Phase-type distributions in population genetics

Hobolth, Asger; Siri-Jégousse, Arno; Bladt, Mogens

Published in:
Theoretical Population Biology

DOI:
10.1016/j.tpb.2019.02.001

Publication date:
2019

Document version
Peer reviewed version

Phase-type distributions in population genetics

Asger Hobolth[1], Arno Siri-Jégousse[2] and Mogens Bladt[3],
1. Aarhus University, Bioinformatics Research Center; asger@birc.au.dk
2. UNAM, IIMAS, Departamento de Probabilidad y Estadística; arno@sigma.iimas.unam.mx
3. University of Copenhagen, Department of Mathematical Sciences; bladt@math.ku.dk

February 18, 2019

Abstract

Probability modelling for DNA sequence evolution is well established and provides a rich framework for understanding genetic variation between samples of individuals from one or more populations. We show that both classical and more recent models for coalescence (with or without recombination) can be described in terms of the so-called phase-type theory, where complicated and tedious calculations are circumvented by the use of matrix manipulations. The application of phase-type theory in population genetics consists of describing the biological system as a Markov model by appropriately setting up a state space and calculating the corresponding intensity and reward matrices. Formulae of interest are then expressed in terms of these aforementioned matrices. We illustrate this procedure by a number of examples: (a) Calculating the mean, (co)variance and even higher order moments of the site frequency spectrum in multiple merger coalescent models, (b) Analysing a sample of DNA sequences from the Atlantic Cod using the Beta-coalescent, and (c) Determining the correlation of the number of segregating sites for multiple samples in the two-locus ancestral recombination graph. We believe that phase-type theory has great potential as a tool for analysing probability models in population genetics. The compact matrix notation is useful for clarification of current models, and in particular their formal manipulation and calculations, but also for further development or extensions.

Key words

Coalescent theory, multiple merger, phase-type theory, recombination, segregating sites, site frequency spectrum.

1 Introduction

Phase-type distributions is a rather general class of distributions for positive random variables which include mixtures and convolutions of exponential distributions [2]. The height and total branch length of the genealogical tree in the basic coalescent model are examples of phase-type distributed random variables. The number of singletons in a sample of DNA sequences is determined by the total length

of branches with one descendant. The total length of branches with one descendant is also phase-type distributed, and more generally the total length of branches with a certain number of descendants is phase-type distributed. The fact that key population genetics quantities are phase-type distributed is useful because properties of phase-type distributed variables are very well understood. In particular, phase-type theory provides explicit expressions of means, variances and higher-order moments in terms of simple manipulations of the rate matrices and vectors that determine the distribution.

In this paper, we demonstrate that important quantities in coalescent models can often be expressed in terms of phase-type distributions. We first consider fundamental quantities such as the height, total branch length and site frequency spectrum in the basic coalescent model (e.g. [26]). Second, we extend from the basic coalescent to the multiple merger coalescent (e.g. [20]), and provide an application of parameter estimation in the Beta-coalescent. We also extend the basic coalescent to a structured coalescent model, namely the seed bank coalescent (e.g. [3]). Third, we extend to a two-locus model with recombination.

The basic coalescent, multiple merger coalescent, structured coalescent, and coalescent with recombination have traditionally required methods tailored to each model. These solutions are often based on computationally intensive or analytically challenging recursion schemes. Phase-type theory, in contrast, preserves a high-level matrix structure of the biological system without the necessity of breaking the matrices into their elements. In all our examples we take advantage of the close connection between coalescent models and phase-type theory. We show that the often complex and difficult-to-derive coalescent formulae are often easy to define and calculate using phase-type theory and matrix notation. We thus provide a unified approach to the calculation of distributions and moments in Markov genealogical models.

The tree height, total branch length, and total length of branches with a certain number of descendants, are distributed according to a univariate phase-type distribution. The joint distribution of total branch length in two neighbouring loci in the coalescent model with recombination are distributed according to a *multivariate* phase-type distribution. The joint distribution of total branch length with e.g. one and two descendants is another example of a multivariate phase-type distributed random vector. We extend the univariate phase-type theory to the multivariate situation. This extension allows us to determine joint moments (e.g. covariances) of entries in the site frequency spectrum or total branch length of the genealogical trees in two neighbouring loci.

The generating function (GF) method advocated in [15] and [16] is similar in spirit to the phase-type theory that we suggest. [15] and [16] also take advantage of the Markovian property of the coalescent models, and the GF methods can be used to analyse rather general demographic models and multiple loci. The GF method can calculate the likelihood of a sample configuration for small sample sizes (up to $n = 6$) and uses a recursive scheme. We focus on summary statistics, and our method applies for sample sizes up to $n = 25$. Furthermore, the recursive procedure in [15] and [16] is substituted by matrix manipulations.

2 Phase–type distributions

The purpose of this section is to provide an exposition of those parts of phase–type theory that we believe are particularly relevant for genealogical models in population genetics. We illustrate the theory by a number of examples from coalescent theory. The phase–type theory presented is mostly well known, and further details may be found in the monograph [2], which will also serve as our main

reference throughout.

We follow the standard notational conventions from phase–type theory which makes it easy to distinguish between matrices, row vectors, column vectors, and their elements. Matrices are written in bold majuscules (e.g. $\boldsymbol{S}$ and $\boldsymbol{\Lambda}$), column vectors in bold, *Roman* minuscules (e.g. $\boldsymbol{s}$ and $\boldsymbol{t}$) while row vectors are bold, *Greek* minuscules (e.g. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$). Elements of vectors and matrices are denoted by their corresponding minuscule letters (e.g. $\boldsymbol{\alpha} = (\alpha_i)_i$ and $\boldsymbol{S} = \{s_{ij}\}_{i,j}$). Dimensions are usually not explicitly stated unless needed. Throughout we let $\boldsymbol{I}$ denote the identity matrix, $\boldsymbol{e} = (1,1,\ldots,1)'$ the (column) vector of ones, $\boldsymbol{e}_i = (0,\ldots,0,1,0,\ldots,0)'$ (1 on the ith position) the ith unit (column) vector, whereas $\mathbf{0}$ may either denote the zero vector or zero matrix.

2.1 Definition and examples

We now proceed to a formal definition of a phase–type distributed random variable. Consider a Markov jump process (continuous time Markov chain) $\{X_t\}_{t\geq 0}$ with finite state-space $\{1,2,...,p,p+1\}$, where states $1,...,p$ are transient and state $p+1$ is absorbing; in a genealogical context state $p+1$ is usually the MRCA. This means that $\{X_t\}_{t\geq 0}$ has an intensity (rate) matrix $\boldsymbol{\Lambda}$ of the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{1}$$

and we refer to the $p \times p$ sub-matrix of rates between the transient states, $\boldsymbol{S} = \{s_{ij}\}_{i,j=1,...,p}$, as a *sub-intensity* matrix, the p-dimensional column vector $\boldsymbol{s} = (s_i)_{i=1,...,p}$ as an *exit rate* vector (since its elements are the intensities for jumping to the absorbing state), and finally $\mathbf{0}$ is a p-dimensional row vector of zeros. The assumption of states $1,...,p$ being transient means that eventually the process will jump to the absorbing state. Since rows sum to zero in intensity matrices (i.e. $\boldsymbol{\Lambda e} = \mathbf{0}$), row sums are non–positive (zero or negative) in sub–intensity matrices. Furthermore, from $\boldsymbol{\Lambda e} = \mathbf{0}$ we get $\boldsymbol{s} = -\boldsymbol{Se}$. Hence the exit rate vector $\boldsymbol{s}$ is implicitly know from the specification of the sub-intensity matrix $\boldsymbol{S}$.

Assume that $\{X_t\}_{t\geq 0}$ begins in a transient state and let $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_p)$ where $\alpha_i = \mathbb{P}(X_0 = i)$, $i = 1,...,p$. Then $\boldsymbol{\alpha e} = \sum_{i=1}^{p} \alpha_i = 1$ and $\boldsymbol{\alpha}$ is a probability vector on the set of transient states $E = \{1,2,\ldots,p\}$. Often $\boldsymbol{\alpha} = (1,0,0,...,0)$, i.e. we start in state 1 and progress through the transient states $2,...,p$ before absorption in the MCRA.

The transition matrix of the Markov process, $\boldsymbol{P^t} = \{p_{ij}^t\}_{i,j=1,...,p+1}$, where $p_{ij}^t = \mathbb{P}(X_t = j | X_0 = i)$, can be calculated as the matrix exponential of the intensity matrix scaled by the time constant t, i.e.

$$\boldsymbol{P^t} = e^{\boldsymbol{\Lambda} t} = \sum_{n=0}^{\infty} \frac{\boldsymbol{\Lambda}^n t^n}{n!}.$$

By using the fact that $\boldsymbol{s} = -\boldsymbol{Se}$ it is easily proved that

$$\boldsymbol{P^t} = \begin{pmatrix} e^{\boldsymbol{S}t} & \boldsymbol{e} - e^{\boldsymbol{S}t}\boldsymbol{e} \\ \mathbf{0} & 1 \end{pmatrix}. \tag{2}$$

The restriction of $\boldsymbol{P^t}$ to the transient states set E is therefore $\exp(\boldsymbol{S}t)$. Hence for $i,j \in \{1,2,...,p\}$, p_{ij}^t equals the (i,j)th element of $\exp(\boldsymbol{S}t)$ which we write as $(\exp(\boldsymbol{S}t))_{ij}$. This simple observation provides the backbone for almost all deductions of explicit formulae in phase–type theory. From this

we can for example calculate the distribution of X_t. To this, simply observe that

$$\mathbb{P}(X_t = j) = \sum_{i=1}^{p} \mathbb{P}(X_t = j | X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i=1}^{p} \alpha_i p_{ij}^t = \sum_{i=1}^{p} \alpha_i \left(e^{\boldsymbol{S}t}\right)_{ij}.$$

In matrix notation we have

$$(\mathbb{P}(X_t = 1), ..., \mathbb{P}(X_t = p)) = \boldsymbol{\alpha} e^{\boldsymbol{S}t}. \tag{3}$$

We say that $\boldsymbol{\alpha} e^{\boldsymbol{S}t}$ is the *defective* distribution of X_t on $\{1, 2, ..., p\}$ because the probabilities do not sum to one due to the possibility of having been absorbed prior to time t.

We are now in position to formally define a phase–type distributed random variable.

Definition 2.1 (Phase-type distribution)**.** *The time until absorption*

$$\tau = \inf\{t > 0 : X_t = p + 1\}$$

is said to have a phase-type distribution of order p with phase-space $E = \{1, 2, \ldots, p\}$, initial distribution $\boldsymbol{\alpha}$ and sub-intensity (generator) matrix $\boldsymbol{S}$, and we write

$$\tau \sim PH_p(\boldsymbol{\alpha}, \boldsymbol{S}).$$

The exit rate vector will always be denoted by a bold minuscule letter corresponding to the letter for the generator, here $\boldsymbol{s}$.

In order to be able to formulate a specific genealogical model in terms of a Markov process with an absorbing state (MCRA), it is instructive to observe its sample path. Let $0 = S_0 < S_1 < S_2 < \ldots$ denote the jump times (e.g. coalescence times) of $\{X_t\}_{t \geq 0}$ and $T_n = S_n - S_{n-1}$, $n = 1, 2, \ldots$, the corresponding inter-arrival times (e.g. time between two consecutive coalescence times). Furthermore we define the discrete time process $Y_n = X_{S_n}$, $n = 0, 1, ...$, which keeps track of the states visited (e.g. the sample size after the nth coalescence). Then $\{Y_n\}_{n \in \mathbb{N}}$ is a Markov chain on $\{1, 2, ..., p + 1\}$ with transition probability matrix $\boldsymbol{Q} = \{q_{ij}\}$, say, and is referred to as the *embedded* Markov chain. Conditionally on $Y_{n-1} = i$, $T_n = S_n - S_{n-1}$ has an exponential distribution with parameter $\boldsymbol{\Lambda}_{ii} = -\lambda_{ii} = \lambda_i$. For $i \neq j$, $i, j = 1, ..., p$, the relation between q_{ij} and λ_{ij} is given by $\lambda_{ij} = \lambda_i q_{ij}$, which suggests the important interpretation

$$\lambda_{ij} dx = \text{probability of a jump from } i \text{ to } j \text{ during a small time interval } [x, x + dx). \tag{4}$$

In Figure 1 we illustrate a sample path of a general Markov jump process with intensity matrix (1) generating a phase-type distribution. The initial state is chosen according to $\boldsymbol{\alpha}$. Given initiation in $Y_0 = X_0 = 3$, the time until the first jump, S_1, will then be exponentially distributed with intensity $\lambda_3 = -\lambda_{33} = -s_{33} > 0$. The process then jumps to a state $j \neq 3$, with probability $q_{3j} = \lambda_{3j}/\lambda_3 = -s_{3j}/s_{33}$ or to the absorbing state with probability $q_{3,p+1} = \lambda_{3,p+1}/\lambda_3 = -s_3/s_{33}$.

