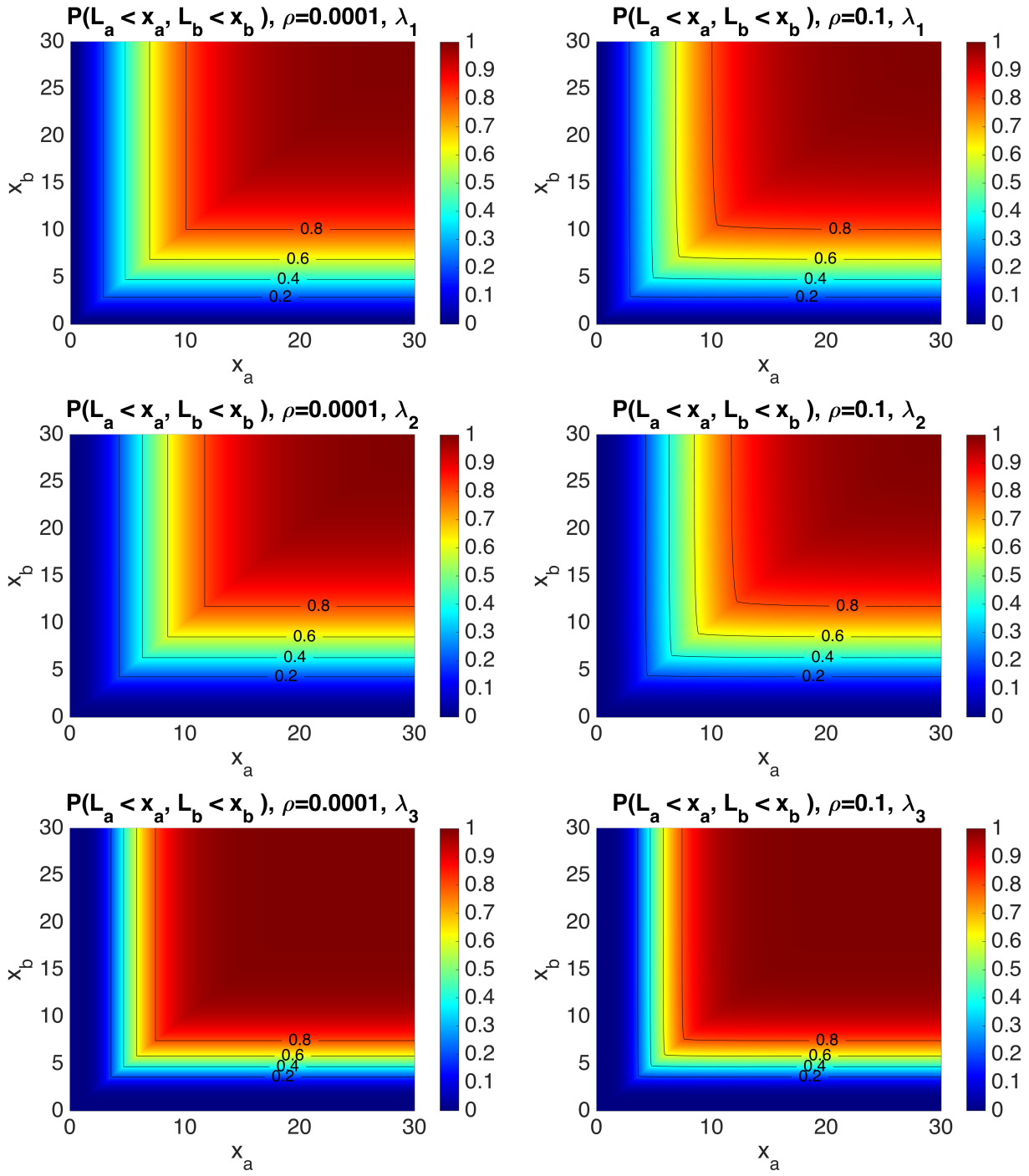
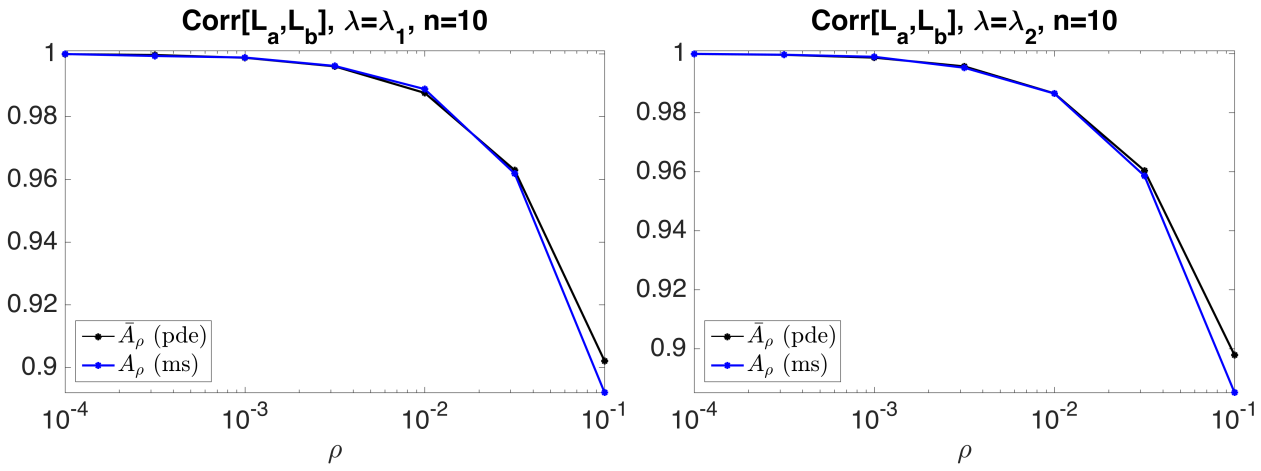


Figure 8: The joint CDF $\mathbb{P}\{\mathcal{L}^a \leq x, \mathcal{L}^b \leq y\}$ for exponentially growing populations (λ_1, first row), the bottleneck scenario (λ_2, middle row), and a constant population size (λ_3, last row). The recombination rate is set to $\rho = 0.0001$ in the first column, and $\rho = 0.1$ in the second column. Again, we use $n = 10$.



(a) Exponential growth scenario (λ_1).

(b) Bottleneck scenario (λ_2).

Figure 9: The correlation between the total tree length $\mathcal{L}^a$ and $\mathcal{L}^b$, under different population size histories for varying values of ρ. The black lines show the values computed using our method under $\bar{A}^\rho$, and the blue lines show value estimated from coalescent simulation under A^ρ using the popular tool `ms`.

recombination rates between neighboring sites are on the order of 10^{-3}.

5 Discussion

In this paper, we presented a novel computational framework to compute the marginal and joint CDF of the total tree length in populations with variable size. To our knowledge, these distributions have not been addressed in the literature before, especially in populations of variable size. We introduced a system of linear hyperbolic PDEs and showed that the requisite CDFs can be obtained from the solution of this system. We introduced a numerical algorithm to compute the solution of this system based on the method of characteristics and demonstrated its accuracy in a wide range of biologically relevant scenarios.

The numerical algorithm that we introduced is an upstream-method that computes the requisite solutions step-wise on a grid. We presented the algorithm for a regular, equidistantly spaced grid. We used the trapezoidal rule for the integration steps in the method, and also used linear interpolation to interpolate values that do not fall onto the specified grid. We introduced the regular grid, and the fairly simple integration and interpolation schemes for ease of exposition, and arrived at an efficient and accurate algorithm. Using higher order interpolation and integration schemes, combined with adaptive grids that have more points in regions where the coalescent-rate function is large will most certainly increase the accuracy. However, such higher order schemes come with additional computational cost. This opens numerous avenues for future research to optimize the balance between accuracy and efficiency that is required in the respective application.

