Genealogies of Samples from Stochastic Populations and Biodiversity Models

Michael Jones
Department of Statistics
The University of Auckland

Supervisor: Simon Harris
2023

Abstract

This project in probability theory will investigate stochastic models used for dating phylogenetic trees. Markov Birth Death processes will be used to derive models for determining past speciation times of extant species. The projects aim to explain and rigorously prove results from Tanja Gernhard's *The Conditioned Reconstructed Process* [5]. We considered many different results. Such as the expectation of the kth speciation time, probability density function for speciation times and their corresponding cumulative distribution functions. We considered density functions for time between events and many others. We also considered some of these results conditional on the time of origin, or unconditional on time of origin, instead assuming a uniform prior.

Contents

Chapter 1

Background Information - Stochastic Processes

This report derives analytical results on phylogenetics, using a stochastic process as a basis. Prior to our analysis, it is important to understand several forms of stochastic processes. These include, Markov Chains, Branching Processes, Birth Death Processes and Poisson Point Processes.

1.1 Markov Chains

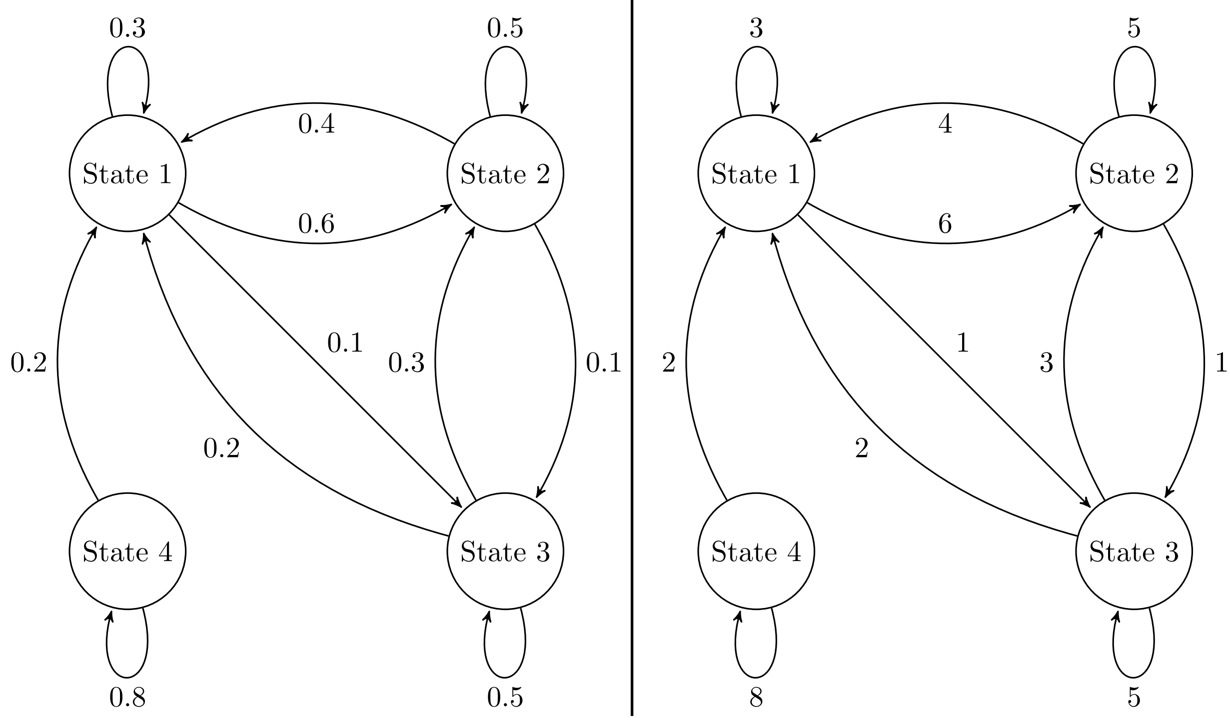
A Markov Chain is a statistical process where a 'particle' moves from one state to another independently of any past moves and with a defined probability. We use a 'particle' as a generic term to describe any object of the Markov Chain. It is best understood with a diagram, see diagram 1.1:

In the Figure 1.1a, a 'particle' would move from state-state by the defined probability. In the discrete case a move would occur at time values that are integers, i.e $1, 2, 3, \ldots$. Where time is continuous, a 'particle' would move at a rate, rather than a probability. This rate is Poisson distributed and the time-between movements is exponentially distributed. See figure 1.1b. In our derivation we will use the continuous markov chain as a basis, this is a more realistic model.

From Markov Chains we can derive the Q-Matrix, which is a matrix that examines the transitions between states. Often we derive the equilibrium distribution of a Markov Chain, however this is not done in this report. Instead of we derive the probability of being in some state after a specified time t. This is done in Chapter 3.

Figure 1.1: Markov Chains

(a) Discrete Chain (b) Continuous Chain



1.2 Continuous Branching Processes

The Branching Process is a form of stochastic process that is often used when examining a species population. We use them in this report to count the number of unique species at a given time.

As a basic model, we can assume we have a species that lives for a random amount of time with an exponential distribution then goes extinct. We also assume that a species lives for an exponential amount of time then speciates, becoming two species.

In general, a Branching Process will have an offspring probability distribution, dictating how many individuals will result from one species. This could take any integer values, $0, 1, 2, \ldots$. In our case the offspring distribution will take values 0 or 2, with some probability. I.e. after some time a species will either turn into zero species (extinction) or into two species (speciation).

A basic model is in figure 1.2, where we have denoted extinctions in red.

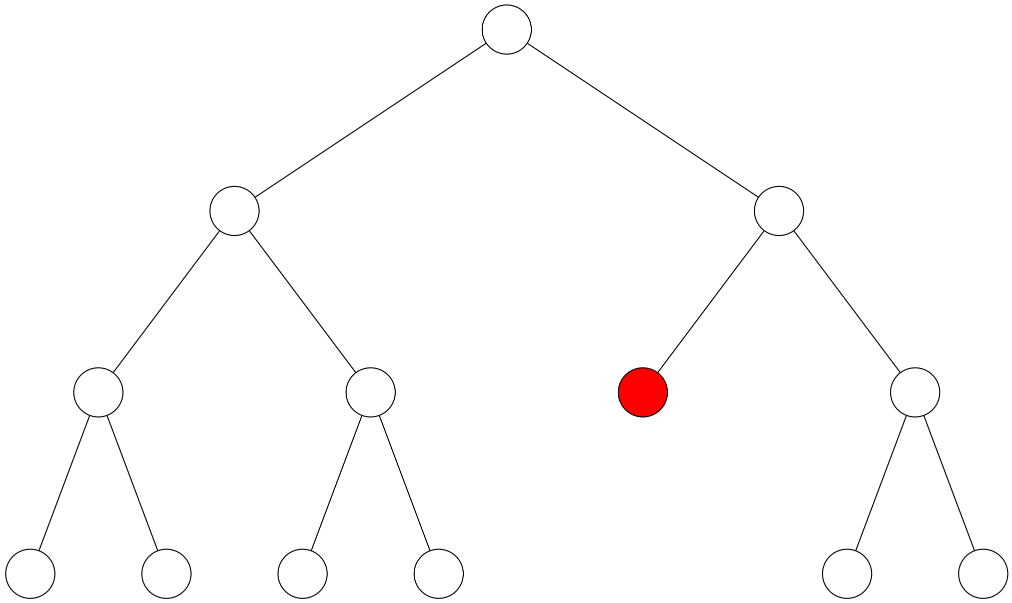
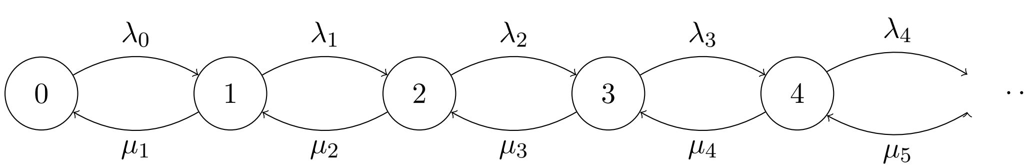


Figure 1.2: A Branching Process, where an extinction is denoted in red [11].

1.3 Birth and Death Processes

The Birth and Death process is a unique continuous time Markov process, whereby the state can only move up or down by one step. This model is similar to the Branching Process, population can only increase or decrease by one at each unique event. This model is useful for our research in unique species because in reality the number of unique species will only increase or decrease by one at a specific point in time. For example, it is rare that a species will evolve into three unique species (increasing the number of species by two). It is even rarer that two different species will have either a speciation or extinction event at the exact same time.

Birth and death processes typically have the form below:



Whereby λ_i for $i = 1, 2, 3, \dots$ is the rate a species will evolve, increasing the number of unique species by one unit. μ_i for $i = 1, 2, 3, \dots$ is the rate a species will go extinct, decreasing the number of unique species by one unit.

We assume if a species is in state i, the waiting time until the population increases is exponentially distributed with rate $\frac{1}{\lambda_i}$. The waiting time until the population decreases is exponentially distributed with rate $\frac{1}{\mu_i}$. The time between birth and deaths (inter-arrival times) are independent [5]. These assumptions are common with all Markov Chains.

Typically when examining a species population, we might make the values $\lambda_0 = 0$ or $\lambda_1 = 0$, as to reflect the population can not keep growing once there is less than two individuals of the species that exist, i.e. they become extinct.

In this report we will set $\lambda_i = i\lambda$ and $\mu_i = i\mu$. This is to reflect a constant rate of extinction and constant rate of a speciation per individual species. Note that this implies $\lambda_0 = 0$.

1.4 Poisson Point Processes

A counting process of $N(t)$ where $t = 0, 1, 2, 3, ..$ is a Poisson process of rate λ if its initial state is 0 ($N(0) = 0$). It's increments of the process are independent, for example N(t) and N(t-1) are independent. Finally, for $s \geq 0$ and $t > 0$, $N(s + t) - N(s) \sim Po(\lambda t)$ (Poisson distribution). [14]. A subset of the Poisson process follows a Poisson distribution of rate λt, where t is the length of time of the subset.

Poisson point processes play a vital role in our genealogical studies, as they allow us to model reconstructed evolutionary trees and estimate the timing of speciation events - that is, when a single species evolved and diverged into two distinct species. By using point processes to represent the tree structure, we gain valuable insights into the patterns of diversification and evolution within a given group of species.

Graphically we can see how a Poisson process can reflect an evolutionary tree in Figure 2.1. This is explained in more detail later, but essentially, each tree splitting is a point in the Poisson point process. Each split happens at a different time and the waiting time until the next split is exponentially distributed with rate $\lambda > 0$.

Representing evolutionary trees as Poisson point processes are important because finding the distribution of the point process gives us the distribution of the evolutionary tree [2].

Chapter 2

Phylogenetics - Basics

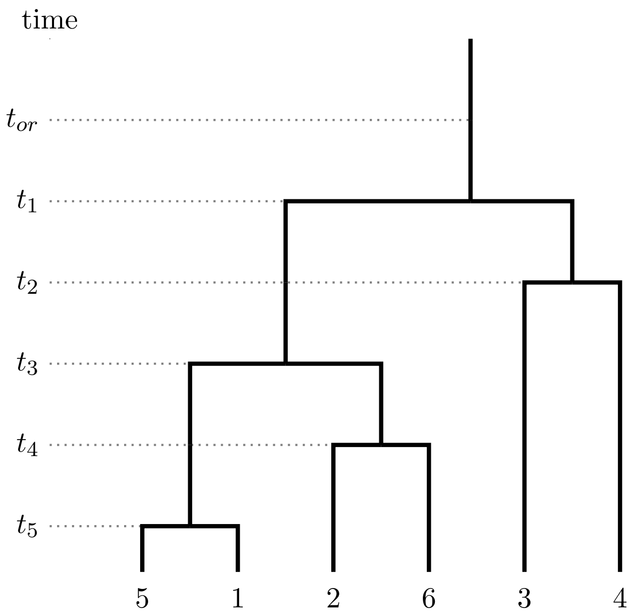
Phylogenetics is the field of studying the evolutionary history and relationships among current living species. In this paper, we employ birth and death processes, as well as Poisson point processes, to advance our understanding of phylogenetics [9]. Specifically, we explore a range of topics, including the time of speciation events, the time of origin for a group of species, the expected time of the k-th speciation events, various properties of speciation times. [5].The results given are all present in Gernhard's *Conditioned Reconstructed Process* [5], our aim is to rigorously prove these results. adding in extra details as necessary which are omitted in the original paper.

The basis for this research is using a conditioned birth death process. We condition on n extant species. Essentially, at the beginning of the birth-death process we know the number of individuals at the end of the process, the present [5]. Occasionally we also condition on the time of origin t_{or}. The time at which the tree starts, assuming the tree ends at time 0, the present. In our research we explore results conditioning on t_{or}, that is t_{or} is some known value. Alternatively, we explore results where t_{or} is unknown. When t_{or} is unknown, we assign it a uniform prior between 0 and ∞.

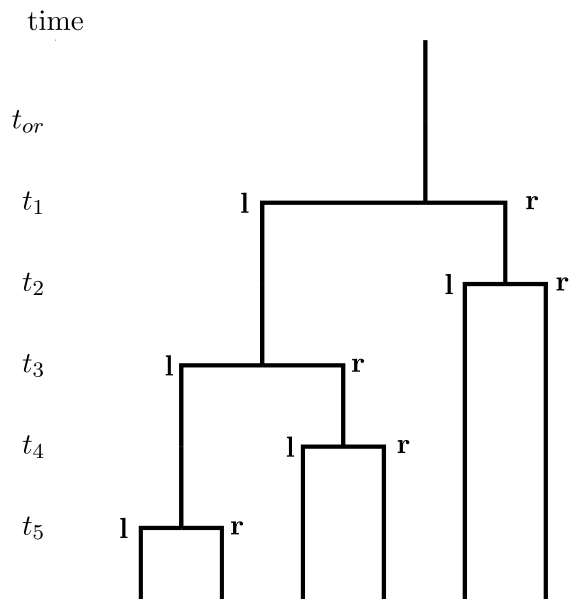
We make use of the result that under a Birth Death process, the probability of each arrangement of speciation events in a tree, has an equal probability. See Corollary 2.2.1 below. Therefore we can imagine our results as if we simulated many different trees under a birth death process, and extracted the distribution of our topics of interest. In this report, we find these results analytically, instead of using simulation.

2.1 Reconstructed Trees - Orientated Trees - Poisson Process

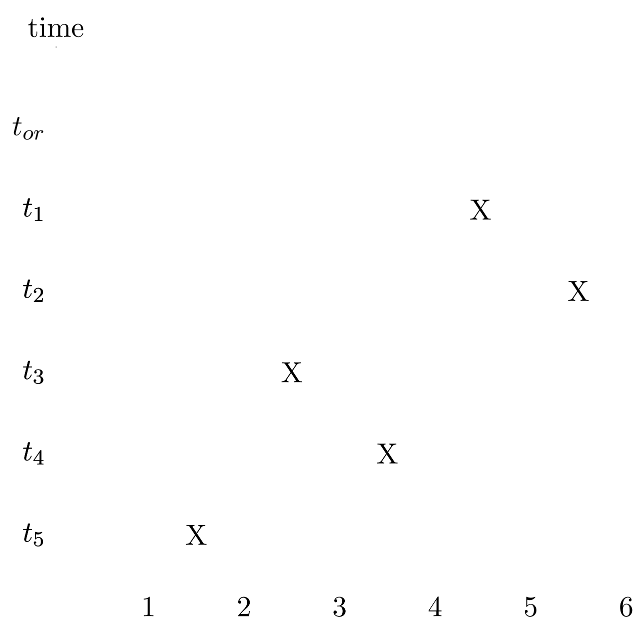
In this section we will describe how we can transition from a Reconstructed Tree to an Orientated Tree to a Poisson Process. This is illustrated in Figure 2.1, this figure is often referenced as a visual aid. Demonstrating the relationship between these figures is significant because representing our tree as a Poisson Process, for example, helps us generate analytical results. To begin our analysis,



(a) Reconstructed Tree



(b) Rank Orientated Tree



(c) Poisson Point Process Representation

Figure 2.1: Illustration of how we go from a Reconstructed Tree to a Rank Orientated Tree, then to a Poisson Point Process

let's take a closer look at a reconstructed tree shape as shown in Figure 2.1a, from which we've removed extinct lineages.

To get from a reconstructed tree, we need to follow a process as dictated by Tanja Gernhard [5]. Initially we start with an evolutionary tree, we then remove extinct lineages and randomly assign the node leaves, see Figure 2.1a. In this figure the node leaves 1 to 6 are randomly assigned. We are then left with a reconstructed tree.

To get to a rank reconstructed tree, order the interior vertices based on time of speciation. This gives us a rank function for the reconstructed tree, in Figure 2.1b, the ranks are (1,2,3,4,5). To make the reconstructed tree a rank oriented tree, randomly label the two daughter lineages for each interior vertex as "l" and "r" as seen in Figure 2.1b.

We can define a Poisson Point Process to reflect our rank orientated tree. First draw number the amount of extant species on the horizontal axis, in our example we have 6 extant species, so add 1,2,3,4,5,6 on the horizontal axis (see Figure 2.1c. Next, pick the number of extant species minus one (i.e. 6 - 1 = 5), to be at the location $\left(i + \frac{1}{2}, s_i\right)$, where s_i is the time of speciation and $\left(i + \frac{1}{2}\right)$ is the point on the horizontal axis. This gives us a Poisson point process.

2.2 Bijection Between Rank Orientated Trees and the Corresponding Poisson Process

In this section we aim to prove that there is a bijection between rank orientated trees, as seen in Figure 2.1b, and the matching Poisson Process, see Figure 2.1c. We prove this result because it indicates that we can represent our lineage tree as a poisson process and the results derived from the poisson process connect to lineages on the tree.

Lemma 2.1 (+)**.** *There is a Bijection Between Rank Orientated Trees (2.1b) and the Corresponding Poisson Process (see Figure 2.1c)*

(See Gernhard [5] Lemma 2.1)

Proof. As expressed by Tanja Gernhard [5], there is a bijection between the rank orientated trees of age t_{or} and the Poisson Point Process of age t_{or}. In this context, "injective" means that each oriented tree of age t_{or} corresponds to a unique point process of age t_{or}. "Surjective" means that for each point process of age t_{or}, there exists at least one oriented tree of age t_{or} that corresponds to it.

To see why this is true, consider this process. Draw the orientated tree we have in Figure 2.1b, adding your labels l for left and r for right at each speciation event. These points of speciation are also the points of the Poisson Process, as seen in Figure 2.1c.

Starting from the oriented tree, we can uniquely identify the location of each speciation event, which corresponds to a point in the point process. Conversely, given a point process, we can reconstruct the corresponding oriented tree by locating the speciation events and determining the orientation of the branches at each interior vertex. Therefore, the mapping between oriented trees and point processes is both injective and surjective, which means it is a bijection. □

2.3 Regardless of conditioning on t_{or}, the probability of producing a rank orientated tree with n leaves - under a constant birth death process - is equal

This process is telling us a very significant result: if we observe a sample of extant species, and we assume that they have evolved according to a constant rate birth and death process, then any of the possible ranked oriented trees on n leaves could have given rise to the observed sample with equal probability.

Theorem 2.2 (+)**.** *Regardless of conditioning on t_{or}, the probability of producing a rank orientated tree with n leaves - under a constant birth death process - is equal eg.*

(See Gernhard [5], Theorem 2.3)

Proof. We can prove this statement by examining the rank orientated tree "upside down", using Figure 2.2 as an aid.

Suppose we have n extant species, in the Figure 2.2 we have 6. Randomly, one is selected and it gets labelled l, then another species is selected and it gets labelled r. We have $n(n-1)$ possible combinations for the first coalescent event. Note the total number of combinations is not n choose 2, $\binom{n}{2} = \frac{n(n-1)}{2}$. Because we are not selecting two species from a list of n and then labelling them. We are first selecting one species from n, this has n possible permutations, then we label it l. Next we select anther species from what is now $n-1$ possible options, then we label to r. Mathematically we are computing two permutations of one object as opposed to one combination of two objects. For the total number of species groupings we multiple the two together, which gives $n(n-1)$, see the maths below:

Let ${}^{n}P_c$ be the number of permutations of c objects from n total objects

$$\text{Total Combinations } = ({}^{n}P_1)({}^{n-1}P_1) = \frac{n!}{(n-1)!}\frac{(n-1)!}{(n-2)!} = n(n-1)$$

The two leaves we select to first coalesce are then replaced by their most common ancestor, for example this would be at time point t_5 in figure 2.2. Now we are left with $n-1$ species. Therefore for the next coalescing species, repeat the process. Select one species from $n-1$, label it l. Select another species from $n-2$, label it r. The total number of combinations for this event is:

$$\text{Total Combinations } = ({}^{n-1}P_1)({}^{n-2}P_1) = \frac{(n-1)!}{(n-2)!}\frac{(n-2)!}{(n-3)!} = (n-1)(n-2)$$

At the next step we have $n-2$ species that can coalesce, by induction, this will give us $(n-2)(n-3)$ combinations. We keep repeating this process until all species have coalesced, leaving one common ancestor, as seen in Figure 2.2 at t_{or}. We can get the total number of different way species can coalesce by multiplying the number combinations together for each species coalsecing event together, which gives us the following result:

$$\begin{aligned}
& n(n-1) \times (n-1)(n-2) \times (n-2)(n-3) \times \ldots \times (3)(2) \times (2)(1) \\
\iff & n(n-1)(n-2)\ldots(3)(2) \times (n-1)(n-2)(n-3)\ldots(2)(1) \\
\iff & n!(n-1)!
\end{aligned}$$

This implies we have $n!(n-1)!$ possible rank orientated trees with n leaves specified and each of these trees are equally likely. Therefore the probability of having a specific rank orientated tree with n leaf labels is:

$$\frac{1}{n!(n-1)!}$$

But, this is not the probability of observing a ranked tree itself, we could have identical trees for example just with different leaf labelling. So, conditioning on the fact there are $n!$ different possibility of rearranging visually the same tree with n leaves. The probability of observing a specific rank orientated tree is:

$$\frac{1}{(n-1)!}$$

Clearly this indicates every possible rank orientated tree has the same probability when generated under a constant rate birth death process and of size n. Further implying the distribution of rank orientated trees is uniform.

□

This following result adds to the theorem.

Corollary 2.2.1 (+). *Permutations of the speciation events in the Poisson Process (see Figure 2.1c) of the Birth-Death Process have equal probability*

(See Gernhard [5] Theorem 2.3)

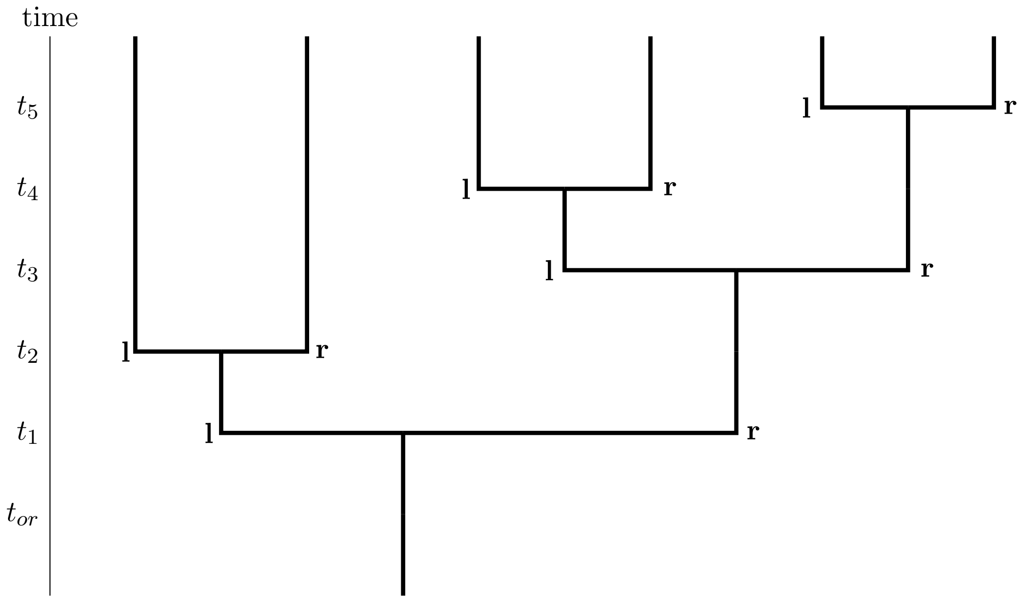


Figure 2.2: Upside down rank orientated tree

Proof. This is theorem is telling us that under a Poisson process of the Birth and Death process each permutation, that is each arrangement of speciation events in a particular order, has an equal probability of occurring. We can prove this using previous theorems. Firstly, we have proven that we have a bijection between the rank orientated trees and the corresponding Poisson process with $n-1$ speciation points, see Lemma 2.1. All Poisson processes will induce a rank Orientated tree. Therefore if we permute each Poisson process, each permutation will produce a unique rank orientated tree. These rank orientated trees follow a uniform distribution (see Theorem 2.2, therefore the probability of each rank orientated tree is equal and the probability of each corresponding Poisson process is equal. □

Chapter 3

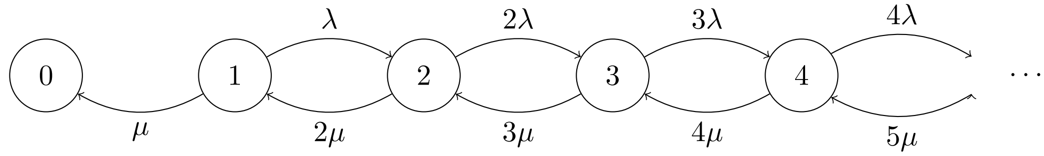
Obtaining the probability of having some n species after some time t has elapsed, $p_n(t)$

The next step is to obtain the probability of having some n species after some time t has elapsed, $p_n(t)$. We find this value assuming a birth-death process model where the birth rate is λ, the death rate is μ and we start in state 1. Obtaining $p_n(t)$ is important because it helps us derive both the speciation times for speciation events and the probability density for the origin time. We derive these in following chapters.

We follow from Equation (1) in Gernhard [5], but she gives only the solution to $p_n(t)$, providing no proof. We provide detailed calculations and clear steps in how to fully derive $p_n(t)$, evident in Lemmas and a final theorem below.

Please note that in the following section we set time t, as to be the time elapsed and we assume the model starts in time 0. This conflicts with our definition in previous sections, which define the origin at t, and the present (end) as 0.

The basic model is below:



Before finding $p_n(t)$ we obtain the following Lemmas which will be used in our main proof.