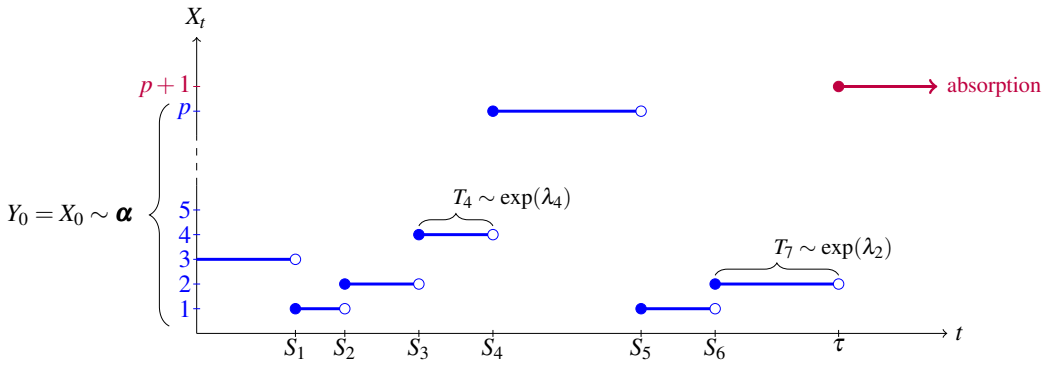

Figure 1: A Markov process with p transient states (blue), one absorbing state (purple), times of jumps $S_1 < S_2, \ldots$ and time to absorption τ. The filled and empty circles indicate that the process is assumed continuous from the right. The embedded chain $Y_n = X_{S_n}$ here takes the values $Y_0 = 3$, $Y_1 = 1$, $Y_2 = 2$, $Y_3 = 4$ etc. Holding times between jumps, $T_n = S_n - S_{n-1}$, are exponentially distributed with a parameter which depends on Y_{n-1} only.

Example 2.2 (Generalized Erlang)**.** Let $T_1, T_2, ..., T_n$ be independent random variables with $T_i \sim \text{Exp}(\lambda_i)$ for some $\lambda_i > 0$, $i = 1, ..., n$. Then we say that $\tau = T_1 + \cdots + T_n$ has a generalized Erlang distribution with parameters $\lambda_1, ..., \lambda_n$ and order n. If $\lambda_1 = \cdots = \lambda_n = \lambda$ then we say that τ has an Erlang distribution with parameter λ and order n, which will be denoted by $\text{Er}_n(\lambda)$.

Generalized Erlang distributions are phase-type distributions (Figure 2). Here, the process initiates in state 1 with probability 1 and jumps to state 2 with probability 1 after time $T_1 \sim \exp(\lambda_1)$. Continuing this way, from state $n-1$ the process jumps to state n with probability 1 and remains in this state for the time $T_n \sim \exp(\lambda_n)$. From here it jumps to the absorbing state. Thus the time τ it takes the process to reach the absorbing state $n+1$ is exactly the sum of the exponentially distributed random variables. A phase-type representation is given by

$$\boldsymbol{\alpha} = (1, 0, 0, \ldots, 0), \quad \boldsymbol{S} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_n \end{pmatrix}.$$

Since it does not matter in which order we sum the random variables in $S = T_1 + \cdots + T_n$ we could have chosen any other permutation of $\lambda_1, ..., \lambda_n$. Thus phase-type representations are *not* unique for a given distribution.

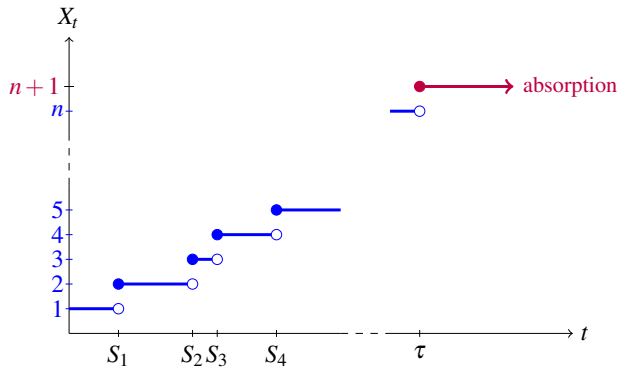

Figure 2: A phase-type representation of the convolution of n exponential distributions. Here $S_i = T_1 + T_2 + \cdots + T_i$ is the time of the ith jump where $T_i \sim \exp(\lambda_i)$ and $T_1, \ldots, T_n$ are independent.

Example 2.3 (Kingman's n–coalescent)**.** Consider independent $T_i \sim \exp(\lambda_i)$ where $\lambda_i = \binom{i}{2} = i(i-1)/2$, $i = n, ..., 2$. The tree height (time to the most recent common ancestor) is given by

$$\tau_n = T_n + \cdots + T_2$$

and the total branch length is

$$\mathscr{L}_n = nT_n + (n-1)T_{n-1} + \cdots + 2T_2.$$

The tree height is then phase-type distributed $\tau_n \sim \mathrm{PH}_{n-1}(\boldsymbol{\pi}, \boldsymbol{T})$ with $\boldsymbol{\pi} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{pmatrix} -n(n-1)/2 & n(n-1)/2 & 0 & \cdots & 0 \\ 0 & -(n-1)(n-2)/2 & (n-1)(n-2)/2 & \cdots & 0 \\ 0 & 0 & -(n-2)(n-3)/2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

Since $T_i \sim \exp(\lambda_i)$ implies $iT_i \sim \exp(\lambda_i/i) = \exp(i-1)$ we see that also $\mathscr{L}_n$ has a phase-type distribution with representation $\mathrm{PH}_{n-1}(\boldsymbol{\pi}, \boldsymbol{S})$, where

$$\boldsymbol{S} = \frac{1}{2} \begin{pmatrix} -(n-1) & n-1 & 0 & \cdots & 0 \\ 0 & -(n-2) & n-2 & \cdots & 0 \\ 0 & 0 & -(n-3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

Example 2.4 (Λ–coalescent)**.** The Λ-coalescent, introduced independently by Pitman [19] and Sagitov [21], defines a class of exchangeable coagulation processes including various useful models in population genetics. Its dynamics are characterized by a finite measure Λ on $[0,1]$. When the process has b lineages, each subset of k lineages merges at a rate

$$\lambda_{b,k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx), \quad k = 2, ..., b. \tag{5}$$

The dynamics of Kingman's coalescent is obtained by taking $\Lambda = \delta_0$, the unit mass at zero, leading to binary mergers only. In Figure 3 we show the five possible unlabelled Λ-coalescent topologies for a sample of size $n = 4$.

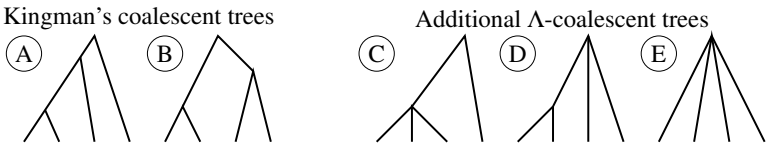

Figure 3: The five possible Λ-coalescent topologies for four sequences.

In the general case, the height of the tree of a sample of size n is phase-type distributed $\mathrm{PH}_{n-1}(\boldsymbol{\alpha}, \boldsymbol{S})$ with $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{S} = \begin{pmatrix} -g_n & g_{n,2} & g_{n,3} & \cdots & g_{n,n-1} \\ 0 & -g_{n-1} & g_{n-1,2} & \cdots & g_{n-1,n-2} \\ 0 & 0 & -g_{n-2} & \cdots & g_{n-2,n-3} \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & -g_2 \end{pmatrix} \tag{6}$$

where $g_{i,k} = \binom{i}{k} \lambda_{i,k}$ for $i = 2, \ldots, n$ and $k = 2, \ldots, i$, and $g_i = \sum_{k=2}^{i} g_{i,k}$ (notice that $g_2 = 1$).

Example 2.5 (Psi coalescent)**.** A class of special interest is the Psi-coalescent that appears as the genealogical process of Moran models with highly skewed offspring distribution [8]. Rare reproduction events implies that the progeny of an individual can replace a proportion $\psi \in (0, 1)$ of the individuals in the next generation. Here the probability measure Λ is the unit mass in ψ. This dynamics gives the transition rates

$$\lambda_{b,k} = \psi^{k-2}(1-\psi)^{b-k}.$$

Note that we deviate from the original model of [8] by a constant ψ^2 so that we obtain the Kingman's coalescent as $\psi \to 0$.

Example 2.6 (Beta-coalescent)**.** Another class of interest is the Beta-coalescent that appears as the limit genealogical process of stable Galton-Watson populations [22] and has been applied to marine populations [1]. Here the probability measure Λ is that of a $\mathrm{Beta}(2-\alpha, \alpha)$ distribution with $1 \leq \alpha < 2$, i.e.

$$\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1} dx.$$

This model results in the transition rates

$$\lambda_{b,k} = \frac{\beta(k-\alpha, b-k+\alpha)}{\beta(\alpha, 2-\alpha)}, k = 2, ..., b, \tag{7}$$

where β is the Beta function. For $\alpha \to 2$ we recover the Kingman coalescent, whereas the case $\alpha = 1$ is known as the Bolthausen-Sznitman coalescent, which appears as the genealogical model of populations under strong selection [5, 18, 23].

Example 2.7 (Seed bank coalescent)**.** In this example we consider a genealogical process appearing in peripatric metapopulations [14] and seed-bank models [3]. In this model lineages can be active (continent or plants) or inactive (islands or seeds) and they switch from one state to the other at a fixed rate. When they are active, lineages coalesce according to the dynamics of the Kingman coalescent. More precisely, let c be the rate for an active branch to inactivate and let K be the rate for an inactive branch to re-activate. Transition rates can then be defined in the following way. Let $\lambda_{i,j} = (i-j)(i-j-1)/2 + (i-j)c + jK$, $j = 0,\ldots,i$. Then define the following $(i+1)\times(i+1)$ matrices

$$\boldsymbol{\Lambda}(i) = \begin{pmatrix} -\lambda_{i,0} & ic & 0 & 0 & \cdots & 0 & 0 \\ K & -\lambda_{i,1} & (i-1)c & 0 & \cdots & 0 & 0 \\ 0 & 2K & -\lambda_{i,2} & (i-2)c & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\!\vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\lambda_{i,i-1} & c \\ 0 & 0 & 0 & 0 & \cdots & iK & -\lambda_{i,i} \end{pmatrix}$$

and $(i+1)\times i$ matrices

$$\boldsymbol{D}(i) = \begin{pmatrix} i(i-1)/2 & 0 & 0 & \cdots & 0 \\ 0 & (i-1)(i-2)/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\!\vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

The subintensity matrix for the height of the coalescent tree can then be represented as

$$\begin{pmatrix} \boldsymbol{\Lambda}(n) & \boldsymbol{D}(n) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}(n-1) & \boldsymbol{D}(n-1) & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}(n-2) & \boldsymbol{D}(n-2) & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots\!\vdots & \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Lambda}(2) \end{pmatrix}.$$

The matrix $\boldsymbol{\Lambda}(i)$ gives the transition rates when the whole system starts and remains with total size i, and the matrix $\boldsymbol{D}(i)$ gives the transition rates when the whole system loses an element (by coalescence) starting from total size i. Row j of $\boldsymbol{\Lambda}(i)$, $j = 1,...,i+1$, corresponds to the case where out of the remaining i branches, $j-1$ of them are presently inactive.

3 Review of Phase–type theory

3.1 Basic distributional properties

In the previous section we described how to express various genealogical models in terms of their sub–intensity matrices. We now show how to manipulate these matrices in order to determine key properties of the models. In the following we may think of $\boldsymbol{S}$ as a sub–intensity matrix from any of the previous examples. Let $\tau \sim \mathrm{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$ and let $\{X_t\}_{t\geq 0}$ denote its underlying Markov jump

process. We may think of τ as being the tree height and X_t as the number of lineages present at time t in a $(p+1)$–coalescent model.

From (3) we know that $X_t \sim \boldsymbol{\alpha}\exp(\boldsymbol{S}t)$, and since the event $\{\tau > t\}$ is identical to the event that absorption has not yet occurred, which is the same as the event $X_t \in \{1,2,...,p\}$, i.e. X_t takes one of the values $1,2,...,p$, we have that

$$\mathbb{P}(\tau > t) = \sum_{i=1}^{p} \mathbb{P}(X_t = i) = \sum_{i=1}^{p} (\boldsymbol{\alpha} e^{\boldsymbol{S}t})_i = \boldsymbol{\alpha} e^{\boldsymbol{S}t} \boldsymbol{e},$$

where we recall that $\boldsymbol{e}$ is the column vector of ones. Hence we have derived a formula for the distribution function F for τ, namely

$$F(t) = 1 - \mathbb{P}(\tau > t) = 1 - \boldsymbol{\alpha} e^{\boldsymbol{S}t} \boldsymbol{e}.$$

The density f of τ can then be deduced by a neat probabilistic argument as follows. From (3),

$$\mathbb{P}(X_u = i) = \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}_i,$$

and since $s_i du$ is the probability of jumping from state i to the absorbing state $p+1$ during $[u, u+du)$ we have

$$f(u)du = \sum_{i=1}^{p} \mathbb{P}(X_{u+du} = p+1 | X_u = i) \mathbb{P}(X_u = i) = \sum_{i=1}^{p} \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{e}_i s_i du,$$

and we get

$$f(u) = \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{s}.$$

The Laplace transform for τ can then be calculated as

$$\begin{aligned} L_\tau(t) &= \int_0^\infty e^{-tx} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} dx \\ &= \boldsymbol{\alpha} \left(\int_0^\infty e^{-(t\boldsymbol{I} - \boldsymbol{S})x} dx \right) \boldsymbol{s}. \end{aligned}$$

Here we have used that $e^{(\boldsymbol{A}+\boldsymbol{B})x} = e^{\boldsymbol{A}x} e^{\boldsymbol{B}x}$ when the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ commute ($\boldsymbol{AB} = \boldsymbol{BA}$), and that $\boldsymbol{I}$ commutes with $\boldsymbol{S}$.