Moreover, for reasons of computational efficiency, we introduced an approximation $\bar{A}^\rho$ to the regular ancestral process with recombination A^ρ, were we limited the number of recombination events, and computed the joint CDF under this approximate process. We demonstrated that this approximation is accurate for a large range of relevant recombination rates. It is straightforward to increase the number of recombination events in the approximation to gain additional accuracy, but computing the joint CDF under the regular ancestral process is desirable. The reason for the computational efficiency is the fact that the generator matrix for $\bar{A}^\rho$ is triangular. However, we will show in an upcoming paper that Proposition A.14 also holds for non-triangular matrices, and the numerical scheme can be adjusted accordingly to compute the CDF

under the regular process.

Another research direction is to use our novel framework to study higher order correlations between trees at multiple loci. On the one hand, this could again be correlations between the total tree lengths, but is also conceivable to include the distribution of other summary statistics of the genealogical trees. Our framework is flexible enough to compute the distribution of multiple path integrals along the trajectories of a given Markov chain. Thus, to implement these additions, one needs to define and implement an appropriate ancestral process and compute suitable integrals along the trajectories.

In this paper, we studied the ancestral process in a single panmictic population. However, in recent years, researchers have gathered an increasing amount of genomic datasets that contain individuals from multiple sub-populations, and studied historical events like migration or population subdivision using these datasets. In light of these studies, it is important to augment our framework to compute joint CDFs of the total tree length in structured populations with complex migration histories. Again, this can be done by introducing suitable ancestral processes and suitable integrals along their trajectories.

Acknowledgements

We thank Yun S. Song for numerous helpful discussions that sparked many of the ideas presented in this paper. This research is supported in part by a National Institutes of Health grant R01-GM094402 (M.S.). We also thank Robin Young for helpful suggestions and fruitful discussions relevant to the design of the numerical scheme for the hyperbolic system present in this paper.

Appendix

A Path-integrals of Markov chains

Since the marginal and joint distributions of the tree length can be obtained by integrating a certain function of the ancestral processes, we now consider distributions of path integrals for Markov chains defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will be using the following assumptions throughout the section.

(A1) $\{X(t,\omega), t \in \mathbb{R}_+, \omega \in \Omega\}$ is a regular jump Markov process with values in $\mathcal{S}_n = \{1, 2, \dots, n\}$, satisfying

$$\mathbb{P}\{X(t+h) = j | X(t) = i\} = q_{ij}(t)h + o(h)\,, \quad i, j \in \mathcal{S}_n \quad \text{as} \quad h \to 0^+.$$

We assume that the trajectories of $t \to X(\cdot, \omega)$ are right-continuous.

(A2) The infinitesimal generator $Q(t) = \{q_{ij}(t)\}_{i=1^n}$ is conservative, that is

$$q_i(t) := -q_{ii}(t) = \sum_{i \neq j} q_{ij}\,,$$

and satisfies $Q \in C(\mathbb{R}_+; M^{n\times n}) \bigcap L^\infty(\mathbb{R}_+; M^{n\times n})$. In addition, for each $i \in \mathcal{S}_n$ either $q_i(t) = 0$ for all $t \geq 0$ or $q_i(t) > 0$ for all $t \geq 0$. In the latter case, we require $\int_0^\infty q_i(s)\, ds = \infty$.

Definition A.1. *Let $X(t)$ satisfy (A1)-(A2). Given a function $v : \mathcal{S}_n \to \mathbb{R}$ we define a path-integral over the interval $[0, t]$ by*

$$L^v(t, \omega) = \int_0^t v(X(s, \omega))\, ds\,, \quad t \in \mathbb{R}_+.$$

Definition A.2. *Let $X(t)$ satisfy (A1)-(A2) and $v : \mathcal{S}_n \to \mathbb{R}$ be some real-valued function defined on the state space. We define a distribution vector-function associated with $\big(X(t), L^v(t)\big)$ by*

$$\mathbf{F}^v = \big(F_1^v, \dots, F_n^v\big) : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+^n \quad \textit{with} \quad F_k^v(x,t) = \mathbb{P}\Big\{X(t) = k\,, L^v(t) \leq x\Big\}\,, \;\; k \in \{1, \dots, n\}\,. \tag{A.1}$$

Remark A.3. Since $X(t)$ is a regular jump process and hence it is separable. Thus, for each $t_0 \in \mathbb{R}_+$ the random variable $L_v(t_0, \cdot)$ is well-defined and $\mathcal{F}$-measurable, and for each $\omega_0 \in \Omega$ the map $t \to L(t, \omega_0)$ is Lipschitz continuous. This in turn implies that the process $L(t)$ is measurable and separable (see Chapter 12 of Koralov and Sinay (2007) and Chapter 2 of Doob (1953)).

Proposition A.4 (Locally Lipschitz)**.** *Let (A1)-(A2) hold. Let $v : \mathcal{S}_n \to \mathbb{R}$ and $\mathbf{F}^v$ be defined by* (A.1)*. Suppose that $\mathbf{F}^v$ is Lipschitz continuous in some open $\mathcal{U} \subset \mathbb{R} \times (0, \infty)$. Then*

$$\partial_t \mathbf{F}^v(x,t) + V \partial_x \mathbf{F}^v(x,t) = \mathbf{F}^v(x,t) Q(t) \quad a.e.\ in\ \ \mathcal{U}\,, \tag{A.2}$$

where $V = \texttt{diag}(v_1, \dots, v_n)$ and $v_k = v(k)$.

Proof. Let us suppress the subscript v in our calculations. Let $\widetilde{\mathcal{U}}$ denote the set of all points in $\mathcal{U}$ at which $\mathbf{F}$ is differentiable. Since $\mathbf{F}$ is Lipschitz continuous in $\mathcal{U}$, Rademachers' Theorem (Federer, 1969) implies that $\mathbf{F}$ is Lebesgue almost surely differentiable in $\mathcal{U}$ and therefore $\mathcal{U} \backslash \widetilde{\mathcal{U}}$ is of Lebesgue measure zero. Take any $k \in \mathcal{S}_n$. Fix any $(x,t) \in \widetilde{\mathcal{U}}$. For any $\varepsilon > 0$ we have

$$F_k(x + \varepsilon v_k, t+\varepsilon) - F_k(x,t) = \Big(\sum_{i=1}^n \mathbb{P}\{X(t) = i\,, X(t+\varepsilon) = k\,, L(t+\varepsilon) \le x + \varepsilon v_k\} \Big) - F_k(x,t). \tag{A.3}$$

Consider first the terms with $i \neq k$. By Lemma A.7

$$\begin{aligned}
&\frac{1}{\varepsilon} \mathbb{P}\big\{X(t+\varepsilon) = k, X(t) = i\big\} F_i(x + \varepsilon v_k - M\varepsilon, t) \\
&\quad \le \frac{1}{\varepsilon} \mathbb{P}\big\{X(t) = k, X(t+\varepsilon) = i\,, L(t+\varepsilon) \le x + \varepsilon v_k\big\} \mathbb{P}\{X(t) = i\} \\
&\quad \le \frac{1}{\varepsilon} \mathbb{P}\big\{X(t+\varepsilon) = k, X(t) = i\big\} F_i(x + \varepsilon v_k - m\varepsilon, t)
\end{aligned}$$

where $m = \min_{j \in \mathcal{S}_n} v(j)$ and $M = \max_{j \in \mathcal{S}_n} v(j)$. Since Q is continuous we must have