Lemma 3.1. *Using the kolmogorov equations we obtain the derivatives of* $p_n(t)$ *as:*

$$\begin{aligned}
p_0'(t) &= \mu p_1(t)\\
p_1'(t) &= 2\mu p_2(t) - (\lambda+\mu)p_1(t)\\
p_n'(t) &= (n-1)\lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t) - n(\lambda+\mu)p_n(t), \quad \textit{for } (n>0)
\end{aligned}$$

With the initial condition that we start in state 1:

$$p_n(0) = \begin{cases} 1 & (n=1) \\ 0 & \textit{otherwise} \end{cases} \tag{3.1}$$

Proof. Firstly let our Markov chain be defined as X_t and let the values $t, h, n > 0$, we will first examine $p_n(t+h)$, that is our Markov process after some time $t+h$. Furthermore assume we initially start off in State 1, that is with one lineage (as seen visually in previous Figures 2.1).

$$\begin{aligned}
p_n(t+h) &:= \mathbb{P}(X_{t+h} = n | X_0 = 1)\\
&= \mathbb{P}(X_{t+h} = n | X_t = n)\mathbb{P}(X_t = n | N_0 = 1)\\
&\quad + \mathbb{P}(X_{t+h} = n | X_t = n-1)\mathbb{P}(X_t = n-1 | N_0 = 1)\\
&\quad + \mathbb{P}(X_{t+h} = n | X_t = n+1)\mathbb{P}(X_t = n+1 | N_0 = 1)\\
&\quad + o(h)\\
&= \{1 - (\lambda_n + \mu_n)h + o(h)\} \times p_n(t)\\
&\quad + \{\lambda_{n-1}h + o(h)\} \times p_{n-1}(t)\\
&\quad + \{\mu_{n+1}h + o(h)\} \times p_{n+1}(t)\\
&\quad + o(h)\\
&= p_n(t) + \{-(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t)\}h + o(h)
\end{aligned}$$

Remembering that the definition of a limit is:

$$\begin{aligned}
p'_n(t) &= \lim_{h\to 0}\frac{p_n(t+h)-p_n(t)}{h}\\
&= \lim_{h\to 0}\frac{p_n(t)+\{-(\lambda_n+\mu_n)p_n(t)+\lambda_{n-1}p_{n-1}(t)+\mu_{n+1}p_{n+1}(t)\}h-p_n(t)}{h}\\
&= \lim_{h\to 0} -(\lambda_n+\mu_n)p_n(t)+\lambda_{n-1}p_{n-1}(t)+\mu_{n+1}p_{n+1}(t)\\
&= -(\lambda_n+\mu_n)p_n(t)+\lambda_{n-1}p_{n-1}(t)+\mu_{n+1}p_{n+1}(t)\\
&= -(n\lambda+n\mu)p_n(t)+(n-1)\lambda p_{n-1}(t)+(n+1)\mu p_{n+1}(t)
\end{aligned}$$

$$\begin{aligned}
\implies\ p'_0(t) &= \mu p_1(t)\\
p'_1(t) &= 2\mu p_2(t)-(\lambda+\mu)p_1(t)\\
p'_n(t) &= (n-1)\lambda p_{n-1}(t)+(n+1)\mu p_{n+1}(t)-n(\lambda+\mu)p_n(t), \quad \text{for } (n>0)
\end{aligned}$$

□

We also need to think about using a probability generating function to find $p_n(t)$. For example see Bailey [3] and also Grimmett and Stirzaker [8].

Lemma 3.2. *By defining $\phi(z,t)=\sum_{n=0}^{\infty} z^n p_n(t)$, where $\phi(z,t)$ is the probability generating function of $p_n(t)$, we get the following result:*

$$\frac{\partial\phi(z,t)}{\partial t}=(\lambda z-\mu)(z-1)\frac{\partial\phi(z,t)}{\partial z} \tag{3.2}$$

Proof. Note that, since we differentiate power series term by term within radius of convergence, $|z|<1$.

$$\begin{aligned}
\phi(z,t) &= \sum_{n=0}^{\infty} z^n p_n(t)\\
\frac{\partial}{\partial t}\phi(z,t) &= \sum_{n=0}^{\infty} z^n p'_n(t)
\end{aligned}$$

Then using Lemma 3.1 we obtain:

$$\frac{\partial}{\partial t}\phi(z,t)=\sum_{n=0}^{\infty} z^n((n-1)\lambda p_{n-1}(t)+(n+1)\mu p_{n+1}(t)-n(\lambda+\mu)p_n(t))$$

$$= \lambda \sum_{n=0}^{\infty} z^n (n-1) p_{n-1}(t) + \mu \sum_{n=0}^{\infty} z^n (n+1) p_{n+1}(t) - (\lambda + \mu) \sum_{n=0}^{\infty} z^n n p_n(t)$$

$$= \lambda \sum_{n=0}^{\infty} z^n (n-1) p_{n-1}(t) + \mu \sum_{n=0}^{\infty} z^n (n+1) p_{n+1}(t) - (\lambda + \mu) \sum_{n=0}^{\infty} z^n n p_n(t)$$

$$= ① + ② + ③$$

Now considering the first term

$$
\begin{aligned}
① \quad \lambda \sum_{n=0}^{\infty} z^n (n-1) p_{n-1}(t) &= \lambda z^2 \sum_{n=0}^{\infty} z^{n-2} (n-1) p_{n-1}(t) \\
&= \lambda z^2 \sum_{n=0}^{\infty} \frac{\partial}{\partial z} z^{n-1} p_{n-1}(t) \\
&= \lambda z^2 \frac{\partial}{\partial z} \sum_{n=1}^{\infty} z^{n-1} p_{n-1}(t) \\
&= \lambda z^2 \frac{\partial}{\partial z} \sum_{n=0}^{\infty} z^n p_n(t) \\
&= \lambda z^2 \frac{\partial \phi(z,t)}{\partial z}
\end{aligned}
$$

Now considering the second term

$$
\begin{aligned}
② \quad \mu \sum_{n=0}^{\infty} z^n (n+1) p_{n+1}(t) &= \mu \sum_{n=0}^{\infty} \frac{\partial}{\partial z} z^{n+1} p_{n+1}(t) \\
&= \mu \frac{\partial}{\partial z} \sum_{n=1}^{\infty} z^n p_n(t) \\
&= \mu \frac{\partial}{\partial z} (\sum_{n=0}^{\infty} z^n p_n(t) - \mu z^0 p_0(t)) \\
&= \mu \frac{\partial}{\partial z} \sum_{n=0}^{\infty} z^n p_n(t) - \frac{\partial}{\partial z} \mu p_0(t) \\
&= \mu \frac{\partial \phi(z,t)}{\partial z}
\end{aligned}
$$

Finally, considering the third term we have

$$
\begin{aligned}
③ \quad (\lambda + \mu) \sum_{n=0}^{\infty} z^n n p_n(t) &= (\lambda + \mu) z \sum_{n=0}^{\infty} z^{n-1} n p_n(t) \\
&= (\lambda + \mu) z \sum_{n=0}^{\infty} \frac{\partial}{\partial z} z^n p_n(t)
\end{aligned}
$$

$$= (\lambda + \mu) z \frac{\partial \phi(z,t)}{\partial z}$$

Hence combining these results together we have found

$$\frac{\partial \phi(z,t)}{\partial t} = \lambda z^2 \frac{\partial \phi(z,t)}{\partial z} + \mu \frac{\partial \phi(z,t)}{\partial z} - (\lambda + \mu) z \frac{\partial \phi(z,t)}{\partial z}$$
$$= (\lambda z^2 - (\lambda + \mu) z + \mu) \frac{\partial \phi(z,t)}{\partial z}$$
$$= (\lambda z - \mu)(z - 1) \frac{\partial \phi(z,t)}{\partial z}$$

□

Using this previous Lemma 3.2 we can derive the probability generating function representation of $p_n(t)$.

Lemma 3.3. *The probability generating representation for $p_n(t)$ is as follows:*

$$\phi(z,t) = \sum_{n=0}^{\infty} z^n p_n(t) = \frac{\mu e^{(\lambda-\mu)t}(\frac{z-1}{\lambda z - \mu}) - 1}{\lambda e^{(\lambda-\mu)t}(\frac{z-1}{\lambda z - \mu}) - 1}$$

Proof. Based on a technique demonstrated by Bailey [3] we can solve the partial differential equations obtained in Lemma 3.2. This helps us obtain $\phi(z,t)$ with the following process. We follow several keys steps:

Step 1: Simplify our partial differential equations and identify subsidiary equations:

$$\frac{\partial}{\partial t}\phi(z,t) = (\lambda z - \mu)(z-1)\frac{\partial}{\partial z}\phi(z,t)$$
$$\iff \frac{\partial}{\partial t}\phi(z,t) - (\lambda z - \mu)(z-1)\frac{\partial}{\partial z}\phi(z,t) = 0$$

Our subsidiary equations are as follows

$$\frac{dt}{1} = \frac{dz}{-(\lambda z - \mu)(z-1)} = \frac{d\phi}{0}$$

Step 2: Simplify Subsidiary equations

Firstly we have (by combining 1 and 3): $dt = \frac{d\phi}{0}$

$\implies \phi$ is a constant

We also have (by combining 1 and 2): $dt = \frac{dz}{-(\lambda z - \mu)(z-1)}$

Step 3: Complete the integral to solve for t

$$dt = \frac{dz}{-(\lambda z - \mu)(z-1)}$$
$$\implies t = \int \frac{-1}{(\lambda z - \mu)(z-1)} dz$$

We can simplify the integral by completing partial fractions

Let $\frac{1}{(\lambda z - \mu)(z-1)} = \frac{A}{\lambda z - \mu} + \frac{B}{z-1} \implies 1 = A(z-1) + B(\lambda z - \mu)$

Letting z = 1 $\implies 1 = B(\lambda - \mu) \implies B = \frac{1}{\lambda - \mu}$

Letting z = $\frac{\mu}{\lambda} \implies 1 = A(\frac{\mu}{\lambda} - 1) \iff 1 = A(\frac{\mu - \lambda}{\lambda}) \implies A = \frac{\lambda}{\mu - \lambda} = \frac{-\lambda}{\lambda - \mu}$

Thus, we have: $A = \frac{-\lambda}{\lambda - \mu}, \quad B = \frac{1}{\lambda - \mu}$

Therefore our integral to solve for t becomes

$$t = \int \left(\frac{\lambda}{\lambda - \mu} \frac{1}{\lambda z - \mu} + \frac{1}{\lambda - \mu} \frac{1}{z-1} \right) dz \iff t = \frac{1}{\lambda - \mu} \left(\int \frac{\lambda}{\lambda z - \mu} dz \right) + \int \left(\frac{1}{z-1} \right) dz$$

$$\iff t(\lambda - \mu) = \frac{\lambda}{\lambda} log|\lambda z - \mu| log|z-1| + C \iff t(\lambda - \mu) = log \left| \frac{\lambda z - \mu}{z - 1} \right| + C$$

$$\iff \frac{t(\lambda - \mu)}{log|\frac{\lambda z - \mu}{z-1}|} = C \iff exp \left\{ \frac{t(\lambda - \mu)}{log|\frac{\lambda z - \mu}{z-1}|} \right\} = C$$

$$\iff e^{(\lambda - \mu)t} \frac{z-1}{\lambda z - \mu} = C$$

Step 4: Identify an arbitrary function to give the most general solution

We can let $e^{(\lambda - \mu)t} \dfrac{z-1}{\lambda z - \mu} = C$ be our arbitrary function, therefore we have:
Our arbitrary function: $\phi(z,t) = \Phi \left\{ e^{(-\lambda - \mu)t} \dfrac{z-1}{\lambda z - \mu} \right\}$

Step 5: Identify the initial condition and complete the partial differential equations to ultimately identify $\phi(z,t)$

First identify the initial condition for: $\phi(z,t) = \displaystyle\sum_{n=0}^{\infty} z^n p_n(t)$

$$\implies \phi(z,0) = \sum_{n=0}^{\infty} z^n p_n(0) \iff \phi(z,0) = 1 + zp_1(0) + z^2 p_2(0) + z^3 p_3(0) \ldots$$

Given we start our phylogenetic tree with one species, our initial state is 1, therefore:
$\phi(z,0) = 0 + z(1) + 0 + 0 \ldots$
$\implies \phi(z,0) = z$

$$\implies \phi(z,0) = \Phi\left\{e^{(-\lambda-\mu)0}\frac{z-1}{\lambda z-\mu}\right\} = z$$
$$\iff \phi(z,0) = \Phi\left\{\frac{z-1}{\lambda z-\mu}\right\} = z$$

Then set $V = \dfrac{z-1}{\lambda z-\mu}$

$$\iff V(\lambda z-\mu) = z-1 \iff V\lambda z - V\mu - z = -1 \iff z(V\lambda-1) = V\mu-1 \iff z = \frac{V\mu-1}{V\lambda-1}$$

Now we have $V = \dfrac{z-1}{\lambda z-\mu}$ and $z = \dfrac{V\mu-1}{V\lambda-1}$

Plugging these values into the initial condition $\phi(z,0)$ gives

$\phi(z,0) = \Phi\left\{\dfrac{z-1}{\lambda z-\mu}\right\} = z \iff \phi(z,0) = \Phi(V) = \dfrac{V\mu-1}{V\lambda-1}$

Now we can let V $= e^{(\lambda-\mu)t}\dfrac{z-1}{\lambda z-\mu}$ to get:

$$\implies \phi(z,t) = \Phi\left\{e^{(\lambda-\mu)t}\frac{z-1}{\lambda z-\mu}\right\} = \frac{\mu e^{(\lambda-\mu)t}(\frac{z-1}{\lambda z-\mu})-1}{\lambda e^{(\lambda-\mu)t}(\frac{z-1}{\lambda z-\mu})-1}$$

□

Using both Lemma 3.1, Lemma 3.2 and Lemma 3.3 we obtain the following theorem for $p_n(t)$

Theorem 3.4 (+). *$p_n(t)$, the probability of having n species after time t is as follows.*

$$p_0(t) = \frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}; \quad p_1(t) = \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}; \tag{3.3}$$
$$p_n(t) = (\lambda/\mu)^{n-1} p_1(t) p_0(t)^{n-1} \quad \textit{for } n > 1 \tag{3.4}$$

With the initial condition that we start in state 1:

$$p_n(0) = \begin{cases} 1 & \textit{if } n = 1 \\ 0 & \textit{otherwise} \end{cases}$$

(See Gernhard [5] Equation (1))

Proof. We have obtained $\phi(z,t)$ from Lemma 3.3, now we can simplify the function, expand in powers of z, then equate coefficients to obtain $p_n(t)$

For simplicity let $b = e^{(\lambda-\mu)t}$

$$\phi(z,t) = \frac{\mu b \frac{z-1}{\lambda z-\mu} - 1}{\lambda b \frac{z-1}{\lambda z-\mu} - 1} \quad \text{Then we can simplify accordingly}$$

$$= \frac{\mu b(z-1) - \lambda z + \mu}{\lambda b(z-1) - \lambda z + \mu}$$

$$= \frac{\mu b z - \mu b - \lambda z + \mu}{\lambda b z - \lambda b - \lambda z + \mu}$$

$$= \frac{\mu - \mu b - z(\lambda - \mu b)}{\mu - \lambda b - z(\lambda - \lambda b)} \times \frac{-1}{-1}$$

$$= \frac{\mu b - \mu - z(\mu b - \lambda)}{\lambda b - \mu - z\lambda(b-1)}$$

$$= \frac{\mu(b-1) - z(\mu b - \lambda)}{\lambda b - \mu} \frac{1}{1 - z\frac{\lambda(b-1)}{\lambda b-\mu}}$$

Then expanding in powers of z gives

$$\phi(z,t) = \left(\frac{\mu(b-1) - z(\mu b - \lambda)}{\lambda b - \mu}\right) \times \left\{1 + \left(z\frac{\lambda(b-1)}{\lambda b - \mu}\right) + \left(z\frac{\lambda(b-1)}{\lambda b - \mu}\right)^2 + \ldots \right.$$
$$\left. + \left(z\frac{\lambda(b-1)}{\lambda b - \mu}\right)^n + \ldots\right\}$$
$$= \left(\frac{\mu(b-1)}{\lambda b - \mu} - \frac{z(\mu b - \lambda)}{\lambda b - \mu}\right) \times \left\{1 + z\left(\frac{\lambda(b-1)}{\lambda b - \mu}\right) + z^2\left(\frac{\lambda(b-1)}{\lambda b - \mu}\right)^2 + \ldots \right.$$
$$\left. + z^n\left(\frac{\lambda(b-1)}{\lambda b - \mu}\right)^n + \ldots\right\}$$

Then multiplying together and simplifying gives

$$\phi(z,t) = \frac{\mu(b-1)}{\lambda b - \mu} - \frac{z(\mu b - \lambda)}{\lambda b - \mu} + z\frac{\lambda(b-1)}{\lambda b - \mu}\frac{\mu(b-1)}{\lambda b - \mu} - z^2\frac{\lambda(b-1)}{\lambda b - \mu}\frac{(\mu b - \lambda)}{\lambda b - \mu}$$
$$+ z^2\frac{\lambda^2(b-1)^2}{(\lambda b - \mu)^2}\frac{\mu(b-1)}{\lambda b - \mu} - z^3\ldots$$

Then by matching coefficients of powers of z we have

$$\phi(z,t) = \frac{\mu(b-1)}{\lambda b - \mu} + z\left(\frac{\lambda(b-1)}{\lambda b - \mu}\frac{\mu(b-1)}{\lambda b - \mu} - \frac{(\mu b - \lambda)}{\lambda b - \mu}\right)$$
$$+ z^2\left(\frac{\lambda^2(b-1)^2}{(\lambda b - \mu)^2}\frac{\mu(b-1)}{\lambda b - \mu} - \frac{\lambda(b-1)}{\lambda b - \mu}\frac{(\mu b - \lambda)}{\lambda b - \mu}\right) + z^3\ldots$$

Generalizing for all powers and simplifying we get

$$\phi(z,t) = \frac{\mu(b-1)}{\lambda b-\mu} + \ldots + z^n \left(\frac{\lambda^n (b-1)^n}{(\lambda b-\mu)^n} \frac{\mu(b-1)}{\lambda b-\mu} - \frac{\lambda^{n-1}(b-1)^{n-1}}{(\lambda b-\mu)^{n-1}} \frac{\mu b-\lambda}{\lambda b-\mu} \right) + \ldots$$
$$= \frac{\mu(b-1)}{\lambda b-\mu} + \ldots + z^n \left(\frac{\lambda^{n-1}(b-1)^{n-1}}{(\lambda b-\mu)^{n-1}} \right) \left(\frac{\lambda(b-1)}{\lambda b-\mu} \frac{\mu(b-1)}{\lambda b-\mu} - \frac{\mu b-\lambda}{\lambda b-\mu} \right) + \ldots$$
$$= \frac{\mu(b-1)}{\lambda b-\mu} + \ldots + z^n \left(\frac{\lambda^{n-1}(b-1)^{n-1}}{(\lambda b-\mu)^{n-1}} \right) \left(1 - \frac{\lambda(b-1)}{\lambda b-\mu} \right) \left(1 - \frac{\mu(b-1)}{\lambda b-\mu} \right) \ldots$$

Equating coefficients of $\phi(z,t) = \sum_{n=0}^{\infty} z^n p_n(t)$ gives:

$$p_n(t) = \left(1 - \frac{\lambda(b-1)}{\lambda b-\mu} \right) \left(1 - \frac{\mu(b-1)}{\lambda b-\mu} \right) \left(\frac{\lambda(b-1)}{\lambda b-\mu} \right)^{n-1} \quad \text{for n } \geq 1$$

and $p_0(t) = \dfrac{\mu(b-1)}{\lambda b-\mu}$

Then using Proposition 3.4.1 to simplify, we have:

$$p_0(t) = \frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}$$
$$p_1(t) = \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$
$$p_n(t) = (\lambda/\mu)^{n-1} p_1(t) p_0(t)^{n-1} \quad \text{as desired}$$

□

Proposition 3.4.1. *We have the following simplification for $p_0(t), p_1(t)$ and $p_n(t)$*

$$\text{if } p_0(t) = \frac{\mu(b-1)}{\lambda b-\mu} \text{ and } p_n(t) = \left(1 - \frac{\lambda(b-1)}{\lambda b-\mu} \right) \left(1 - \frac{\mu(b-1)}{\lambda b-\mu} \right) \left(\frac{\lambda(b-1)}{\lambda b-\mu} \right)^{n-1} \quad \text{for } n \geq 1$$

Letting $b = e^{(\lambda-\mu)t}$ *gives*

$$p_0(t) = \frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}; \quad p_1(t) = \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}; \quad p_n(t) = (\lambda/\mu)^{n-1} p_1(t) p_0(t)^{n-1}$$

Proof. Firstly, we simplify for $p_0(t)$

$$p_0(t) = \frac{\mu(b-1)}{\lambda b-\mu}$$

Plugging back in b, where b $= e^{(\lambda-\mu)t}$ gives:

(3.5)

$$p_0(t) = \frac{\mu(e^{(\lambda-\mu)t}-1)}{\lambda e^{(\lambda-\mu)t}-\mu} \times \frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}$$
$$= \frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}} \quad \text{as desired} \tag{3.6}$$

Then we can simplify for $p_1(t)$

$$p_n(t) = \left(1-\frac{\lambda(b-1)}{\lambda b-\mu}\right)\left(1-\frac{\mu(b-1)}{\lambda b-\mu}\right)\left(\frac{\lambda(b-1)}{\lambda b-\mu}\right)^{n-1} \quad \text{for n } \geq 1$$

Then plugging in for $n=1$

$$p_1(t) = \left(1-\frac{\lambda(b-1)}{\lambda b-\mu}\right)\left(1-\frac{\mu(b-1)}{\lambda b-\mu}\right)\left(\frac{\lambda(b-1)}{\lambda b-\mu}\right)^{0}$$

Multiplying by 1, helps us get the desired simplification below

$$= \left(1-\frac{\lambda(e^{(\lambda-\mu)t}-1)}{\lambda e^{(\lambda-\mu)t}-\mu}\right)\left(1-\frac{\mu(e^{(\lambda-\mu)t}-1)}{\lambda e^{(\lambda-\mu)t}-\mu}\right) \times \frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}$$
$$= \left(1-\frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)\left(1-\frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right) \tag{3.7}$$

Next we expand the brackets and simplify to get the following

$$= 1-\frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}} - \frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}} + \frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}$$
$$= \frac{\lambda-\mu e^{-(\lambda-\mu)t}-\mu(1-e^{-(\lambda-\mu)t})-\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}} + \frac{\lambda\mu(1-e^{-(\lambda-\mu)t})(1-e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$
$$= \frac{\lambda-\mu e^{-(\lambda-\mu)t}-\mu+\mu e^{-(\lambda-\mu)t}-\lambda+\lambda e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}} + \frac{\lambda\mu(1-2e^{-(\lambda-\mu)t}+e^{-2(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$

We can multiply the left fraction by $\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}} = 1$ to get a common denominator

$$= \frac{\lambda e^{-(\lambda-\mu)t}-\mu}{\lambda-\mu e^{-(\lambda-\mu)t}} + \frac{\lambda\mu(1-2e^{-(\lambda-\mu)t}+e^{-2(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$

Then by expanding and cancelling like terms we have

$$= \frac{(\lambda e^{-(\lambda-\mu)t}-\mu)(\lambda-\mu e^{-(\lambda-\mu)t})+\lambda\mu(1-2e^{-(\lambda-\mu)t}+e^{-2(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$
$$= \frac{\lambda^2 e^{-(\lambda-\mu)t}-\lambda\mu-\lambda\mu e^{-2(\lambda-\mu)t}+\mu^2 e^{-(\lambda-\mu)t}+\lambda\mu-2\lambda\mu e^{-(\lambda-\mu)t}+\lambda\mu e^{-2(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}$$

Finally we isolate $e^{-(\lambda-\mu)t}$ to get the desired result

$$= \frac{\lambda^2 e^{-(\lambda-\mu)t} + 2\lambda\mu e^{-(\lambda-\mu)t} + \mu^2 e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^2}$$

$$= \frac{e^{-(\lambda-\mu)t}(\lambda^2 + 2\lambda\mu + \mu^2)}{(\lambda - \mu e^{-(\lambda-\mu)t})^2}$$

$$= \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^2} \quad \text{as desired}$$

When simplifying $p_n(t)$ we firstly multiply by $\left(\frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}\right)^{n-1} = 1$, to get the desired result

$$p_n(t) = \left(1 - \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(1 - \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(\frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)^{n-1} \times \left(\frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}\right)^{n-1}$$

$$= \left(1 - \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(1 - \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(\frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} \times \frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}\right)^{n-1}$$

Again multiplying by $\frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}} = 1$ gives

$$= \left(1 - \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(\frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}\right)\left(1 - \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}\right)\left(\frac{e^{-(\lambda-\mu)t}}{e^{-(\lambda-\mu)t}}\right)\left(\frac{\lambda(1 - e^{-(\lambda-\mu)t})}{\lambda - \mu e^{-(\lambda-\mu)t}}\right)^{n-1}$$

Which simplifies to the following

$$= \left(1 - \frac{\lambda(1 - e^{-(\lambda-\mu)t})}{\lambda - \mu e^{-(\lambda-\mu)t}}\right)\left(1 - \frac{\mu(1 - e^{-(\lambda-\mu)t})}{\lambda - \mu e^{-(\lambda-\mu)t}}\right)\left(\frac{\lambda(1 - e^{-(\lambda-\mu)t})}{\lambda - \mu e^{-(\lambda-\mu)t}}\right)^{n-1}$$

Now see the equation for $p_1(t)$ above, in line 3.7 above, we use this representation of $p_1(t)$ below

$$= p_1(t)\left((\lambda/\mu) \times \frac{\mu(1 - e^{-(\lambda-\mu)t})}{\lambda - \mu e^{-(\lambda-\mu)t}}\right)^{n-1}$$

Finally plugging in $p_0(t)$ from Equation 3.7 from the beginning of the proof gives us the desired result

$$= (\lambda/\mu)^{n-1} p_1(t) p_0(t)^{n-1} \quad \text{completing the proof}$$

□

Chapter 4

Obtaining the probability density of speciation times [5]

In this chapter our aim is to derive the probability density function and corresponding cumulative distribution function for speciation times. This is one of our most significant results. We also prove that speciation times are independent and identically distributed (*iid*). These two results, help us prove future theorems.

We first prove the following lemmas below, to supplement the theorems.

Let τ be the orientated tree with n leaves as seen in Figure 2.1b. Let $x_1 > x_2 > \ldots > x_{n-1}$ be the order of the $n-1$ speciation events by time as denoted by the X's in Figure 2.1c. We only go up to x_{n-1} speciation events because we are conditioning on the fact there are n extant species, for that to occur we need $n-1$ speciation events.

As indicated by Gernhard [5], the x_i, where $i = \{1, 2, \ldots, n-1\}$ are the order statistics of the s_i. Where s_i is the time of the speciation event, also denoted by X's in Figure 2.1c.