The eigenvalues for $\boldsymbol{S} - t\boldsymbol{I}$ are on the form $\lambda - t$, where λ is an eigenvalue for $\boldsymbol{S}$. Since eigenvalues for $\boldsymbol{S}$ all have strictly negative real parts (see [2], p.134), the eigenvalues for $\boldsymbol{S} - t\boldsymbol{I}$ all have strictly negative real parts whenever t is larger than or equal zero. In particular, $\boldsymbol{S} - t\boldsymbol{I}$ is invertible for $t \geq 0$ since the determinant of a matrix equals the product of its eigenvalues. Using that

$$\int e^{\boldsymbol{A}x} dx = \boldsymbol{A}^{-1} e^{\boldsymbol{A}x}$$

we get

$$L_\tau(t) = \boldsymbol{\alpha}(t\boldsymbol{I} - \boldsymbol{S})^{-1} \boldsymbol{s}. \tag{8}$$

From the Laplace transform we obtain the moments of τ by differentiation and evaluation in zero

$$\mu_n = \mathbb{E}(\tau^n) = \boldsymbol{\alpha}(-\boldsymbol{S}^{-1})^n \boldsymbol{e} = \boldsymbol{\alpha} \boldsymbol{U}^n \boldsymbol{e}, \tag{9}$$

where $\boldsymbol{U} = -\boldsymbol{S}^{-1}$. The matrix $\boldsymbol{U} = \{u_{ij}\}$ is the so-called *Green* matrix and its elements have the following interpretation: u_{ij} equals the expected time the process $\{X_t\}$ spends in state j prior to absorption given that $X_0 = i$. From this interpretation we can also obtain the formula for μ_1 without using the Laplace transform.

Example 3.1. Using the fomula $\mu_i = i!\boldsymbol{\alpha}(-\boldsymbol{S})^i\boldsymbol{e}$ we calculate the two first moments ($i = 1,2$) of the total tree height in the Kingman's, Psi– and Beta–coalescent models.

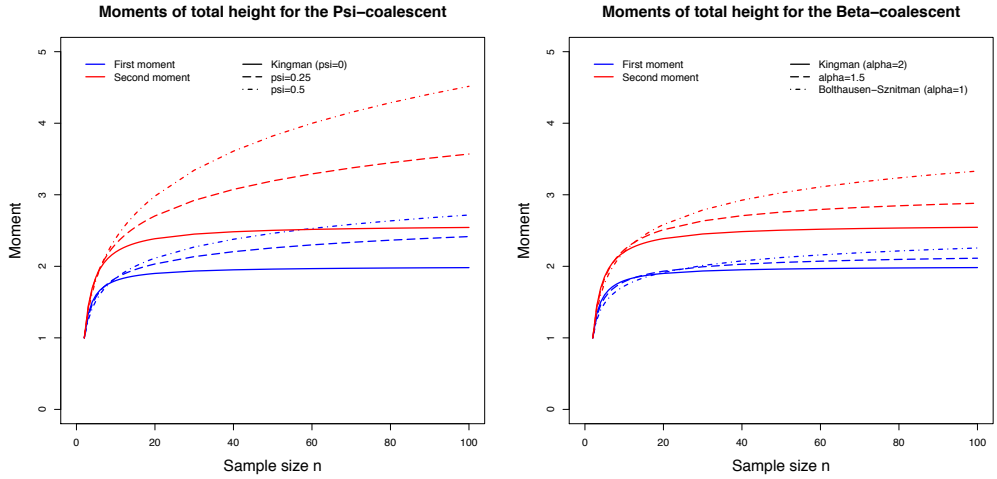

Figure 4: Two first moments of total tree height as a function of sample size. Left: Kingman's and Psi–coalescent model. Right: Kingman's and Beta-coalescent model.

3.2 Transformations using rewards

Let $\tau \sim \text{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$, $\{X_t\}_{t\geq 0}$ its underlying Markov jump process and $\boldsymbol{r} = (r(1), \ldots, r(p)) \in \mathbb{R}_+^p$ a vector of non-negative numbers (reward rates). We then define the total reward Y earned before time τ as

$$Y = \int_0^\tau r(X_t)dt. \tag{10}$$

If $r(i) \neq 0$ and $T \sim \exp(\lambda_i)$ is a holding time in state i, then the reward earned during this holding time is simply $r(i) \cdot T \sim \exp(\lambda_i / r(i))$. Hence, if all $r(i) \neq 0$ and $\boldsymbol{\Delta}(\boldsymbol{r})$ denotes the diagonal matrix with $\boldsymbol{r}$ on the diagonal, we have that

$$Y \sim \text{PH}_p(\boldsymbol{\alpha}, \boldsymbol{\Delta}^{-1}(\boldsymbol{r})\boldsymbol{S}).$$

In the context of genealogical trees, consider an n–coalescent model in which the tree height has a phase–type distribution $\text{PH}_{n-1}(\boldsymbol{\pi}, \boldsymbol{T})$ like in the Examples 2.3 to 2.6. Then the total branch length will have a phase–type distribution $\text{PH}_{n-1}\big(\boldsymbol{\pi}, \boldsymbol{\Delta}^{-1}(\boldsymbol{r})\boldsymbol{T}\big)$, where $\boldsymbol{r} = (n, n-1, ..., 2)$. This follows immediately from a reward consideration.

As we shall see in relation to the site frequency spectrum for the multiple merger coalescent in Section 4, and the coalescent with recombination in Section 5, some rewards may be zero. Then the non-zero rewards earned during holding times will still be exponentially distributed obtained by scaling with the appropriate reward, but the embedded chain of the new phase-type distribution will change since going from one state with positive reward to another positive-reward state can take place via transitions in zero-reward states.

Define $E^+ = \{i \in E : r(i) > 0\}$ and $E^0 = \{i \in E : r(i) = 0\}$ and decompose accordingly the vector $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^0)$ and transition matrix $\boldsymbol{Q}$ of the embedded chain $\{Y_n\}_{n\in\mathbb{N}}$

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}^{++} & \boldsymbol{Q}^{+0} \\ \boldsymbol{Q}^{0+} & \boldsymbol{Q}^{00} \end{pmatrix}.$$

Let $d = |E^+|$ be the number of elements in E^+ and define

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+} \\ \boldsymbol{\pi} &= \boldsymbol{\alpha}^+ + \boldsymbol{\alpha}^0(\boldsymbol{I} - \boldsymbol{Q}^{00})^{-1}\boldsymbol{Q}^{0+}. \end{aligned}$$

Then $\boldsymbol{P} = \{p_{ij}\}_{i,j=1,\ldots,d}$ is the transition matrix of the Markov chain which is obtained from $\{Y_n\}_{n\in\mathbb{N}}$ at times when $Y_n \in E^+$. This follows by noticing that the (i,j)'th element of $\boldsymbol{Q}^{+0}(\boldsymbol{Q}^{00})^n\boldsymbol{Q}^{0+}$ is the probability of going from i to j by first making a transition to a state in E^0, remaining in E^0 for the next n jumps, and finally jumping from a state in E^0 to j, and since

$$\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} = \sum_{m=0}^{\infty} (\boldsymbol{Q}_{00})^m.$$

With a similar argument, π_i gives the probability that a Markov process starts earning rewards from state $i \in E^+$, which can either happen by $X_0 = i \in E^+$ or by $X_0 \in E^0$ and eventually returning to E^+. Since there in general exists the possibility of never entering E^+ if the process is started in E^0, there will in general be an atom (point mass) at zero of size $\pi_{d+1} = 1 - \boldsymbol{\pi}\boldsymbol{e}$. Hence we have proved the following:

Theorem 3.2 ([2], p. 164)**.** *The random variable Y of (10) is a mixture distribution of a point mass at 0 of size $\pi_{d+1} = 1 - \boldsymbol{\pi}\boldsymbol{e}$ and a phase-type distribution with representation $PH_d(\boldsymbol{\pi}, \boldsymbol{T}^*)$ where $\boldsymbol{T}^* = \{t^*_{ij} : (i,j) \in E^+\}$ is given by*

$$t^*_{ij} = -\frac{s_{ii}}{r(i)}p_{ij} \text{ for } i \neq j \text{ and } t^*_{ii} = \frac{s_{ii}}{r(i)}(1 - p_{ii}).$$

Example 3.3 (Seed bank model continued)**.** In the seed bank model the mutation rate can be inferred from the total branch length of the active part of the coalescent (because mutations only occur out of the seed bank). To this end, we use the reward vector $(n, n-1, ..., 1, 0, n-1, n-2, ..., 0, ..., 2, 1, 0)$. In the peripatric model, the population is separated in continent and islands, hence they can mutate at both stages. The total number of mutations is in this case related to the total branch length, and in this case the reward vector is $(n, \ldots, n, n-1, \ldots, 3, 2, 2, 2)$. Results on expected heights and branch lengths are summarized in Figure 5. Moreover, it is interesting to consider the total number of mutations as the sum of continental mutations and island mutations. This problem can be studied in the multivariate phase-type framework.

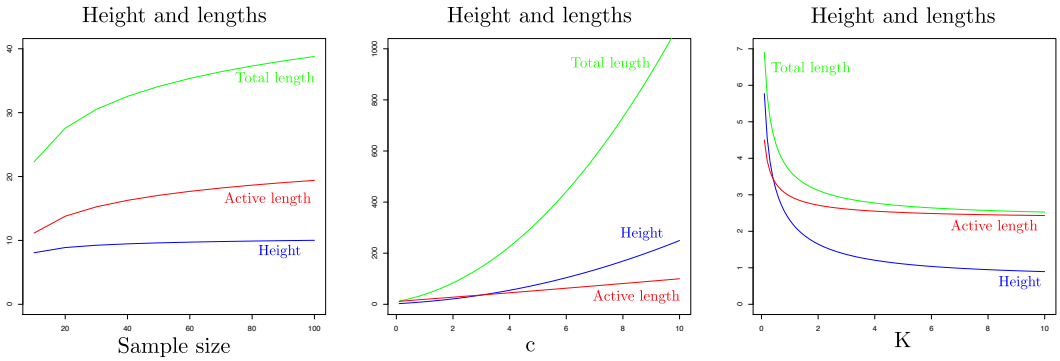

Figure 5: Expected height, active branch length and total branch length of the peripatric/seed-bank coalescent with respect to the sample size n, active branch rate c and inactive branch rate K. Left: $c = K = 1$, Middle: $n = 100$ and $K = 1$. Right: $n = 100$ and $c = 1$.

3.3 Multivariate phase-type distributions

Let $\tau \sim \text{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$ and let $\{X_t\}_{t\geq 0}$ denote the underlying Markov jump process which generates τ. Let m be a positive integer and let $\boldsymbol{R} = \{R_{ij}\}$ be a $p \times m$ matrix of non-negative rewards. Each column j of $\boldsymbol{R}$ may be considered to be a function $r_j : \{1, 2, \ldots, p\} \to \mathbb{R}_+$ defined by $r_j(i) = R_{ij}$. Then we define

$$Y_j = \int_0^\tau r_j(X_t)dt = \int_0^\tau R_{X_t,j}\, dt,$$

and say that the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ has a multivariate phase-type distribution parametrized by $\boldsymbol{\alpha}$, $\boldsymbol{S}$, and $\boldsymbol{R}$, and write $\boldsymbol{Y} \sim \text{MPH}^*_p(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$.

For example, we may consider the joint distribution of the times that the process $\{X_t\}_{t\geq 0}$ has spent in different (possibly overlapping) subsets of the state-space prior to absorption. This will generate a multivariate phase-type distribution based on rewards which are either zero or one (see Figure 6 for an example).

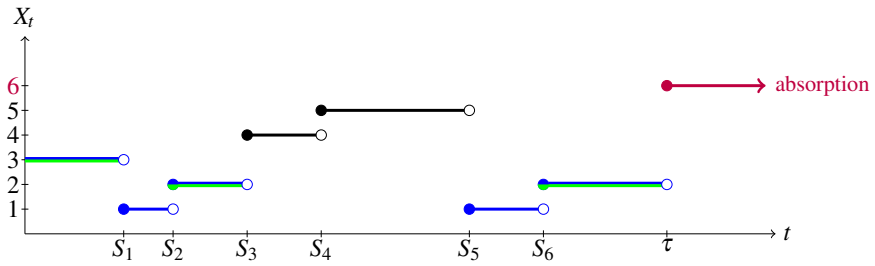

Figure 6: A Markov process with 5 transient states (blue, green and black) and one absorbing state (purple). The total time Y_1 spent in states 2 and 3 prior to absorption (green) and the total time Y_2 spent in states $1, 2, 3$ (blue) prior to absorption defines a bivariate phase-type distribution.

The joint distribution of $\boldsymbol{Y}$ can be expressed in a compact form in terms of the joint Laplace transform.