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \mathbb{P}\big\{X(t+\varepsilon) = i, X(t) = k\big\} = q_{ki}(t) \mathbb{P}\{X(t) = k\}$$

and hence, employing the continuity of F, we conclude

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \mathbb{P}\Big\{X(t) = i, X(t+\varepsilon) = k,\ L(t+\varepsilon) \le x + \varepsilon v_k\Big\} = q_{ki}(t) F_i(x,t)\,. \tag{A.4}$$

Next consider the case $i = k$. By Lemma A.8 we have

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \Big(\mathbb{P}\big\{X(t) = k, X(t+\varepsilon) = k,\ L(t+\varepsilon) \le x + \varepsilon v_k\big\} - q_k(t) F_k(x,t) \Big) = -q_k(t) F_k(x,t)\,. \tag{A.5}$$

Combining (A.3) with (A.4) and (A.5) we obtain

$$\begin{aligned}
&\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \big(F_k(x + \varepsilon v_k, t+\varepsilon) - F_k(x,t) \big) \\
&\qquad = -q_k(t) F_k(x,t) + \sum_{i \neq k} q_{ik}(t) F_i(x,t) = \big(F(x,t) Q(t) \big)_k\,.
\end{aligned} \tag{A.6}$$

Since $\mathbf{F}$ is differentiable at $(x,t) \in \widetilde{\mathcal{U}}$ and the map $\varepsilon \to (x + \varepsilon v_k, t + \varepsilon)$ is differentiable with the image contained in $\mathcal{U}$ for sufficiently small ε, the chain rule is applicable (see Rudin (1976)[Theorem 9.15]) and we conclude

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \Big(F_k(x + \varepsilon v_k, t+\varepsilon) - F_k(x,t) \Big) = \frac{d}{d\varepsilon} F(x + \varepsilon v_k, t + \varepsilon) \Big|_{\varepsilon = 0} = \partial_t F_k(x,t) + v_k \partial_x F_k(x,t).$$

Since both $k \in \mathcal{S}_n$ and $(x,t) \in \widetilde{\mathcal{U}}$ were arbitrary, (A.6) imply (A.2). □

We next show that $\mathbf{F}^v$ as $t \to 0^+$ has certain continuity properties.

Proposition A.5 (Initial values)**.** *Let (A1)-(A2) hold. Let $v : \mathcal{S}_n \to \mathbb{R}$ be nonnegative and $\mathbf{F}^v$ be defined by* (A.1)*. Then for any $x \neq 0$*

$$\lim_{t\to 0^+} F_k^v(x,t) = \mathbb{1}_{\{x\geq 0\}}\mathbb{P}\{X(0)=k\} = F_k^v(x,0)\,.$$

Moreover, the convergence is uniform in the x-variable on $\mathbb{R}\backslash(-\delta,\delta)$ for any $\delta > 0$.

Proof. First, we note that $L(0) = 0$ for all $\omega \in \Omega$ and hence $F_k^v(x,0) = \mathbb{1}_{\{x\geq 0\}}\mathbb{P}\{X(0)=k\}$.

Now take any $\delta > 0$. Observe that

$$m^v t \leq L(t) = \int_0^t v(X(s))\,ds \leq M^v t$$

where $m^v = \min_{j\in\mathcal{S}_n} v(j)$ and $M^v = \max_{j\in\mathcal{S}_n} v(j)$. Thus for all $0 < t < \delta(1+\max(|m^v|,|M^v|))^{-1}$ and every x such that $|x| > \delta$ we have

$$\begin{aligned} F_k(x,t) &= \mathbb{1}_{\{x\geq 0\}}\mathbb{P}\{X(t)=k, L(t)\leq x\} \\ &= \mathbb{1}_{\{x\geq 0\}}\mathbb{P}\{X(t)=k\} \to \mathbb{1}_{\{x\geq 0\}}P\{X(0)=k\} \quad \text{as} \quad t\to 0^+ \end{aligned}$$

and this proves uniform convergence on the set $|x| > \delta$. □

Remark A.6. *From Proposition A.5 it follows that the 'initial values' of $\mathbf{F}^v$ are discontinuous. Since the system* (A.2) *a is linear hyperbolic system, discontinuities present at time $t = 0$ will travel in space as time t increases and therefore $\mathbf{F}^v$ is not C^1 or even continuous. Nevertheless, one can show that* (A.2) *holds in a weaker sense. To do that one needs to employ the notion of weak solutions, and we will pursue this avenue in an upcoming paper. However, for certain type of state space functions v, relevant to our application, one can show additional regularity properties of $\mathbf{F}^v$. We provide more details in Appendix A.1.*

Lemma A.7. *Let $X(t)$ satisfy (A1)-(A2), $v : \mathcal{S}_n \to \mathbb{R}$, and let $m^v = \min_{j\in\mathcal{S}_n} v(j)$ and $M^v = \max_{j\in\mathcal{S}_n} v(j)$. For any $i,k \in \mathcal{S}_n$, any $\varepsilon > 0$, and each $x \in \mathbb{R}$*

$$\begin{aligned} &\mathbb{P}\big\{X(t)=i,\, X(t+\varepsilon)=k\big\}F_i^v(x-M^v\varepsilon,t) \\ &\quad\leq \mathbb{P}\big\{X(t)=i, X(t+\varepsilon)=k\,, L^v(t+\varepsilon)\leq x\big\}\mathbb{P}\{X(t)=i\} \\ &\quad\leq \mathbb{P}\big\{X(t)=i\,, X(t+\varepsilon)=k\big\}F_i^v(x-m^v\varepsilon,t)\,. \end{aligned}$$

Proof. We suppress the subscript v in our calculations. Take any $\varepsilon > 0$. Observe that

$$L(t) + m\varepsilon \leq L(t+\varepsilon) = L(t) + \int_t^{t+\varepsilon} v(X(s))\,ds \leq L(t) + M\varepsilon$$

and therefore

$$\begin{aligned} &\mathbb{P}\big\{X(t)=i,\, X(t+\varepsilon)=k,\, L(t)\leq x - M\varepsilon\big\} \\ &\quad\leq \mathbb{P}\big\{X(t)=i,\, X(t+\varepsilon)=k,\, L(t+\varepsilon)\leq x\big\} \\ &\quad\leq \mathbb{P}\big\{X(t)=i,\, X(t+\varepsilon)=k,\, L(t)\leq x - m\varepsilon\big\}\,. \end{aligned}$$

Let $y \in \mathbb{R}$. Suppose that $0 < F_i(y,t)$. Observe that for any $t_0 > 0$ the path-integral $L(t_0)$ is fully determined by $\big(X(t), t\in[0,t_0)\big)$. This fact (and the separability of the process) enables us to use Markov property and we obtain

$$\begin{aligned} &\mathbb{P}\big\{X(t)=i,\, X(t+\varepsilon)=k,\, L(t)\leq y\big\} \\ &= \mathbb{P}\big\{X(t+\varepsilon)=k | X(t)=i, L(t)\leq y\big\}\mathbb{P}\big\{X(t)=i, L(t)\leq y\big\} \\ &= \mathbb{P}\big\{X(t+\varepsilon)=k | X(t)=i\big\}F_i(y,t)\,. \end{aligned}$$

This together with the previous inequality finishes the proof. □

Lemma A.8. *Let $X(t)$ satisfy (A1)-(A2) and $v : \mathcal{S}_n \to \mathbb{R}$. For every $(x,t) \in \mathbb{R} \times \mathbb{R}_+$ and each $k \in \mathcal{S}_n$*