Now our goal is to first obtain the density of the ordered speciation events given n extant and setting the first speciation event to t_1, so condition on $x_1 = t_1$. Remember that $0 < t_{n-1} < t_{n-2} < \ldots < t_2 < t_1 < t_{or}$. When we are estimating speciation times we set the present as 0 and we count backwards to the origin time. We have $0 < x_i < t_{or}$ for all $x_i : i = \{1, 2, \ldots, n-1\}$.

Lemma 4.1 (+). *The ordered speciation times of $x_1, \ldots, x_{n-1}$ given n extant species and $t_1 = t$ is*

$$g(x_2, \ldots, x_{n-1} | t_1 = t, n) = (n-2)! \prod_{i=2}^{n-1} \mu \frac{p_1(x_i)}{p_0(t)}$$

(See Gernhard [5] Section 2)

Proof. From *Human Evolutionary Trees* by Elizabeth Alison Thompson (1975) [1] we are given that the ordered speciation times are approximately

$$g(x_2, \dots, x_{n-1}|t_1 = t, n) = (n-2)! \prod_{i=2}^{n-1} \lambda p_1(x)$$

Then we can make the following simplification to match Gernhards notation

Note that when origin time t is of relatively large or of a 'realistic value'

We get $\dfrac{\lambda - \mu e^{-(\lambda-\mu)t}}{\lambda(1 - e^{-(\lambda-\mu)t})} = 1$, this implies

$$\begin{aligned} g(x_2, \dots, x_{n-1}|t_1 = t, n) &= (n-2)! \prod_{i=2}^{n-1} \lambda p_1(x_i) \times \frac{\lambda - \mu e^{-(\lambda-\mu)t}}{\lambda(1 - e^{-(\lambda-\mu)t})} \\ &= (n-2)! \prod_{i=2}^{n-1} \mu p_1(x_i) \times \frac{\lambda - \mu e^{-(\lambda-\mu)t}}{\mu(1 - e^{-(\lambda-\mu)t})} \\ &= (n-2)! \prod_{i=2}^{n-1} \mu \frac{p_1(x_i)}{p_0(t)} \quad \text{as desired} \end{aligned}$$

□

Lemma 4.2 (+)**.** *We generate the same results whether we condition on $t_1 = t$ or $t_{or} = t$, where t_1 is the time of the first speciation event and t_{or} is the origin of the tree.*

(See Gernhard [5] Section 2)

Proof. We have conditioned on the fact that $t_1 = t$, that is first speciation event is equal to time t. We generate the same results if we condition time t as equal to the origin time t_{or}. This is described in Gernhard [5]. They argue that, suppose we have a tree where the most recent common ancestor was at time t_1. Then the speciation event at this time produces two separate trees, $\mathscr{F}_n$ and $\mathscr{F}_m$, which have n and m extant species respectively. Now, since after this speciation event each tree will evolve independently on one another. And the origin of each tree is t_1, yet it still has the speciation density indicated above. Therefore we can condition on both t_1 and t_{or} and have the same result. Completing the Lemma. □

4.1 Density and CDF for General λ and μ

Having obtained, $p_n(t)$, we can use this result to obtain the density for any speciation time as described by Gernhard in *The Conditioned Reconstructed Process* [5]. The following result is obtained:

Theorem 4.3 (+)**.** *The Density of Speciation Times is as follows for general* λ *and* μ

$$f(s|t_{or} = t) = \begin{cases} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) & \textit{if } s \leq t \\ 0 & \textit{otherwise} \end{cases} \tag{4.1}$$

(See Gernhard [5] Theorem 2.5)

Proof. We can derive the unordered speciation times, $f(s_2, \ldots, s_{n-1}|t_1 = t, n)$, by multiplying the density for the ordered speciation times from Lemma 4.1 and the probability of observing that specific ordering , or that specific tree τ. The probability density function of τ is derived in section 2.3.

Therefore we have:

$$\begin{aligned} f(s_2, \ldots, s_{n-1}|t_1 = t, n) &= g(x_1, \ldots, x_{n-1}|t_1 = t, n) f(\tau|t_1 = t, n) \\ &= (n-2)! \prod_{i=2}^{n-1} \mu \frac{p_1(s_i)}{p_0(t)} \frac{1}{(n-2)!} \\ &= \prod_{i=2}^{n-1} \mu \frac{p_1(s_i)}{p_0(t)} \end{aligned} \tag{4.2}$$

In section 2.2.1 we proved that each permutation of speciation events of, our $s_1, \ldots, s_{n-1}$ have equal probability. Therefore, and given independence from the memoryless property of Poisson processes, we can derive that:

$$\begin{aligned} f(s_i|t_1 = t, n) = \mu \frac{p_1(s_i)}{p_0(t)} &= \mu \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s_i}}{(\lambda-\mu e^{-(\lambda-\mu)s_i})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\mu(1-e^{-(\lambda-\mu)t})}\right) \\ &= \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s_i}}{(\lambda-\mu e^{-(\lambda-\mu)s_i})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) \end{aligned}$$

Then using Lemma 4.2 we get the desired result that

$$f(s|t_{or} = t) = \begin{cases} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) & \text{if } s \leq t \\ 0 & \text{otherwise} \end{cases}$$

□

Corollary 4.3.1 (+)**.** *Speciation times, as indicated by the probability distribution in Theorem 4.3, are independent, and identically distributed (iid).*

Proof. We have already proved in Theorem 4.3 that the ordered speciation times are distributed with density function $f(s_2, \ldots, s_{n-1}|t_1 = t, n) = \prod_{i=2}^{n-1} \mu \frac{p_1(s_i)}{p_0(t)}$, see equation 4.2. Given this is a product of all speciation times, it is evident that speciation times are independent, as independence means multiply. We also already proved speciation times are identically distributed in Theorem 4.3. Therefore, speciation times are iid, as desired. □

Theorem 4.4 (+)**.** *The cumulative distribution function of speciation times is as follows for general λ and μ*

$$F(s|t_{or} = t) = \begin{cases} \left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) & \textit{if } s \leq t \\ 1 & \textit{otherwise} \end{cases} \tag{4.3}$$

(See Gernhard [5] Theorem 2.5)

Proof. We can find the cumulative distribution function $F(s_i|t_{or} = t, n)$ by integrating $f(s_i|t_{or} = t, n)$ from Equation 4.1

$$F(s_i|t_{or}=t,n) = \int_0^{s_i} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x}}{(\lambda-\mu e^{-(\lambda-\mu)x})^2} \right) \left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right) dx$$

$$= \frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \int_0^{s_i} \frac{e^{-(\lambda-\mu)x}}{(\lambda-\mu e^{-(\lambda-\mu)x})^2} dx$$

Apply u subsitution: Let u $= \lambda - \mu e^{-(\lambda-\mu)x}$

$$\implies \frac{du}{dx} = -(\lambda-\mu)(-\mu e^{-(\lambda-\mu)x})$$

$$\iff \frac{du}{dx} = (\lambda-\mu)(\mu e^{-(\lambda-\mu)x})$$

$$\implies dx = du \frac{1}{(\lambda-\mu)(\mu e^{-(\lambda-\mu)x})}$$

We also need to change the bounds of the definite integral

First letting $x = 0 \implies$ Lower Bound $= \lambda - \mu e^{-(\lambda-\mu)0}$

$\implies$ Lower Bound $= \lambda - 1$

Next letting $x = s_i \implies$ Upper Bound $= \lambda - \mu e^{-(\lambda-\mu)s_i}$

$$\implies F(s_i|t_1=t|n) = \frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \int_{\lambda-1}^{\lambda-\mu e^{-(\lambda-\mu)s_i}} \left(\frac{e^{-(\lambda-\mu)x}}{u^2} \right) \left(\frac{1}{(\lambda-\mu)(\mu e^{-(\lambda-\mu)x})} \right) du$$

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \int_{\lambda-1}^{\lambda-\mu e^{-(\lambda-\mu)s_i}} \frac{1}{u^2} du$$

Now solving the integral

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \left[\frac{-1}{u} \right]_{\lambda-1}^{\lambda-\mu e^{-(\lambda-\mu)s_i}}$$

Then plugging in the bounds

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \left(\frac{-1}{\lambda-\mu e^{-(\lambda-\mu)s_i}} - \frac{-1}{\lambda-\mu} \right)$$

Now simplifying to get the desired result

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \left(\frac{1}{\lambda-\mu} - \frac{1}{\lambda-\mu e^{-(\lambda-\mu)s_i}} \right)$$

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \left(\frac{\lambda-\mu e^{-(\lambda-\mu)s_i}-\lambda+\mu}{(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s_i})} \right)$$

$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1}{(\lambda-\mu)\mu} \right) \left(\frac{(\mu)(1-e^{-(\lambda-\mu)s_i})}{(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s_i})} \right)$$
$$= \left(\frac{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)t})}{1-e^{-(\lambda-\mu)t}} \right) \left(\frac{1-e^{-(\lambda-\mu)s_i}}{(\lambda-\mu)^2(\lambda-\mu e^{-(\lambda-\mu)s_i})} \right)$$
$$= \left(\frac{1-e^{-(\lambda-\mu)s_i}}{\lambda-\mu e^{-(\lambda-\mu)s_i}} \right) \left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)$$

Hence we have the following formula for the distribution function for speciation times of $s_i \geq 0$:

$$F(s_i|t_{or}=t,n) = \left(\frac{1-e^{-(\lambda-\mu)s_i}}{\lambda-\mu e^{-(\lambda-\mu)s_i}} \right) \left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)$$

□

4.2 Density and CDF for $\lambda = \mu$

The previously proved functions (Equations 4.1, 4.3) only exist for $\lambda \neq \mu$. If $\lambda = \mu$, we are dividing $f(s|t_{or}=t,n)$ by zero as, $\lambda - \mu e^{-(\lambda-\mu)s} = 0$. Because $f(s|t_{or}=t,n)$ does not exist in this case, this implies $F(s|t_{or}=t,n)$ does not exist either. Therefore we need to derive them.

Gernhard [5] gives the following theorem in section 2.1.2 with no proof, we provide the proof below.

Theorem 4.5 (+)**.** *The probability density function for speciation times in the critical case, $\lambda = \mu$, is*

$$f(s|t_{or}=t) = \begin{cases} \left(\frac{1}{(1+\lambda s)^2}\right)\left(\frac{1+\lambda t}{t}\right) & \text{if } s \leq t \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

(See Gernhard [5] Section 2.1.2)

Proof. We derive the density function in the critical case using limits

$$\lim_{\mu\to\lambda} f(s|t_{or}=t,n) = \lim_{\mu\to\lambda} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2} \right) \left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)$$

$$= \lim_{\mu\to\lambda} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2} \right) \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)$$

Using the approximation that $e^{-\epsilon} \sim 1-\epsilon$ for $\epsilon \to 0$

$$\implies \lim_{\mu\to\lambda} f(s|t_{or}=t,n) = \lim_{\mu\to\lambda} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2} \right) \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)$$

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

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

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

Plugging in λ for μ gives us an indeterminate result for both limits
as we get zero for both numerator and denominators

However, we can apply L'Hopitals: $\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x))}{g'(x)}$ when $f(c)=0=g(c)$

$$\implies \lim_{\mu\to\lambda} f(s|t_{or}=t,n) = \lim_{\mu\to\lambda} \left(\frac{\frac{\partial}{\partial\mu}(\lambda-\mu)^2(1-\lambda s+\mu s)}{\frac{\partial}{\partial\mu}(\lambda-\mu+\lambda\mu s-\mu^2 s)^2} \right) \lim_{\mu\to\lambda} \left(\frac{\frac{\partial}{\partial\mu}(\lambda-\mu+\lambda\mu t-\mu^2 t)}{\frac{\partial}{\partial\mu}(\lambda-\mu)t} \right)$$

$= ① + ②$ (We split the limits and solve separately

Solve the first limit ① $= \lim_{\mu\to\lambda} \left(\frac{\frac{\partial}{\partial\mu}(\lambda-\mu)^2(1-\lambda s+\mu s)}{\frac{\partial}{\partial\mu}(\lambda-\mu+\lambda\mu s-\mu^2 s)^2} \right)$

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

The limit is still indeterminate so we can apply L'Hopitals again and simplify

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

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

Finally, we let $\mu=\lambda$ to get

$$① = \frac{2(1-\lambda s+\lambda s)-2(\lambda-\lambda)}{2(-1+\lambda s-2\lambda s)(-1+\lambda s-2\lambda s)+2(\lambda-\lambda+\lambda\lambda s-\lambda^2 s)(-2s)}$$

$$= \frac{2}{2(-1-\lambda s)(-1-\lambda s)}$$

$$= \frac{1}{(1+\lambda s)^2}$$

Now we can solve the second limit using L'Hopitals again

$$\text{Let } ② = \lim_{\mu\to\lambda}\left(\frac{\frac{\partial}{\partial\mu}(\lambda-\mu+\lambda\mu t-\mu^2 t)}{\frac{\partial}{\partial\mu}(\lambda-\mu)t}\right)$$
$$= \lim_{\mu\to\lambda}\left(\frac{-1+\lambda t-2\mu t}{(-1)t}\right)$$

Then letting $\mu = \lambda$ gives

$$② = \frac{-1+\lambda t-2\lambda t}{(-1)t}$$
$$= \frac{-1(1+\lambda t)}{(-1)t}$$
$$= \frac{1+\lambda t}{t}$$

Combining ① and ② we get the desired result

$$\implies \lim_{\mu\to\lambda} f(s|t_{or}=t,n) = \left(\frac{1}{(1+\lambda s)^2}\right)\left(\frac{1+\lambda t}{t}\right)$$

□

Theorem 4.6 (+)**.** *The cumulative distribution function for speciation times in the critical case, $\lambda = \mu$, is*

$$F(s|t_{or}=t) = \begin{cases} \left(\frac{1+\lambda t}{t}\right)\left(\frac{s}{1+\lambda s}\right) & \text{if } s \le t \\ 1 & \text{otherwise} \end{cases} \tag{4.5}$$

(See Gernhard [5] section 2.1.2)

Proof. We can determine the cumulative distribution function at $\lambda = \mu$ by integrating $f(s|t_{or}=t)$ from Equation 4.4

Use integration of the density to find the cumulative distribution function

$$F(s|t_{or}=t,n) = \int_0^s \left(\frac{1}{(1+\lambda x)^2}\right)\left(\frac{1+\lambda t}{t}\right)dx$$
$$= \left(\frac{1+\lambda t}{t}\right)\int_0^s \left(\frac{1}{(1+\lambda x)^2}\right)dx$$

Now solving the integral using u substitution

Let u $= 1 + \lambda x \implies \dfrac{du}{dx} = \lambda \implies dx = \dfrac{du}{\lambda}$ We also need to change the bounds

Lower Bound $= 1 + \lambda(0) = 1$; Upper Bound $= 1 + \lambda s$

Now plugging in u, du and the bounds

$$
\begin{aligned}
F(s|t_{or} = t, n) &= \left(\frac{1+\lambda t}{t}\right) \int_1^{1+\lambda s} \left(\frac{1}{u^2}\frac{du}{\lambda}\right) \\
&= \left(\frac{1+\lambda t}{\lambda t}\right) \int_1^{1+\lambda s} \left(\frac{1}{u^2}\right) du \\
&= \left(\frac{1+\lambda t}{\lambda t}\right) \left[\frac{-1}{u}\right]_1^{1+\lambda s}
\end{aligned}
$$

Solving the integral and simplifies gives the desired result

$$
\begin{aligned}
&= \left(\frac{1+\lambda t}{\lambda t}\right) \left(\frac{-1}{1+\lambda s} - \frac{-1}{1}\right) \\
&= \left(\frac{1+\lambda t}{\lambda t}\right) \left(1 - \frac{1}{1+\lambda s}\right) \\
&= \left(\frac{1+\lambda t}{\lambda t}\right) \left(\frac{\lambda s}{1+\lambda s}\right) \\
&= \left(\frac{1+\lambda t}{t}\right) \left(\frac{s}{1+\lambda s}\right)
\end{aligned}
$$

□

4.3 Density and CDF for $\mu = 0$ (Yule Case)

Gernhard [5] gives the following theorem in section 2.1.1 with no proof, we provide the proof below.

Theorem 4.7 (+). *In the Yule case ($\mu = 0$) we have the following probability density function for a speciation times:*

$$
f(s|t_{or} = t) = \begin{cases} \left(\frac{\lambda e^{-\lambda s}}{1-e^{-\lambda t}}\right) & \text{if } s \leq t \\ 0 & \text{otherwise} \end{cases} \tag{4.6}
$$

(See Gernhard [5] section 2.1.1)

Proof. We can prove this result by setting $\mu = 0$ in the general formula, see Equations 4.1

$$
\begin{aligned}
f(s|t) &= \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) \\
&= \left(\frac{(\lambda)^2 e^{-(\lambda)s}}{(\lambda)^2}\right)\left(\frac{\lambda}{1-e^{-(\lambda)t}}\right) \\
&= \left(\frac{(\lambda)^2 e^{-(\lambda)s}}{(\lambda)^2}\right)\left(\frac{\lambda}{1-e^{-(\lambda)t}}\right) \\
&= \left(\frac{\lambda e^{-\lambda s}}{1-e^{-\lambda t}}\right)
\end{aligned}
$$

□

Theorem 4.8 (+). *In the Yule case ($\mu = 0$) we have the following probability density function for a speciation times:*

$$
F(s|t_{or} = t) = \begin{cases} \left(\frac{1-e^{-\lambda s}}{1-e^{-\lambda t}}\right) & \text{if } s \leq t \\ \\ 1 & \text{otherwise} \end{cases} \tag{4.7}
$$

(See Gernhard [5] section 2.1.1)

Proof. We prove this result by integrating Equation 4.6 for $f(s|t_{or} = t)$

$$
\begin{aligned}
F(s|t) &= \int_0^s \left(\frac{\lambda e^{-\lambda x}}{1-e^{-\lambda t}}\right) dx \\
&= \left[\frac{e^{-\lambda x}}{1-e^{-\lambda t}}\right]_0^s \\
&= \frac{1-e^{-\lambda s}}{1-e^{-\lambda t}}
\end{aligned}
$$

□

Chapter 5

Including the time of origin t_{or}

So far, we have introduced evolutionary trees and illustrated how to estimate speciation times through probability density functions and cumulative distribution functions. Given we have conditioned on t_{or}, the origin of the tree, it is important to determine a density function for the origin time. We can do this through a bayesian framework, by finding the conditional density for the origin time, with a uniform prior for t_{or}. Once we find the density function, it helps us to remove our condition on t for further calculations. We remove t by weighting for it and integrating it out, see the law of total probability.

Gernhard [5] provides the probability density function for t_{or} in Theorem 3.2, and a sketch proof. We provide the complete proof to verify the density function, which is mostly omitted in Gernhard [5].

Theorem 5.1 (+)**.** *The probability density function for the origin time t_{or}, assuming a uniform prior from 0 to ∞ is as follows:*

$$q_{or}(t|n) = \mathbb{P}(t_{or} = t|X_0 = n) = \begin{cases} n\lambda^n(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} & \textit{if } t \geq 0 \\ 0 & \textit{otherwise} \end{cases} \tag{5.1}$$

(See Gernhard [5] Theorem 3.2)

Proof. A Bayesian framework follows the format of $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$, where $\mathbb{P}(A)$ is our prior, $\mathbb{P}(B|A)$ is our likelihood and $\mathbb{P}(B)$ is our data. In our case we would set:

- $\mathbb{P}(A|B) = \mathbb{P}(t_{or} = t|X_t = n)$. That is the density for the origin time, given at the end of the process there are n species.
- $\mathbb{P}(B|A) = \mathbb{P}(X_t = n|t_{or} = t)$, in the previous section we derived this value, $p_n(t)$, this was proved in Theorem 3.4.

- $\mathbb{P}(A) = \mathbb{P}(t_{or} = t)$, this is the uniform prior we have assigned to t_{or}.
- Finally $\mathbb{P}(B) = \mathbb{P}(X_t = n) = \int_0^\infty \mathbb{P}(X_t = n|t_{or} = t)\mathbb{P}(t_{or} = t)$, by the law of total probability.

Under this framework we can find the density for this origin time.

Note: Remember that $\mathbb{P}(X_t = n|t_{or} = t)$ reflects the number of species after a total time elapsed of t. $t_{or} = t$ is setting the origin time to t, where the present time is 0. So the total time elapsed will still be t. Given t_{or} has an improper prior from 0 to ∞. We will assume that $t_{or} \sim U(0, N)$ then take the limit as $N \to \infty$.

$$\mathbb{P}(t_{or} = t|X_t = n) = \frac{\mathbb{P}(X_t = n|t_{or} = t)\mathbb{P}(t_{or} = t)}{\mathbb{P}(X_t = n)}$$

In the next two steps we simplify the denominator using the law of total probability and we assume $t_{or} \sim U(0, N)$

$$\begin{aligned}
\mathbb{P}(t_{or} = t|X_t = n) &= \frac{\lim_{N\to\infty} \mathbb{P}(X_t = n|t_{or} = t)\mathbb{P}(t_{or} = t)}{\lim_{N\to\infty} \int_0^N \mathbb{P}(X_t = n|t_{or} = t)\mathbb{P}(t_{or} = t)dt} \\
&= \frac{\lim_{N\to\infty} \mathbb{P}(X_t = n|t_{or} = t)\frac{1}{N}\mathbb{1}_{t\in(0,N)}}{\lim_{N\to\infty} \int_0^N \mathbb{P}(X_t = n|t_{or} = t)\frac{1}{N}\mathbb{1}_{t\in(0,N)}dt}
\end{aligned}$$

Remove terms that aren't t from the integration

$$= \frac{\lim_{N\to\infty} \frac{1}{N}\mathbb{1}_{t\in(0,N)}\mathbb{P}(X_t = n|t_{or} = t)}{\lim_{N\to\infty} \frac{1}{N}\mathbb{1}_{t\in(0,N)} \int_0^N \mathbb{P}(X_t = n|t_{or} = t)dt}$$

Dividing over terms $1/N$ and $\mathbb{1}_{t\in(0,N)}$, removes all limit related terms therefore we can remove the limit

$$\begin{aligned}
&= \frac{\mathbb{P}(X_t = n|t_{or} = t)}{\int_0^\infty \mathbb{P}(X_t = n|t_{or} = t)dt} \\
&= \frac{p_n(t)}{\int_0^\infty p_n(t)} \quad \text{by definition} \\
&= \frac{\left(\frac{\lambda}{\mu}\right)^{n-1} p_1(t)[p_0(t)]^{n-1}}{\int_0^\infty \left(\frac{\lambda}{\mu}\right)^{n-1} p_1(t)[p_0(t)]^{n-1}dt} \quad \text{by plugging in values from Theorem 3.4} \\
&= \frac{p_1(t)[p_0(t)]^{n-1}}{\int_0^\infty p_1(t)[p_0(t)]^{n-1}dt} \quad \text{by dividing over common values}
\end{aligned}$$

$$= \frac{p_1(t)[p_0(t)]^{n-1}}{\int_0^\infty \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)\left(\frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{n-1} dt}$$

$$= \frac{p_1(t)[p_0(t)]^{n-1}}{\int_0^\infty \mu^{n-1}(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n-1+2}} dt} \quad \text{by multiplying out}$$

Then we can multiply by $\frac{n}{n} = 1$ to get

$$\mathbb{P}(t_{or} = t | X_t = n) = \frac{p_1(t)[p_0(t)]^{n-1}}{\mu^{n-1}\int_0^\infty (\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} dt} \times \frac{n}{n}$$

$$= \frac{np_1(t)[p_0(t)]^{n-1}}{\mu^{n-1}\int_0^\infty n(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} dt}$$

Then using Proposition 5.1.1 below we have

$$\mathbb{P}(t_{or} = t | X_t = n) = \frac{np_1(t)[p_0(t)]^{n-1}}{\mu^{n-1}\left[\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n\right]_0^\infty}$$

Then we can plug in bounds of the integration

$$= \frac{np_1(t)[p_0(t)]^{n-1}}{\mu^{n-1}\left[\left(\frac{1-e^{-(\lambda-\mu)\infty}}{\lambda-\mu e^{-(\lambda-\mu)\infty}}\right)^n - \left(\frac{1-1}{\lambda-\mu}\right)^n\right]}$$

Then assuming $0 < \mu < \lambda$ so $e^{-(\lambda-\mu)\infty} \to 0$ we have

$$= \frac{np_1(t)[p_0(t)]^{n-1}}{\mu^{n-1}\left(\frac{1}{\lambda}\right)^n}$$

Next plugging in values from Theorem 3.4, then simplifying

$$= \frac{n\lambda^n \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)\left(\frac{\mu(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{n-1}}{\mu^{n-1}}$$

$$= \frac{n\lambda^n \mu^{n-1} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{n-1}}{\mu^{n-1}}$$

$$= n\lambda^n \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{n-1}$$

$$= n\lambda^n(\lambda-\mu)^2 \left(\frac{e^{-(\lambda-\mu)t}\left(1-e^{-(\lambda-\mu)t}\right)^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2(\lambda-\mu e^{-(\lambda-\mu)t})^{n-1}}\right)$$

$$= n\lambda^n(\lambda-\mu)^2 \frac{\left(1-e^{-(\lambda-\mu)t}\right)^{n-1} e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}$$

Using Gernhard's notation [5]

$$q_{or}(t|n) = n\lambda^n(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} \quad \text{as desired}$$

□

It remains to prove the following proposition

Proposition 5.1.1. *The following integral has the anti-derivative:*

$$\int n(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} dt = \left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n$$

Proof. We can solve by taking the derivative of the solution, $\frac{d}{dt}\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n$ and checking to see if we get the original integral

We can use the quotient rule (with the chain rule too) to solve the derivative and simplify:

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right]^n = n\left(\frac{f(x)}{g(x)}\right)^{n-1} \times \left(\frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}\right)$$

$$\implies \frac{d}{dt}\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n = n\left(\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{n-1}$$

$$\times\left(\frac{(\lambda-\mu e^{-(\lambda-\mu)t})((\lambda-\mu)e^{-(\lambda-\mu)t})-((\lambda-\mu)\mu e^{-(\lambda-\mu)t})(1-e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)$$

$$= n\left(\frac{(1-e^{-(\lambda-\mu)t})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n-1}}\right)$$

$$\times\left(\frac{(\lambda-\mu)(e^{-(\lambda-\mu)t})(\lambda-\mu e^{-(\lambda-\mu)t})-(\lambda-\mu)(e^{-(\lambda-\mu)t})(\mu)(1-e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right)$$

$$= n\left(\frac{(1-e^{-(\lambda-\mu)t})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}\right)(\lambda-\mu)(e^{-(\lambda-\mu)t})\left(\lambda-\mu e^{-(\lambda-\mu)t}-\mu+\mu e^{-(\lambda-\mu)t}\right)$$

$$= n\left(\frac{(1-e^{-(\lambda-\mu)t})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}\right)(\lambda-\mu)(e^{-(\lambda-\mu)t})(\lambda-\mu)$$

$$= n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)t})^{n-1}(e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}$$

This is the desired result

□

We have found the density function for the origin time given n extant species after time t, then next step is to find the distribution function $Q_{or}(t|n)$. This is solved by taking the integral of the density function $q_{or}(t|n)$

The following theorem for the cumulative distribution is also in Gernhard [5] Corollary 3.3, with a sketch proof. We provide the full detailed proof below.