Theorem 3.4 (Theorem 8.1.2 in [2]). *Let $\boldsymbol{Y} \sim MPH^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $\langle \cdot, \cdot \rangle$ denote the usual dot product. Then for any vector $\boldsymbol{\theta} \geq \boldsymbol{0}$, the joint Laplace transform $L_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \mathbb{E}(\exp(-\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle))$ is given by*

$$L_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \boldsymbol{\alpha} \left(\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta}) - \boldsymbol{S} \right)^{-1} \boldsymbol{s}. \tag{11}$$

In general it is not possible to provide explicit formulae for the joint density function or distribution functions, however, in some important special cases it is possible to derive strikingly simple expressions (see e.g. Section 8.1 of [2]). Of special interest are means, variances and covariances between elements of $\boldsymbol{Y}$. Let $\boldsymbol{R}_{\cdot i}$ denotes the ith column of $\boldsymbol{R}$ and recall $\boldsymbol{U} = -\boldsymbol{S}^{-1}$ is the Green matrix. Then we have that

$$\begin{aligned} \mathbb{E}(Y_i) &= \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{R}_{\cdot i} & (12) \\ \mathbb{E}(Y_i Y_j) &= \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{U} \boldsymbol{R}_{\cdot j} + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{U} \boldsymbol{R}_{\cdot i} & (13) \end{aligned}$$

for all i, j (including $i = j$), and from which we can calculate the covariance by the well known formula

$$\mathrm{Cov}(Y_i, Y_j) = \mathbb{E}(Y_i Y_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j). \tag{14}$$

Higher order moments (see Theorem 8.1.5 of [2]) can be calculated by the formula

$$\mathbb{E}\left(\prod_{j=1}^{p} Y_j^{h_j} \right) = \boldsymbol{\alpha} \sum_{\ell=1}^{h!} \left(\prod_{i=1}^{h} U \boldsymbol{\Delta}\left(\boldsymbol{R}_{\cdot \sigma_\ell(i)} \right) \right) \boldsymbol{e}, \tag{15}$$

where $h = \sum_{j=1}^{n} h_j$ and $\sigma_\ell(i)$ is the index value for entrance ℓ of the ith ordered permutation of the indices. For example, if we want to calculate $\mathbb{E}(Y_i Y_j Y_k)$ for i, j, k all different, then we consider all ordered permutations of (i, j, k) which amounts to (i, j, k),(i, k, j),(j, i, k),(j, k, i), (k, i, j) and (k, j, i) resulting in the formula

$$\begin{aligned} \mathbb{E}(Y_i Y_j Y_k) &= \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{e} + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{e} \\ &\quad + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{e} + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{e} \\ &\quad + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{e} + \boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot k}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot j}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i}) \boldsymbol{e}. \end{aligned} \tag{16}$$

For $\mathbb{E}(Y_i^2 Y_j Y_k)$ we would have to consider permutations of (i, i, j, k) and summing expressions on the form

$$\boldsymbol{\alpha} \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i_1}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i_2}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i_3}) \boldsymbol{U} \boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i_4}) \boldsymbol{e},$$

where two among the i_1, i_2, i_3, i_4 are identical to i while among the remaining two one equals j and the other equals k.

3.4 Discrete phase–type distributions and the number of segregating sites

A discrete phase–type distribution is defined very similar to the continuous case, where the Markov jump process is simply replaced by a Markov chain. Thus we consider a Markov chain $\{X_n\}_{n \in \mathbb{N}}$ on a state–space $\{1, 2, ..., p, p+1\}$, where $1, 2, ..., p$ are transient and $p+1$ absorbing. The transition matrix for $\{X_n\}_{n \in \mathbb{N}}$ is hence on the form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix},$$

where $\boldsymbol{T}$ is now a *sub-transition* matrix and $\boldsymbol{t}$ the *exit probability* vector. The initial distribution $\boldsymbol{\pi}$ is again supposed to be concentrated on $\{1,...,p\}$ so that the support for

$$\tau_d = \inf\{n \geq 1 : X_n = p+1\}$$

is the positive natural numbers (not including zero). With arguments entirely similar to the continuous case it is easy to prove that τ_d has density

$$\mathbb{P}(\tau_d = n) = \boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t},$$

and distribution function

$$\mathbb{P}(\tau_d \leq n) = 1 - \boldsymbol{\pi}\boldsymbol{T}^{n}\boldsymbol{e}.$$

For discrete distributions it is mostly the probability generating function which is of interest which amounts to

$$\mathbb{E}(z^{\tau_d}) = \sum_{n=1}^{\infty} \mathbb{P}(\tau_d = n)z^n = z\boldsymbol{\pi}(\boldsymbol{I} - z\boldsymbol{T})^{-1}\boldsymbol{t} = \boldsymbol{\pi}\left(z^{-1}\boldsymbol{I} - \boldsymbol{T}\right)^{-1}\boldsymbol{t},$$

and from which we obtain the *factorial moments*,

$$\mathbb{E}(\tau_d(\tau_d - 1)\cdots(\tau_d - k + 1)) = k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}(\boldsymbol{I} - \boldsymbol{T})^{-k}\boldsymbol{e}.$$

In the discrete case the moments $\mathbb{E}(\tau_d^n)$ are not directly available but can be deduced from the factorial moments. For a detailed account on discrete phase–type distributions we refer to [2], pp. 29–36.

Discrete phase–type distributions appear in a natural way as the distribution of the number of segregating sites. Consider a genealogical model where the total branch length $\mathscr{L} \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ and let the mutation rate at the locus be $\lambda = \theta/2$. Then the number of segregating sites S has a conditional distribution given $\mathscr{L} = x$ which is Poisson with parameter λx. Thus

$$\begin{aligned}
\mathbb{P}(S = n) &= \int_0^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x}\boldsymbol{\pi}e^{\boldsymbol{T}x}\boldsymbol{t}dx \\
&= \frac{\lambda^n}{n!}\int_0^{\infty} x^n e^{-\lambda x} f(x)dx \\
&= \frac{\lambda^n}{n!}(-1)^n \frac{\partial^n}{\partial\lambda^n} L(\lambda) \\
&= \frac{\lambda^n}{n!}(-1)^n(-1)^n n!\boldsymbol{\pi}(\lambda\boldsymbol{I} - \boldsymbol{T})^{-(n+1)}\boldsymbol{t} \\
&= \lambda^{-1}\boldsymbol{\pi}(\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-(n+1)}\boldsymbol{t},
\end{aligned}$$

where $L(s) = \boldsymbol{\pi}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t}$ is the Laplace transform for $\mathscr{L}$. Now

$$(\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1} = \boldsymbol{I} + \lambda^{-1}(\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1}\boldsymbol{T}$$

from which

$$\boldsymbol{I} - (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1} = (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1}(-\lambda^{-1}\boldsymbol{T}). \tag{17}$$

The matrix $\boldsymbol{R}(\lambda) = (\lambda\boldsymbol{I} - \boldsymbol{T})^{-1}$ is called the *resolvent* of $\boldsymbol{T}$, and

$$\boldsymbol{R}(\lambda) = (\lambda\boldsymbol{I} - \boldsymbol{T})^{-1} = \int_0^{\infty} e^{-\lambda x}e^{\boldsymbol{T}x}dx,$$

or, in terms of individual elements,

$$r_{ij}(\lambda) = \int_0^\infty e^{-\lambda x} q_{ij}(x) dx,$$

where $\boldsymbol{Q}(x) = \exp(\boldsymbol{T}x)$ is a sub–transition matrix, whose (i,j)'th element $q_{ij}(x)$ is the probability of going from state i to state j in time x , where i and j are transient states. It follows that

$$\lambda r_{ij}(\lambda) = \int_0^\infty \lambda e^{-\lambda x} q_{ij}(x) dx.$$

Thus the (i,j)'th element of

$$\lambda \boldsymbol{R}(\lambda) = (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1}$$

the probability that the Markov process underlying the phase–type distribution goes from state i to state j at an exponentially distributed random time with parameter λ. In particular, we conclude that

$$\boldsymbol{P} = (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1} \tag{18}$$

is a sub–transition matrix. Now from

$$\boldsymbol{p} = \boldsymbol{e} - \boldsymbol{P}\boldsymbol{e} = \big(\boldsymbol{I} - (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T}^{-1})\big)\boldsymbol{e}$$

and (17), we have then proven the following theorem.

Theorem 3.5. *Let $\mathscr{L} \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ and the mutation rate at the locus be $\lambda = \theta/2$. Then the number of segregating sites S has a density given by*

$$\mathbb{P}(S = n) = \boldsymbol{\pi}\boldsymbol{P}^n\boldsymbol{p}, \tag{19}$$

where $\boldsymbol{P} = (\boldsymbol{I} - \lambda^{-1}\boldsymbol{T})^{-1}$ and $\boldsymbol{p} = \boldsymbol{e} - \boldsymbol{P}\boldsymbol{e}$, i.e.

$$S + 1 \sim DPH_p(\boldsymbol{\pi}, \boldsymbol{P}).$$

The reason for adding one to S is that the support for discrete phase–type distributions is on the natural numbers excluding zero (immediate absorption is not possible). It is of no practical relevance at all but should be kept in mind when applying standard formlae from discrete phase–type theory.

Example 3.6 (Beta-coalescent continued)**.** Take $n = 5$, $\alpha = 1.5$ and $\theta = 2$. Then the matrix $\boldsymbol{S}$ in (6) is

$$\boldsymbol{S} = \begin{pmatrix} -6.5625 & 5.46875 & 0.78125 & 0.234375 \\ 0 & -4.375 & 3.75 & 0.5 \\ 0 & 0 & -2.5 & 2.25 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

so that the tree–height is phase–type distributed $\mathrm{PH}_4(\boldsymbol{e}_1, \boldsymbol{S})$ while the total branch length is phase–type distributed $\mathrm{PH}_4(\boldsymbol{e}_1, \boldsymbol{T})$ where

$$\boldsymbol{T} = \begin{pmatrix} 1/5 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix} \boldsymbol{S} = \begin{pmatrix} -1.31250000 & 1.09375000 & 0.156250000 & 0.004687500 \\ 0 & -1.09375000 & 0.937500000 & 0.125000000 \\ 0 & 0 & -0.833333313 & 0.750000000 \\ 0 & 0 & 0 & -0.500000000 \end{pmatrix}.$$

Then

$$\boldsymbol{P} = \left(\boldsymbol{I} - \frac{2}{\theta}\boldsymbol{T}\right)^{-1} = \begin{pmatrix} 0.432432443 & 0.225897521 & 0.152370825 & 0.108523726 \\ 0.00000000 & 0.477611929 & 0.244233400 & 0.161917701 \\ 0.00000000 & 0.00000000 & 0.545454562 & 0.272727311 \\ 0.00000000 & 0.00000000 & 0.00000000 & 0.666666687 \end{pmatrix}$$

and we get that

$$\mathbb{P}(S=m) = (1,0,0,0) \begin{pmatrix} 0.432432443 & 0.225897521 & 0.152370825 & 0.108523726 \\ 0.00000000 & 0.477611929 & 0.244233400 & 0.161917701 \\ 0.00000000 & 0.00000000 & 0.545454562 & 0.272727311 \\ 0.00000000 & 0.00000000 & 0.00000000 & 0.666666687 \end{pmatrix}^m \begin{pmatrix} 0.080775499 \\ 0.116236985 \\ 0.181818128 \\ 0.333333313 \end{pmatrix}.$$

The point probabilities are given in Table 1 and plotted in Figure 7.

m	$\mathbb{P}(S=m)$
0	0.080775499
1	0.125065953
2	0.141926453
3	0.137703091
5	0.100281194
7	0.060273220
9	0.032527771
10	0.023270819
11	0.016433677
12	0.011484630
13	0.007958241
14	0.005476677
15	0.003747694
16	0.002552703
17	0.001732148
18	0.001171687
19	0.007905333

Table 1: Density of S.

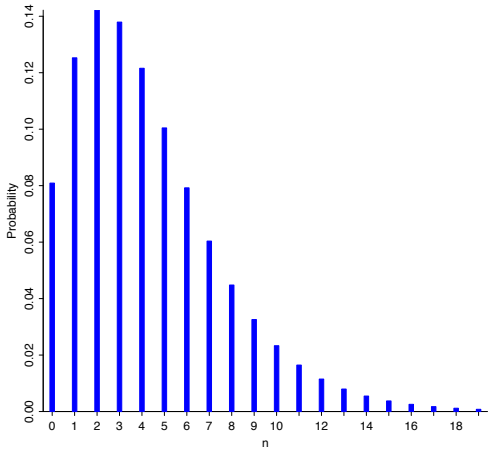

Figure 7: Density of the number of segregating sites for the Beta-coalescent with $\alpha = 1.5$.

4 Coalescent theory without recombination

In order to study the site frequency spectrum we need to introduce an appropriate state-space and a corresponding reward matrix. For a sample of size n, we represent the states by a vector $\boldsymbol{a} = (a_1, a_2, ..., a_n)$ where a_i denotes the number of branches with i descendants. The state-space is thus given by

$$\left\{\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{Z}_+^n \ : \ \sum_{i=1}^{n} i a_i = n\right\}.$$

This representation is similar to the summary of a sample of DNA sequences used for the infinite alleles model in Ewens' sampling formula. For Kingman's coalescent the possible transitions are

$$(a_1, \ldots, a_n) \to (a_1, \ldots, a_i - 1, \ldots, a_j - 1 \ldots, a_{i+j} + 1, \ldots, a_n)$$

with rate $a_i a_j$ for $a_i, a_j \geq 1$, and

$$(a_1, \ldots, a_n) \to (a_1, \ldots, a_i - 2, \ldots, a_{2i} + 1, \ldots, a_n)$$

with rate $\binom{a_i}{2}$ for $a_i \geq 2$. The row in the reward matrix corresponding to a state $\boldsymbol{a} = (a_1, ..., a_n)$ is given by $(a_1, ..., a_{n-1})$ because a_1 is the number of branches with one descendant, a_2 is the number of branches with two descendants etc. (see also Table 2).