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \Big| \mathbb{P}\Big\{ X(t) = k, X(t+\varepsilon) = k, L^v(t+\varepsilon) \le x + \varepsilon v_k \Big\} - \exp\Big(- \int_t^{t+\varepsilon} q_k(s)\, ds \Big) F_k(x,t) \Big| = 0$$

Proof. Due to (A1) the process $X(t)$ is separable and therefore (see (Karlin, 1981, p. 146))

$$\mathbb{P}\Big\{ X(s) = k, s \in [t, t+\varepsilon] \Big\} = \exp\Big\{ - \int_t^{t+\varepsilon} q_i(s)\, ds \Big\} \mathbb{P}\big\{ X(t) = k \big\}. \tag{A.7}$$

Suppose now that $F_k(x,t) > 0$. Then, using Markov property, we obtain

$$\begin{aligned}
&\mathbb{P}\big\{ X(s) = k\,, s \in [t, t+\varepsilon]\,, L(t+\varepsilon) \le x + \varepsilon v_k \big\} \\
&= \mathbb{P}\big\{ X(s) = k\,, s \in [t, t+\varepsilon] \,|X(t) = k, L(t) \le x \big\} \mathbb{P}\Big\{ X(t) = k, L(t) \le x \Big\} \\
&= \mathbb{P}\big\{ X(s) = k\,, s \in [t, t+\varepsilon] | X(t) = k \big\} F_k(x,t) = \exp\Big(- \int_t^{t+\varepsilon} q_k(s)\, ds \Big) F_k(x,t)\,.
\end{aligned}$$

If $F_k(x,t) = 0$ then the first and the last terms in the above identity are zero. Thus, employing (A.7) and recalling that Q is continuous we conclude

$$\begin{aligned}
&\frac{1}{\varepsilon} \Big| \mathbb{P}\Big\{ X(t) = k, X(t+\varepsilon) = k, L(t+\varepsilon) \le x + \varepsilon v_k \Big\} - \exp\Big(- \int_t^{t+\varepsilon} q_k(s)\, ds \Big) F_k(x,t) \Big| \\
&\quad \le \Big| \frac{1}{\varepsilon} \Big(\mathbb{P}\big\{ X(t) = k, X(t+\varepsilon) = k \big) - P(X(t) = k\} \Big) \\
&\qquad\qquad - \frac{1}{\varepsilon} \Big(\exp\Big(- \int_t^{t+\varepsilon} q_i(s)\, ds \Big) - 1 \Big) \mathbb{P}\big\{ X(t) = k \big\} \Big| \to 0 \quad \text{as} \quad \varepsilon \to 0^+.
\end{aligned}$$

□

A.1 Path-integrals for monotone state space functions

Hyperbolic systems of partial differential equations admit in general solutions that are not classical even if the initial (or boundary data) is smooth. Typically there are two distinct classes of solutions: strong solutions, which are Lipschitz continuous (see Dafermos (2010)), and weak solutions, which allow for discontinuities. Here, we will be using the first type of solutions.

Definition A.9. *Let $\mathcal{U} \subset \mathbb{R} \times \mathbb{R}_+$ be open and let $A, B \in L^\infty(\mathcal{U}; \mathbb{R}^{n \times n})$. We say that $u(x,t) : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}^n$ is a strong solution of*

$$\partial_t u(x,t) + \partial_x \big(A(x,t) u \big) = B(x,t) u(x,t) \quad \text{in} \quad \mathcal{U} \subset \mathbb{R}, \tag{A.8}$$

if u is Lipschitz continuous in $\mathcal{U}$, and the equation (A.8) *holds for Lebesgue almost all points (x,t) in $\mathcal{U}$.*

Remark A.10. *By the Rademacher' Theorem a function $u(x,t)$ that is Lipschitz continuous in an open domain $\mathcal{U}$ is Lebesgue almost sure differentiable in $\mathcal{U}$. In fact, its pointwise partial derivatives, which exist almost everywhere, coincide with its corresponding weak partial derivatives (see Evans (2010)).*

Remark A.11. *Regularity of solutions to hyperbolic problems depends on both the initial (or boundary) data and the domain itself. For linear hyperbolic problems as long as the initial data is smooth and the domain has a smooth boundary one may expect a solution to be (locally) smooth. Typically one studies the solutions to a hyperbolic problem on the domain $\mathcal{U} = \mathbb{R}_+ \times \mathbb{R}$ with initial data $u_0(x)$ at $t = 0$ (Cauchy problem). The initial data for the vector of probabilities $\mathbf{F}^v$ (studied in $\mathcal{U}$) are unfortunately discontinuous (which is shown below). To avoid unnecessary difficulties, in Proposition A.14 we split the upper-half plane into two regions*

$\mathcal{U}_l$ and $\mathcal{U}_r$. In $\mathcal{U}_r$ the values of $\mathbf{F}^v$ admit a simpler form while in $\mathcal{U}_r$ the vector $\mathbf{F}^v$ is obtained via solving a linear hyperbolic system with smooth initial data. We note that components of $\mathbf{F}^v$ are in general merely Lipschitz continuous in $\mathcal{U}_r$. This is not surprising for two reasons. Firstly, the domain is singular because it has a 'corner' and the discontinuities of the derivatives of $\mathbf{F}^v$ originating at $(x,t)=(0,0)$ travel along the corresponding characteristics. Secondly, the vector $\mathbf{F}^v$ solves the same system of equations in the domain $\mathcal{U}$ with discontinuous initial data and hence it is in general not smooth.

Remark A.12. *In Proposition A.14 we assume that the generator $Q(t)$ of a process $X(t)$ is upper triangular and that $X(0)=1$. In that setup, after each transition the process $X(t)$ increases the value of its state. This setup is more conveninent when dealing with general processes on a finite state space. To apply Proposition A.14 to the ancestral process $A(t)$ one must renumerate its states as follows $n\to 1$, $(n-1)\to 2$, $\dots$, $1\to n$.*

Definition A.13. *Let (A1)-(A2) hold. Let $v:\mathcal{S}_n=\{1,2,\dots,n\}\to\mathbb{R}$. We say that v is monotone along the process $X(t)$ if the map $t\to v(X(t,\omega))$ is either nonincreasing for $\mathbb{P}$-almost all $\omega\in\Omega$ or nondecreasing for $\mathbb{P}$-almost all $\omega\in\Omega$.*

Proposition A.14. *Let $X(t)$ satisfy (A1)-(A2), $X(0)=1$, and $Q(t)$ be upper triangular. Let $v:\mathcal{S}_n\to\mathbb{R}$ be monotone along $X(t)$. Suppose that $t\to v(X(t))$ is non-increasing on Ω and that*

$$m^v=\min_{s\in\mathcal{S}}(v(s))\ <\ M^v=\max_{s\in\mathcal{S}}(v(s))\,.$$

Then $\mathbf{F}^v$ defined by (A.1) *has the following properties:*

(i) For each $i\in\mathcal{S}_n$ with $v(i)=M^v$, $F_i^v(x,t)=\mathbb{1}_{\{x\geq M^v t\}}\mathbb{P}(X(t)=i)\,.$

(ii) For each $i\in\mathcal{S}_n$, with $v(i)<M^v$, $F_i^v(x,t)$ is Lipschitz continuous on $\mathbb{R}\times[0,\infty)$.