Theorem 5.2 (+). *The cumulative distribution function for the origin time t_{or}, assuming a uniform prior from 0 to ∞ is as follows:*

$$Q_{or}(t|n) = \mathbb{P}(t_{or} \leq t | X_0 = n) = \begin{cases} \left(\frac{\lambda(1-e^{-(\lambda-\mu)t}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

(See Gernhard [5] Corollary 3.3)

Proof.

$$\begin{aligned} Q_{or}(t|n) &= \mathbb{P}(t_{or} \leq t | X_t = n) \\ &= \int_0^t \mathbb{P}(t_{or} = x | X_t = n) dx \\ &= \int_0^t n\ \lambda^n (\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)x})^{n-1} e^{-(\lambda-\mu)x}}{(\lambda-\mu e^{-(\lambda-\mu)x})^{n+1}} dx \end{aligned}$$

We can again solve by finding the anti derivative
Using Proposition 5.2.1 to solve the integral

$$\begin{aligned} \implies Q_{or}(t) &= \left[\left(\frac{\lambda(1-e^{-(\lambda-\mu)x})}{\lambda-\mu e^{-(\lambda-\mu)x}}\right)^n\right]_0^t \\ &= \left(\frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n - \left(\frac{\lambda(1-e^{-(\lambda-\mu)0})}{\lambda-\mu e^{-(\lambda-\mu)0}}\right)^n \\ &= \left(\frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n - \left(\frac{\lambda(1-1)}{\lambda-\mu}\right)^n \\ &= \left(\frac{\lambda(1-e^{-(\lambda-\mu)t})}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^n \quad \text{as desired} \end{aligned}$$

□

It remains to solve the following proposition

Proposition 5.2.1. *The following integral solves to:*

$$\int n\lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)x})^{n-1}e^{-(\lambda-\mu)x}}{(\lambda-\mu e^{-(\lambda-\mu)x})^{n+1}}dx = \left(\frac{\lambda(1-e^{-(\lambda-\mu)x})}{\lambda-\mu e^{-(\lambda-\mu)x}}\right)^n$$

Proof. Again, we prove this by taking the derivative of the solution, $\frac{d}{dx}\left(\frac{\lambda(1-e^{-(\lambda-\mu)x})}{\lambda-\mu e^{-(\lambda-\mu)x}}\right)^n$ and checking if we get the original integral. We first differentiate by using the quotient rule and chain rule

$$\frac{d}{dx}\left(\frac{\lambda(1-e^{-(\lambda-\mu)x})}{\lambda-\mu e^{-(\lambda-\mu)x}}\right)^n = n\left(\frac{\lambda(1-e^{-(\lambda-\mu)x})}{\lambda-\mu e^{-(\lambda-\mu)x}}\right)^{n-1}$$
$$\times\frac{(\lambda-\mu e^{-(\lambda-\mu)x})\lambda(\lambda-\mu)e^{-(\lambda-\mu)x}-\mu(\lambda-\mu)e^{-(\lambda-\mu)x}\lambda(1-e^{-(\lambda-\mu)x})}{(\lambda-\mu e^{-(\lambda-\mu)x})^2}$$

Then we can expand the first half and isolate common terms in the second half

$$= n\frac{\lambda^{n-1}(1-e^{-(\lambda-\mu)x})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)x})^{n-1}}$$
$$\times\lambda(\lambda-\mu)e^{-(\lambda-\mu)x}\frac{((\lambda-\mu e^{-(\lambda-\mu)x})-\mu(1-e^{-(\lambda-\mu)x}))}{(\lambda-\mu e^{-(\lambda-\mu)x})^2}$$

Then by expanding the brackets in the second fraction

$$= n\frac{\lambda^{n-1}(1-e^{-(\lambda-\mu)x})^{n-1}}{((\lambda-\mu e^{-(\lambda-\mu)x})^{n-1})}$$
$$\times\lambda(\lambda-\mu)e^{-(\lambda-\mu)x}\frac{(\lambda-\mu e^{-(\lambda-\mu)x}-\mu+\mu e^{-(\lambda-\mu)x})}{(\lambda-\mu e^{-(\lambda-\mu)x})^2}$$

Next we can simplify the second fraction to get

$$= n\frac{\lambda^{n-1}(1-e^{-(\lambda-\mu)x})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)x})^{n-1}}\lambda(\lambda-\mu)e^{-(\lambda-\mu)x}\frac{(\lambda-\mu)}{(\lambda-\mu e^{-(\lambda-\mu)x})^2}$$

Finally, by matching common terms we get the desired result

$$= n\ \lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)x})^{n-1}e^{-(\lambda-\mu)x}}{(\lambda-\mu e^{-(\lambda-\mu)x})^{n+1}}$$

□

Chapter 6

Density for Time of K-th Speciation Events

Another useful analytical result to derive is the density for the time of the kth speciation event. In a previous section we derived the density for the time of a single speciation event $f(s|t_{or} = t)$. We also found the times are independent and identically distributed for all speciation events. Therefore the natural next step is to find the density for the 2nd, 3rd, ..., kth speciation events.

We have two possibilities when considering this, we either condition on a known origin time $t_{or} = t$ or we can assume t_{or} is unknown, in this case we do not condition on t_{or}.

Using Gernhard's [5] notation, let $\mathscr{A}_{n,t}^{k}$ be the time of the kth speciation given n extant species and a known origin time t, where $\mathscr{A}$ is the reconstructed evolutionary tree. Let $\mathscr{A}_{n}^{k}$ be the time of the kth speciation event and assuming the origin time, t_{or}, is unknown, assuming a uniform prior across the reals.

6.1 Known origin time t_{or}

In this section we derive both the probability density function and cumulative distribution function. We choose to also derive the cumulative distribution function because it is useful in deriving the expectation of the kth speciation time in further sections, see Chapter 7.

6.1.1 Probability Density Function

The following theorem is present in Gernhard [5] Equation (6), however there is no proof provided. We provide the theorem and the proof below.

Theorem 6.1 (+). *The probability density function for the kth speciation time, given a known*

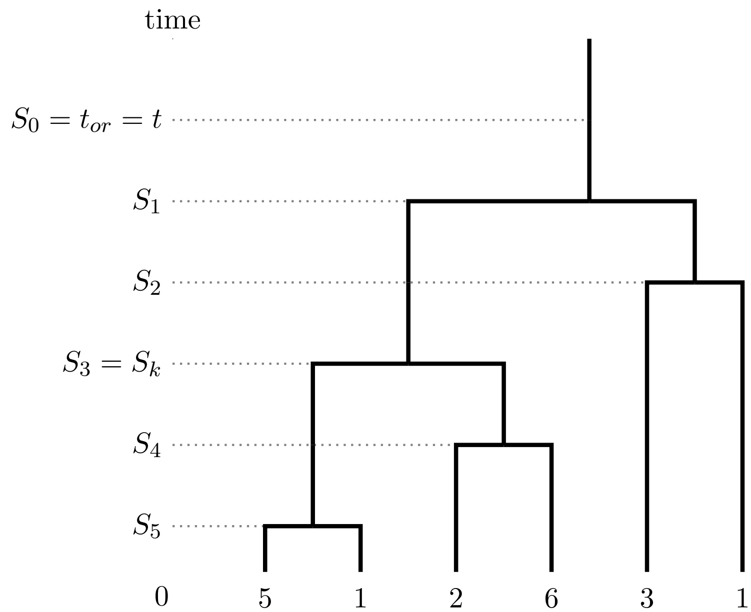


Figure 6.1: Reconstructed Tree

origin time is:

$$f_{\mathscr{A}^k_{n,t}}(s) = \begin{cases} \binom{n-1}{n-k}(n-k)F(s|t)^{n-k-1}(1-F(s|t))^{k-1}f(s|t) & \textit{if } s \leq t \\ 0 & \textit{otherwise} \end{cases} \tag{6.1}$$

This result is applicable in the general case, the critical case $\lambda = \mu$ and in the Yule case $\mu = 0$. See equations 4.1 and 4.4 for $f(s|t)$. See equations 4.3 and 4.5 for $F(s|t)$.

(See Gernhard [5] Equation (6))

Proof. Our goal is to find $f_{\mathscr{A}^k_{n,t}}(s)$ where s is the time of speciation event, in another form $\mathbb{P}(\mathscr{A}^k = s|t_{or} = t, n)$, where $s \leq t$. Let (uppercase) S_i denote our random speciation events and s be the time they occur. Intuitively the density for the k-th speciation event will be the joint probability that speciation events $S_{n-1}, S_{n-2}, \ldots, S_{n-(k+1)}$ occur after the kth event, and events $S_{k-1}, \ldots, S_2, S_1, S_0$ occur before k, where S_0 is the origin time. This might seem confusing hence a graph might help, see Figure 6.1, note the time-axis of the illustration. If we let $k = 3$, then random speciation times S_0, S_1, S_2 occur before, and times S_4, S_5 occur after.

We also have to take into account all the possible orderings of S_i for all $i = \{0, 1, 2, \ldots, n-1\}$. Firstly, there are $n-1$ speciation events and the S_i's are iid, but we exclude S_0 because it is fixed. Therefore we have $n-1$ permutations of all the S_i, so we have to multiply our probability by $(n-1)!$.

But we also need to account for the probability we observe the group of S's that occur before k and the group of $S's$ that occur after k. Therefore we divide by $(n-(k+1))!$, which is all the possible permutations of speciation events that occur after k. We also divide by $(k-1)!$ which is all the permutations of the speciation events that occur before k.

We must multiple our probability by $\binom{n-1}{k+1}$ to reflect all possible ordering of S's. We can simplify this using Proposition 6.1.1, $\binom{n-1}{k+1} = \binom{n-1}{n-k}(n-k)$:

Now we can determine the density for the k-th speciation time when $t_{or} = t$ is fixed given n extant species. This conclusion is valid for the all λ and μ.

$$
\begin{aligned}
f_{\mathscr{A}^k_{n,t}}(s) &= \binom{n-1}{n-k}(n-k) \times \mathbb{P}(\mathscr{A}^k = s|t_{or} = t, n) \quad \text{then letting } C = \binom{n-1}{n-k}(n-k) \\
&= C \times \mathbb{P}(S_{n-1}, S_{n-2}, \ldots, S_{k+1} < s; S_k = s; S_{k-1}, S_{k-2}, \ldots, S_1 > s|t_{or} = t, n) \\
&= C \times \mathbb{P}(S_{n-1} < s, S_{n-2} < s, \ldots, S_{k+1} < s; S_k = s; S_{k-1} > s, S_{k-2} > s, \\
&\qquad\qquad \ldots S_1 > s|t_{or} = t, n)
\end{aligned}
$$

By independence of speciation times, the S_i's (see Corollary 4.3.1), we have the following

$$
f_{\mathscr{A}^k_{n,t}}(s) = C \prod_{i=k+1}^{n-1} \mathbb{P}(S_i < s|t_{or} = t, n) \times \mathbb{P}(S_k = s|t_{or} = t, n) \times \prod_{i=1}^{k-1} \mathbb{P}(S_i > s|t_{or} = t, n)
$$

Then by identical distribution of speciation times, also see Corollary 4.3.1, we have

$$
\begin{aligned}
f_{\mathscr{A}^k_{n,t}}(s) &= C \prod_{i=k+1}^{n-1} \mathbb{P}(S < s|t_{or} = t, n) \times \mathbb{P}(S_k = s|t_{or} = t, n) \times \prod_{i=1}^{k-1} \mathbb{P}(S > s|t_{or} = t, n) \\
&= C \times \mathbb{P}(S < s)^{n-1-(k+1)+1} \mathbb{P}(S_k = s) \mathbb{P}(S > s)^{k-1-1+1} \\
&= C \times \mathbb{P}(S < s)^{n-k-1} \mathbb{P}(S_k = s) \mathbb{P}(S > s)^{k-1}
\end{aligned}
$$

Given speciation times are continuous, we can say

$$
f_{\mathscr{A}^k_{n,t}}(s) = C \times F(S|t_{or} = t, n)^{n-k-1} f(s|t_{or} = t, n)(1 - F(S|t_{or} = t, n)^{k-1}
$$

Finally using Gernhards [5] Notation, we get the final result

$$
\begin{aligned}
f_{\mathscr{A}^k_{n,t}}(s) &= \binom{n-1}{n-k}(n-k) \times F(s|t)^{n-k-1}(1 - F(s|t))^{k-1} f(s|t) \\
&= \binom{n-1}{n-k}(n-k) F(s|t)^{n-k-1}(1 - F(s|t))^{k-1} f(s|t)
\end{aligned}
$$

□

It remains the prove the proposition below.

Proposition 6.1.1. *The factorial is simplified to the following*

$$\frac{(n-1)!}{(n-(k+1)!(k-1)!} = \binom{n-1}{n-k}(n-k)$$

Proof.

$$\binom{n-1}{k+1} = \frac{(n-1)!}{(n-(k+1)!(k-1)!}$$

Next we multiple by $\frac{(n-k)}{(n-k)} = 1,$ to simplify

$$= \frac{(n-1)!}{(n-k-1)!(k-1)!} \times \frac{(n-k)}{(n-k)}$$
$$= \frac{(n-1)!(n-k)}{(n-k)!(k-1)!}$$
$$= \frac{(n-1)!(n-k)}{(n-k)!(n-1-n+k)}$$
$$= \frac{(n-1)!(n-k)}{(n-k)!(n-1-(n-k))!}$$
$$= \binom{n-1}{n-k}(n-k)$$

□

6.1.2 Cumulative Distribution Function

In this section we find the cumulative distribution. This result is important for later theorems. Gernhard [5], provides the result in Equation (7), but no proof. Below, we prove the result in full details.

Theorem 6.2 (+)**.** *The cumulative distribution function for the kth speciation time, given a known origin time is:*

$$F_{\mathscr{A}^k_{n,t}}(s) = \begin{cases} \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-i-1}(1-F(s|t))^i & \text{if } s \le t \\ 1 & \text{otherwise} \end{cases} \tag{6.2}$$

(See Gernhard [5] Equation (7))

Proof. Next we find the cumulative distribution function for the kth speciation event, $F_{\mathscr{A}^k_{n,t}}$. We only find the CDF when we condition on t, because we need to find the Expectation of the kth speciation time later on in Chapter 7.1.

We can think about finding $F_{\mathscr{A}_{n,t}^k}$ intuitively. We also use Theorem 9.17 from *Einführung in die Wahrscheinlichkeitstheorie und Statistik* by Herold Dehling and Beate Haupt [4] as inspiration. Their theorem states that for k order statistics where $X_1 \leq X_2 \leq \ldots \leq X_n$, where X's are independent identically distributed with a common CDF, $F(x)$. The the distribution function of the kth order statistic is

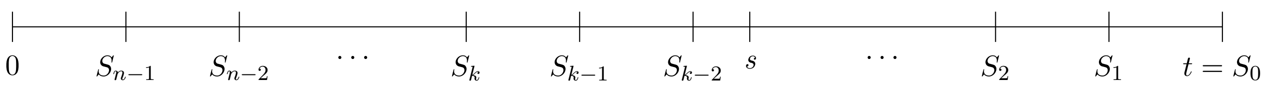
$$F_{(k)}(s) = \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}$$

Our result will be somewhat similar, however note that for speciation times S, our order is inverted,

$$S_{n-1} \leq S_{n-2} \leq \ldots \leq S_1 \leq S_0$$

where S_i is the ith speciation time, and S_0 is the origin. The ordering is like this because S_{n-1} is closer to the present time, 0.

In order to find the cumulative distribution function for the kth speciation time we need to think of it more intuitively. We can rewrite the probability, $\mathbb{P}(\mathscr{A}_{n,t}^k \leq s)$ as the probability that the number of events that occurs after s is less than or equal to $k-1$. For example if $k = 6$, and 6 events occur after s then its impossible S_6 occured before s. Alternatively, if only $k - 1 = 5$ events occurred after s, then S_6 must've occurred before. See the image below.



$k - 3$ events occurred after s in the image below, therefore the kth event must've occurred before s.

We can rewrite our target probability as follows, first define Z_{n-1} as a discrete random variable that measures the number of speciation events that occur after s. This implies

$$Z_{n-1} \sim \text{Binomial}(n - 1, 1 - F(s|t))$$

There are $n - 1$ total speciation events, so $n - 1$ total trials. We are testing whether the events occurred after s, which is just $\mathbb{P}(S > s) = 1 - F(s|t)$, see Equations 4.3 and 4.5. Next we can write this formula using the Binomial distribution.

$$\begin{aligned}
\mathbb{P}(Z_{n-1} = i) &= \binom{n-1}{i} F(s|t)^{n-1-i} (1 - F(s|t))^i \\
\implies \mathbb{P}(Z_{n-1} \leq k - 1) &= \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-1-i} (1 - F(s|t))^i \\
\implies F_{\mathscr{A}_{n,t}^k} &= \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-1-i} (1 - F(s|t))^i
\end{aligned}$$

$\square$

Checking the Result

We can check that our $F_{\mathscr{A}^k_{n,t}}(s)$ is correct by taking its derivative and seeing if its equal to $f_{\mathscr{A}^k_{n,t}}(s)$.

$$\frac{d}{ds}F_{\mathscr{A}^k_{n,t}} = \frac{d}{ds}\sum_{i=0}^{k-1}\binom{n-1}{i}F(s|t)^{n-i-1}(1-F(s|t))^i$$

$$= \sum_{i=0}^{k-1}\binom{n-1}{i}\frac{d}{ds}F(s|t)^{n-i-1}(1-F(s|t))^i$$

Then applying the derivative inside the sum

$$= \sum_{i=0}^{k-1}\binom{n-1}{i}\left[(n-i-1)F(s|t)^{n-i-2}F'(s|t)(1-F(s|t))^i - F(s|t)^{n-i-1}i(1-F(s|t))^{i-1}F'(s|t)\right]$$

Because $F'(s|t) = f(s|t)$ we get the following

$$= \sum_{i=0}^{k-1}\binom{n-1}{i}\left[(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^i - F(s|t)^{n-i-1}i(1-F(s|t))^{i-1}\right]f(s|t)$$

Then we can expand to get

$$= f(s|t)\sum_{i=0}^{k-1}\binom{n-1}{i}(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^i$$
$$- \binom{n-1}{i}iF(s|t)^{n-i-1}(1-F(s|t))^{i-1}$$

Expanding the factorial and cancelling terms gives

$$= f(s|t)\sum_{i=0}^{k-1}\binom{n-1}{i}(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^i$$
$$- \frac{(n-1)!i}{i!(n-i-1)!}F(s|t)^{n-i-1}(1-F(s|t))^{i-1}$$

Then multiplying by $\frac{n-i}{n-i} = 1$ we get

$$= f(s|t)\sum_{i=0}^{k-1}\binom{n-1}{i}(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^i$$
$$- \frac{(n-1)!}{(i-1)!(n-i-1)!}\cdot\frac{n-i}{n-i}F(s|t)^{n-i-1}(1-F(s|t))^{i-1}$$

$$= f(s|t)\sum_{i=0}^{k-1}\binom{n-1}{i}(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^{i}$$
$$-\frac{(n-1)!}{(i-1)!(n-i)!}(n-i)F(s|t)^{n-i-1}(1-F(s|t))^{i-1}$$
$$= f(s|t)\sum_{i=0}^{k-1}\binom{n-1}{i}(n-i-1)F(s|t)^{n-i-2}(1-F(s|t))^{i}$$
$$-\binom{n-1}{i-1}(n-i)F(s|t)^{n-i-1}(1-F(s|t))^{i-1}$$

If you examine the sum closely, it is evident it is a telescopic sum

The previous LHS term is subtracted by the following RHS term, which are equal

For example take $i = 5$ and 6

$$\implies \ldots\binom{n-1}{5}(n-5-1)F(s|t)^{n-5-2}(1-F(s|t))^{5} - \ldots - \binom{n-1}{6-1}(n-6)F(s|t)^{n-6-1}(1-F(s|t))^{6-1}$$
$$\iff \ldots\binom{n-1}{5}(n-6)F(s|t)^{n-7}(1-F(s|t))^{5} - \ldots - \binom{n-1}{5}(n-6)F(s|t)^{n-7}(1-F(s|t))^{5}$$
$$\iff \ldots 0 \ldots \quad \text{indicating its a telescopic sum}$$

$\implies$ Now the LHS term previous is cancelled by the RHS term following

So we take first and last sum entries, $i = 0$ and $i = k-1$

But the first LHS term is cancelled and the last RHS term is cancelled

$$\frac{d}{ds}F_{\mathscr{A}^k_{n,t}} = f(s|t)\left(-\binom{n-1}{0-1}(n-0)F(s|t)^{n-0-1}(1-F(s|t))^{0-1}\right)$$
$$+ f(s|t)\binom{n-1}{k-1}(n-(k-1)-1)F(s|t)^{n-(k-1)-2}(1-F(s|t))^{k-1}$$
$$= -f(s|t)\binom{n-1}{-1}nF(s|t)^{n-1}(1-F(s|t))^{-1}$$
$$+ f(s|t)\binom{n-1}{k-1}(n-k+1-1)F(s|t)^{n-k+1-2}(1-F(s|t))^{k-1}$$
$$= -f(s|t)\binom{n-1}{-1}nF(s|t)^{n-1}(1-F(s|t))^{-1}$$
$$+ f(s|t)\binom{n-1}{k-1}(n-k)F(s|t)^{n-k-1}(1-F(s|t))^{k-1}$$

We can't choose -1 objects from n-1 objects, so we set: $\binom{n-1}{-1} = 0$

$$\frac{d}{ds}F_{\mathscr{A}^k_{n,t}}(s) = \binom{n-1}{k-1}(n-k)F(s|t)^{n-k-1}(1-F(s|t))^{k-1}f(s|t)$$
$$= f_{\mathscr{A}^k_{n,t}}(s) \quad \text{as desired, see Equation: 6.1}$$

We have proved that $\frac{d}{ds}F_{\mathscr{A}_{n,t}^k}(s) = f_{\mathscr{A}_{n,t}^k}(s)$. Note this result is consistent with Theorem 9.17 in *Einführung in die Wahrscheinlichkeitstheorie und Statistik* [4], and completes proof.

6.2 Unknown Origin Time, t_{or}

Previously, we conditioned that t_{or} is known and equal to some t. Now we can suppose t_{or} is random and unknown, in this instance we can calculate $f_{\mathscr{A}_n^k}(s)$. Using the the law of total probability: $\mathbb{P}(A|C) = \sum_i \mathbb{P}(A|B_i, C)\mathbb{P}(B_i|C)$. The theorems for $f_{\mathscr{A}_n^t}(s)$ are derived. They are also present in Gernhards *The Conditioned Reconstructed Process* [5].

In section we don't find the cumulative distribution function because it is not useful for any further calculations.

6.2.1 General λ and μ

The following theorem for General λ and μ, was provided by Gernhard [5] in Theorem 4.1, with a minimal proof. We provide a complete proof below.

Theorem 6.3 (+). *The probability density function for the kth speciation time for a General λ and μ given an unknown origin time, t_{or}, that we assume is $U(0, \infty)$, is:*

$$f_{\mathscr{A}_n^k}(s) = \begin{cases} (k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}} & \text{if } s \leq t \\ 0 & \text{otherwise} \end{cases} \tag{6.3}$$

(See Gernhard [5] Theorem 4.1)

Proof.