Example 4.1. Consider Kingman's coalescent with $n = 4$. In Figure 8 we show the state space and possible transitions.

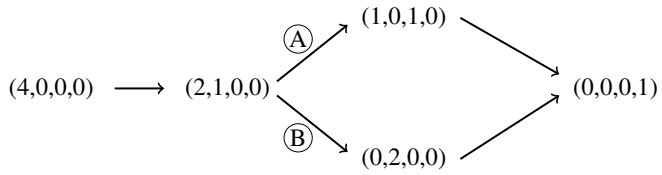

Figure 8: A flow diagram for the case of four sequences in the Kingman's coalescent, where circles refer to the topologies from Figure 3.

The intensity and reward matrices are given in Table 2.

State		Intensity matrix					Reward $\boldsymbol{R}$			Number of
Type	Index	1	2	3	4	5	$\boldsymbol{R}_{\cdot 1}$	$\boldsymbol{R}_{\cdot 2}$	$\boldsymbol{R}_{\cdot 3}$	branches
$(4,0,0,0)$	1	$-\binom{4}{2}$	$\binom{4}{2}$	0	0	0	4	0	0	4
$(2,1,0,0)$	2	0	-3	1	2	0	2	1	0	3
$(0,2,0,0)$	3	0	0	-1	0	1	0	2	0	2
$(1,0,1,0)$	4	0	0	0	-1	1	1	0	1	2
$(0,0,0,1)$	5	0	0	0	0	0	0	0	0	1

Table 2: Intensity matrix for Kingman's coalescent and reward matrix for calculating the site frequency spectrum for $n = 4$ sequences.

The elements of each row in $\boldsymbol{R}$ correspond to the number of branches with one, two or three descendants. The row sums of the reward matrix equals the number of branches, except for the last absorbing state where only one lineage is present.

We now provide an algorithm for generating the general state-space and corresponding transition rates.

Algorithm 4.2. *The state-space is determined as follows. The transition*

$$\boldsymbol{a} = (a_1, \ldots, a_n) \to \boldsymbol{b} = (b_1, \ldots, b_n)$$

is possible if the vector $\boldsymbol{c} = (c_1, \ldots, c_n) = \boldsymbol{b} - \boldsymbol{a}$ *fulfils the three conditions*

(i) $\sum_{i=1}^{n} c_i \mathbf{1}_{\{c_i > 0\}} = 1$ *(one new branch is created)*

(ii) $\sum_{i=1}^{n} c_i \mathbf{1}_{\{c_i<0\}} = -2$ *(two branches are merged)*

(iii) $\sum_{i=1}^{n} i c_i = 0$ *(balance equation on the sample size).*

The transition rates between the states are

$$S_{\boldsymbol{ab}} = \lambda_{\sum_{i=1}^{n} a_i, -\sum_{i=1}^{n} c_i \mathbf{1}_{\{c_i<0\}}} \prod_{i:c_i<0} \binom{a_i}{-c_i}. \tag{20}$$

We start with $\boldsymbol{a} = (n, 0, ..., 0)$ and identify the remaining states subsequently. In Figure 9 we show the state space and possible transitions for the general Λ-coalescent.

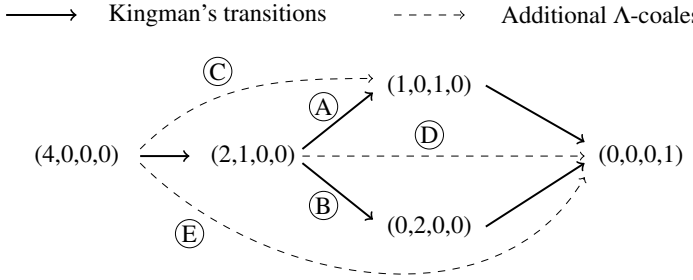

Figure 9: Flow diagram for the case of four sequences in the Λ-coalescent model. The numbers in the circles refer to the topologies in Figure 3.

For a general Λ-coalescent process, mutations on branches with one descendant give rise to singletons in the site frequency spectrum, while mutations with two or three descendants give rise to doubletons, tripletons and so on in the site frequency spectrum. The quantities

$$Y_i = \int_0^{\tau} \boldsymbol{R}_{X_t,i}\, dt, \;\; i = 1, 2, 3, ..., n-1,$$

are the total branch lengths where a mutation is shared by exactly i samples. If the mutation rate is $\theta/2$, then the expected site frequency spectrum (SFS) is given by

$$\mathbb{E}(\xi_i) = \frac{\theta}{2}\mathbb{E}(Y_i),$$

where $\mathbb{E}(Y_i)$ is given by (12). Covariances are given by

$$\mathrm{Cov}(\xi_i, \xi_j) = \frac{\theta^2}{4}\mathrm{Cov}(Y_i, Y_j),$$

for which we use (13) and (14).

Example 4.3. Here we consider the variance, covariance and expected site frequency spectrum for the Psi-coalescent (Figure 10) and the Beta-coalescent (Figure 11). The parameter ψ gives the proportion of lineages that merge at each jump. Hence, after the first jump we have one block of size ψn. This

explains the bumps in the SFS entry in index ψn. For example, we have a bump at index $0.75 \cdot 20 = 15$ for the green curve, and a bump at $0.5 \cdot 20 = 10$ for the blue curve.

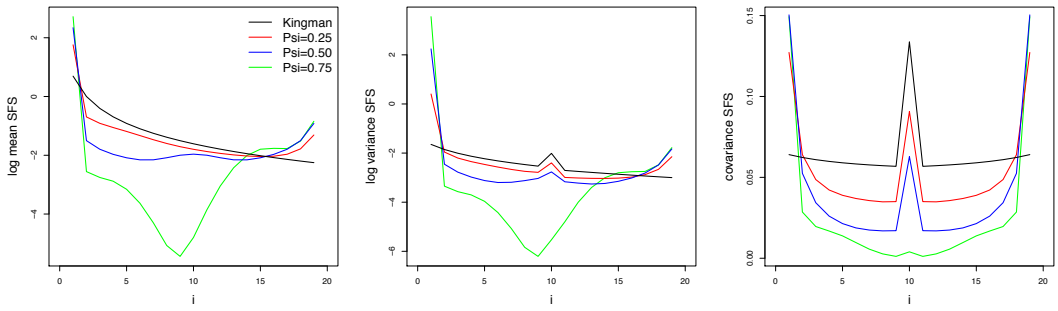

Figure 10: Psi-coalescence for $\psi = 0.25$ (red), $\psi = 0.5$ (blue), $\psi = 0.75$ (green) compared to Kingman's coalescent (black) with a sample size of $n = 20$. Left: Logarithm of the expected SFS. Middle: Logarithm of the variances of the SFS. Right: Anti–diagonal values for the covariance matrix of the SFS.

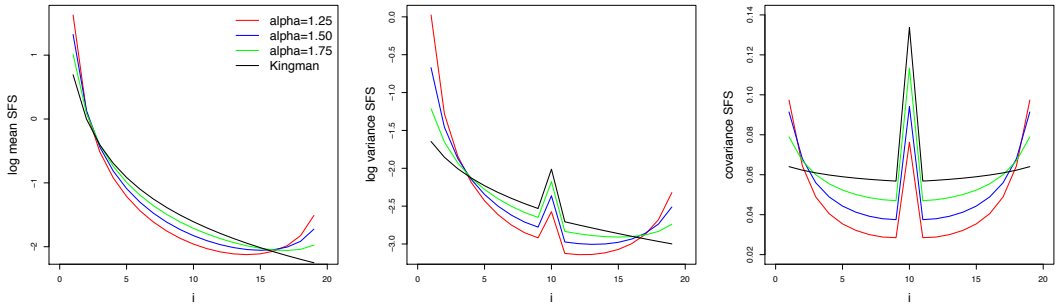

Figure 11: Beta-coalescence for $\alpha = 1.25$ (red), $\alpha = 1.50$ (blue), $\alpha = 1.75$ (green) and Kingman's coalescent (black) with a sample size of $n = 20$. The left, middle and right plot show the same summary statistics as in Figure 10.

Concerning the covariances we only plot the anti-diagonal entries of the covariance matrix as in [6] p.56. Our results for the mean, variance and covariance agree with those obtained by the recursive formulae presented in [1]. Higher order moments can also be calculated using (15). For example, [12] use the 3rd order moments to derive analytical results for the bias of Tajima's D and other neutrality tests. Their derivation essentially reduces to calculating the formula (16).

4.1 Application: Analysis of the Faroe Island Atlantic Cod data

We now consider the Faroe Island Atlantic Cod mitochondrial data from [24]. The data is illustrated in the left plot in Figure 12. The data consists of $n = 74$ sequences, the number of haplotypes is 41,

and the data has $S = 44$ segregating sites. The folded site frequency spectrum is a vector of length $n/2 = 37$ and is given by

$$\eta = (23, 10, 1, 0, 1, 0, 2, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, \ldots, 0).$$

Recall that if the unfolded SFS is given by ξ_i, $i = 1, \ldots, n-1$, then the folded is given by $\eta_i = \xi_i + \xi_{n-i}$ for $i = 1, \ldots, n/2 - 1$ and $\eta_{n/2} = \xi_{n/2}$.

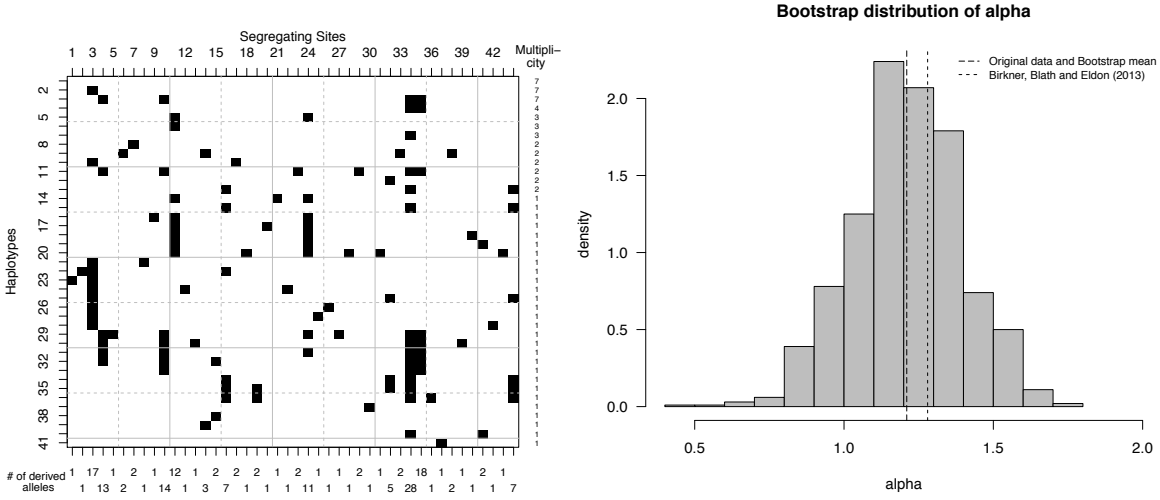

Figure 12: Left: The segregating sites matrix and corresponding multiplicity for the Faroe Island Atlantic Cod data. The data consists of 74 sequences, 41 haplotypes, and 44 segregating sites. Right: Bootstrap distribution of the α parameter in the Beta-coalescent. Kingman's coalescent corresponds to $\alpha = 2$, and this parameter value is not reasonable for this data. We find the parameter value 1.21 for the original data, and a bootstrap mean of 1.21. These values are very similar to the pseudolikelihood parameter estimate of 1.28 in [1].

This data was analysed using both the Psi- and the Beta-coalscent by [1]. They found that the Beta-coalescent was the most appropriate, and therefore we also focus on the Beta-coalescent. Recall from Example 2.6 that the Beta-coalescent is parametrized by α with parameter space $0 < \alpha < 2$.

[1] estimate α in the Beta-coalescent as the minimizer of an objective function that measures the difference between the scaled observed and scaled expected values of the site frequency spectrum. [1] derive recursions for the expected site frequency spectrum that apply for very large sample sizes (in the order 10.000). Denote the observed scaled folded site frequency spectrum $\zeta_i = \eta_i / S$, where $S = \sum_{i=1}^{n/2} \eta_i$ is the number of segregating sites, and denote the expected scaled folded site frequency spectrum r_i. Then [1] consider the Eucledian distance function $\left\{ \sum_{i=1}^{n/2} (\zeta_i - r_i)^2 \right\}^{1/2}$ and the negative pseudolikelihood distance $-\sum_{i=1}^{n} \zeta_i \log r_i$.

Unfortunately, our method for calculating the expected SFS suffers from a state space explosion. For samples of size $n = (5, 10, 15, 20, 25, 30)$ the state space is of size (7,42,176,627,1958,5604). We need to invert a matrix the size of the state space, and this means that we are limited to a sample size of at most 25. We therefore applied a sub-sampling procedure for the Faroe Island Atlantic Cod data.