(iii) $\mathbf{F}^v$ is a strong solution of

$$\partial_t\mathbf{F}^v(x,t)+V\partial_v\mathbf{F}^v(x,t)=\mathbf{F}^v(x,t)Q(t) \tag{A.9}$$

in the open regions

$$\mathcal{U}_l=\Big\{(x,t):x<M^vt,t>0\Big\}\quad\text{and}\quad\mathcal{U}_r=\Big\{(x,t):x>M^vt,t>0\Big\}$$

and satisfies for each $x\in\mathbb{R}$ and $t\in(0,\infty)$

$$\begin{aligned}\lim_{(\bar x,\bar t)\in\mathcal{U}_l\to(M^vt,t)}F_i^v(\bar x,\bar t)&=\mathbb{1}_{\{v(i)<M^v\}}\mathbb{P}\{X(t)=i\}\\ \lim_{(\bar x,\bar t)\in\mathcal{U}_r\to(M^vt,t)}F_i^v(\bar x,\bar t)&=\mathbb{P}\{X(t)=i\}\\ \lim_{\bar t\to0^+}F_i^v(x,\bar t)&=\mathbb{1}_{\{i=1,x>0\}}.\end{aligned} \tag{A.10}$$

Proof. Let Δ denote the set of absorbing states of the process $X(t)$. Since Q is upper triangular, $n\in\Delta$ and thus Δ is not empty.

Take any $i\in\mathcal{S}_n$ with $v(i)=M^v$. Since v is monotone along the process we conclude that

$$L(t,\omega)=\int_0^t X(s,\omega)\,ds=M^vt\quad\text{for all}\quad\omega\in\{\tilde\omega:X(t,\tilde\omega)=i\}$$

and this yields (i).

Recall next that for a time-inhomogeneous Markov processes $X(t)$ (under the assumptions (A1)-(A2)) the jumping times $T_1,T_2,T_3,\dots$ of $X(t)$ satisfy $\mathbb{P}\big\{T_1>\alpha\big\}=\exp\big(-\int_0^\alpha q_1(s)\,ds\big)$ and for $k\geq2$

$$\mathbb{P}\Big\{T_k>t+\alpha\Big|\,T_{k-1}=t,X(T_{k-1})=i\Big\}=\exp\Big(-\int_t^{t+\alpha}q_i(s)\,ds\Big)\,. \tag{A.11}$$

Take any $i \in \mathcal{S}_n$ with $v(i) < M^v$, in which case $i > 1$. Since $Q(t)$ is upper triangular, each trajectory of the process has at most $n-1$ jumps before it enters into the absorbing set Δ. Thus we obtain

$$\Big\{\omega : X(\cdot,\omega) \text{ enters the state } i\Big\} = \bigcup_{k=1}^{n-1} \Omega_k^{(i)}\,, \quad \Omega_k^{(i)} = \Big\{\omega :\, X(\cdot,\omega) \text{ enters the state } i \text{ on the } k\text{-th jump}\Big\}\,.$$

We next denote $T_0 = 0$, $s_0 = 1$, $s_i^{(k)} = (s_1, s_2, \ldots, s_{k-1}, s_k = i) \in \mathcal{S}_n^k$, with $k \geq 1$, and

$$A(s_i^{(k)}) = \Big\{\omega : X(T_1) = s_1, \ldots, X(T_{k-1}) = s_{k-1}, X(T_k) = s_k = i\Big\} \subset \Omega_k^{(i)}\,.$$

First, suppose that $i \notin \Delta$. Using the above partitioning we write

$$\begin{aligned} F_i^v(x,t) &= \mathbb{P}(X(t) = i, L(t) < x) \\ &= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\Big\{A(s_i^{(k)}), T_k < t < T_{k+1}, L(t) < x\Big\} \\ &= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\big\{A(s_i^{(k)}), T_k < t < T_{k+1}, \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x - v(i)t\Big\}. \end{aligned} \tag{A.12}$$

We now show that F_i^v is Lipschitz continuous. To this end, consider the function

$$G(x,t;s_i^{(k)}) = \mathbb{P}\Big\{A(s_i^{(k)}), T_k < t < T_{k+1}, \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x\Big\}.$$

Observe that $G(x,t;s_i^{(k)})$ is well-defined for $(x,t) \in \mathbb{R}^2$. Moreover, since $i \notin \Delta$, the assumption (A2) implies that the process after entering the state i leaves this state in finite time $\mathbb{P}$-almost surely. Thus $\Omega_k^{(i)} \subset \{T_k < \infty\} \subset \{T_{k+1} < \infty\}$ and therefore

$$\begin{aligned} G(x,t;s_i^{(k)}) &= \mathbb{P}\Big\{A(s^{(k)}), T_k < t, \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x\Big\} \\ &\qquad - \mathbb{P}\Big\{A(s^{(k)}), T_{k+1} < t, \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x\Big\} \\ &=: G_1(x,t;s_i^{(k)}) - G_2(x,t;s_i^{(k)})\,. \end{aligned}$$

Using (A.11) and induction, one can show that for each $r \in \mathcal{S}_n$ and $k \geq 1$

$$\mathbb{P}\{A(s_r^{(k)}), T_{k+1} < z\} = \int_{-\infty}^{z} f_{k+1}(\alpha; s_r^{(k)})\, d\alpha \tag{A.13}$$

where $f_{k+1}(\cdot\,; s_r^{(k-1)})$ is a globally bounded function. Then, setting

$$V_k := \Big\{\omega : X(\cdot,\omega) \text{ has at least } k \text{ jumps}\Big\}$$

and summing up over all $s_r^{(k)} \in \mathcal{S}_n^k$ we obtain

$$\sum_{r \in \mathcal{S}_n} \mathbb{P}\{A(s_r^{(k)}), T_{k+1} < z\} = \mathbb{P}\{V_k, T_{k+1} < z\} = \int_{-\infty}^{z} \bar{f}_{k+1}(s)\, ds\,, \quad k \geq 1\,, \tag{A.14}$$

where $\bar{f}_{k+1}$ is some globally bounded function (which in general is not a probability density). In view of (A.11), the representation (A.14) holds also for T_1 with $V_0 = \Omega$. Thus, we conclude that the map $z \to \mathbb{P}\{V_{k-1}, T_k < z\}$ is globally Lipschitz for each $k \geq 1$.

Since v is monotone along the process, for each $s_i^{(k)}$ we have

$$v(1) = M^v \geq v(s_1) \cdots \geq v(s_k) = v(i)\,.$$

By assumption $v(i) < M^v$ and hence there exists $k_0 \in \{1, \ldots, k\}$ such that $v(s_{k_0-1}) - v(s_{k_0}) > 0$, which guarantees that not all terms in the nonnegative sum $\sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j))$ vanish. Then, in view of the fact that the event $A(s_i^{(k)})$ does not depend on the (x,t)-variable and that $A(s_i^{(k)}) \bigcap V_j = A(s_i^{(k)})$ for $j \leq k$, using (A.14) and induction, we conclude that

$$\mathbb{P}\Big\{A(s_i^{(k)}), \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x\Big\} = \int_{-\infty}^{x} \tilde{f}_k(s)\, ds\,, \quad k \geq 1\,, \tag{A.15}$$

for some globally bounded function $\tilde{f}_k$. This implies that the expression on the left-hand side of (A.15) is globally Lipschitz as a function of x. Combining (A.13), (A.14) and (A.15) and using the definition of the Lipschitz continuity we conclude that $G_1(x,t;s_i^{(k)})$ and $G_2(x,t;s_i^{(k)})$ are globally Lipschitz and hence $G(x,t;s_i^{(k)})$ is as well.