$f_{\mathscr{A}_n^k}(s) = \mathbb{P}(\mathscr{A}^k = s|n)$

Then using the law of total probability, noting that $s \leq t_{or} \leq \infty$

$$= \int_s^\infty \mathbb{P}(\mathscr{A}^k = s|t_{or} = t, n)\mathbb{P}(t_{or} = t|n)dt$$

$$= \int_s^\infty f_{\mathscr{A}_{n,t}^k}(s)q_{or}(t|n)dt$$

Then using the equation for origin time from Equation 5.1 and $f_{\mathscr{A}_{n,t}^k}(s)$, from Equation 6.1

$$f_{\mathscr{A}_n^k}(s) = \int_s^\infty \binom{n-1}{n-k}(n-k)F(s|t)^{n-k-1}(1-F(s|t))^{k-1}f(s|t)$$

$$\times\, n\ \lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}dt$$

Then using equations for $F(s|t)$ and $f(s|t)$ under General λ, μ (see Equations: 4.1 and 4.3)

$$f_{\mathscr{A}_n^k}(s) = \int_s^\infty \binom{n-1}{n-k}(n-k)\left\{\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\right\}^{n-k-1}$$
$$\times\left\{1-\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\right\}^{k-1}$$
$$\times\left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)$$
$$\times\left(n\ \lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}\right)dt$$

Next we can simplify by expanding where able

$$f_{\mathscr{A}_n^k}(s) = \frac{(n-1)!(n-k)}{(n-k)!(n-1-n+k)!}\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)^{n-k-1}\left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)n\ \lambda^n(\lambda-\mu)^2$$
$$\times\int_s^\infty\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-k-1}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\left(\frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$
$$\times\left\{\frac{(\lambda-\mu e^{-(\lambda-\mu)s})(1-e^{-(\lambda-\mu)t})-(1-e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)s})(1-e^{-(\lambda-\mu)t})}\right\}^{k-1}dt$$

Then by multiplying the factorial by 1 and simplifying
the terms inside curly brackets we get

$$f_{\mathscr{A}_n^k}(s) = n\frac{(n-1)!}{(n-k-1)!(k-1)!}\cdot\frac{k(k+1)}{k(k+1)}\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)^{n-k-1}\left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)$$
$$\times\lambda^n(\lambda-\mu)^2\int_s^\infty\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-k}\left(\frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$
$$\times\left\{\frac{\lambda-\lambda e^{-(\lambda-\mu)t}-\mu e^{-(\lambda-\mu)s}+\mu e^{-(\lambda-\mu)(s+t)}}{(\lambda-\mu e^{-(\lambda-\mu)s})(1-e^{-(\lambda-\mu)t})}\right.$$
$$\left.\frac{-\lambda+\mu e^{-(\lambda-\mu)t}+\lambda e^{-(\lambda-\mu)s}-\mu e^{-(\lambda-\mu)(s+t)}}{(\lambda-\mu e^{-(\lambda-\mu)s})(1-e^{-(\lambda-\mu)t})}\right\}^{k-1}dt$$

We can then remove subtracted terms inside the curly brackets and isolate λ and μ

$$f_{\mathscr{A}_n^k}(s) = k(k+1)\frac{n!}{(n-k-1)!(k+1)!}\lambda^n(\lambda-\mu)^4\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)^{n-k-1}\left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)$$

$$\times \int_s^\infty \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}}\right)^{n-k} \left(\frac{(1 - e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$

$$\times \left\{\frac{\lambda(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t}) + \mu(e^{-(\lambda-\mu)t} - e^{-(\lambda-\mu)s})}{(\lambda - \mu e^{-(\lambda-\mu)s})(1 - e^{-(\lambda-\mu)t})}\right\}^{k-1} dt$$

Then turn early fractions into a binomial coefficient

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^4 \left(\frac{1 - e^{-(\lambda-\mu)s}}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^{n-k-1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})^2}\right)$$

$$\times \int_s^\infty \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}}\right)^{n-k} \left(\frac{(1 - e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$

$$\times \left\{\frac{\lambda(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t}) - \mu(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})}{(\lambda - \mu e^{-(\lambda-\mu)s})(1 - e^{-(\lambda-\mu)t})}\right\}^{k-1} dt$$

Next isolate common terms in curly brackets and extract non-t terms from the integral

$$f_{\mathscr{A}_n^k}(s) = k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^4 \left(\frac{1 - e^{-(\lambda-\mu)s}}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^{n-k-1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})^2}\right)$$

$$\times \int_s^\infty \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}}\right)^{n-k} \left(\frac{(1 - e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$

$$\times \left\{\frac{(\lambda-\mu)(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})}{(\lambda - \mu e^{-(\lambda-\mu)s})(1 - e^{-(\lambda-\mu)t})}\right\}^{k-1} dt$$

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^4(\lambda-\mu)^{k-1}(1 - e^{-(\lambda-\mu)s})^{n-k-1} e^{-(\lambda-\mu)s}$$

$$\times \left(\frac{1}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^{n-k-1} \left(\frac{1}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^2 \left(\frac{1}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^{k-1}$$

$$\times \int_s^\infty \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}}\right)^{n-k} \left(\frac{(1 - e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{n+1}}\right)$$

$$\times \left\{\frac{(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})}{(1 - e^{-(\lambda-\mu)t})}\right\}^{k-1} dt$$

Next multiply out terms in the integral

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3} e^{-(\lambda-\mu)s}(1 - e^{-(\lambda-\mu)s})^{n-k-1}$$

$$\times \left(\frac{1}{(\lambda - \mu e^{-(\lambda-\mu)s})}\right)^{n-k-1+k-1+2}$$

$$\times \int_s^\infty \frac{(\lambda - \mu e^{-(\lambda-\mu)t})^{n-k}(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})^{k-1}}{(1-e^{-(\lambda-\mu)t})^{n-k}(\lambda - \mu e^{-(\lambda-\mu)t})^{n+1}(1-e^{-(\lambda-\mu)t})^{k-1}} dt$$

Then we move the denominator to like terms inside the numerator inside the integral

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$
$$\times \int_s^\infty (\lambda - \mu e^{-(\lambda-\mu)t})^{n-k-n-1}(1-e^{-(\lambda-\mu)t})^{n-1-n+k-k+1}$$
$$\times e^{-(\lambda-\mu)t}(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})^{k-1} dt$$
$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$
$$\times \int_s^\infty (\lambda - \mu e^{-(\lambda-\mu)t})^{-(k+1)}(1-e^{-(\lambda-\mu)t})^{0}e^{-(\lambda-\mu)t}(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})^{k-1} dt$$

Then multiplying the integral by $\left(\frac{e^{(\lambda-\mu)s}}{e^{(\lambda-\mu)s}}\right)^{k-1} = 1$ for simplication

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$
$$\times \int_s^\infty \frac{e^{-(\lambda-\mu)t}(e^{-(\lambda-\mu)s} - e^{-(\lambda-\mu)t})^{k-1}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{k+1}} \times \left(\frac{e^{(\lambda-\mu)s}}{e^{(\lambda-\mu)s}}\right)^{k-1} dt$$
$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$
$$\times \int_s^\infty \frac{e^{-(\lambda-\mu)t}(e^{-(\lambda-\mu)s}e^{(\lambda-\mu)s} - e^{-(\lambda-\mu)t}e^{(\lambda-\mu)s})^{k-1}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{k+1}} \frac{1}{(e^{(\lambda-\mu)s})^{k-1}} dt$$

Next we remove non-t terms from the integral to get

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n} e^{-(\lambda-\mu)s(k-1))}$$
$$\times \int_s^\infty \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{k+1}} dt$$

Then we combine like terms to get the folllwing

$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)(k-1+1)s}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$
$$\times \int_s^\infty \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda - \mu e^{-(\lambda-\mu)t})^{k+1}} dt$$
$$= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)ks}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n}$$

$$\times \int_s^\infty \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}}dt$$

Then we use Proposition 6.3.1 and 6.3.2 to solve the integral to get

$$\begin{aligned} f_{\mathscr{A}_n^k}(s) &= k(k+1)\binom{n}{k+1}\lambda^n(\lambda-\mu)^{k+3}\frac{e^{-(\lambda-\mu)ks}(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^n} \times \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\lambda^{-k} \\ &= (k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}} \end{aligned}$$

as desired

□

It remains to prove the following propositions, these are important and we're not proved rigorously in Gernhard [5], so we proved them below.

Proposition 6.3.1. *The integral has the following anti-derivative*

$$\int \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}}dt = \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left(\frac{1-e^{-(\lambda-\mu)(t-s)}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^k$$

Proof. We solve this proof by differentiating the result and checking to see if we get the integral back

$$\frac{d}{dt}\frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left(\frac{1-e^{-(\lambda-\mu)(t-s)}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^k$$

First differentiate using the chain and quotient rule

$$\begin{aligned} &= \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}k\left(\frac{1-e^{(\lambda-\mu)(t-s)}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^{k-1} \\ &\times \frac{(\lambda-\mu)e^{-(\lambda-\mu)(t-s)}(\lambda-\mu e^{-(\lambda-\mu)t})-(1-e^{-(\lambda-\mu)(t-s)})\mu(\lambda-\mu)e^{-(\lambda-\mu)t})}{(\lambda-\mu e^{-(\lambda-\mu)t})^2} \end{aligned}$$

Then we simplify accordingly to get the desired result

$$\begin{aligned} &= \left(\frac{e^{-(\lambda-\mu)s}k(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k-1}(\lambda-\mu e^{-(\lambda-\mu)t})^2}\right) \\ &\times \left\{(\lambda-\mu)(e^{-(\lambda-\mu)(t-s)}(\lambda-\mu e^{-(\lambda-\mu)t})-\mu(1-e^{-(\lambda-\mu)(t-s)})e^{-(\lambda-\mu)t})\right\} \end{aligned}$$

Next expand the curly brackets to get

$$= \frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}}$$

$$\times \left\{\lambda e^{-(\lambda-\mu)(t-s)} - \mu e^{-(\lambda-\mu)(2t-s)} - \mu e^{-(\lambda-\mu)t} + \mu e^{-(\lambda-\mu)(2t-s)}\right\}$$

Then remove subtracted terms

$$= \frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} \cdot (\lambda e^{-(\lambda-\mu)(t-s)} - \mu e^{-(\lambda-\mu)t})$$

$$= \frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} \cdot (\lambda e^{-(\lambda-\mu)t}e^{(\lambda-\mu)s} - \mu e^{-(\lambda-\mu)t})$$

Next multiply by $(\lambda e^{-(\lambda-\mu)t}e^{(\lambda-\mu)s} - \mu e^{-(\lambda-\mu)t}) = 1$ for simplification purposes

$$= \frac{e^{-(\lambda-\mu)s}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} \cdot (e^{-(\lambda-\mu)t})(\lambda e^{(\lambda-\mu)s} - \mu) \times \frac{e^{-(\lambda-\mu)s}}{e^{-(\lambda-\mu)s}}$$

Then simplify more to get the desired result

$$= \frac{e^{-(\lambda-\mu)s}e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}(\lambda-\mu e^{-(\lambda-\mu)s})}{e^{-(\lambda-\mu)s}(\lambda-\mu e^{-(\lambda-\mu)s})(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}}$$

$$= \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} \quad \text{as desired}$$

□

Proposition 6.3.2. *Given Proposition 6.3.1 the following definite integral solves to:*

$$\int_s^{\infty} \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} dt = \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left[\left(\frac{1}{\lambda}\right)^k\right]$$

Proof. Here we solve the definite integral using the anti-deriative use found in Proposition 6.3.1

$$\int_s^{\infty} \frac{e^{-(\lambda-\mu)t}(1-e^{-(\lambda-\mu)(t-s)})^{k-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{k+1}} dt = \left[\left(\frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\right)\left(\frac{1-e^{-(\lambda-\mu)(t-s)}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^k\right]_0^{\infty}$$

$$= \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left[\left(\frac{1-e^{-(\lambda-\mu)(t-s)}}{\lambda-\mu e^{-(\lambda-\mu)t}}\right)^k\right]_0^{\infty}$$

$$= \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left[\left(\frac{1-e^{-(\lambda-\mu)(\infty-s)}}{\lambda-\mu e^{-(\lambda-\mu)\infty}}\right)^k - \left(\frac{1-e^{-(\lambda-\mu)(s-s)}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)^k\right]$$

Then given $0 < s < t < \infty$ noting that a speciation time cannot be infinity

$$= \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left[\left(\frac{1}{\lambda}\right)^k - \left(\frac{1-1}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)^k\right]$$

$$= \frac{e^{-(\lambda-\mu)s}}{k(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)s})}\left[\left(\frac{1}{\lambda}\right)^k\right]$$

□

6.2.2 Special Case: Yule ($\mu = 0$)

We add the following theorem because it is significant in theory and for later calculations, it is also exlcuded from Gernhard [5].

Theorem 6.4. *In the special case where $\mu = 0$, we have the following probability density function*

$$f_{\mathscr{A}_n^k}(s) = \begin{cases} (k+1)\binom{n}{k+1}\lambda\frac{(e^{\lambda s}-1)^{n-k-1}}{e^{\lambda sn}} & \text{if } s \le t \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

Proof. Letting $\mu = 0$, using the result from Theorem 6.3 for $f_{\mathscr{A}_n^k}(s)$ and some mathematical simplification we get the following result

$$\begin{aligned}
f_{\mathscr{A}_n^k}(s) =&(k+1)\binom{n}{k+1}\lambda^{n-k}\lambda^{k+2}e^{-\lambda(k+1)s}\frac{(1-e^{-\lambda s})^{n-k-1}}{\lambda^{n+1}} \\
=&(k+1)\binom{n}{k+1}\frac{\lambda^{n+2}}{\lambda^{n+1}}e^{-\lambda(k+1)s}(1-e^{-\lambda s})^{n-k-1}\times\frac{e^{\lambda sn}}{e^{\lambda sn}} \\
=&(k+1)\binom{n}{k+1}\lambda\frac{(1-e^{-\lambda s})^{n-k-1}e^{-\lambda s(k+1)}(e^{\lambda s})^n}{e^{\lambda sn}} \\
=&(k+1)\binom{n}{k+1}\lambda\frac{(1-e^{-\lambda s})^n(e^{\lambda s})^n(1-e^{-\lambda s})^{-(k+1)}(e^{\lambda s})^{-(k+1)}}{e^{\lambda sn}} \\
=&(k+1)\binom{n}{k+1}\lambda\frac{((1-e^{-\lambda s})(e^{\lambda s}))^n((1-e^{-\lambda s})(e^{\lambda s}))^{-(k+1)}}{e^{\lambda sn}} \\
=&(k+1)\binom{n}{k+1}\lambda\frac{(e^{\lambda s}-1)^n(e^{\lambda s-1})^{-(k+1)}}{e^{\lambda sn}} \\
=&(k+1)\binom{n}{k+1}\lambda\frac{(e^{\lambda s}-1)^{n-k-1}}{e^{\lambda sn}}
\end{aligned}$$

$\square$

6.2.3 Special Case: $\mu \to \lambda$

This theorem is present in Gernhard [5] Remark 4.2, but is missing some details, so we add them in below for a more complete proof. In particular we compute the limit step by step.

Theorem 6.5 (+). *In the special case where $\mu = \lambda$, we have the following probability density function*

$$f_{\mathscr{A}_n^k}(s) = \begin{cases} (k+1)\binom{n}{k+1}\lambda^{n-k}\frac{s^{n-k-1}}{(1+\lambda s)^{n+1}} & \text{if } s \le t \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

(See Gernhard [5] Remark 4.2)

Proof. We need to determine $f_{\mathscr{A}_n^k}(s)$ as $\mu \to \lambda$. We can find this using limits.

$$\begin{aligned}\lim_{\mu\to\lambda} f_{\mathscr{A}_n^k}(s) &= \lim_{\mu\to\lambda}(k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}} \\ &= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{k+2}e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}}\end{aligned}$$

First using the identity as $\epsilon \to 0 \implies e^{-\epsilon} \sim 1 - \epsilon + o(\epsilon)$ where $0(\epsilon)$ is of smaller order than ϵ

Meaning $\dfrac{o(\epsilon)}{\epsilon} \to 0$ as $\epsilon \downarrow 0$

This implies as $e^{-(\lambda-\mu)s} \to 1 \implies e^{-(\lambda-\mu)s} \sim 1-(\lambda-\mu)s$

We input this result below, then expand and simplify accordingly

$$\begin{aligned} &= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{k+2}(1-(\lambda-\mu)s)^{k+1}\frac{(1-(1-(\lambda-\mu)s))^{n-k-1}}{(\lambda-\mu(1-(\lambda-\mu)s))^{n+1}} \\ &= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{k+2}(1-\lambda s+\mu s)^{k+1}\frac{(\lambda s-\mu s)^{n-k-1}}{(\lambda-\mu(1-\lambda s+\mu s))^{n+1}} \\ &= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{k+2}(1-\lambda s+\mu s)^{k+1}\frac{(\lambda-\mu)^{n-k-1}s^{n-k-1}}{(\lambda-\mu+\mu\lambda s-\mu^2 s)^{n+1}}\end{aligned}$$

Here we isolate common terms in the fraction to get

$$\begin{aligned} &= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{k+2+n-k-1}(1-\lambda s+\mu s)^{k+1} \\ &\quad\times\frac{s^{n-k-1}}{(\lambda(1+\mu s)-\mu(1+\mu s))^{n+1}}\end{aligned}$$

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}(\lambda-\mu)^{n+1}(1-\lambda s+\mu s)^{k+1}\frac{s^{n-k-1}}{(\lambda-\mu)^{n+1}(1+\mu s)^{n+1}}$$

Then we get common terms on the numerator and denominator which we divide out

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}\frac{(\lambda-\mu)^{n+1}}{(\lambda-\mu)^{n+1}}\frac{s^{n-k-1}(1-\lambda s+\mu s)^{k+1}}{(1+\mu s)^{n+1}}$$

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\lim_{\mu\to\lambda}\frac{s^{n-k-1}(1-\lambda s+\mu s)^{k+1}}{(1+\mu s)^{n+1}}$$

Then evaluating the limit gives

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\frac{s^{n-k-1}(1-\lambda s+\lambda s)^{k+1}}{(1+\lambda s)^{n+1}}$$

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\frac{s^{n-k-1}}{(1+\lambda s)^{n+1}}$$

□

Chapter 7

Expected Speciation Times

In the previous sections we calculated the density function for the speciation times and the kth speciation time. The natural next step is to determine their expectation.

7.1 Tree origin is known, t_{or} is fixed

In the following theorems and proofs we assume that tree origin is known, i.e. t_{or} is fixed.

7.1.1 Solution for $0 < \mu < \lambda$ (General Case)

The following theorem is identified in Gernhard [5] Theorem 5.1, there is a sketch proof provided, below I provide the complete proof with extra details.

Theorem 7.1 (+)**.** *The expectation for the kth speciation time,* $\mathbb{E}\left(\mathscr{A}_{n,t}^k\right)$ *where* $0 < \mu < \lambda$ *is*

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}}\right)^{n-j-1}$$
$$\times\left[g(j) + \sum_{l=1}^{n-j-1}\sum_{m=0}^{l-1}\binom{n-j-1}{l}\binom{l-1}{m}(-1)^m \times \frac{\lambda^{l-1-m}}{(\lambda-\mu)\mu^l}h(j,m)\right] \tag{7.1}$$

Where

$$g(j) = \frac{1}{(\lambda-\mu)\lambda^{n-j-1}}\left[ln\left(\frac{\lambda e^{(\lambda-\mu)t} - \mu}{\lambda-\mu}\right) - \sum_{m=1}^{n-j-2}\binom{n-j-2}{m}\right.$$
$$\left.\times\frac{\mu^m}{m}(\lambda e^{(\lambda-\mu)t} - \mu^{-m}) - (\lambda-\mu)^{-m}\right] \tag{7.2}$$

$$and\ h(j,m) = \begin{cases} ln\left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{\lambda-\mu}\right) & if\ m+j+1-n=-1 \\ \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{m+j+2-n} - (\lambda-\mu)^{m+j+2-n}}{m+j+2-n} & else \end{cases} \tag{7.3}$$

(See Gernhard [5] Theorem 5.1)

Proof. We determine the expectation as follows for $0 < \mu < \lambda$:

$$\begin{aligned}
\mathbb{E}(\mathscr{A}^k_{n,t}) &= \int_0^t s f_{\mathscr{A}^k_{n,t}}(s) ds \\
&= \int_0^t \left(s f_{\mathscr{A}^k_{n,t}}(s) + F_{\mathscr{A}^k_{n,t}}(s) - F_{\mathscr{A}^k_{n,t}}(s) \right) ds \quad \text{This is true by definition} \\
&= \int_0^t \left(s f_{\mathscr{A}^k_{n,t}}(s) + F_{\mathscr{A}^k_{n,t}}(s) \right) ds - \int_0^t F_{\mathscr{A}^k_{n,t}}(s) ds \\
&= \int_0^t \left(s F'_{\mathscr{A}^k_{n,t}}(s) + F_{\mathscr{A}^k_{n,t}}(s) \right) - \int_0^t F_{\mathscr{A}^k_{n,t}}(s) ds
\end{aligned}$$

As: $\dfrac{d}{dx}(xF(x)) = xF'(x) + F(x)$ We have the following

$$\begin{aligned}
&= \left[s F_{\mathscr{A}^k_{n,t}}(s) \right]_0^t - \int_0^t F_{\mathscr{A}^k_{n,t}}(s) ds \\
&= \left[t F_{\mathscr{A}^k_{n,t}}(t) \right] - \int_0^t F_{\mathscr{A}^k_{n,t}}(s) ds
\end{aligned}$$

Then $F_{\mathscr{A}^k_{n,t}}(s)$ must be 1 because $\mathscr{A}^k_{n,t}$ takes values $\in (0,t)$

$$\begin{aligned}
\mathbb{E}(\mathscr{A}^k_{n,t}) &= t(1) - \int_0^t F_{\mathscr{A}^k_{n,t}}(s) ds \\
&= t - \int_0^t \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-i-1} (1 - F(s|t))^i ds
\end{aligned}$$

Then we flip the sign in by multiplying by $(-1)^i$

$$= t - \int_0^t \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-i-1} (F(s|t) - 1)^i (-1)^i ds$$

Then by using the binomial theorem for: $(F(s|t)-1)^i = \sum_{j=0}^{i} \binom{i}{j} F(s|t)^{n-j} (-1)^j$

$$= t - \int_0^t \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-i-1} \sum_{j=0}^{i} \binom{i}{j} F(s|t)^{i-j} (-1)^j (-1)^i ds$$

Then we can combine like terms

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t F(s|t)^{n-i-1+i-j}ds$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t F(s|t)^{n-j-1}ds$$

Then using Equation: 4.3 for $F(s|t)$ gives

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t \left\{\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\right\}^{n-j-1} ds$$

Then we can remove non-t terms from the integration and simplify

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-j-1}\int_0^t \left(\frac{(1-e^{-(\lambda-\mu)s})^{n-j-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) ds$$

Below we multiply by (-1) to flip terms inside the numerator

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-j-1}$$

$$\times\int_0^t \left(\frac{(-1+e^{-(\lambda-\mu)s})^{n-j-1}(-1)^{n-j-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) ds$$

Then using the binomial theorem for: $(e^{-(\lambda-\mu)s}-1)^{n-j-1} = \sum_{l=0}^{n-j-1}\binom{n-j-1}{l}e^{-(\lambda-\mu)s^l}(-1)^{n-j-1-l}$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-j-1}$$

$$\times\int_0^t \sum_{l=0}^{n-j-1}\binom{n-j-1}{l}\left(\frac{e^{-(\lambda-\mu)s^l}(-1)^{n-j-1-l}(-1)^{n-j-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) ds$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\sum_{l=0}^{n-j-1}\binom{n-1}{i}\binom{i}{j}\binom{n-j-1}{l}(-1)^{i+j+l}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-j-1}$$

$$\times\int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) ds$$

Then given we are summing over some $l \geq 0$, we can split the sum for $l = 0$ and $l > 0$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j+l}\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)^{n-j-1}$$

$$\times \left[\sum_{l=0}^{n-j-1} \binom{n-j-1}{l} \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds \right]$$

Below we split this sum into two parts

$$= t - \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{n-1}{i} \binom{i}{j} (-1)^{i+j+l} \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}} \right)^{n-j-1}$$

$$\times \left[\sum_{l=0}^{n-j-1} \binom{n-j-1}{0} \int_0^t \left(\frac{e^{-(\lambda-\mu)0s}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds \right.$$

$$\left. + \sum_{l=0}^{n-j-1} \binom{n-j-1}{l} \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds \right]$$

$$= t - \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{n-1}{i} \binom{i}{j} (-1)^{i+j+l} \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}} \right)^{n-j-1}$$

$$\times \sum_{l=0}^{n-j-1} \int_0^t \left(\frac{1}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds$$

$$\left[+ \sum_{l=1}^{n-j-1} \binom{n-j-1}{l} \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds \right]$$

Then using Proposition's 7.1.2 and 7.1.3 to solve the two integrals means we get

$$\implies \mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{n-1}{i} \binom{i}{j} (-1)^{i+j} \left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{1 - e^{-(\lambda-\mu)t}} \right)^{n-j-1}$$

$$\times \left[g(j) + \sum_{l=1}^{n-j-1} \binom{n-j-1}{l} (-1)^l \frac{1}{(\lambda-\mu)\mu^l} \sum_{m=0}^{l-1} \binom{l-1}{m} (-1)^{l+m} \lambda^{l-1-m} h(j,m) \right]$$

Where

$$g(j) = \frac{1}{(\lambda-\mu)\lambda^{n-j-1}} \left[\ln\left(\frac{\lambda e^{(\lambda-\mu)t} - \mu}{\lambda 0 \mu} \right) - \sum_{m=1}^{n-j-2} \binom{n-j-2}{m} \right.$$

$$\left. \times \frac{\mu^m}{m} (\lambda e^{(\lambda-\mu)t} - \mu^{-m}) - (\lambda-\mu)^{-m} \right]$$

$$\text{and } h(j,m) = \begin{cases} \ln\left(\frac{\lambda - \mu e^{-(\lambda-\mu)t}}{\lambda-\mu} \right) & \text{if } m+j+1-n = -1 \\ \frac{(\lambda - \mu e^{-(\lambda-\mu)t})^{m+j+2-n} - (\lambda-\mu)^{m+j+2-n}}{m+j+2-n} & \text{else} \end{cases}$$

This completes the proof

□

The natural next step is to prove the cited propositions in the above proof

Proposition 7.1.1. *The integral solves to the following*

$$h(j,m) = \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx$$
$$= \begin{cases} ln\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\lambda-\mu}\right) & \textit{if } m+j+1-n=-1 \\ \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{m+j+2-n}-(\lambda-\mu)^{m+j+2-n}}{m+j+2-n} & \textit{else} \end{cases}$$

Proof.

We need to solve the integral: $\displaystyle\int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx$

First assume $m+j+1-n=-1$

$$\implies \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx = \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \frac{1}{x} dx$$
$$= [\ln(x)]_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}}$$
$$= \ln(\lambda-\mu e^{-(\lambda-\mu)t}) - \ln(\lambda-\mu)$$
$$= \ln\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\lambda-\mu}\right)$$

Now assuming $m+j+1-n \neq 1$

$$\implies \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx = \left[\frac{x^{m+j+2-n}}{m+j+2-n}\right]_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}}$$
$$= \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{m+j+2-n} - (\lambda-\mu)^{m+j+2-n}}{m+j+2-n}$$

$$\implies h(j,m) = \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx$$
$$= \begin{cases} \ln\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\lambda-\mu}\right) & \text{if } m+j+1-n=-1 \\ \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{m+j+2-n}-(\lambda-\mu)^{m+j+2-n}}{m+j+2-n} & \text{else} \end{cases}$$

□

Proposition 7.1.2. *Suppose $l > 0$, then the following integral simplifies to*

$$\int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds = \frac{1}{(\lambda-\mu)\mu^l} \sum_{m=0}^{l-1} \binom{l-1}{m} (-1)^m \lambda^{l-1-m} h(j,m)$$

$$Where\ h(j,m) = \begin{cases} ln\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{\lambda-\mu}\right) & if\ m+j+1-n=-1 \\ \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{m+j+2-n}-(\lambda-\mu)^{m+j+2-n}}{m+j+2-n} & else \end{cases}$$

Proof.

We can solve the integral through substitution, let $x = \lambda - \mu e^{-(\lambda-\mu)s}$

$$\begin{aligned} \text{Then this } &\implies \lambda - \mu e^{-(\lambda-\mu)s} = x \\ &\iff \frac{\lambda - x}{\mu} = e^{-(\lambda-\mu)s} \\ &\implies \left(\frac{\lambda - x}{\mu}\right)^l = e^{-(\lambda-\mu)ls} \end{aligned}$$

$$\begin{aligned} \text{We also get} \frac{dx}{ds} &= \mu(\lambda-\mu)e^{-(\lambda-\mu)s} \\ &\implies ds = \frac{dx}{\mu(\lambda-\mu)e^{-(\lambda-\mu)s}} \end{aligned}$$

Now that we've got our terms to subsitute, we can place them in, giving

$$\implies \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds$$

$$= \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \left(\frac{\lambda - x}{\mu}\right)^l \left(\frac{1}{x^{n-j-1}}\right) \frac{dx}{\mu(\lambda-\mu)e^{-(\lambda-\mu)s}}$$

The next steps are to simplify as much as possible

$$= \frac{1}{\mu(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \left(\frac{\lambda - x}{\mu}\right)^l \left(\frac{1}{x^{n-j-1}}\right) \frac{1}{e^{-(\lambda-\mu)s}} dx$$

$$= \frac{1}{\mu(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \left(\frac{\lambda - x}{\mu}\right)^l \left(\frac{1}{x^{n-j-1}}\right) \left(\frac{\lambda - x}{\mu}\right)^{-1} dx$$

$$= \frac{1}{\mu(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \left(\frac{\lambda - x}{\mu}\right)^{l-1} \left(\frac{1}{x^{n-j-1}}\right) dx$$

$$= \frac{1}{\mu(\lambda-\mu)\mu^{l-1}} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \left(\frac{(\lambda-x)^{l-1}}{x^{n-j-1}} \right) dx$$

Then since we've assumed that $l > 0 \implies (l-1) \geq 0$ we can use the binomial theorem for:

$$(\lambda-x)^{l-1} = \sum_{m=0}^{l-1} \binom{l-1}{n} \lambda^{l-1-m}(-x)^m \quad \text{Note that } l \in \mathbb{Z}$$

$$= \frac{1}{(\lambda-\mu)\mu^l} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \sum_{m=0}^{l-1} \binom{l-1}{m} \lambda^{l-1-m}(-x)^m \frac{1}{x^{n-j-1}} dx$$

$$= \frac{1}{(\lambda-\mu)\mu^l} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} \sum_{m=0}^{l-1} \binom{l-1}{m} \lambda^{l-1-m}(-1)^m x^m x^{j+1-n} dx$$

Now we can move terms from the integration that are not x

$$= \frac{1}{(\lambda-\mu)\mu^l} \sum_{m=0}^{l-1} \binom{l-1}{m} (-1)^m \lambda^{l-1-m} \int_{\lambda-\mu}^{\lambda-\mu e^{-(\lambda-\mu)t}} x^{m+j+1-n} dx$$

Then using Proposition 7.1.1 get the following result for $l > 0$

$$\implies \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds = \frac{1}{(\lambda-\mu)\mu^l} \sum_{m=0}^{l-1} \binom{l-1}{m} (-1)^m \lambda^{l-1-m} h(j,m)$$

□

Proposition 7.1.3. *Suppose $l = 0$, then the following integral simplifies to*

$$g(j) = \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds = \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \Bigg[ln \left(\frac{\lambda e^{(\lambda-\mu)t}-\mu}{\lambda-\mu} \right) - \sum_{m=1}^{n-j-2} \binom{n-j-2}{m} \frac{\mu^m}{m} ((\lambda e^{(\lambda-\mu)t}-\mu)^{-m} - (\lambda-\mu)^{-m}) \Bigg]$$

Proof.