In particular we sampled 20 sequences from the data 2000 times. For each sample we calculated the scaled site frequency spectrum, and finally we calculated the average scaled site frequency spectrum. We then applied the pseudolikelihood method from [1]. We obtained an estimated value of α of 1.21, which is very similar to the value of 1.28 obtained by [1].

In order to understand the uncertainty in the parameter value we carried out a bootstrap procedure. Each bootstrap sample was obtained by sampling a new set of 44 segregating sites uniformly at random with replacement from the original columns of segregating sites. For each sample we estimated α using the same method as for the original data. In Figure 12 we show the histogram of the parameter values based on 1000 bootstrap samples. The mean of the bootstrap values is almost identical to the value of the original data, but more importantly the bootstrap distribution is far away from the Kingman's coalescent of $\alpha = 2$.

5 Ancestral graph with recombination

In this section we show how multivariate phase-type theory applies as a model for the distribution of branch length and can be used to express expected summaries for mutation patterns at different loci. We begin with a sample of size $n = 2$ and then extend to larger sample sizes.

5.1 Sample size two

Recall the ancestral recombination graph for two loci and two samples originally presented in [25], summarized as Figure 7.7 in [26], and recently discussed in detail in [11]. For reference the graph is reproduced here in Figure 13. The filled circles represent material ancestral to the sample, and the crosses indicate that the most common ancestor has been found. The lines between the circles or crosses indicate if the ancestral material is present on the same chromosome. The starting state is state 1 at present day with two samples from the same chromosome.

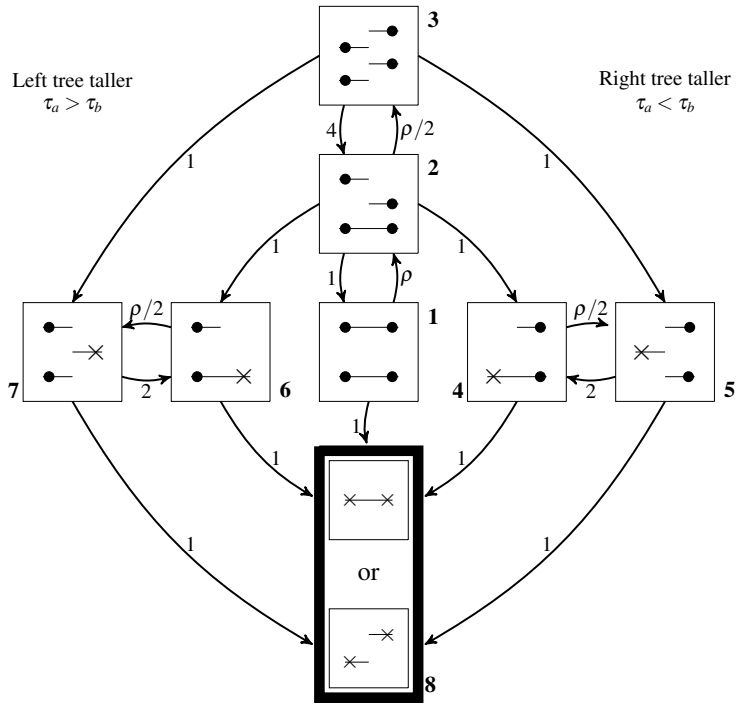

Figure 13: Flow diagram for the two-locus ancestral recombination graph.

The time when both loci have found their common ancestor is $\mathrm{PH}_7(\boldsymbol{\alpha}, \boldsymbol{S})$ distributed with $\boldsymbol{\alpha} = (1,0,0,0,0,0,0)$ and

$$\boldsymbol{S} = \left(\begin{array}{ccc|cc|cc} -1-\rho & \rho & 0 & 0 & 0 & 0 & 0 \\ 1 & -3-\rho/2 & \rho/2 & 1 & 0 & 1 & 0 \\ 0 & 4 & -6 & 0 & 1 & 0 & 1 \\ \hline 0 & 0 & 0 & -1-\rho/2 & \rho/2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -3 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -1-\rho/2 & \rho/2 \\ 0 & 0 & 0 & 0 & 0 & 2 & -3 \end{array}\right), \quad \boldsymbol{s} = \left(\begin{array}{c} 1 \\ 0 \\ 0 \\ \hline 1 \\ 1 \\ \hline 1 \\ 1 \end{array}\right). \tag{21}$$

We observe that $\boldsymbol{S}$ has the natural block structure partitioning

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \boldsymbol{S}_{13} \\ \boldsymbol{0} & \boldsymbol{S}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{33} \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \\ \boldsymbol{s}_3 \end{pmatrix},$$

as already indicated in (21). Also note the highly symmetric structure of the partitioning where $\boldsymbol{S}_{12} = \boldsymbol{S}_{13}$, $\boldsymbol{S}_{22} = \boldsymbol{S}_{33}$ and $\boldsymbol{s}_2 = \boldsymbol{s}_3$.

If the Markov jump process underlying $\mathrm{PH}_7(\boldsymbol{\alpha}, \boldsymbol{S})$ exits to the absorbing state from state 1, then the height τ_a of the left tree and the height τ_b of the right tree are the same, i.e. $\tau_a = \tau_b$ with the common height being phase-type distributed with representation $\mathrm{PH}_3(\boldsymbol{\alpha}_1, \boldsymbol{S}_{11})$ where $\boldsymbol{\alpha}_1 = (1,0,0)$.

The common distribution of $\tau_a = \tau_b$ then has density

$$f(x) = (1,0,0)\exp\left\{\begin{pmatrix} -1-\rho & \rho & 0 \\ 1 & -3-\rho/2 & \rho/2 \\ 0 & 4 & -6 \end{pmatrix} x\right\}\begin{pmatrix}1\\0\\0\end{pmatrix}. \tag{22}$$

The density for equal tree height is shown in the left plot in Figure 14. This is a defective distribution since second and third exit rates are set to zero prohibiting the process to jump to the left or right states of the diagram so

$$-\boldsymbol{S}_{11}\boldsymbol{e} \neq \boldsymbol{s}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}.$$

The missing mass is exactly the probability of these events occurring and amounts to

$$1 - \int_0^\infty f(x)dx = 1 - \boldsymbol{\alpha}_1(-\boldsymbol{S}_{11})^{-1}\boldsymbol{s}_1.$$

The density function (22) can also be evaluated explicitly, i.e. expressed in terms of polynomials and exponentials involving ρ and x. However, this expression is lengthy and messy since the eigenvalues of the intensity matrix are not particularly nice functions. On the other hand, for specific numeric values of ρ, the numeric calculation of (22) is straightforward and efficient. Thus there seems to be no reason for pursuing a non-matrix representation of (22) in practice.

Now let us consider the case where $\tau_a \neq \tau_b$. Assume that $x = \tau_a < y = \tau_b$. Then the right tree is taller, and we must exit from states $\{1,2,3\}$ to $\{4,5\}$ at time x. The 3-dimensional row vector

$$\boldsymbol{\alpha}_1 e^{\boldsymbol{S}_{11}x}$$

contains the probabilities of being in state 1, 2 or 3 when exiting while the 2-dimensional row vector

$$\boldsymbol{\pi}_1 = \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_{11}x}\boldsymbol{S}_{12}$$

contains the probabilities that states 4 and 5 are entered. Thus $\boldsymbol{\pi}_1$ serves as the initial (defective) distribution of entering states $\{4,5\}$, and the remaining time spent in states $\{4,5\}$ prior to absorption is hence phase-type distributed $\mathrm{PH}_2(\boldsymbol{\pi}_1, \boldsymbol{S}_{22})$. Hence we conclude that the joint density for (τ_a, τ_b), $f_{(\tau_a,\tau_b)}(x,y)$, for the case of $x < y$ is

$$f_{(\tau_a,\tau_b)}(x,y) = \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_{11}x}\boldsymbol{S}_{12}e^{\boldsymbol{S}_{22}(y-x)}\boldsymbol{s}_2, \quad x < y.$$

Similarly, for the case of $x > y$ we get that

$$f_{(\tau_a,\tau_b)}(x,y) = \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_{11}y}\boldsymbol{S}_{13}e^{\boldsymbol{S}_{33}(x-y)}\boldsymbol{s}_2, \quad x > y,$$

and since $\boldsymbol{S}_{22} = \boldsymbol{S}_{33}$ and $\boldsymbol{S}_{12} = \boldsymbol{S}_{13}$ we get that the two densities are identical.

We can perform a reduction of the state-space. The exit rates are

$$\boldsymbol{s}_2 = \boldsymbol{s}_3 = \begin{pmatrix}1\\1\end{pmatrix},$$

and therefore the phase-type distributions corresponding to the states $\{4,5\}$ and $\{6,7\}$ are both exponential distributions with rate 1. Thus the direct inter-action between states 4 and 5 (respectively 6 and 7) has no practical effect and we can reduce $\boldsymbol{S}$ to

$$\tilde{\boldsymbol{S}} = \left(\begin{array}{ccc|c|c} -1-\rho & \rho & 0 & 0 & 0 \\ 1 & -3-\rho/2 & \rho/2 & 1 & 1 \\ 0 & 4 & -6 & 1 & 1 \\ \hline 0 & 0 & 0 & -1 & 0 \\ \hline 0 & 0 & 0 & 0 & -1 \end{array}\right), \quad \tilde{\boldsymbol{s}} = \left(\begin{array}{c} 1 \\ 0 \\ 0 \\ \hline 1 \\ \hline 1 \end{array}\right). \tag{23}$$

The corresponding joint densities are then given by

$$f_{(\tau_a,\tau_b)}(x,y) = (1,0,0)\exp\left\{\begin{pmatrix} -1-\rho & \rho & 0 \\ 1 & -3-\rho/2 & \rho/2 \\ 0 & 4 & -6 \end{pmatrix} x\right\}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} e^{-(y-x)} \tag{24}$$

for $x < y$ and vice versa for $x > y$. The density is illustrated in the right plot in Figure 14.

In [26] it is noted that the inter-actions between states $4,5$ and $6,7$ are not needed. This remark results in the reduction

$$\boldsymbol{S}_{22} = \boldsymbol{S}_{33} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

where the two states are preserved instead of collapsing them into a single one as in our case where $\boldsymbol{S}_{22} = \boldsymbol{S}_{33} = \{-1\}$. This representation of course results in the same joint density as above.

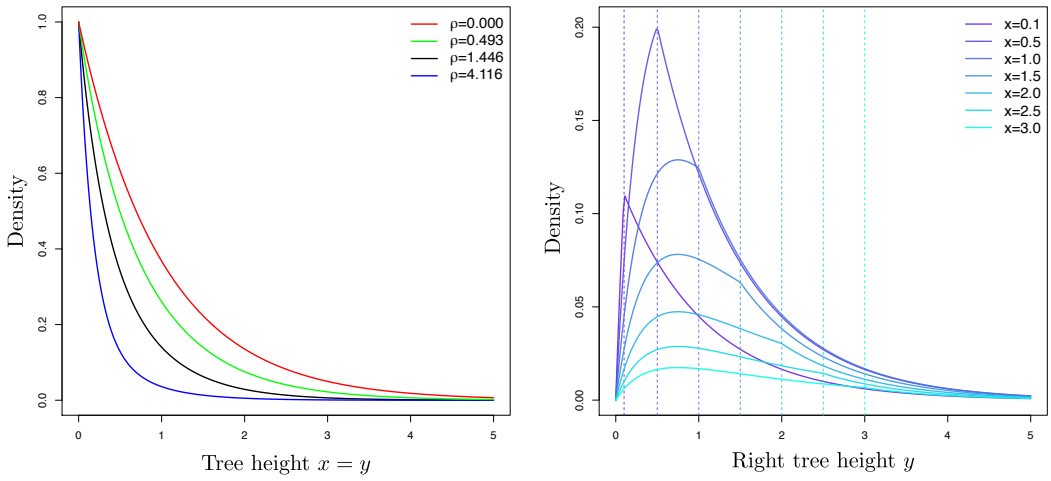

Figure 14: Left: The density for equal tree height (22) for values of $\rho \in \{0, 0.493, 1.446, 4.116\}$. The density integrates to $(1, 0.75, 0.50.25)$ for these values of ρ such that, for e.g. $\rho = 1.446$, the probability for the two tree heights being equal is 0.5. Right: The joint density (24) for a right tree height y for various values of a left tree height x.

5.2 General sample size

In Figure 15 we recapitulate Figure 7.5 page 226 in [26] and introduce the notation. Four linked sequences have evolved back in time according to the ancestral recombination graph. We are interested in the joint distribution of the total branch length $\mathscr{L}_a$ in locus a and the total branch length $\mathscr{L}_b$ in locus b. This process was recently studied using a rather complex hyperbolic system of partial differential equations [17] . We avoid labelling the sequences and consider the number of sequences K_{ab} with ancestral material in both loci, the number of sequences K_a with ancestral material in locus a only, and the number of sequences K_b with ancestral material in locus b only.