Recall that any Lipschitz continuous function in two variables composed with a linear map is also Lipschitz continuous. Thus $\bar{G}(x,t;s_i^{(k)}) := G(B(x,t);s_i^{(k)})$, where $B(x,t) = (x - v(i)t, t)$, is globally Lipschitz. In (A.12) each of the terms in the sum is one of the functions $\bar{G}(x,t;s_i^{(k)})$. Hence F_i which is restricted to $(x,t) \in \mathbb{R} \times [0,\infty)$ is globally Lipschitz on this domain.

We next assume that $i \in \Delta$. Observe that

$$\Big\{T_k < \infty, X(T_k) = i\Big\} \subset \Big\{T_{k+1} = \infty\Big\}$$

and therefore

$$\begin{aligned} F_i^v(x,t) &= \mathbb{P}\{X(t) = i, L(t) < x\} \\ &= \sum_{k=1}^{n-1} \sum_{s_i^{(k)} \in \mathcal{S}^k} \mathbb{P}\Big\{A(s_i^{(k)}), T_k < t, \sum_{j=1}^{k} T_j(v(s_{j-1}) - v(s_j)) < x - v(i)t\Big\}. \end{aligned}$$

Using an analogous approach (to the one in the case $i \notin \Delta$) one can show that each term in the above expression is globally Lipschitz continuous. This yields (ii).

From (i) and (ii) it follows that $\mathbf{F}^v$ is Lipschitz continuous in the open regions $\mathcal{U}_l$ and $\mathcal{U}_r$. Then, by Proposition A.5 we conclude that $\mathbf{F}^v$ is a strong solution of (A.9) in both $\mathcal{U}_l$ and $\mathcal{U}_r$. The boundary conditions (A.10) follow directly from the definition of $\mathbf{F}^v$. This proves (iii).

□

Remark A.15. *Since the components F_i^v, with $v(i) < M^v$, are Lipschitz continuous in the upper half plane and those F_i^v, with $v(i) = M^v$, are discontinuous along the line $x = M^v t = v(i)t$ which is a characteristic for the i-th equation it follows (see Renardy and Rogers (2004)) that $\mathbf{F}^v$ is a weak solution to the Cauchy problem comprising of* (A.9) *in the upper half-plane and (discontinuous) initial data given by $F_i^v(x,0) = \mathbb{1}_{\{x \geq 0, i=1\}}$.*

B Numerical Schemes

B.1 Upstream Numerical Scheme for Single-Locus Case

Here we present a numerical algorithm for computing solutions to the system (3.5). The numerical scheme is an upstream scheme based on the method of characteristics. In particular, the numerical scheme we develop makes use of the integral representation formulas (3.10), (3.11).

To define a grid in the (t,x)-space suitable for computation, choose $x_{\max}$, the maximum value that the CDF $\mathbb{P}\{\mathcal{L} \leq x_{\max}\}$ should be computed for. Due to Lemma 3.1, the relation $\mathbb{P}\{\mathcal{L} \leq x\} = F_1(t_{\max}, x)$ holds for all $x \leq x_{\max}$, with $t_{\max} := \frac{x_{\max}}{2}$. Thus $t_{\max}$ is set as the maximal gridpoint for t. In addition to the maximum gridpoints, choose small step sizes Δt and Δx. The number of gridpoints in the t dimension is then given by $M := \lceil \frac{t_{\max}}{\Delta t} \rceil + 1$, and the set of gridpoints is given as

$$T := \big\{0, \Delta t, 2\Delta t, \ldots, (M-1)\Delta t, \min(M\Delta t, t_{\max})\big\}. \tag{B.1}$$

For each point T_i, define a grid in the x-dimension as

$$X_i := \big\{0, \Delta x, \ldots, \min(U\Delta x, nT_i)\big\} \cup \big\{2\Delta t + \bar{X}_{i-1}, 3\Delta t + \bar{X}_{i-1}, \ldots, n\Delta t + \bar{X}_{i-1}\big\} \cup \big\{2T_i, 3T_i, \ldots, nT_i\big\}, \tag{B.2}$$

with $U = \lceil \frac{nT_i}{\Delta x} \rceil$ and $\bar{X}_{i-1} := \max(X_{i-1})$. Furthermore, set $U_i := |X_i|$. The same grid will be used for all $k \in \{1, \ldots, n\}$. The points $k\Delta t + \bar{X}_{i-1}$ and kT_i are added for numerical stability reasons, to improve the accuracy of the interpolation we will perform in subsequent steps.

Now fix $i \in \{0, \ldots, M\}$ and $k \in \{1, \ldots, n\}$, and assume that $F_\ell(T_{i-1}, X_{i-1,j})$ has been computed for all $\ell \in \{1, \ldots, n\}$ and $X_{i-1,j} \in X_{i-1}$. Furthermore, assume that $F_\ell(T_i, X_{i,j})$ has been computed for all $\ell \in \{k+1, \ldots, n\}$ and $X_{i,j} \in X_i$. Under these assumptions, $F_k(T_i, X_{i,j})$ can be computed for all $X_{i,j} \in X_i$ as follows. If $X_{i,j} < v(k)T_i$, then

$$F_k(T_i, X_{i,j}) = 0. \tag{B.3}$$

If $X_{i,j} = nT_i$, the maximal value of X_i, then

$$F_k(T_i, X_{i,j}) = \mathbb{P}\Big\{A\big(T_i\big) = k\Big\}. \tag{B.4}$$

The values on the right-hand side can be pre-computed for all k and $T_i \in T$ by solving the ODE (2.3) numerically. In the general case, note that the characteristic of F_k that goes through the point $(T_i, X_{i,j})^\top$ and the boundary $x = nt$ intersect at the point $(T_x, nT_x)^\top$, with $T_x := \frac{X_{i,j} - v(k)T_i}{n - v(k)}$. Thus, define

$$(X^\downarrow_{i,j}, T^\downarrow_{i,j})^\top := \begin{cases} (T_{i-1}, X_{i,j} - v(k)\Delta t)^\top, & \text{if } T_x < T_{i-1}, \\ (T_x, nT_x)^\top, & \text{otherwise,} \end{cases}$$

the projection of $(T_i, X_{i,j})^\top$ back along the corresponding characteristic to the previous time-slice T_{i-1}, or onto the boundary $x = nt$, whichever has the larger t-component. This backward projection step is illustrated in Figure 10(a). Then, according to equation (3.10)

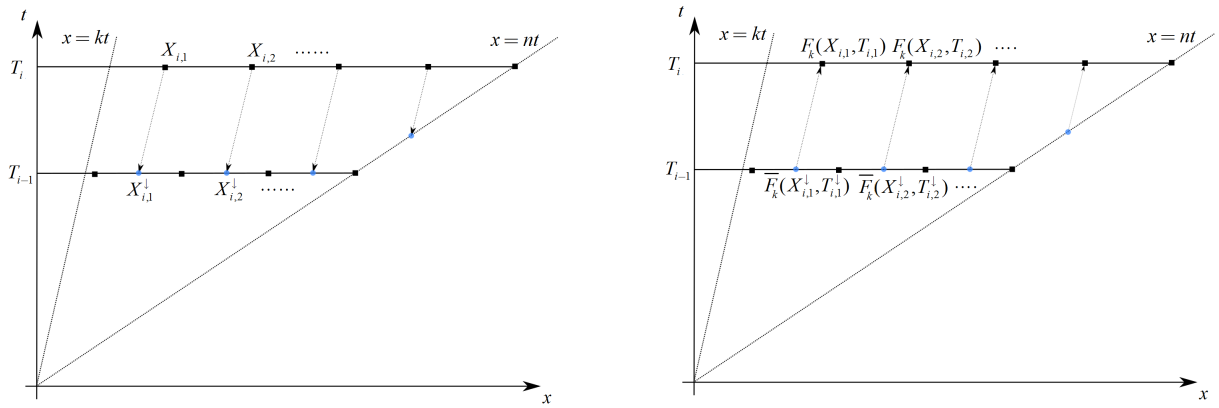
$$F_k(T_i, X_{i,j}) = e^{-(H_k^{(1)}(T_i) - H_k^{(1)}(T^\downarrow_{i,j}))} \left(\int_{T^\downarrow_{i,j}}^{T_i} g_k^{(1)}(\alpha) e^{(H_k^{(1)}(\alpha) - H_k^{(1)}(T^\downarrow_{i,j}))} d\alpha + F_k(X^\downarrow_{i,j}, T^\downarrow_{i,j}) \right) \tag{B.5}$$

holds.