Now we need to solve $g(j) = \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds$ for $l = 0$

Letting $l = 0 \implies \int_0^t \left(\frac{e^{-(\lambda-\mu)ls}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds = \int_0^t \left(\frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}} \right) ds$

We can solve through substitution, let $x = \lambda e^{(\lambda-\mu)s} - \mu$

Then we get $\dfrac{dx}{ds} = \lambda(\lambda-\mu)e^{(\lambda-\mu)s}$

$$\implies ds = \frac{dx}{\lambda(\lambda-\mu)e^{-(\lambda-\mu)s}}$$

Now we plug this in to our integral

$$\implies \int_0^t \left(\frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) ds$$

$$= \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \left(\frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1}}\right) \frac{dx}{\lambda(\lambda-\mu)e^{(\lambda-\mu)s}}$$

We multiply by 1 for simplication purposes

$$= \frac{1}{\lambda(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \left(\frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n-j-1} e^{-(\lambda-\mu)s}}\right) dx \times \left(\frac{e^{(\lambda-\mu)s}}{e^{(\lambda-\mu)s}}\right)^{n-j-1}$$

Cancellation leads to the following

$$= \frac{1}{\lambda(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \left(\frac{e^{(\lambda-\mu)(n-j-1)s}}{\{(\lambda-\mu e^{-(\lambda-\mu)s})(e^{(\lambda-\mu)s})\}^{n-j-1} e^{-(\lambda-\mu)s}}\right) dx$$

$$= \frac{1}{\lambda(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \left(\frac{e^{(\lambda-\mu)(n-j-2)s}}{(\lambda e^{(\lambda-\mu)s}-\mu)^{n-j-1}}\right) dx \quad \text{given } e^{-(\lambda-\mu)s}e^{(\lambda-\mu)s} = 1$$

And given $x = \lambda e^{(\lambda-\mu)s} - \mu$

$$\iff \frac{x+\mu}{\lambda} = e^{(\lambda-\mu)s}$$

$$\implies e^{(\lambda-\mu)(n-j-2)s} = \left(\frac{x+\mu}{\lambda}\right)^{n-j-2}$$

$$\implies g(j) = \frac{1}{\lambda(\lambda-\mu)} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \left(\frac{x+\mu}{\lambda}\right)^{n-j-2} \left(\frac{1}{x^{n-j-1}}\right) dx$$

$$= \frac{1}{\lambda(\lambda-\mu)\lambda^{n-j-2}} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} (x+\mu)^{n-j-2} \left(\frac{1}{x^{n-j-1}}\right) dx$$

Then we can apply the binomial theorem for: $(x+\mu)^{n-j-2} = \displaystyle\sum_{m=0}^{n-j-2} \binom{n-j-2}{m} x^{n-j-2-m}\mu^m$

$$\implies g(j) = \frac{1}{(\lambda-\mu)\lambda^{n-j-1}} \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \sum_{m=0}^{n-j-2} \binom{n-j-2}{m} x^{n-j-2-m}\mu^m \left(\frac{1}{x^{n-j-1}}\right) dx$$

$$= \frac{1}{(\lambda-\mu)\lambda^{n-j-1}} \sum_{m=0}^{n-j-2} \binom{n-j-2}{m} \mu^m \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{n-j-2-m-n+j+1} dx$$

$$= \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \sum_{m=0}^{n-j-2} \binom{n-j-2}{m} \mu^m \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx$$

Now we need to solve the integral $\int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx$

First if m = 0 $\implies \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx = \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} \frac{1}{x} dx$

$$= \ln(x)_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu}$$

$$= \ln(\lambda e^{(\lambda-\mu)t}-\mu) - \ln(\lambda-\mu)$$

$$= \ln\left(\frac{\lambda e^{(\lambda-\mu)t}-\mu}{\lambda-\mu}\right)$$

Then solving for $m > 0$

$$\implies \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx = \left[\frac{x^{-m}}{-m}\right]_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu}$$

$$= \frac{(\lambda e^{(\lambda-\mu)t}-\mu)^{-m}}{-m} - \frac{(\lambda-\mu)^{-m}}{-m}$$

Now we plug this back into $g(j)$

$$\implies g(j) = \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \sum_{m=0}^{n-j-2} \binom{n-j-2}{m} \mu^m \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx$$

$$= \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \left[\binom{n-j-2}{0} \mu^0 \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-0-1} dx\right.$$

$$\left.+ \sum_{m=1}^{n-j-2} \binom{n-j-2}{m} \mu^m \int_{\lambda-\mu}^{\lambda e^{(\lambda-\mu)t}-\mu} x^{-m-1} dx\right]$$

$$= \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \left[\ln\left(\frac{\lambda e^{(\lambda-\mu)t}-\mu}{\lambda-\mu}\right)\right.$$

$$\left.+ \sum_{m=1}^{n-j-2} \binom{n-j-2}{m} \mu^m \frac{(\lambda e^{(\lambda-\mu)t}-\mu)^{-m}}{-m} - \frac{(\lambda-\mu)^{-m}}{-m}\right]$$

Then solving the integral means we get the following

$$= \frac{1}{\lambda^{n-j-1}(\lambda-\mu)} \left[\ln\left(\frac{\lambda e^{(\lambda-\mu)t}-\mu}{\lambda-\mu}\right)\right.$$

$$+\sum_{m=1}^{n-j-2}\binom{n-j-2}{m}\frac{\mu^m}{-m}((\lambda e^{(\lambda-\mu)t}-\mu)^{-m}-(\lambda-\mu)^{-m})\Bigg]$$

Then some simplification gives us the desired result

$$\implies g(j)=\frac{1}{\lambda^{n-j-1}(\lambda-\mu)}\Bigg[\ln\left(\frac{\lambda e^{(\lambda-\mu)t}-\mu}{\lambda-\mu}\right)$$
$$-\sum_{m=1}^{n-j-2}\binom{n-j-2}{m}\frac{\mu^m}{m}((\lambda e^{(\lambda-\mu)t}-\mu)^{-m}-(\lambda-\mu)^{-m})\Bigg]$$

□

7.1.2 Yule Case: $\mu = 0$

The natural next step is to find the expectation for the k-th speciation time in the Yule case. We can use Gernhard's in *Stochastic properties of generalised Yule models* as a guide [7]. We add important details to her proof.

Theorem 7.2 (+)**.** *The expectation for the kth speciation time given a known origin time,* $\mathbb{E}\left(\mathscr{A}^k_{n,t}\right)$*, in the Yule case,* $\mu = 0$*:*

$$\mathbb{E}(\mathscr{A}^k_{n,t})=\frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}(1-e^{-\lambda t})^{1-n}}{\lambda(k+i-j)^2}$$
$$\times(1-e^{-\lambda t})^{1-n}\left\{e^{-\lambda jt}-((k+i-j)\lambda t+1)e^{-\lambda(k+i)t}\right\} \tag{7.4}$$

(See Gernhard [5] Theorem 5.1)

Proof.

$$\mathbb{E}(\mathscr{A}^k_{n,t})=\int_0^t sf_{\mathscr{A}^k_{n,t}}(s)ds$$

Then letting $\mu = 0$ we can use Equation: 6.1 for $f_{\mathscr{A}^k_{n,t}}(s)$

$$\implies \mathbb{E}(\mathscr{A}^k_{n,t})=\int_0^t s\binom{n-1}{n-k}(n-k)F(s|t)^{n-k-1}(1-F(s|t))^{k-1}f(s|t)ds$$
$$=\binom{n-1}{n-k}(n-k)\int_0^t sF(s|t)^{n-k-1}(-1)^{k-1}(F(s|t)-1)^{k-1}f(s|t)ds$$

Then plugging the values from Equations: 4.1 and 4.3 for $f(s|t)$ and $F(s|t)$, respectively

$$\implies \mathbb{E}(\mathscr{A}^k_{n,t}) = \int_0^t s\binom{n-1}{n-k}(n-k)\left\{\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\right\}^{n-k-1}$$
$$\times\left\{1-\left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)\right\}^{k-1}\left(\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right)ds$$

Then letting $\mu = 0$ because we are examining the Yule model

$$= \int_0^t s\binom{n-1}{n-k}(n-k)\left\{\left(\frac{1-e^{-\lambda s}}{\lambda}\right)\left(\frac{\lambda}{1-e^{-\lambda t}}\right)\right\}^{n-k-1}$$
$$\times\left\{1-\left(\frac{1-e^{-\lambda s}}{\lambda}\right)\left(\frac{\lambda}{1-e^{-\lambda t}}\right)\right\}^{k-1}\left(\frac{\lambda^2e^{-\lambda s}}{\lambda^2}\right)\left(\frac{\lambda}{1-e^{-\lambda t}}\right)ds$$

The next step is to expand powers where possible

$$= \int_0^t s\binom{n-1}{n-k}(n-k)\left(\frac{1-e^{-\lambda s}}{1-e^{-\lambda t}}\right)^{n-k-1}\left\{1-\left(\frac{1-e^{-\lambda s}}{1-e^{-\lambda t}}\right)\right\}^{k-1}\left(\frac{\lambda e^{-\lambda s}}{1-e^{-\lambda t}}\right)ds$$

Then we expand the term $\left\{1-\left(\frac{1-e^{-\lambda s}}{1-e^{-\lambda t}}\right)\right\}^{k-1}$ accordingly

$$= \binom{n-1}{n-k}(n-k)(1-e^{-\lambda t})^{(-1)n-k-1}(1-e^{-\lambda t})^{-1}$$
$$\times\int_0^t s(1-e^{-\lambda s})^{n-k-1}\left(\frac{(1-e^{-\lambda t})-(1-e^{-\lambda s})}{1-e^{-\lambda t}}\right)^{k-1}(\lambda e^{-\lambda s})ds$$

Then we remove the denominator and match terms with same value

$$= \binom{n-1}{n-k}(n-k)(1-e^{-\lambda t})^{1+k-n}(1-e^{-\lambda t})^{(-1)}(1-e^{-\lambda t})^{(-1)(k-1)}$$
$$\times\int_0^t s(1-e^{-\lambda s})^{1+k-n}(1-e^{-\lambda t}-1+e^{-\lambda s})^{k-1}(\lambda e^{-\lambda s})ds$$
$$= \binom{n-1}{n-k}(n-k)(1-e^{-\lambda t})^{1+k-n-1-k+1}$$
$$\times\int_0^t s(1-e^{-\lambda s})^{1+k-n}(1-e^{-\lambda t}-1+e^{-\lambda s})^{k-1}(\lambda e^{-\lambda s})ds$$
$$= \binom{n-1}{n-k}(n-k)(1-e^{-\lambda t})^{1-n}\int_0^t s\lambda e^{-\lambda s}(1-e^{-\lambda s})^{1-n}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$

Then we can simplify $(n-k)\binom{n-1}{n-k} = \frac{(n-k)(n-1)!}{(n-k)!(n-1-(n-k)))!}$

$$= \frac{(n-k)(n-1)!}{(n-k)!(k-1)!} = \frac{(n-1)!}{(n-k-1)!(k-1)!} = \frac{k(n-1)!}{k!(n-1-k)!} = k\binom{n-1}{k}$$

Adding in this simplification gives the following result

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = k\binom{n-1}{k}(1-e^{-\lambda t})^{n-k+1}\int_0^t s\lambda e^{-\lambda s}(1-e^{-\lambda s})^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$

Next we multiply by 1, to simplify the integral

$$= k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}$$
$$\times \int_0^t s\lambda e^{-\lambda s}(1-e^{-\lambda s})^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds \times \left(\frac{e^{\lambda s}}{e^{\lambda s}}\right)^{n-k-1}$$
$$= k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}\int_0^t s\lambda \frac{e^{-\lambda s}}{e^{\lambda(n-k-1)s}}(e^{\lambda s}-1)^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$
$$= k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}\int_0^t s\lambda e^{-\lambda s}e^{-\lambda(n-k-1)s}(e^{\lambda s}-1)^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$
$$= k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}\int_0^t s\lambda e^{-\lambda s(1+n-k-1)}(e^{\lambda s}-1)^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$
$$= k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}\int_0^t s\lambda e^{-\lambda(n-k)s}(e^{\lambda s}-1)^{n-k-1}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$

Then we can use binomial expansion around: $(e^{\lambda s}-1)^{n-k-1} = \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}e^{\lambda s^{n-k-1-i}}(-1)^i$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = k\binom{n-1}{k}(1-e^{-\lambda t})^{1-n}$$
$$\times \int_0^t s\lambda e^{-\lambda(n-k)s}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}e^{\lambda s^{n-k-1-i}}(-1)^i(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$

Next we match common terms

$$= k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}(1-e^{-\lambda t})^{1-n}\lambda(-1)^i$$
$$\times \int_0^t se^{-\lambda(n-k)s}e^{\lambda s^{n-k-1-i}}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$
$$= k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}(1-e^{-\lambda t})^{1-n}\lambda(-1)^i$$
$$\times \int_0^t se^{\lambda(n-k-1-i-n-k)s}(e^{-\lambda s}-e^{-\lambda t})^{k-1}ds$$
$$= k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}(1-e^{-\lambda t})^{1-n}\lambda(-1)^i$$

$$\times \int_0^t se^{-\lambda(i+1)s}(e^{-\lambda s} - e^{-\lambda t})^{k-1} ds$$

Then we can use binomial expansion around: $(e^{-\lambda s} - e^{-\lambda t})^{k-1} = \sum_{j=0}^{k-1} \binom{k-1}{j} e^{-\lambda s^{k-1-j}} (-e^{-\lambda t})^j$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} (1-e^{-\lambda t})^{1-n} \lambda(-1)^i e^{-\lambda(k-1)t}$$

$$\times \int_0^t se^{-\lambda(i+1)s} \sum_{j=0}^{k-1} \binom{k-1}{j} e^{-\lambda s^{k-1-j}} (-e^{-\lambda t})^j ds$$

Then we simplify accordingly, matching terms where we can

$$= k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} (1-e^{-\lambda t})^{1-n} \lambda(-1)^i e^{-\lambda(k-1)t}$$

$$\times \int_0^t se^{-\lambda(i+1)s} \sum_{j=0}^{k-1} \binom{k-1}{j} e^{-\lambda s^{k-1-j}} (-1)^j e^{-\lambda t^j} ds$$

$$= k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \sum_{j=0}^{k-1} \binom{n-k-1}{i} \binom{k-1}{j} \lambda(-1)^{i+j} (1-e^{-\lambda t})^{1-n} e^{-\lambda jt}$$

$$\times \int_0^t se^{-\lambda(i+1)s} e^{-\lambda(k-1-j)s} ds$$

Combine $e^{-\lambda}$ terms for an easier integration

$$= k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \sum_{j=0}^{k-1} \binom{n-k-1}{i} \binom{k-1}{j} \lambda(-1)^{i+j} (1-e^{-\lambda t})^{1-n} e^{-\lambda jt}$$

$$\times \int_0^t se^{-\lambda(i+1+k-1-j)s} ds$$

$$= k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \sum_{j=0}^{k-1} \binom{n-k-1}{i} \binom{k-1}{j} \lambda(-1)^{i+j} (1-e^{-\lambda t})^{1-n} e^{-\lambda jt}$$

$$\times \int_0^t se^{-\lambda(k+i-j)s} ds$$

Then using Proposition 7.2.1 to solve the integral gives

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = k\binom{n-1}{k} \sum_{i=0}^{n-k-1} \sum_{j=0}^{k-1} \binom{n-k-1}{i} \binom{k-1}{j} \lambda(-1)^{i+j} (1-e^{-\lambda t})^{1-n} e^{-\lambda jt}$$

$$\times \frac{1}{\lambda^2(k+i-j)^2} \left\{ 1 - e^{-\lambda(k+i-j)t} - t\lambda(k+i-j) e^{\lambda(k+i-j)t} \right\}$$

The final steps are to simplify, so we match Gernhard's [5] notation

$$= \frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}(1-e^{-\lambda t})^{1-n}}{\lambda(k+i-j)^2}$$
$$\times e^{-\lambda jt}\left\{1-e^{-\lambda(k+i-j)t}-t\lambda(k+i-j)e^{\lambda(k+i-j)t}\right\}$$
$$= \frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}}{\lambda(k+i-j)^2}$$
$$\times (1-e^{-\lambda t})^{1-n}\left\{e^{-\lambda jt}-e^{-\lambda(k+i-j)t}e^{-\lambda jt}-t\lambda(k+i-j)e^{\lambda(k+i-j)t}e^{-\lambda jt}\right\}$$
$$= \frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}(1-e^{-\lambda t})^{1-n}}{\lambda(k+i-j)^2}$$
$$\times (1-e^{-\lambda t})^{1-n}\left\{e^{-\lambda jt}-e^{-\lambda(k+i)t}-t\lambda(k+i-j)e^{\lambda(k+i)t}\right\}$$
$$= \frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}(1-e^{-\lambda t})^{1-n}}{\lambda(k+i-j)^2}$$
$$\times (1-e^{-\lambda t})^{1-n}\left\{e^{-\lambda jt}-(1+t\lambda(k+i-j))e^{-\lambda(k+i)t}\right\}$$
$$= \frac{k\binom{n-1}{k}\sum_{i=0}^{n-k-1}\sum_{j=0}^{k-1}\binom{n-k-1}{i}\binom{k-1}{j}(-1)^{i+j}(1-e^{-\lambda t})^{1-n}}{\lambda(k+i-j)^2}$$
$$\times (1-e^{-\lambda t})^{1-n}\left\{e^{-\lambda jt}-((k+i-j)\lambda t+1)e^{-\lambda(k+i)t}\right\}$$

This result completes the proof for the expectation of the k-th speciation time in the Yule case. Remaining propositions are solved below. □

Proposition 7.2.1. *The solution to the definite integral is as follows*

$$\int_0^t se^{-\lambda(k+i-j)s}ds = \frac{1}{\lambda^2(k+i-j)^2}\left\{1-e^{-\lambda(k+i-j)t}-t\lambda(k+i-j)e^{\lambda(k+i-j)t}\right\}$$

Proof.

Now we can solve the integral by using integration by parts: $\int udv = uv - \int vdu$

Letting $u = s \implies du = 1$ and $dv = e^{-\lambda(k+i-j)s} \implies v = \dfrac{-e^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}$

$$\implies \int_0^t se^{-\lambda(k+i-j)s}ds = \left[\frac{-se^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}\right]_0^t - \int_0^t \frac{-e^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}ds$$

$$= \int_0^t \frac{e^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}ds - \left[\frac{se^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}\right]_0^t$$

Then by solving the given integral

$$= \left[\frac{-e^{-\lambda(k+i-j)s}}{\lambda^2(k+i-j)^2} - \frac{se^{-\lambda(k+i-j)s}}{\lambda(k+i-j)}\right]_0^t$$

After inputting the required bounds, we just simply accordingly

$$= \frac{1}{\lambda^2(k+i-j)^2}\left[-e^{-\lambda(k+i+j)s} - s\lambda(k+i-j)e^{-\lambda(k+i-j)s}\right]_0^t$$

$$= \frac{1}{\lambda^2(k+i-j)^2}\left\{-e^{-\lambda(k+i-j)t} - t\lambda(k+i-j)e^{\lambda(k+i-j)t} + 1 + 0\right\}$$

$$= \frac{1}{\lambda^2(k+i-j)^2}\left\{1 - e^{-\lambda(k+i-j)t} - t\lambda(k+i-j)e^{\lambda(k+i-j)t}\right\} \quad \text{as desired}$$

□

7.1.3 Equal Case: $\lambda = \mu$

We can also determine the expectation of the k-th speciation event for the critical case, $\lambda = \mu$. This theorem is also in Theorem 5.1 by Gernhard [5], except we prove rigorously below, adding many details that are necessary.

Theorem 7.3 (+)**.** *The expectation for the kth speciation time given a known origin time,* $\mathbb{E}\left(\mathscr{A}_{n,t}^k\right)$*, in the Critical case,* $\mu = \lambda$*:*

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$
$$\times\left[\lambda t - (n-j-1)ln(1+\lambda t) + \sum_{l=2}^{n-j-1}\binom{n-j-1}{l}(-1)^l\frac{(1+\lambda t)^{-l+1}-1}{1-l}\right] \tag{7.5}$$

(See Gernhard [5] Theorem 5.1)

Proof.

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = \int_0^t sf_{\mathscr{A}_{n,t}^k}(s)ds$$

We first apply a method used before whereby we split and partly solve the integral

$$= \int_0^t \Big(s f_{\mathscr{A}_{n,t}^k}(s) + F_{\mathscr{A}_{n,t}^k}(s) - F_{\mathscr{A}_{n,t}^k}(s) \Big)\, ds$$

$$= \int_0^t \Big(s f_{\mathscr{A}_{n,t}^k}(s) + F_{\mathscr{A}_{n,t}^k}(s) \Big)\, ds - \int_0^t F_{\mathscr{A}_{n,t}^k}(s) ds$$

$$= \int_0^t \Big(s F'_{\mathscr{A}_{n,t}^k}(s) + F_{\mathscr{A}_{n,t}^k}(s) \Big) - \int_0^t F_{\mathscr{A}_{n,t}^k}(s) ds$$

As: $\dfrac{d}{dx}\left(xF(x)\right) = xF'(x) + F(x)$

$$= \Big[s F_{\mathscr{A}_{n,t}^k}(s) \Big]_0^t - \int_0^t F_{\mathscr{A}_{n,t}^k}(s) ds$$

$$= \Big[t F_{\mathscr{A}_{n,t}^k}(t) \Big] - \int_0^t F_{\mathscr{A}_{n,t}^k}(t) ds$$

Noting that $\mathscr{A}_{n,t}^k$ takes values between $(0, t)$, implying $F_{\mathscr{A}_{n,t}^k}(s) = 1$

$$= t(1) - \int_0^t F_{\mathscr{A}_{n,t}^k}(s) ds$$

Then using Equation: 6.2 for $F_{\mathscr{A}_{n,t}^k}(s)$ gives

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \int_0^t \sum_{i=0}^{k-1} \binom{n-1}{i} F(s|t)^{n-1-i}(1 - F(s|t))^i ds$$

$$= t - \sum_{i=0}^{k-1} \binom{n-1}{i} \int_0^t F(s|t)^{n-1-i}(-1)^i (F(s|t) - 1)^i ds$$

Then binomial expansion around $(F(s|t) - 1)^i$ gives $(F(s|t) - 1)^i = \sum_{j=0}^{i} \binom{i}{j} F(s|t)^{i-j}(-1)^j$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1} \binom{n-1}{i} \int_0^t F(s|t)^{n-1-i}(-1)^i \sum_{j=0}^{i} \binom{i}{j} F(s|t)^{i-j}(-1)^j ds$$

Move the sum outside of the integral

$$= t - \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{n-1}{i} \binom{i}{j} (-1)^{i+j} \int_0^t F(s|t)^{n-1-i} F(s|t)^{i-j} ds$$

Then we match common terms to get

$$= t - \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{n-1}{i} \binom{i}{j} (-1)^{i+j} \int_0^t F(s|t)^{n-1-i+i-j} ds$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t F(s|t)^{n-j-1}ds$$

Then plugging in Equation: 4.5 for $F(s|t)$ where $\lambda = \mu$ we get

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t \left\{\left(\frac{1+\lambda t}{t}\right)\left(\frac{s}{1+\lambda s}\right)\right\}^{n-j-1} ds$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\int_0^t \left(\frac{1+\lambda t}{t}\right)^{n-j-1}\left(\frac{s}{1+\lambda s}\right)^{n-j-1} ds$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\int_0^t \left(\frac{s}{1+\lambda s}\right)^{n-j-1} ds$$

We can then solve the integral using substitution

Then letting $x = 1+\lambda s \implies \dfrac{dx}{ds} = \lambda \implies dx = \dfrac{ds}{\lambda}$

and if $x = 1+\lambda s \implies s = \dfrac{x-1}{\lambda}$

Furthermore we need to plug in the bounds, lower bound: $1+\lambda(0) = 1$, upper bound: $1+\lambda t$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\int_1^{1+\lambda t} \left(\frac{x-1}{x\lambda}\right)^{n-j-1}\frac{dx}{\lambda}$$

Now we simplify the integral so we can do binomial expansion

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}(-1)^{i+j}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\int_1^{1+\lambda t} \left(\frac{x-1}{x}\right)^{n-j-1}\frac{dx}{\lambda^{n-j}}$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\int_1^{1+\lambda t} \left(\frac{x-1}{x}\right)^{n-j-1}dx$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\int_1^{1+\lambda t} \left(1-\frac{1}{x}\right)^{n-j-1}dx$$

Then using binomial expansion around: $\left(1-\dfrac{1}{x}\right)^{n-j-1} = \displaystyle\sum_{l=0}^{n-j-1}\binom{n-j-1}{l}1^{n-j-1-l}\left(-\frac{1}{x}\right)^{l}$

$$\mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$

$$\times \int_1^{1+\lambda t}\sum_{l=0}^{n-j-1}\binom{n-j-1}{l}1^{n-j-1-l}\left(-\frac{1}{x}\right)^{l}dx$$

Now we have to isolate terms from the integral

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$

$$\times \int_1^{1+\lambda t}\sum_{l=0}^{n-j-1}\binom{n-j-1}{l}(-1)^l x^{-l}dx$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}\sum_{l=0}^{n-j-1}\binom{n-j-1}{l}(-1)^l$$

$$\times \int_1^{1+\lambda t} x^{-l}dx$$

In these next steps we separate $l = 0$ and $l = 1$ from the sum and solve separately

We do this because these integrals have unique solutions for whether $l = 1, l = 1, l > 1$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$

$$\times\left[\binom{n-j-1}{0}(-1)^0\int_1^{1+\lambda t}x^{-0}dx + \binom{n-j-1}{1}(-1)^1\int_1^{1+\lambda t}x^{-1}dx\right.$$

$$\left.+\sum_{l=2}^{n-j-1}\binom{n-j-1}{l}(-1)^l\int_1^{1+\lambda t}x^{-l}dx\right]$$

$$= t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$

$$\times\left[\int_1^{1+\lambda t}dx - (n-j-1)\int_1^{1+\lambda t}\frac{1}{x}dx + \sum_{l=2}^{n-j-1}\binom{n-j-1}{l}(-1)^l\int_1^{1+\lambda t}x^{-l}dx\right]$$

Then we can solve the integrals using Proposition 7.3.1

$$\implies \mathbb{E}(\mathscr{A}_{n,t}^k) = t - \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{n-1}{i}\binom{i}{j}\frac{(-1)^{i+j}}{\lambda^{n-j}}\left(\frac{1+\lambda t}{t}\right)^{n-j-1}$$

$$\times\left[\lambda t - (n-j-1)\ln(1+\lambda t) + \sum_{l=2}^{n-j-1}\binom{n-j-1}{l}(-1)^l\frac{(1+\lambda t)^{-l+1}-1}{1-l}\right]$$

This completes the proof for $\mathbb{E}(\mathscr{A}_{n,t}^k)$ when $\lambda = \mu$ $\square$

Proposition 7.3.1. *The following integrals solve to:*

$$\text{First} \int_1^{1+\lambda t} dx = [x]_1^{1+\lambda t} = 1 + \lambda t - 1 = \lambda t$$

$$\text{Next} \int_1^{1+\lambda t} \frac{1}{x} dx = [ln(x)]_1^{1+\lambda t} = ln(1+\lambda t) - ln(1) = ln(1+\lambda t)$$

$$\text{Finally} \int_1^{1+\lambda t} x^{-l} dx = \left[\frac{x^{-l+1}}{-l+1}\right]_1^{1+\lambda t} = \frac{(1+\lambda t)^{-l+1}}{1-l} - \frac{1^{-l+1}}{1-l} = \frac{(1+\lambda t)^{-l+1} - 1}{1-l}$$

7.2 Tree origin is unknown, t_{or} is random

Now we derive the expectation for the kth speciation time, assuming that the tree age is unknown, we will be deriving $\mathbb{E}(\mathscr{A}_n^k)$. We assign t_{or} a uniform prior across the positive real numbers, $t_{or} \sim U(0, \infty)$.