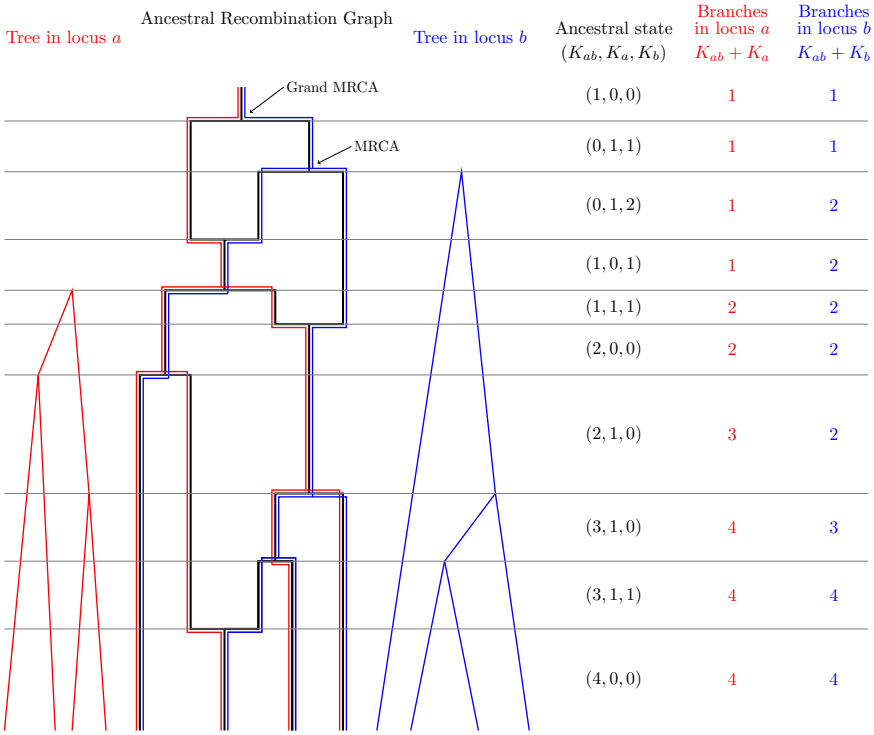

Figure 15: Ancestral recombination graph (in black) for two loci and four sequences and the corresponding trees in the left locus (in red) and right locus (in blue). The figure is adapted from Figure 7.5 in [26].

Define the state of the ancestral recombination graph at time t to be $A(t) = (K_{ab}(t), K_a(t), K_b(t))$. The number of branches in the two loci at time t is then $L_a(t) = K_{ab}(t) + K_a(t)$ and $L_b(t) = K_{ab}(t) + K_b(t)$. The time to the most recent common ancestor (the tree height) in each locus is given by

$$\tau_a = \inf\{t \geq 0 : L_a(t) = 1\} \text{ and } \tau_b = \inf\{t \geq 0 : L_b(t) = 1\}.$$

The total branch length in each locus is

$$\mathscr{L}_a = \int_0^{\tau_a} L_a(t)dt \text{ and } \mathscr{L}_b = \int_0^{\tau_b} L_b(t)dt.$$

Similarly as for two samples we want to study the joint distribution of $(\mathscr{L}_a, \mathscr{L}_b)$ as a function of the recombination rate ρ.

The ancestral process for two loci and a sample of n unlabelled sequences has a state-space given by triplets (k_{ab}, k_a, k_b) where entries are non-negative integers with $k_{ab} + \max\{k_a, k_b\} \leq n$ and with triplets $(0, k_a, 0)$ for $0 \leq k_a \leq n$ and $(0, 0, k_b)$ for $0 \leq k_b \leq n$ removed. The grand MRCA $(1, 0, 0)$ is defined to be the absorbing state because at that time all the ancestral sequences have found common ancestry.

The rates between the states are given by

$$\boldsymbol{Q} = \boldsymbol{Q}^c + \frac{\rho}{2}\boldsymbol{Q}^r, \tag{25}$$

where the transitions that correspond to coalescent events are

$$\begin{aligned} q^c_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a,k_b)} &= \binom{k_{ab}}{2} \\ q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a-1,k_b)} &= \binom{k_a}{2} + k_{ab}k_a \\ q^c_{(k_{ab},k_a,k_b),(k_{ab},k_a,k_b-1)} &= \binom{k_b}{2} + k_{ab}k_b, \end{aligned}$$

and the transitions that correspond to recombination events are

$$q^r_{(k_{ab},k_a,k_b),(k_{ab}-1,k_a+1,k_b+1)} = k_{ab}.$$

Example 5.1. Consider the case $n = 4$. In Figure 16 we illustrate the state space and the rates between states. The intensity matrix is indexed in the order of $(K_{ab} + K_a, K_{ab} + K_b)$ such that we begin with the 9 blocks

$$(4,4), (4,3), (4,2), (3,4), (3,3), (3,2), (2,4), (2,3), (2,2),$$

where both loci have at least two lineages. The next 3 blocks are $(4,1), (3,1), (2,1)$, where the tree in locus b is finished. Then we have the 3 blocks $(1,4), (1,3), (1,2)$ where the tree in locus a is finished. The final block $(1,1)$ is the overall absorbing state. In a block-partitioned form we write the intensity

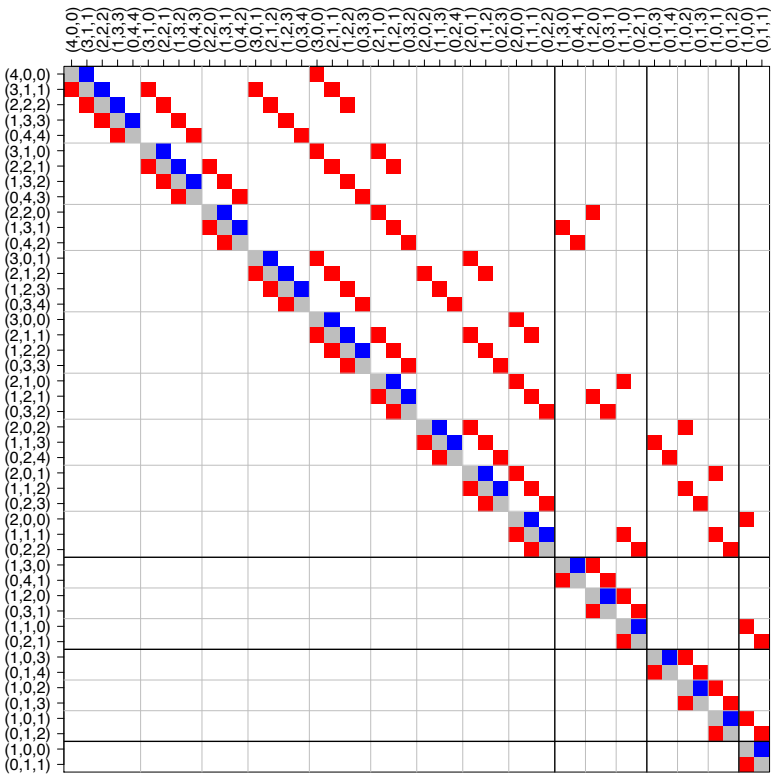

Figure 16: State-space and rates for four sequences. Red entries in the rate matrix correspond to coalescent events and blue entries in the rate matrix correspond to recombination events.

matrix as follows:

$$
\left(\begin{array}{ccccccccc|ccc|ccc|c}
\mathbf{A}_{44}^{44} & \mathbf{A}_{44}^{43} & \mathbf{0} & \mathbf{A}_{44}^{34} & \mathbf{A}_{44}^{33} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{A}_{43}^{43} & \mathbf{A}_{43}^{42} & \mathbf{0} & \mathbf{A}_{43}^{33} & \mathbf{A}_{43}^{32} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{A}_{42}^{42} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{42}^{32} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{42}^{41} & \mathbf{A}_{42}^{31} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{34}^{34} & \mathbf{A}_{34}^{33} & \mathbf{0} & \mathbf{A}_{34}^{24} & \mathbf{A}_{34}^{23} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{33}^{33} & \mathbf{A}_{33}^{32} & \mathbf{0} & \mathbf{A}_{33}^{23} & \mathbf{A}_{33}^{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{32}^{32} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{32}^{22} & \mathbf{0} & \mathbf{A}_{32}^{31} & \mathbf{A}_{32}^{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{24}^{24} & \mathbf{A}_{24}^{23} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{24}^{14} & \mathbf{A}_{24}^{13} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{23}^{23} & \mathbf{A}_{23}^{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{23}^{13} & \mathbf{A}_{23}^{12} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{22}^{22} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{22}^{21} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{22}^{12} & \mathbf{A}_{22}^{11} \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{41}^{41} & \mathbf{A}_{41}^{31} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{31}^{31} & \mathbf{A}_{31}^{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{21}^{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{21}^{11} \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{14}^{14} & \mathbf{A}_{14}^{13} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{13}^{13} & \mathbf{A}_{13}^{12} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{12}^{12} & \mathbf{A}_{12}^{11} \\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{array}\right).
$$

$$
= \begin{pmatrix} \boldsymbol{S}_{ab} & \boldsymbol{S}_{ab}^{a} & \boldsymbol{S}_{ab}^{b} & \boldsymbol{S}_{ab}^{0} \\ \mathbf{0} & \boldsymbol{S}_{a} & \mathbf{0} & \boldsymbol{S}_{a}^{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{S}_{b} & \boldsymbol{S}_{b}^{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},
$$

where the block matrices are defined in the obvious way.

If $\tau_a > \tau_b$ it is because there is a transition from the $\boldsymbol{S}_{ab}$ block to the red square $\boldsymbol{S}_a$, the transition of which is performed by the matrix $\boldsymbol{S}^a_{ab}$ in the red rectangle. From $\boldsymbol{S}_a$ the remaining time is phase-type distributed with exit rate vector $\boldsymbol{S}^0_a$, denoted by the red rectangle at the level of $\boldsymbol{S}_a$. The situation where $\tau_b > \tau_a$ is entirely symmetrical. The density for (τ_a, τ_b) is hence given by

$$f_{(\tau_a,\tau_b)}(x,y) = \begin{cases} \boldsymbol{e}_1' e^{\boldsymbol{S}_{ab} y} \boldsymbol{S}^a_{ab} e^{\boldsymbol{S}_a (x-y)} \boldsymbol{S}^0_a & \text{for } x > y \\ \boldsymbol{e}_1' e^{\boldsymbol{S}_{ab} x} \boldsymbol{S}^0_{ab} & \text{for } x = y \\ \boldsymbol{e}_1' e^{\boldsymbol{S}_{ab} x} \boldsymbol{S}^b_{ab} e^{\boldsymbol{S}_b (y-x)} \boldsymbol{S}^0_b & \text{for } y > x \end{cases} \tag{26}$$

where $\boldsymbol{e}_1' = (1, 0, \ldots, 0)$ because the first state (indexed by (4,0,0)) is the starting state.

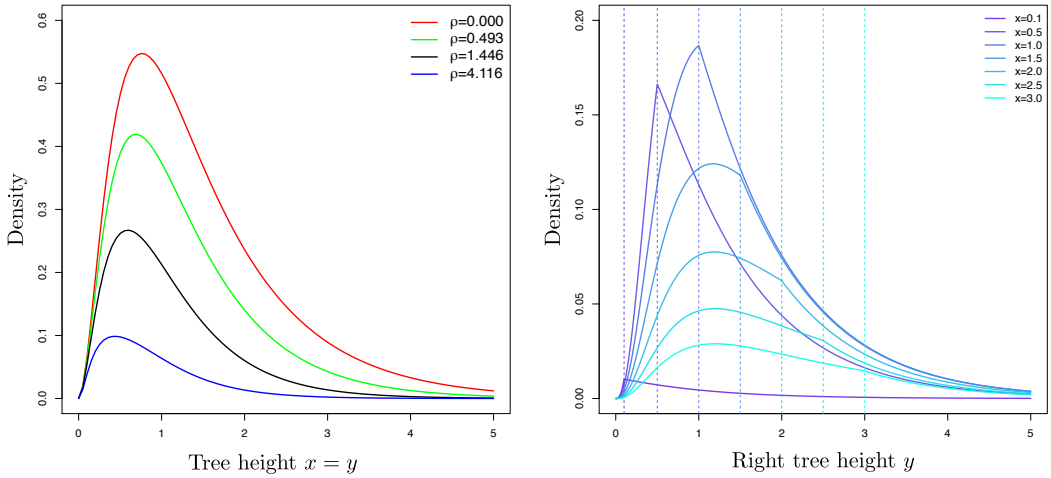

Figure 17: Left: The density for equal tree height (26) for values of $\rho \in \{0, 0.493, 1.446, 4.116\}$. Right: The joint density (26) for a right tree height y for various values of a left tree height x.

Next we consider the total branch lengths. The reward matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \left(\begin{array}{cc} 4\boldsymbol{e} & 4\boldsymbol{e} \\ 4\boldsymbol{e} & 3\boldsymbol{e} \\ 4\boldsymbol{e} & 2\boldsymbol{e} \\ 3\boldsymbol{e} & 4\boldsymbol{e} \\ 3\boldsymbol{e} & 3\boldsymbol{e} \\ 3\boldsymbol{e} & 2\boldsymbol{e} \\ 2\boldsymbol{e} & 4\boldsymbol{e} \\ 2\boldsymbol{e} & 3\boldsymbol{e} \\ 2\boldsymbol{e} & 2\boldsymbol{e} \\ \hline 4\boldsymbol{e} & \boldsymbol{0} \\ 3\boldsymbol{e} & \boldsymbol{0} \\ 2\boldsymbol{e} & \boldsymbol{0} \\ \hline \boldsymbol{0} & 4\boldsymbol{e} \\ \boldsymbol{0} & 3\boldsymbol{e} \\ \boldsymbol{0} & 2\boldsymbol{e} \end{array}\right),$$

where $\boldsymbol{e}$ are column vectors of ones, and $\boldsymbol{0}$ are column vectors of zero, all of appropriate dimensions. Then

$$(\mathscr{L}_a, \mathscr{L}_b) \sim \text{MPH}^*(\boldsymbol{e}_1', \boldsymbol{S}, \boldsymbol{R}),$$

where

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{ab} & \boldsymbol{S}_{ab}^a & \boldsymbol{S}_{ab}^b \\ \boldsymbol{0} & \boldsymbol{S}_a & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{ab}^b \end{pmatrix}.$$

While the joint Laplace transform and (cross) moments have explicit forms, this is in general not the case for the densities and distribution functions in the MPH* class, and the case of $(\mathscr{L}_a, \mathscr{L}_b)$ presents such an example.