The right-hand side of the equation (B.5) can now be computed using two approximations. Note that the point $(T^\downarrow_{i,j}, X^\downarrow_{i,j})^\top$ is in general not on the grid X_i, and thus $F_k(T^\downarrow_{i,j}, X^\downarrow_{i,j})$ has not been pre-computed. If $T_x < T_{i-1}$, the point is equal to $(T_{i-1}, X_{i,j} - v(k)\Delta t)^\top$. In that case, identify the two grid points in X_{i-1} that are closest to $X_{i,j} - v(k)\Delta t$ to the right and ~~to the~~ left. Then let $\bar{F}_k(T^\downarrow_{i,j}, X^\downarrow_{i,j})$ be the linear interpolation between the values of F_k at those gridpoints. If $T_x \geq T_{i-1}$, then the point is given by $(T^\downarrow_{i,j}, X^\downarrow_{i,j})^\top = (T_x, nT_x)^\top$ and is located on the boundary. Thus

$$F_k(X^\downarrow_{i,j}, T^\downarrow_{i,j}) = \mathbb{P}\Big\{A\big(T_x\big) = k\Big\},$$

which is also not pre-computed. However, $\mathbb{P}\Big\{A\big(T_{i-1}\big) = k\Big\}$ and $\mathbb{P}\Big\{A\big(T_i\big) = k\Big\}$ have been pre-computed, and thus set $\bar{F}_k(T^\downarrow_{i,j}, X^\downarrow_{i,j})$ as the linear interpolation between these two values.



(a) Projection the grid points backwards along the characteristics from time layer T_i to T_{i-1}.

(b) Propagating the interpolated values of the function F_k via numerical integration.

Figure 10: The back-tracing and propagation step of the upstream numerical scheme to compute F_k at all points of the grid.

The second approximation is to compute the integral on the right-hand side of equation (B.5) using the trapezoidal rule. Thus, the values of F_k on the grid can be computed using

$$F_k(T_i, X_{i,j}) = \frac{\Delta t}{2}\left(g_k^{(1)}(T_i) + e^{-(H_k^{(1)}(T_i)-H_k^{(1)}(T^{\downarrow}_{i,j}))} g_k^{(1)}(T^{\downarrow}_{i,j})\right) + e^{-(H_k^{(1)}(T_i)-H_k^{(1)}(T^{\downarrow}_{i,j}))}\bar{F}_k(T^{\downarrow}_{i,j}, X^{\downarrow}_{i,j}) + o(\Delta t^3) + o(\Delta x^2) \tag{B.6}$$

The terms $g_k(\cdot)$ depend on values of F_ℓ with $k < \ell \leq n$ that might not have been pre-computed on the grid either. However, the same interpolation schemes as for $\bar{F}_k$ can be applied. At this stage it is important though to strictly set F_ℓ to 0 if it should be 0 according to equation (3.7). Lastly, the values for $H_k^{(1)}(\cdot)$ can either be obtained using analytic formulas in equation (3.11) for certain classes of coalescent-speed functions we will consider (e.g. piece-wise constant), or by computing the requisite integrals using the trapezoidal rule, which can be done incrementally. The integration step of out numerical scheme is illustrated in Figure 10(b).

These equations lead naturally to a dynamic programming algorithm to compute F_k on the specified grid. To this end, iterate through the values $T_i \in T$ in increasing order. For each T_i, iterate through $k \in \{1, 2, \ldots, n\}$ in decreasing order, starting with $k = n$. Then, for each fixed T_i and k, $F_k(T_i, X_{i,j})$ can be computed for every $X_{i,j} \in X_i$ using equations (B.3), (B.4), and (B.6). The order of iteration guarantees that all necessary quantities have been pre-computed. This dynamic program can be employed to compute F_k on the specified grid for all k. Due to Lemma 3.1, the relation

$$\mathbb{P}\{\mathcal{L} \leq X_{M,j}\} = F_1(T_M, X_{M,j})$$

holds, which yields the values of the CDF $\mathbb{P}\{\mathcal{L} \leq x\}$ on the specified grid X_M.

B.2 Upstream Numerical Scheme for Two-Locus Case

In the two-locus case, we can compute F_s efficiently on a chosen grid similar to the marginal case. To this end, we again choose $x_{\max} = y_{\max}$, set $t_{\max} := \frac{1}{n} x_{\max}$, and choose step sizes Δt and $\Delta x = \Delta y$. Then, define the grid T as in definition (B.1), $M = |T|$ and for each T_i, define X_i as in definition (B.2). Furthermore, set $Y_i := X_i$ and $U_i := |Y_i|$. Thus, we use the regular grid $X_i \times Y_i$ in the (x, y)-space.

Now fix T_i and $s \in \bar{\mathcal{S}}^\rho$, and assume that $F_{s'}(T_{i-1}, X_{i-1,j}, Y_{i-1,\ell})$ has been computed for all $s' \in \bar{\mathcal{S}}^\rho$, $X_{i-1,j} \in X_{i-1}$, and $Y_{i-1,\ell} \in Y_{i-1}$. Furthermore, assume that $F_{s'}(T_i, X_{i,j}, Y_{i,\ell})$ has been computed for all s'

with $s \prec s'$, $X_{i,j} \in X_i$, and $Y_{i,\ell} \in Y_i$. To compute $F_s(T_i, X_{i,j}, Y_{i,\ell})$, first check using equation (3.13) whether the requisite point lies on the boundary, or in the zero region. The values on the boundary according to equation (3.13) are computed as time-dependent CDFs of marginal integrals along trajectories of the process $\bar{A}^\rho$, and thus they can be computed using exactly the same procedure as detailed in Section B.1, replacing A by $\bar{A}^\rho$. In the interior region, applying the trapezoidal rule to the solution of the first-order ODE, for all $X_{i,j} \in X_i$, and $Y_{i,\ell} \in Y_i$ the value of $F_s(T_i, X_{i,j}, Y_{i,\ell})$ can be computed using

$$\begin{aligned} &F_s(T_i, X_{i,j}, Y_{i,\ell}) \\ &\quad = \frac{\Delta t}{2}\left(g_s^{(2)}(T_i) + e^{-(H_s^{(2)}(T_i) - H_s^{(2)}(T_{i,j,\ell}^{\downarrow}))} g_s^{(2)}(T_{i,j,\ell}^{\downarrow}) \right) + e^{-(H_s^{(2)}(T_i) - H_s^{(2)}(T_{i,j,\ell}^{\downarrow}))} \bar{F}_s(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow}) \\ &\qquad + o(\Delta t^3) + o(\Delta x^2) + o(\Delta y^2)\,. \end{aligned}$$

Here

$$(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})^\top := \begin{cases} (T_{i-1}, X_{i,j} - v^a(s)\Delta t, Y_{i,\ell} - v^b(s)\Delta t)^\top, & \text{if } \max(T_x, T_y) < T_{i-1}, \\ \big(T_x, nT_x, Y_{i,\ell} - v^b(s)\cdot(T_i - T_x)\big)^\top, & \text{if } \max(T_{i-1}, T_y) \le T_x, \\ \big(T_y, X_{i,j} - v^a(s)\cdot(T_i - T_y), nT_y\big)^\top, & \text{if } \max(T_x, T_{i-1}) \le T_y, \end{cases}$$

with

$$T_x := \frac{X_{i,j} - v^a(s) T_i}{n - v^a(s)}$$

being the t-coordinate of the point of intersection between the characteristic through the point $(T_i, X_{i,j}, Y_{i,\ell})^\top$ and the boundary $x = nt$, and

$$T_y := \frac{Y_{i,\ell} - v^b(s) T_i}{n - v^a(s)}$$

likewise for the boundary $y = nt$.