We will prove $\mathbb{E}(\mathscr{A}_n^k)$ for the general case, $0 < \mu < \lambda$, for the Yule Case $\mu = 0$ and the critical case $\mu = \lambda$. We will also simulate some of the outcomes below, the R-Code will be available in the Appendix A. All theorems are from Tanja Gernhards *The Conditioned Reconstructed Process* [5].

7.2.1 General Case $0 < \mu < \lambda$

The following theorem is stated in Gernhard [5] Theorem 5.2 and has a brief proof outline, we add details to achieve a more rigorous proof.

We use the following expectation to simulate some results for different ρ in Chapter: 7.3.

Theorem 7.4 (+)**.** *The expectation for the kth speciation time, given an unknown t_{or} and $0 < \mu < \lambda$*

$$\mathbb{E}(\mathscr{A}_n^k) = \frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i} \times \frac{1}{(k+i+1)\rho}\left(\frac{1}{\rho}-1\right)^{k+i} \times \left\{ ln\left(\frac{1}{1-\rho}\right) - \sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{1}{1-\rho}\right)^j\right]\right\} \tag{7.6}$$

(See Gernhard [5] Theorem 5.2)

Proof.

$$\mathbb{E}(\mathscr{A}_n^k) = \int_0^\infty s f_{\mathscr{A}_n^k}(s) ds$$

Then using Equation: 6.3 for $f_{\mathscr{A}_n^k}(s)$

$$\implies \mathbb{E}(\mathscr{A}_n^k) = \int_0^\infty (k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}} s ds$$

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}\int_0^\infty e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}} s ds$$

Now let $C_1 = (k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}$ and $f(s) = e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}}$

Therefore from Proposition 7.4.1 we have:

$$F(s) = \frac{1}{\lambda^{n-k}}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\frac{(-(\lambda-\mu))^{i-1}}{k+i+1}\left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

$$\implies \mathbb{E}(\mathscr{A}_n^k) = C_1\int_0^\infty s f(s) ds$$

$$= C_1\left\{\int_0^\infty (sf(s) + F(s) - F(s))\right\}$$

$$= C_1\left\{\int_0^\infty (sf(s) + F(s))\, ds - \int_0^\infty F(s) ds\right\}$$

$$\implies \mathbb{E}(\mathscr{A}_n^k) = C_1\left\{[sF(s)]_0^\infty - \int_0^\infty F(s) ds\right\}$$

Using Proposition 7.4.4 $\quad [sF(s)]_0^\infty = 0$

$$\implies \mathbb{E}(\mathscr{A}_n^k) = C_1\left\{[sF(s)]_0^\infty - \int_0^\infty F(s) ds\right\}$$

$$= C_1\left\{0 - \int_0^\infty F(s) ds\right\}$$

$$= -C_1\int_0^\infty F(s) ds \quad \text{Now we can find } \int_0^\infty F(s) ds$$

We use results from Proposition 7.4.5 to solve this integral, giving us

$$\implies \mathbb{E}(\mathscr{A}_n^k) = -C_1\frac{(-1)^k}{\lambda^{n-k}}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\frac{(\lambda-\mu)^{i-2}}{k+i+1}\frac{1}{\mu^{k+i+1}}$$

$$\times\left\{-\ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i}\binom{k+i}{j}\lambda^j\frac{(-1)^j}{j}\left[\lambda^{-j} - (\lambda-\mu)^{-j}\right]\right\}$$

Then we just simplify by cancelling and matching terms

$$= -(k+1)\binom{n}{k+1}\lambda^{n-k}(\lambda-\mu)^{k+2}\frac{(-1)^k}{\lambda^{n-k}}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\frac{(\lambda-\mu)^{i-2}}{k+i+1}\frac{1}{\mu^{k+i+1}}$$

$$\times\left\{-\ln\left(\frac{\lambda}{\lambda-\mu}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\lambda^j\frac{(-1)^j}{j}\left[\lambda^{-j}-(\lambda-\mu)^{-j}\right]\right\}$$

$$= -(k+1)\binom{n}{k+1}(-1)^k\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\frac{(\lambda-\mu)^{k+i}}{k+i+1}\frac{1}{\mu^{k+i+1}}$$

$$\times\left\{-\ln\left(\frac{\lambda}{\lambda-\mu}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\lambda^j\frac{(-1)^j}{j}\left[\lambda^{-j}-(\lambda-\mu)^{-j}\right]\right\}$$

To create consistent results with Gernhard 2008 [5], we let $\rho := \frac{\mu}{\lambda}$ and solve

We use the simplification from Proposition 7.4.6

$$\implies \mathbb{E}(\mathscr{A}_n^k) = \frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}$$

$$\times\frac{1}{(k+i+1)\rho}\left(\frac{1}{\rho}-1\right)^{k+i}$$

$$\times\left\{\ln\left(\frac{1}{1-\rho}\right)-\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{1}{1-\rho}\right)^j\right]\right\}$$

This completes the proof

□

All that remains is to prove the following propositions that supplemented the proof above

Proposition 7.4.1. *Suppose we have a density function of the form*

$$f(s) = e^{-(\lambda-\mu)(k+1)s}\frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda-\mu e^{-(\lambda-\mu)s})^{n+1}}$$

Then the corresponding cumulative distribution function is

$$F(s) = \frac{1}{\lambda^{n-k}}\sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\frac{(-(\lambda-\mu))^{i-1}}{k+i+1}\left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Proof.

$$F(s) = \int f(s)ds$$

Next we plug in $f(s)$

$$= \int e^{-(\lambda-\mu)(k+1)s} \frac{(1-e^{-(\lambda-\mu)s})^{n-k-1}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n+1}} ds$$

$$= \int \frac{e^{-(\lambda-\mu)s^{(k+1)}}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n+1}} (1-e^{-(\lambda-\mu)s})^{n-k-1} ds$$

Then multiply by 1 so that we can simplify further

$$= \int \frac{e^{-(\lambda-\mu)s^{(k+1)}}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n+1}} (1-e^{-(\lambda-\mu)s})^{n-k-1} ds \times \frac{e^{-(\lambda-\mu)s^{n-k-1}}(\lambda - \mu e^{-(\lambda-\mu)s})}{e^{-(\lambda-\mu)s^{n-k-1}}(\lambda - \mu e^{-(\lambda-\mu)s})}$$

$$= \int \frac{e^{-(\lambda-\mu)s^{(k+1+n-k-1)}}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n+1-1}} (1-e^{-(\lambda-\mu)s})^{n-k-1} ds \times \frac{1}{e^{-(\lambda-\mu)s^{n-k-1}}(\lambda - \mu e^{-(\lambda-\mu)s})}$$

$$= \int \frac{e^{-(\lambda-\mu)s^{n}}}{(\lambda - \mu e^{-(\lambda-\mu)s})^{n}} \left(\frac{(1-e^{-(\lambda-\mu)s})}{e^{-(\lambda-\mu)s}} \right)^{n-k-1} \times \frac{ds}{(\lambda - \mu e^{-(\lambda-\mu)s})}$$

$$= \int \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})} \right)^{n} \left(\frac{(1-e^{-(\lambda-\mu)s})}{e^{-(\lambda-\mu)s}} \right)^{n-k-1} \frac{ds}{(\lambda - \mu e^{-(\lambda-\mu)s})}$$

Now using integration by substitution: Let $x = \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})} \right)$

Next we have to find the derivative to replace

$$\implies \frac{dx}{ds} = \frac{(\lambda - \mu e^{-(\lambda-\mu)s})(-(\lambda-\mu)e^{-(\lambda-\mu)s}) - (\mu(\lambda-\mu)e^{-(\lambda-\mu)s})(e^{-(\lambda-\mu)s})}{(\lambda - \mu e^{-(\lambda-\mu)s})^2}$$

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

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

$$= \frac{-\lambda(\lambda-\mu)e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})^2}$$

$$\implies ds = -\frac{dx(\lambda - \mu e^{-(\lambda-\mu)s})^2}{\lambda(\lambda-\mu)e^{-(\lambda-\mu)s}}$$

Now we plug these results back into the equation

$$\implies F(s) = \int x^n \left(\frac{1-e^{-(\lambda-\mu)s}}{e^{-(\lambda-\mu)s}} \right)^{n-k-1} \left(\frac{1}{\lambda - \mu e^{-(\lambda-\mu)s}} \right) \left(-\frac{dx(\lambda - \mu e^{-(\lambda-\mu)s})^2}{\lambda(\lambda-\mu)e^{-(\lambda-\mu)s}} \right)$$

Then we can simplify by removing non-x terms from the integration and combining like terms

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^n\left(\frac{1-e^{-(\lambda-\mu)s}}{e^{-(\lambda-\mu)s}}\right)^{n-k-1}\left(\frac{dx(\lambda-\mu e^{-(\lambda-\mu)s})^2}{(\lambda-\mu e^{-(\lambda-\mu)s})e^{-(\lambda-\mu)s}}\right)$$

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^n\left(\frac{1-e^{-(\lambda-\mu)s}}{e^{-(\lambda-\mu)s}}\right)^{n-k-1}\frac{dx(\lambda-\mu e^{-(\lambda-\mu)s})}{x}$$

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}\left(\frac{1-e^{-(\lambda-\mu)s}}{e^{-(\lambda-\mu)s}}\right)^{n-k-1}dx$$

We still have s terms in the equation, we aim to get $e^{-(\lambda-\mu)s}$ in terms of x

Then using $x = \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)$

$$\iff x(\lambda-\mu e^{-(\lambda-\mu)s}) = e^{-(\lambda-\mu)s} \iff \lambda x - \mu x e^{-(\lambda-\mu)s} = e^{-(\lambda-\mu)s}$$

$$\iff \lambda x = e^{-(\lambda-\mu)s} + \mu x e^{-(\lambda-\mu)s}\lambda x = e^{-(\lambda-\mu)s}(1+\mu x)$$

$$\iff e^{-(\lambda-\mu)s} = \frac{\lambda x}{1+\mu x}$$

$$\implies F(s) = -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}\left(\frac{1-\frac{\lambda x}{1+\mu x}}{\frac{\lambda x}{1+\mu x}}\right)^{n-k-1}dx$$

Now plugging this into $F(s)$ gives the following

$$F(s) = -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}\left(\frac{\frac{1+\mu x-\lambda x}{1+\mu x}}{\frac{\lambda x}{1+\mu x}}\right)^{n-k-1}dx$$

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}\left(\frac{1+\mu x-\lambda x}{\lambda x}\right)^{n-k-1}dx$$

Then isolate the x in the numerator

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}\left(\frac{1-(\lambda-\mu)x}{\lambda x}\right)^{n-k-1}dx$$

Bringing the denominator and combining like terms gives the following

$$= -\frac{1}{\lambda(\lambda-\mu)}\int x^{n-1}(1-(\lambda-\mu)x)^{n-k-1}\lambda^{-n+k+1}x^{-n+k+1}dx$$

$$= -\frac{\lambda^{-n+k+1}}{\lambda(\lambda-\mu)}\int x^{n-1-n+k+1}(1-(\lambda-\mu)x)^{n-k-1}dx$$

$$= -\frac{1}{\lambda^{n-k}(\lambda-\mu)}\int x^k(1-(\lambda-\mu)x)^{n-k-1}dx$$

Then using binomial expansion around:

$$(1-(\lambda-\mu)x)^{n-k-1} = \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} 1^{n-k-1-i}(-(\lambda-\mu)x)^i \text{gives the following}$$

$$\implies F(s) = -\frac{1}{\lambda^{n-k}(\lambda-\mu)} \int x^k \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} 1^{n-k-1-i}(-(\lambda-\mu)x)^i dx$$

Now we isolate the integral

$$= -\frac{1}{\lambda^{n-k}(\lambda-\mu)} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \int x^k(-(\lambda-\mu))^i x^i dx$$

$$= -\frac{1}{\lambda^{n-k}(\lambda-\mu)} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} (-(\lambda-\mu))^i \int x^{k+i} dx$$

Solving the integral gives

$$= \frac{1}{\lambda^{n-k}(-(\lambda-\mu))} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} (-(\lambda-\mu))^i \left[\frac{x^{k+i+1}}{k+i+1}\right]$$

Plugging in the bounds gives us the completed proof

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^i}{(-(\lambda-\mu))k+i+1} \left[x^{k+i+1}\right]$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left[x^{k+i+1}\right]$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

□

Proposition 7.4.2. *The limit to infinity for the following function $sF(s)$ is 0*

$$\lim_{s\to\infty} sF(s) = \lim_{s\to\infty} s\frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Proof.

$$\lim_{s\to\infty} sF(s) = \lim_{s\to\infty} s\frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

We then isolate the limit terms in the limit

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \lim_{s\to\infty} s \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \lim_{s\to\infty} \left(\frac{s}{e^{(\lambda-\mu)s}(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Then given continuity of the relevant limit, we can split the limit into two

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\lim_{s\to\infty} \frac{s}{e^{(\lambda-\mu)s}(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\lim_{s\to\infty} \frac{s}{e^{(\lambda-\mu)s}} \cdot \frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Now we can solve these limits separately

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1}$$

$$\times \left(\lim_{s\to\infty} \frac{s}{e^{(\lambda-\mu)s}}\right)^{k+i+1} \left(\lim_{s\to\infty} \frac{1}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Solving the first limit gives the following

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \frac{1}{\lambda} \left(\lim_{s\to\infty} \frac{s}{e^{(\lambda-\mu)s}}\right)^{k+i+1}$$

We can use L'Hopitals for the second limit because the numerator and denominator independently approach infinity, hence we take derivatives, giving

$$= \frac{1}{\lambda^{n-k-1}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\lim_{s\to\infty} \frac{1}{(\lambda-\mu)e^{(\lambda-\mu)s}}\right)^{k+i+1}$$

We solve the relevant limit to get 0

$$= \frac{1}{\lambda^{n-k-1}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \times 0$$

This makes the whole result 0, as desired

$$\implies \lim_{s\to\infty} sF(s) = 0$$

□

Proposition 7.4.3. *The limit to* 0 *for the following function* $sF(s)$ *is* 0

$$\lim_{s\to 0} sF(s) = \lim_{s\to 0} s\frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1} = 0$$

Proof.

Then using Proposition 7.4.1 for $F(s)$ we get the following

$$\lim_{s\to 0} sF(s) = \lim_{s\to 0} s\frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})}\right)^{k+i+1}$$

Next we can plug in 0 for s

$$= 0 \times \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{1}{(\lambda-\mu)}\right)^{k+i+1}$$

$$\implies \lim_{s\to 0} sF(s) = 0$$

□

Proposition 7.4.4. *The solution to the equation* $[sF(s)]_0^\infty$ *is* 0*, where* $F(s)$ *is derived in Proposition 7.4.1 is*

Proof.

$$[sF(s)]_0^\infty = \lim_{s\to\infty} - \lim_{s\to 0}$$

Then using Proposition's 7.4.2 and 7.4.3 to solve the limits

$$\implies [sF(s)]_0^\infty = 0 - 0 = 0$$

□

Proposition 7.4.5. *The solution to the following integral of* $F(s)$ *from Propositions 7.4.1 is*

$$\int_0^\infty F(s)ds = \frac{(-1)^k}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{1}{\mu^{k+i+1}}$$

$$\times \left\{ -ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{(-1)^j}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j}\right] \right\}$$

Proof.

Use the representation of $F(s)$ from 7.4.1 to get

$$\int_0^\infty F(s)ds = \int_0^\infty \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})} \right)^{k+i+1} ds$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \int_0^\infty \left(\frac{e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})} \right)^{k+i+1} ds$$

Then using substitution let $x = \lambda - \mu e^{-(\lambda-\mu)s}$ and $e^{-(\lambda-\mu)s} = \frac{\lambda - x}{\mu}$

We also need to add the new definite integral bounds:

$$x_0 = \lambda - \mu e^{-(\lambda-\mu)0} = \lambda - \mu$$

$$x_\infty = \lambda - \mu e^{-(\lambda-\mu)\infty} = \lambda \text{ given } 0 < \mu < \lambda$$

We finally need to find the relevant derivative to substitute

$$\implies \frac{dx}{ds} = \mu(\lambda-\mu)e^{-(\lambda-\mu)s} \implies ds = \frac{dx}{\mu(\lambda-\mu)e^{-(\lambda-\mu)s}}$$

$$\iff ds = \frac{dx}{\mu(\lambda-\mu)} \cdot \frac{\mu}{\lambda - x} \iff ds = \frac{dx}{(\lambda-\mu)(\lambda-x)}$$

We combine this all to get the following

$$\implies \int_0^\infty F(s)ds = \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-(\lambda-\mu))^{i-1}}{k+i+1} \int_{\lambda-\mu}^{\lambda} \left(\frac{\frac{\lambda-x}{\mu}}{x} \right)^{k+i+1} \frac{dx}{(\lambda-\mu)(\lambda-x)}$$

We need to then simplify to isolate the integral

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(-1)^{i-1}(\lambda-\mu)^{i-1}}{k+i+1} \frac{1}{(\lambda-\mu)}$$

$$\times \int_{\lambda-\mu}^{\lambda} \left(\frac{\lambda-x}{\mu x} \right)^{k+i+1} \frac{1}{(\lambda-x)} dx$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-1}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}(\lambda-\mu)}$$

$$\times \int_{\lambda-\mu}^{\lambda} \left(\frac{\lambda-x}{x} \right)^{k+i+1} \frac{1}{(\lambda-x)} dx$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}} \int_{\lambda-\mu}^{\lambda} \frac{(\lambda-x)^{k+i}}{x^{k+i+1}} dx$$

Then using binomial theorem around: $(\lambda - x)^{k+i} = \sum_{j=0}^{k+i} \binom{k+i}{j} \lambda^j (-x)^{k+i-j}$

$$\implies \int_0^\infty F(s)ds = \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \int_{\lambda-\mu}^{\lambda} \sum_{j=0}^{k+i} \binom{k+i}{j} \lambda^j (-x)^{k+i-j} \frac{1}{x^{k+i+1}} dx$$

We then need to manipulate the x powers in the integral

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \sum_{j=0}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \int_{\lambda-\mu}^{\lambda} \frac{x^{k+i-j}}{x^{k+i+1}} dx$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \sum_{j=0}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \int_{\lambda-\mu}^{\lambda} x^{k+i-j-k-i-1} dx$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \sum_{j=0}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \int_{\lambda-\mu}^{\lambda} x^{-(j+1)} dx$$

By separating the sum into $j = 0, j = 1$ and $j > 1$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ \binom{k+i}{0} \lambda^0 (-1)^{k+i-0} \int_{\lambda-\mu}^{\lambda} x^{-(0+1)} dx + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \right.$$

$$\left. \times \int_{\lambda-\mu}^{\lambda} x^{-(j+1)} dx \right\}$$

Next we can solve the integrals as follows

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \int_{\lambda-\mu}^{\lambda} x^{-1} dx + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \int_{\lambda-\mu}^{\lambda} x^{-(j+1)} dx \right\}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \left[\ln(x)\right]_{\lambda-\mu}^{\lambda} + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \left[\frac{x^{-(j+1)+1}}{-(j+1)+1} \right]_{\lambda-\mu}^{\lambda} \right\}$$

The next step is to solve for the bounds

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \left[\ln(\lambda) - \ln(\lambda-\mu)\right] + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \left[\frac{x^{-j-1+1}}{-j-1+1} \right]_{\lambda-\mu}^{\lambda} \right\}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \left[\frac{x^{-j}}{-j} \right]_{\lambda-\mu}^{\lambda} \right\}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j} \left[\frac{\lambda^{-j}}{-j} - \frac{(\lambda-\mu)^{-j}}{-j} \right] \right\}$$

Simplifying further gives the following

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k+i} \ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j (-1)^{k+i-j+1} \left[\frac{\lambda^{-j}}{j} - \frac{(\lambda-\mu)^{-j}}{j} \right] \right\}$$

Note we can flip the signs of $(-1)^{k+i}$ to $(-1)^{k-i}$ because they produce the same value for all k and i

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$

$$\times \left\{ (-1)^{k-i} \ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{(-1)^{k-i-j+1}}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j} \right] \right\}$$

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}}$$
$$\times \left\{ (-1)^{k-i+1}(-1)^{-1}\ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{(-1)^{k-i+1}(-1)^{-j}}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j}\right] \right\}$$
$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1}}{\mu^{k+i+1}} (-1)^{k-i+1}$$
$$\times \left\{ -\ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{(-1)^{-j}}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j}\right] \right\}$$

We can also flip the sign of $(-1)^{-j}$ to $(-1)^j$ because these two values are equivalent for all j

$$= \frac{1}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{(-1)^{i-1+k-i+1}}{\mu^{k+i+1}} (-1)^{k-i+1}$$
$$\times \left\{ -\ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{-1)^{(j}}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j}\right] \right\}$$

Simplifying further achieves the final result

$$= \frac{(-1)^k}{\lambda^{n-k}} \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \frac{(\lambda-\mu)^{i-2}}{k+i+1} \frac{1}{\mu^{k+i+1}}$$
$$\times \left\{ -\ln\left(\frac{\lambda}{\lambda-\mu}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \lambda^j \frac{(-1)^j}{j} \left[\lambda^{-j} - (\lambda-\mu)^{-j}\right] \right\}$$

□

Proposition 7.4.6. *Letting $\rho := \frac{\mu}{\lambda}$ we can simplify the following result*

$$-(k+1)\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1} \binom{n-k-1}{i} \left(\frac{\lambda-\mu}{\mu}\right)^{k+i} \frac{1}{\mu(k+i+1)}$$
$$\times \left\{ ln\left(\frac{\lambda-\mu}{\lambda}\right) + \sum_{j=1}^{k+i} \binom{k+i}{j} \frac{(-1)^j}{j} \left[\frac{\lambda^{-j} - (\lambda-\mu)^{-j}}{\lambda^{-j}}\right] \right\}$$

$$= \frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}$$
$$\times \frac{1}{(k+i+1)\rho}\left(\frac{1}{\rho}-1\right)^{k+i}$$
$$\times \left\{ ln\left(\frac{1}{1-\rho}\right) - \sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{1}{1-\rho}\right)^j\right]\right\}$$

Proof.

$$-(k+1)\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{\lambda-\mu}{\mu}\right)^{k+i}\frac{1}{\mu(k+i+1)}$$
$$\times\left\{\ln\left(\frac{\lambda-\mu}{\lambda}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[\frac{\lambda^{-j}-(\lambda-\mu)^{-j}}{\lambda^{-j}}\right]\right\}$$

We first find ways to isolate λ and μ

$$= -(k+1)\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{\lambda}{\mu}-1\right)^{k+i}\frac{1}{\mu(k+i+1)}$$
$$\times\left\{\ln\left(1-\frac{\mu}{\lambda}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\frac{(\lambda-\mu)^{-j}}{\lambda^{-j}}\right]\right\}$$

We use the quality that $\dfrac{\lambda}{\mu}=\dfrac{1}{\rho}$

$$= -(k+1)\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{1}{\rho}-1\right)^{k+i}\frac{1}{\lambda\rho(k+i+1)}$$
$$\times\left\{\ln(1-\rho)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{\lambda-\mu}{\lambda}\right)^{-j}\right]\right\}$$
$$= -(k+1)\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{1}{\rho}-1\right)^{k+i}\frac{1}{\lambda\rho(k+i+1)}$$
$$\times\left\{\ln(1-\rho)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(1-\frac{\mu}{\lambda}\right)^{-j}\right]\right\}$$

$$= -\frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{1}{\rho}-1\right)^{k+i}\frac{1}{\rho(k+i+1)}$$
$$\times \left\{\ln(1-\rho)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(1-\frac{\mu}{\lambda}\right)^{-j}\right]\right\}$$
$$= -\frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{1}{\rho}-1\right)^{k+i}\frac{1}{(k+i+1)\rho}$$
$$\times \left\{-\ln\left(\frac{1}{1-\rho}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-(1-\rho)^{-j}\right]\right\}$$
$$= -\frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}\left(\frac{1}{\rho}-1\right)^{k+i}\frac{1}{(k+i+1)\rho}$$
$$\times \left\{-\ln\left(\frac{1}{1-\rho}\right)+\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{1}{1-\rho}\right)^{j}\right]\right\}$$
Finally we generate the desired result by moving the $-$ sign
$$= \frac{(k+1)}{\lambda}\binom{n}{k+1}(-1)^k \sum_{i=0}^{n-k-1}\binom{n-k-1}{i}$$
$$\times \frac{1}{(k+i+1)\rho}\left(\frac{1}{\rho}-1\right)^{k+i}$$
$$\times \left\{\ln\left(\frac{1}{1-\rho}\right)-\sum_{j=1}^{k+i}\binom{k+i}{j}\frac{(-1)^j}{j}\left[1-\left(\frac{1}{1-\rho}\right)^{j}\right]\right\}$$

□

7.2.2 Yule Case, $\mu = 0$

We aim to find $\mathbb{E}(\mathscr{A}_n^k)$, the expectation of the time of the k-th speciation event given n extant species and an unknown origin time t_{or} for the Yule Model.