5.3 Number of segregating sites

Now let S_a and S_b denote the number of segregating sites in locus a and locus b, and let the mutation rates in the two loci be $\theta_a/2$ and $\theta_b/2$. Recall that $S_a|\mathscr{L}_a \sim \text{Pois}(\mathscr{L}_a\theta_a/2)$ and $S_b|\mathscr{L}_b \sim \text{Pois}(\mathscr{L}_b\theta_b/2)$ and $S_a|(\mathscr{L}_a, \mathscr{L}_b)$ is independent of $S_b|(\mathscr{L}_a, \mathscr{L}_b)$ (i.e. S_a and S_b are conditionally independent given $(\mathscr{L}_a, \mathscr{L}_b)$). We get the means, (co)variances and correlation

$$\begin{aligned}
\mathbb{E}[S_a] &= \mathbb{E}\Big[\mathbb{E}[S_a|\mathscr{L}_a]\Big] = \frac{\theta_a}{2}\mathbb{E}[\mathscr{L}_a] \\
\mathbb{E}[S_aS_b] &= \mathbb{E}\Big[\mathbb{E}[S_aS_b|(\mathscr{L}_a, \mathscr{L}_n)]\Big] = \frac{\theta_a}{2}\frac{\theta_b}{2}\mathbb{E}[\mathscr{L}_a\mathscr{L}_b] \\
\text{Cov}[S_a, S_b] &= \mathbb{E}[S_aS_b] - \text{E}[S_a]\text{E}[S_b] = \frac{\theta_a\theta_b}{4}\text{Cov}\big[\mathscr{L}_a, \mathscr{L}_b\big] \\
\text{Corr}[S_a, S_b] &= \frac{\text{Cov}[S_a, S_b]}{\sqrt{\text{Var}[S_a]\text{Var}[S_b]}} = \frac{\text{Cov}[\mathscr{L}_a, \mathscr{L}_b]}{\sqrt{\text{Var}[\mathscr{L}_a] + \frac{2}{\theta_a}\mathbb{E}[\mathscr{L}_a]}\sqrt{\text{Var}[\mathscr{L}_b] + \frac{2}{\theta_b}\mathbb{E}[\mathscr{L}_b]}}.
\end{aligned}$$

Remark 5.2. We emphasize the following six properties of the correlation structure

(i) The correlation is a separable function of θ_a and θ_b.

(ii) The correlation is increasing as a function of θ_a or θ_b.

(iii) $\mathrm{Corr}[S_a, S_b] < \mathrm{Corr}[\mathscr{L}_a, \mathscr{L}_b]$ for any (θ_a, θ_b).

(iv) $\mathrm{Corr}[S_a, S_b] \to \mathrm{Corr}[\mathscr{L}_a, \mathscr{L}_b]$ for $\theta_a \to \infty$ and $\theta_b \to \infty$.

(v) $\mathrm{Corr}[S_a, S_b] \to 0$ for $\theta_a \to 0$ or $\theta_b \to 0$.

(vi) For $\theta_a = \theta_b = \theta$ we have $\mathrm{Var}[S_a] = \mathrm{Var}[S_b]$ and

$$\mathrm{Corr}[S_a, S_b] = \frac{\mathrm{Cov}[S_a, S_b]}{\mathrm{Var}[S_a]} = \frac{\mathrm{Cov}[\mathscr{L}_a, \mathscr{L}_b]}{\mathrm{Var}\left[\mathscr{L}_a\right] + \frac{2}{\theta}\mathrm{E}\left[\mathscr{L}_a\right]}. \tag{27}$$

Example 5.3. In Figure 18 we show the correlation (27) between the number of segregating sites in two loci for sample sizes $n = (2,4,8)$ and mutation rates $\theta = (0.1,1,10)$, and as a function of the recombination rate ρ. For $n = 2$ we recover the well known result $\mathrm{Corr}(\mathscr{L}_a, \mathscr{L}_b) = (\rho + 18)/(\rho^2 + 13\rho + 18)$ (e.g. [26] equation (7.17) page 231).

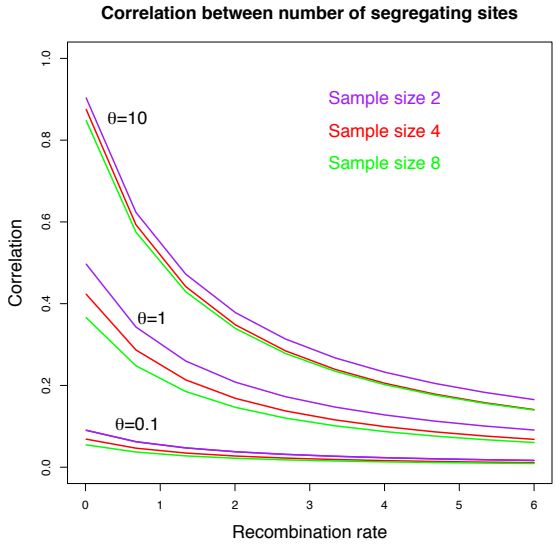

Figure 18: Correlation between the number of segregating sites (27) at two loci for sample sizes $n = (2,4,8)$ and mutation rates $\theta = (0.1,1,10)$.

6 Discussion

In this paper we have provided a unified theory and analysis for a number of coalescent models. We have demonstrated that the class of multivariate phase-type distributions is a useful tool for understanding the behaviour of key population genetics quantities.

In our analysis of the Faroe Island Atlantic Cod data we estimated the α parameter in the Beta-coalescent using a similarity measure between the observed and expected site frequency spectrum. More generally, the focus in this paper has mainly been to quantify the behaviour of summary statistics in various coalescent models. An important research topic could be to develop more principled statistical inference methods for genetic data based on the phase-type framework. Statistical inference in phase-type distributions has traditionally been based on observations of the time of absorption of the stochastic process. The observed genetic data is very different, and thus new methodology is required.

We have demonstrated how phase-type theory is a useful framework for calculating distributions and summary statistics in basic models in population genetics. The coalescent models that we have analysed are time-homogeneous. The structured coalescent analysed in [13] is another example of a time-homogeneous model that can be explored in the phase-type framework. A future research direction could be to extend the analysis to time-inhomogeneous evolutionary models. [17] recently computed the joint distribution of the total branch length in two loci with variable population size. It could be interesting to extend our constant population size analysis in Section 4 and Section 5 to the variable population size model. A first approach could be to consider a piecewise constant population size model, handle each epoch of constant size separately, and finally merge the various epochs. Such an approach requires calculations of moments in end-point conditioned continuous Markov chains, and these can be found using results from [10].

Another important coalescent model is the isolation-with-migration model with multiple populations (e.g. [9]). This model is characterized by times in the past where populations merge, and migration rates between the populations. Statistical inference in this model is very challenging, but [15] and [16] have developed an efficient and general method for likelihood inference using generating functions. [4] recently developed an alternative fast method for fitting a general isolation-with-initial-migration model. The data in [4] is pairs of DNA sequences, and therefore the rate matrix during the migration epoch in the past becomes analytically tractable. Perhaps [4] could constitute a building block for formulating general likelihood-based inference procedures for phase-type distributions based on observed genetic data.

In a recent paper, [7] study the genealogical properties of nested samples in the Beta-coalescent and in a time-changed Kingman coalescent. They study quantities such as (a) the probability that the most recent common ancestor is shared between the complete sample and the subsample within the complete sample, (b) the fraction between the tree height of the subsample and the complete sample, and (c) the fraction between the total internal branches of the subsample and the complete sample. For the latter two fractions, they use a simulation study where they vary the Beta-coalescent parameter from 1 to 2 in steps of 0.1, the complete sample is of size 10000, and the subsample is of size 10, 100 and 1000. We are limited to much smaller sample sizes (at most 25), but for small sample sizes it should be straight forward to determine the joint distribution for the height of the two samples, as well as the joint distribution for the total internal branch length of the two samples. A future research topic could be to investigate extensions to a larger state spaces, or more clever ways of defining the coalescent model.

Acknowledgements

We thank Lars Nørvang Andersen, Johanna Bertl, Svend Nielsen, Paula Tataru and Kai Zeng for discussions and comments, and are grateful to three anonymous reviewers for constructive and helpful suggestions on earlier versions of the manuscript. ASJ is partially supported by CONACyT Grant CB-2014/243068. AH has been funded by the Vienna Science and Technology Fund (WWTF) through project MA16-061.

Supplementary Information

In the Supplementary Information we provide R code for the reproduction of selected figures in the paper: The right plot in Figure 4 (the two first moments of the Beta-coalescent), Figure 11 (the mean and covariance of the SFS for the Beta-coalescent), Figure 14 (the tree height densities for two loci and two samples), Figure 17 (the tree height densities for two loci and four samples), and finally Figure 18 (the correlation between the number of segregating sites in two loci).

References

[1] Matthias Birkner, Jochen Blath and Bjarki Eldon. Statistical Properties of the Site-Frequency Spectrum Associated with Λ-Coalescents. *Genetics*, 195:1037 – 1055, 2013.

[2] Mogens Bladt and Bo Friis Nielsen. *Matrix–exponential distributions in Applied Probabiity.* Springer Verlag, 2017.

[3] Jochen Blath, Adrián González-Casanova, Noemi Kurt and Maite Wilke-Berenguer. A new coalescent for seed-bank models. *Ann. Appl. Probab.*, 26(2):857 – 891, 2016.

[4] Rui J. Costa and Hilde Wilkinson-Herbots. Inference of gene flow in the process of speciation: An efficient maximum-likelihood method for the isolation-with-migration model. *Genetics*, 205:1597–1618, 2017.

[5] Michael Desai, Aleksandra Walczak and Daniel Fisher. Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 193:565 – 585, 2013.

[6] Rick Durrett. *Probability Models for DNA Sequence Evolution.* Second Edition. Springer Verlag, 2008.

[7] Bjarki Eldon and Fabian Freund. Genealogical properties of subsamples in highly fecund populations. *J Stat Phys*, 172: 175–207, 2018.

[8] Bjarki Eldon, John Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172:2621–2633, 2006.

[9] Jody Hey. Isolation with migration models for more than two populations. *Mol. Biol. Evol.*, 27(4): 905-920, 2010.

[10] Asger Hobolth and Jens Ledet Jensen. Summary statistics for endpoint-conditioned continuous-time Markov chains *J. Appl. Probab.*, 48(4):911–924, 2011.

[11] Asger Hobolth and Jens Ledet Jensen. Markovian approximation to the finite loci coalescent with recombination along multiple sequences. *Theor. Pop. Biol.*, 98:48 – 58, 2014.

[12] Alexander Klassmann and Luca Ferretti. The third moments of the site frequency spectrum. *Theor. Pop. Biol.*, 120: 16–28, 2018.

[13] Seiji Kumagai and Marcy K. Uyenoyama. Genealogical histories in structured populations. *Theor. Pop. Biol.*, 102: 3–15, 2015.

[14] Amaury Lambert and Chunhua Ma. The coalescent in peripatric metapopulations. *J. Appl. Probab.*, 52(2):538 – 557, 2015.

[15] Konrad Lohse, Richard J. Harrison and Nicholas H. Barton. A general method for calculating likelihoods under the coalescent process. *Genetics*, 189: 977 – 987, 2011.

[16] Konrad Lohse, Martin Chmelik, Simon H. Martin and Nicholas H. Barton. Efficient strategies for calculating blockwise likelihoods under the coalescent. *Genetics*, 202: 775 – 786, 2016.

[17] Alexey Miroshnikov and Matthias Steinrücken. Computing the joint distribution of the total tree length across loci in populations with variable population size. *Theor. Pop. Biol.*, 118: 1–19, 2017.

[18] Richard Neher and Oskar Hallatschek. Genealogies of rapidly adapting populations. *Proc. Natl. Acad. Sci.*, 110:437–442, 2013.

[19] Jim Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870 – 1902, 1999.

[20] Ori Sargsyan and John Wakeley. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms *Theor. Pop. Biol.*, 74: 104-114, 2008.

[21] Serik Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):116–1125, 1999.

[22] Jason Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, 2003.

[23] Jason Schweinsberg. Rigorous results for a population model with selection II: genealogy of the population. *Electron. J. Probab.*, 22(38):1–54, 2017.

[24] Hlynur Sigurgíslason and Einar Árnason. Extent of mitochondrial DNA sequence variation in Atlantic cod from the Faroe Islands: a resolution of gene genealogy. *Heredity*, **91**, 557-564, 2003.

[25] Katy L. Simonsen and Gary A. Churchill. A Markov Chain Model of Coalescence with Recombination. *Theor. Pop. Biol.*, 52: 43–59, 1997.

[26] John Wakeley. *Coalescent Theory: An Introduction*. W. H. Freeman, 2008.