The points $(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})^\top$ will in general not be on the grid of pre-computed values, and thus the approximation $\bar{F}_s(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})$ has to be used. In the case $\max(T_x, T_y) < T_{i-1}$, this value can be obtained by identifying the four points in $X_{i-1} \times Y_{i-1}$ surrounding $(X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})$, and interpolating the respective values of $F_s(T_{i-1}, \cdot, \cdot)$ linearly. In the case $\max(T_{i-1}, T_y) \le T_x$, the point $(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow})^\top$ is on the boundary $x = nt$, and

$$F_s(T_{i,j,\ell}^{\downarrow}, X_{i,j}^{\downarrow}, Y_{i,\ell}^{\downarrow}) = \mathbb{P}\big\{\bar{A}^\rho(T_x) = s, L^b(T_x) \le Y_{i,\ell} - v^b(s)\cdot(T_i - T_x)\big\}$$

holds. The value of the time-dependent CDF on the right-hand can be obtained as the linear interpolation between the values $\mathbb{P}\big\{\bar{A}^\rho(T_{i-1}) = s, L^b(T_{i-1}) \le Y_{i,\ell} - v^b(s)\Delta t\big\}$ and $\mathbb{P}\big\{\bar{A}^\rho(T_i) = s, L^b(T_i) \le Y_{i,\ell}\big\}$, which we pre-compute (or approximations thereof) using the numerical scheme for the marginal case (see Appendix B.1) on the boundary. By symmetry, the case $\max(T_x, T_{i-1}) \le T_y$ can be handled in the same way. Computing $g_s^{(2)}(\cdot)$ will require some $F_{s'}$ with $s \prec s'$, which can be obtained by similar interpolation procedures, or setting it to zero in the appropriate regions. The values of $H_s^{(2)}(\cdot)$ can be computed according to equation (3.16) analytically or numerically, as before.

Again, we can implement these formulas in an efficient dynamic programming algorithm to compute the values of $F_s(t, x, y)$ on the specified grid for all $s \in \bar{\mathcal{S}}^\rho$, and thus compute

$$\mathbb{P}\{\mathcal{L}^a \le X_{M,j}, \mathcal{L}^b \le Y_{M,\ell}\} = F_{(1,0,0,0)}(t_{\max}, X_{M,j}, Y_{M,\ell}) + F_{(1,0,0,1)}(t_{\max}, X_{M,j}, Y_{M,\ell}),$$

the joint CDF of the total tree length at two linked loci evaluated on the specified grid.

References

Cheng, J. Y. and Mailund, T. (2015). Ancestral population genomics using coalescence hidden markov models and heuristic optimisation algorithms. *Comp. Biol. Chem.*, **57,** 80–92.

Dafermos, C. M. (2010). *Hyperbolic conservation laws in continuum physics*. Springer, New York.

Doob, J. (1953). *Stochastic processes*. Wiley, New York.

Evans, L. (2010). *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society.

Federer, H. (1969). *Geometric measure theory*. Grundlehren der mathematischen Wissenschaften. Springer.

Ferretti, L., Disanto, F., and Wiehe, T. (2013). The effect of single recombination events on coalescent tree height and shape. *PLoS ONE*, **8,**(4) 1–15.

Griffiths, R. C. and Marjoram, P. (1997). An ancestral recombination graph. In *Progress in Population Genetics and Human Evolution*, Donnelly, P. and Tavaré, S., editors, volume 87, pages 257–270. Springer, Berlin.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. *J. Evol. Biol.*, **7,** 1–44.

Hudson, R. R. (2002). Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, **18,**(2) 337–338.

Karlin, S. (1981). *A second course in stochastic processes*. Academic Press, New York.

Keinan, A. and Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, **336,**(6082) 740–743.

Kingman, J. F. (1982). The coalescent. *J. Evol. Biol.*, **13,**(3) 235–248.

Koralov, L. and Sinay, Y. (2007). *Theory of probability and random processes*. Springer, Berlin.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, **475,** 493–496.

Marjoram, P. and Wall, J. D. (2006). Fast "coalescent" simulation. *BMC Genet.*, **7,** 16.

McVean, G. A. T. (2002). A genealogical interpretation of linkage disequilibrium. *Genetics*, **162,**(2) 987–991.

McVean, G. A. and Cardin, N. J. (2005). Approximating the coalescent with recombination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **360,** 1387–93.

Paul, J. S., Steinrücken, M., and Song, Y. S. (2011). An accurate sequentially markov conditional sampling distribution for the coalescent with recombination. *Genetics*, **187,**(4) 1115–1128.

Pfaffelhuber, P., Wakolbinger, A., and Weisshaupt, H. (2011). The tree length of an evolving coalescent. *Probab. Theory Related Fields*, **151,**(3) 529–557.

Polanski, A., Bobrowski, A., and Kimmel, M. (2003). A note on distributions of times to coalescence, under time-dependent population size. *Theor. Popul. Biol.*, **63,**(1) 33–40.

Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLoS Genet.*, **10,**(5) 1–27.

Renardy, M. and Rogers, R. C. (2004). *An Introduction to Partial Differential Equations*. Springer, 2nd edition.

Rudin, W. (1976). *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill.

Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.*, **46,**(8) 919–25.

Sheehan, S., Harris, K., and Song, Y. S. (2013). Estimating variable effective population sizes from multiple genomes: A sequentially markov conditional sampling distribution approach. *Genetics*, **194,**(3) 647–662.

Simonsen, K. L. and Churchill, G. A. (1997). A markov chain model of coalescence with recombination. *Theor. Popul. Biol.*, **52,**(1) 43–59.

Steinrücken, M., Kamm, J., , and Song, Y. (2016). Inference of complex population histories using whole-genome sequences from multiple populations. Preprint at: http://dx.doi.org/10.1101/026591.

Steinrücken, M., Kamm, J. A., and Song, Y. S. (2015). Inference of complex population histories using whole-genome sequences from multiple populations. *bioRxiv*. http://biorxiv.org/content/early/2015/09/16/026591.abstract.

Stroock, D. W. (2008). *An Introduction to Markov Processes (Graduate Texts in Mathematics)*. Springer.

Tavaré, S. and Zeitouni, O. (2004). *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXXI - 2001 (Lecture Notes in Mathematics)*. Springer.

Wakeley, J. (2008). *Coalescent Theory: An Introduction*. W. H. Freeman.

Wiuf, C. and Hein, J. (1999). Recombination as a point process along sequences. *Theor. Popul. Biol.*, **55,** 248–259.