The following theorem is provided in Gernhard [5] Theorem 5.2, but it is not proved. We give a full proof below.

Theorem 7.5 (+). *The expectation for the kth speciation time, given an unknown t_{or} and in the*

Yule Case, $\mu = 0$

$$\mathbb{E}(\mathscr{A}_n^k) = \sum_{i=k+1}^{n} \frac{1}{\lambda i}$$

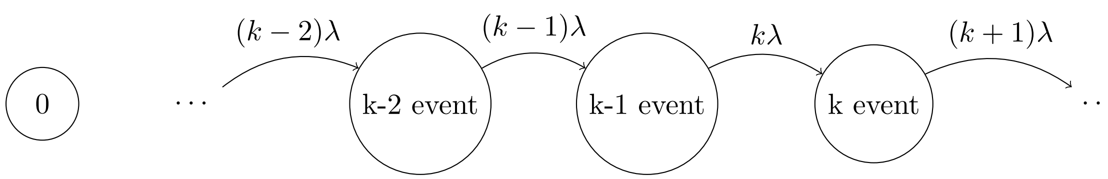
(See Gernhard [5] Theorem 5.2)

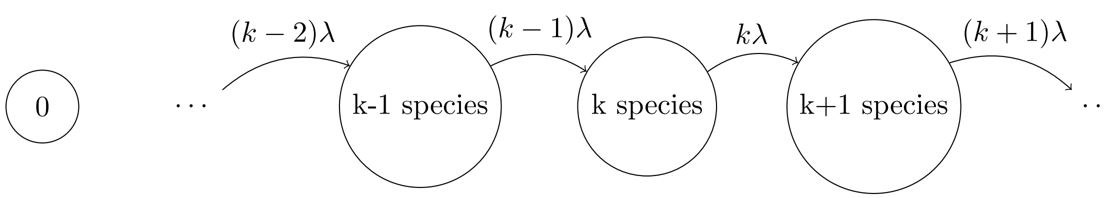
Proof. We do this proof a little bit differently.

In order to find $\mathbb{E}(\mathscr{A}_n^k)$ when using the Yule model and assuming t_{or} is unknown, we don't need to condition on n extant species in the present. This is because once we reach n species, we will know that the model is in the present. This is because of two reasons, firstly, we don't need to have a specified time t to have elapsed till we reach the present. We have assumed the origin time is random, therefore the time until the present is also random. Secondly, under the Yule model, species do not go extinct. Once we reach n species in the model, we will never go below this number. To prevent going above this number once the model reaches n species, we assume it is the present. Hence $\mathscr{A}_n^k = \mathscr{A}^k$.

Now let X_k be the random variable that is the time between the $k-1$ and kth speciation event. This implies $X_k = \mathscr{A}^{k-1} - \mathscr{A}^k$. Note that $\mathscr{A}^k \leq \mathscr{A}^{k-1}$ because $\mathscr{A}^{k-1}$ is further from the present time, 0. Given we are modelling under a Markov Chain Birth Death Process, the rate at which a species, speciates, splitting into two species, is Poisson distributed. Therefore the time between Poisson events, is exponentially distributed. Therefore X_k would have an exponential distribution. We need to find the rate parameter of X_k.

Now remember, we start with one species, therefore after the first speciation event we finish with one species, after the second we finish with three species. Therefore after the kth speciation event, we will finish with $k+1$ species. Therefore the rate at which species split after $k-1$ speciation events, is equivalent to the rate of splits with k species. See the diagrams below.





This implies that the time between the k and $k-1$ speciation events is also the time between having k and $k+1$ species, which is exponentially distributed with rate $k\lambda$. Hence, $X_k \sim \text{Exp}(k\lambda)$.

Next we can rewrite X_k to isolate $\mathscr{A}^k$.

$$
\begin{aligned}
X_k &= \mathscr{A}^{k-1} - \mathscr{A}^k \\
\iff \mathscr{A}^{k-1} &= X_k + \mathscr{A}^k \\
\implies \mathscr{A}^k &= X_{k+1} + \mathscr{A}^{k+1} \\
\iff \mathscr{A}^k &= X_{k+1} + X_{k+2} + \mathscr{A}^{k+2} \\
\iff \mathscr{A}^k &= X + k + 1 + X_{k+2} + X_{k+3} + \mathscr{A}^{k+3}
\end{aligned}
$$

Then by recursion

$$
\begin{aligned}
\implies \mathscr{A}^k &= X_k + X_{k+1} + \ldots + X_n + \mathscr{A}^n \\
\iff \mathscr{A}^k &= X_k + X_{k+1} + \ldots + X_n \quad \text{because there is no } n\text{th speciation event} \\
\iff \mathscr{A}^k &= \sum_{i=k+1}^{n} X_i
\end{aligned}
$$

Now we can find $\mathbb{E}(\mathscr{A}_n^k)$

$$
\begin{aligned}
\mathbb{E}(\mathscr{A}_n^k) &= \mathbb{E}(\mathscr{A}^k) \\
&= \mathbb{E}\left(\sum_{i=k+1}^{n} X_i\right) \\
&= \sum_{i=k+1}^{n} \mathbb{E}(X_i) \quad \text{By linearity of the expectation operator} \\
&= \sum_{i=k+1}^{n} \frac{1}{\lambda i} \quad \text{Because } X_i \sim \text{Exp}(\lambda i)
\end{aligned}
$$

Above is our final answers, which completes the proof. □

7.2.3 Critical Case, $\mu = \lambda$ [6]

The following theorem is also provided in Gernhard [5] Theorem 5.2, but with no proof. A sketch proof is provided in Gernhard *New Analytic Results* [6] Equation (10). We provide a fully detailed proof below.

Theorem 7.6 (+). *The expectation for the kth speciation time, given an unknown origin time t_{or}, in the critical case, $\mu = \lambda$ is*

$$\mathbb{E}(\mathscr{A}_n^k) = \frac{n-k}{\lambda k}$$

(See Gernhard [5] Theorem 5.2)

Proof.

$$\mathbb{E}(\mathscr{A}_n^k) = \int_0^\infty s f_{\mathscr{A}_n^k}(s)$$

Using the formula for $f_{\mathscr{A}_n^k}(s)$ from equation: 6.5

$$\implies \mathbb{E}(\mathscr{A}_n^k) = \int_0^\infty s(k+1)\binom{n}{k+1}\lambda^{n-k}\frac{s^{n-k-1}}{(1+\lambda s)^{n+1}}ds$$

$$= (k+1)\binom{n}{k+1}\lambda^{n-k}\int_0^\infty \frac{s^{n-k}}{(1+\lambda s)^{n+1}}ds$$

$$= (k+1)\binom{n}{k+1}\int_0^\infty \frac{(\lambda s)^{n-k}}{(1+\lambda s)^{n+1}}ds$$

Using substituion let $u = \lambda s \implies \dfrac{du}{ds} = \lambda \implies ds = \dfrac{du}{\lambda}$

Then inputting the bounds $x_0 = \lambda(0) = 0; x_\infty = \lambda(\infty) = \infty$

$$\implies \mathbb{E}(\mathscr{A}_n^k) = (k+1)\binom{n}{k+1}\int_0^\infty \frac{(u)^{n-k}}{(1+us)^{n+1}}\frac{du}{\lambda}$$

$$= (k+1)\binom{n}{k+1}\frac{1}{\lambda}\int_0^\infty \frac{(u)^{n-k}}{(1+us)^{n+1}}du$$

Then using the identity that $\displaystyle\int_0^\infty \frac{x^a}{(1+x)^b}dx = \begin{cases} \frac{1}{(b-a-1)\binom{b-1}{a}} & \text{if } b > a+1 \\ \infty & \text{else} \end{cases}$

This identity is from Lebedew 1973 [10]

$$\implies \mathbb{E}(\mathscr{A}_n^k) = (k+1)\binom{n}{k+1}\left(\frac{1}{(n+1-n+k-1)\binom{n+1-1}{n-k}}\right)$$

for $n+1 > n-k+1$ this is true for all speciation times k

$$= \frac{1}{\lambda}(k+1)\binom{n}{k+1}\left(\frac{1}{k\binom{n}{n-k}}\right)$$

$$= (k+1)\left(\frac{n!}{(k+1)!(n-k-1)!}\right)\left(\frac{1}{\frac{n!}{(n-k)!(n-k+n)!}}\right)\left(\frac{1}{\lambda k}\right)$$

$$= (k+1)\left(\frac{n!}{(k+1)!(n-k-1)!}\right)\left(\frac{1}{\frac{n!}{(k)!(k)!}}\right)\left(\frac{1}{\lambda k}\right)$$

$$
\begin{aligned}
&= (k+1)\left(\frac{n!}{(k+1)!(n-k-1)!}\right)\left(\frac{(n-k)!(k)!}{n!}\right)\left(\frac{1}{\lambda k}\right) \\
&= \left(\frac{(n-k)!(k)!}{k!(n-k-1)!}\right)\left(\frac{1}{\lambda k}\right) \\
&= \left(\frac{(n-k)}{1}\right)\left(\frac{1}{\lambda k}\right) \\
\mathbb{E}(\mathscr{A}_n^k) &= \frac{n-k}{\lambda k} \quad \text{as desired}
\end{aligned}
$$

This completes the proof

□

7.2.4 Special Critical Case Moments, for $\lambda = \mu = 1$

We can also find $\mathbb{E}((\mathscr{A}_n^k)^m)$ the m-th moment for the k-th speciation time where t_{or} is unknown and $\lambda = \mu = 1$

These results are not added in Gernhard's results but are cited from *New Analytic Results* by Gernhard [6] in Corollary 2.2. We provide a proof here also.

Theorem 7.7 (+)**.** *The result for the mth moment for the kth speciation time, where $\lambda = \mu = 1$ and t_{or} is unknown*

$$
\implies \mathbb{E}((\mathscr{A}_n^k)^m) = \begin{cases} \frac{\binom{n+m-k-1}{m}}{\binom{k}{m}} & \text{if } k \geq m \\ \infty & \text{else} \end{cases} \tag{7.7}
$$

(See Gernhard [6] Corollary 2.2)

Proof.

$$
\mathbb{E}((\mathscr{A}_n^k)^m) = \int_0^\infty s^m f_{\mathscr{A}_n^k}(s)
$$

Using the formula for $f_{\mathscr{A}_n^k}(s)$ from equation: 6.5

$$
\implies \mathbb{E}((\mathscr{A}_n^k)^m) = \int_0^\infty s^m (k+1)\binom{n}{k+1}\lambda^{n-k}\frac{s^{n-k-1}}{(1+\lambda s)^{n+1}}ds
$$

Then letting $\lambda = 1$

$$= (k+1)\binom{n}{k+1}1^{n-k}\int_0^\infty \frac{s^{n+m-k-1}}{(1+s)^{n+1}}ds$$
$$= (k+1)\binom{n}{k+1}\int_0^\infty \frac{s^{n-k-2+m}}{(1+s)^{n+1}}ds$$
$$= (k+1)\binom{n}{k+1}\int_0^\infty \frac{s^{n+m-k-1}}{(1+s)^{n+1}}ds$$

Then using Lebedew's identity again [10] $\int_0^\infty \frac{x^a}{(1+x)^b}dx = \begin{cases} \frac{1}{(b-a-1)\binom{b-1}{a}} & \text{if } b > a+1 \\ \infty & \text{else} \end{cases}$

Noting that $n+1 > n+m-k \iff n+1 > n-(k-m)$

Hence we need $k - m > 1 \iff k-1 > m \implies$ we need $k \geq m$ otherwise $\mathbb{E}(\mathscr{A}_n^k) = \infty$

$$\implies \mathbb{E}((\mathscr{A}_n^k)^m) = (k+1)\binom{n}{k+1}\frac{1}{(n+1-n-m+k+1-1)\binom{n+1-1}{n+m-k-1}} \text{ for } k \geq m$$
$$= (k+1)\binom{n}{k+1}\frac{1}{(k-m+1)\binom{n}{n+m-k-1}}$$

Now we simplify the factorials as follows

$$= \frac{(k+1)n!}{(k+1)!(n-k-1)!} \times \frac{(n+m-k-1)!(n-n-m+k+1)!}{(k-m+1)n!}$$
$$= \frac{n!}{k!(n-k-1)!} \times \frac{(n+m-k-1)!(k-m+1)!}{(k-m+1)n!}$$
$$= \frac{n!}{k!(n-k-1)!} \times \frac{(n+m-k-1)!(k-m)!}{n!}$$
$$= \frac{(n+m-k-1)!(k-m)!}{k!(n-k-1)!} \times \frac{m!}{m!}$$
$$= \frac{(n+m-k-1)!}{m!(n-k-1)!} \times \frac{m!(k-m)!}{k!}$$
$$= \binom{n+m-k-1}{m}\frac{1}{\binom{k}{m}}$$
$$= \frac{\binom{n+m-k-1}{m}}{\binom{k}{m}}$$

$$\implies \mathbb{E}((\mathscr{A}_n^k)^m) = \begin{cases} \frac{\binom{n+m-k-1}{m}}{\binom{k}{m}} & \text{if } k \geq m \\ \infty & \text{else} \end{cases}$$

This completes the proof, note it is only valid if $\lambda = \mu = 1$, and t_{or} is unknown [6]. $\square$

7.2.5 Variance: Special Critical Case $\lambda = \mu = 1$ [6]

Given we have calculated the moments, the variance for $\mathscr{A}_n^k$ is a simple calculation.

Theorem 7.8 (+). *The variance for the kth speciation where $\lambda = \mu = 1$ and t_{or} is unknown*

$$Var\left(\mathscr{A}_n^k\right) = \frac{n(n-k)}{k^2(k-1)}$$

(See Gernhard [6] Corollary 2.2)

Proof.

$$\begin{aligned}
\text{Var}\left(\mathscr{A}_n^k\right) &= \mathbb{E}((\mathscr{A}_n^k)^2) - \mathbb{E}(\mathscr{A}_n^k)^2 \quad \text{by definition} \\
&= \frac{\binom{n+2-k-1}{2}}{\binom{k}{2}} - \left(\frac{n-k}{k}\right)^2 \\
&= \frac{\binom{n-k+1}{2}}{\binom{k}{2}} - \left(\frac{n-k}{k}\right)^2
\end{aligned}$$

We simplify the factorials as follows

$$\begin{aligned}
&= \frac{(n-k+1)!}{2!(n-k+1-2)!} \times \frac{2!(k-2)!}{k!} - \frac{(n-k)^2}{k^2} \\
&= \frac{(n-k+1)!}{(n-k-1)!} \times \frac{(k-2)!}{k!} - \frac{(n-k)^2}{k^2} \\
&= \frac{(n-k+1)(n-k)}{k(k-1)} - \frac{(n-k)^2}{k^2} \\
&= \frac{k(n-k+1)(n-k)}{k^2(k-1)} - \frac{(k-1)(n-k)^2}{k^2(k-1)}
\end{aligned}$$

Next we combine denominators by multiplying

$$\begin{aligned}
&= \frac{k(n-k+1)(n-k) - (k-1)(n-k)^2}{k^2(k-1)} \\
&= \frac{(n-k)(k(n-k+1) - (k-1)(n-k))}{k^2(k-1)} \\
&= \frac{(n-k)(kn-k^2+k-kn+k^2+n-k)}{k^2(k-1)}
\end{aligned}$$

Cancelling terms gives the final result

$$= \frac{n(n-k)}{k^2(k-1)}$$

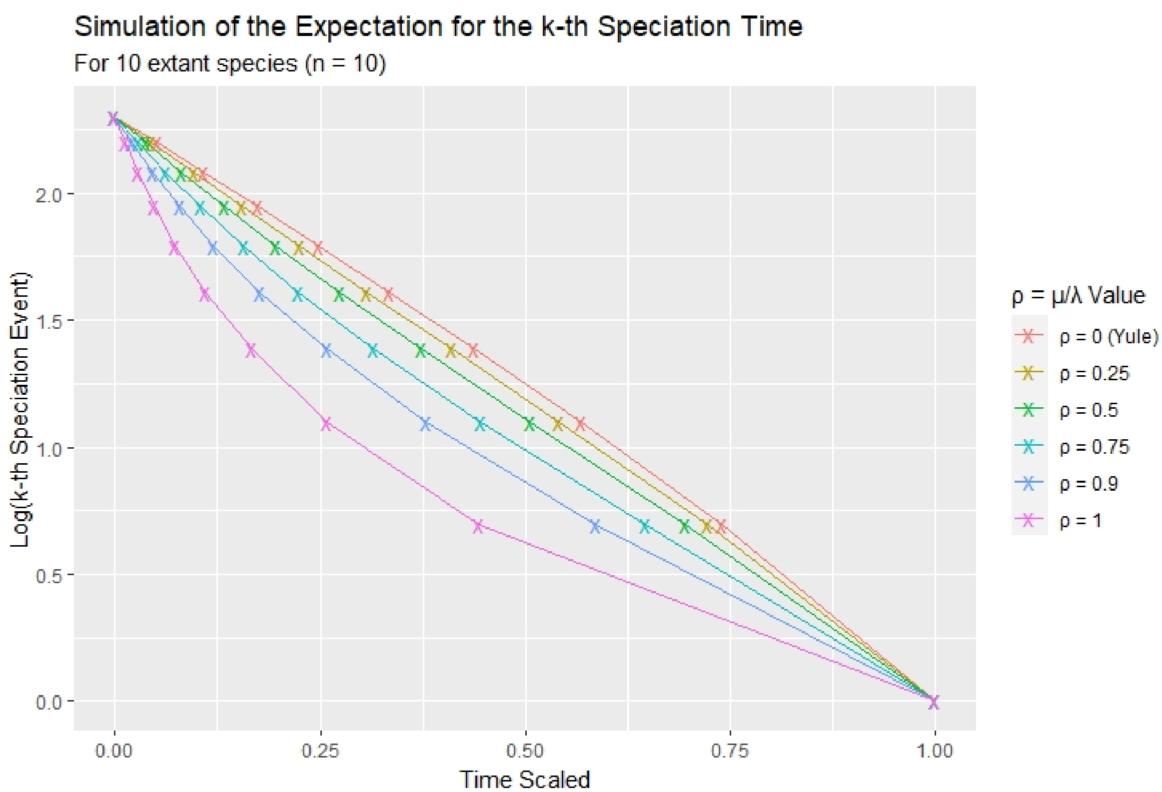
□

7.2.6 Comments

When examining these results it is easy to notice that expected time of the k-th speciation event is decreasing as k increases. This is because the present time is considered 0 and we are counting back in time. Therefore if we have 10 species at present the 10th speciation event will be closer to 0 than the 9th, because it is closer to the present

7.3 Simulation

Now that we've found the expectation of the k-th speciation time, for both unknown t_{or} and known t_{or}, we can simulate a lineage through time plot for unique values of $\rho = \frac{\mu}{\lambda}$ [5]. Assume n, the number of extant species, is equal to 10.



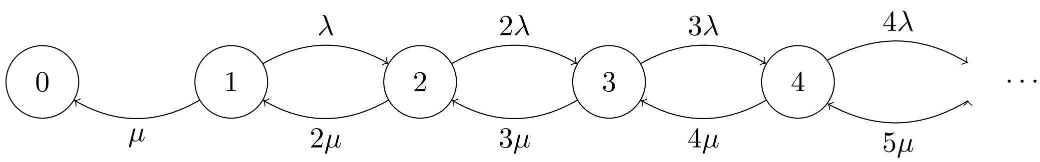
The above graph is a simulation for k-th speciation event across different values of ρ, ($\rho = \{0, 0.25, 0.50, 0.75, 0.9, 1\}$). The present is time 0 and the first speciation event is time 1. We are basically working backwards, counting from the present.

Note that for lower ρ the time-axis is much, much smaller because speciation events happen more frequently. I've scaled all the time-axis to see the difference relatively when speciation events occur. The scaling was done by dividing each speciation event by the time of the first speciation event. Hence all curves finish at time-scale is equal to 1.

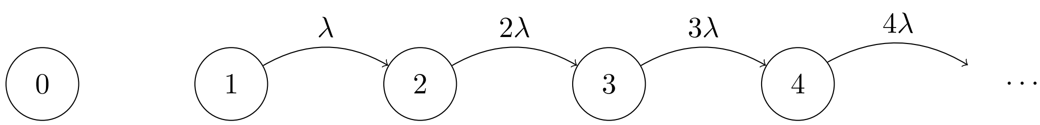
Notice that as we increase ρ, in effect increasing μ, the expected speciation time occurs relatively closer to the present. This is because with a higher μ species go extinct more frequently, hence the process is much slower the beginning, as both λ and μ increase the process occurs relatively more frequently.

See the two figures below comparing the $\rho = 1 \implies \mu = \lambda$ model with the Yule model $\rho = 0 \implies \mu = 0$

$\rho = 1$ **Model**

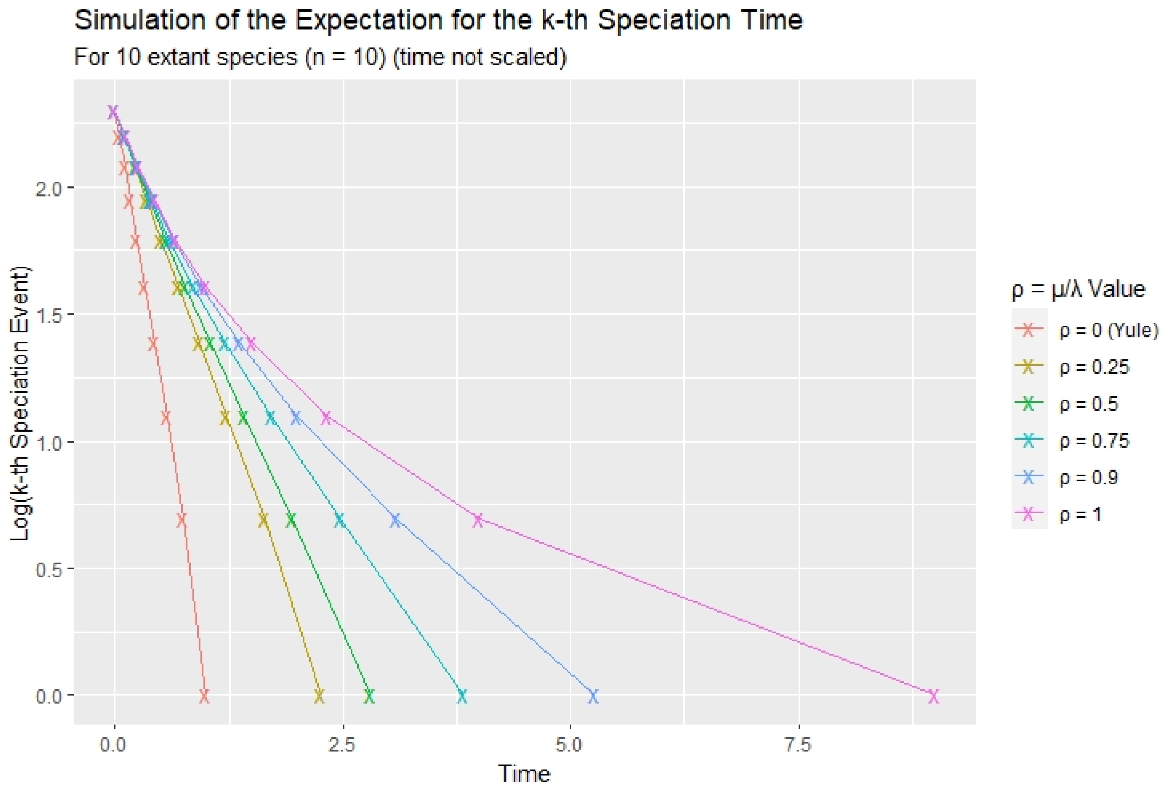


Yule Model



In the $\rho = 1$ model, the process will initially start slower due to extinction events and then increase relatively faster for a higher number of species.

For reference, here is the time unscaled simulation of the expectation of k-th speciation times



Expected speciation times are still later for increased ρ. But the speciation times increase relatively faster for a higher k.

Chapter 8

Further Properties of Speciation Times

In this chapter we examine further properties of speciation times, examining the backwards process, the coalescent process and the probability density function for differences in speciation times. The results included can be found in Gernhard [5] Chapter 6.

8.1 Re-examining the Point Process of the cBDP (conditioned Birth-Death Process)

We have already shown in early chapters regarding the Basics of Phylogenentics, how we can interpret our phylogenetic tree as a Poisson point process, with $n-1$ points that are all iid [5]. This is only true if we condition on the age of the tree. We can not interpret the tree as a Poisson point process if the origin is unknown and we assign a uniform prior.

The following theorem is discussed in Gernhard [5]. We expand the proof to include more direct calculations that are omitted.

Theorem 8.1 (+)**.** *We can not interpret the tree as a Poisson point process if the origin is unknown and we assign a uniform prior onto* t_{or}

(See Gernhard [5] Remark 6.1)

Proof. Recall we obtained the density function for the order statistics of speciation times, $x_1, \ldots, x_{n-1}$, obtained in 4. We got the result

$$f(x|t_1, n) = (n-1)! \prod_{i=1}^{n-1} \mu \frac{p_1(x_i)}{p_0(t)}$$

Using Equations 3.3 and 3.4, this can be written as

$$f(x|t,n) = (n-1)! \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \cdot \frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \tag{8.1}$$

In section 4 we proved that unordered speciation times were iid, by dividing by $(n-1)!$. We will attempt this without conditioning on $t = t_{or}$ and assuming a uniform prior.

$$f(x|n) = \int_{x_1}^{\infty} f(x|t,n) q_{or}(t|n) dt \quad \text{by Law of Total Probability}$$

Then using Equations above and Equation 5.1 for $q_{or}(t|n)$

$$= \int_{x_1}^{\infty} (n-1)! \frac{(1-e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} n\lambda^n (\lambda-\mu)^2$$

$$\times \prod_{i=1}^{n-1} \left\{ \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \cdot \frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right\} dt$$

We simplify to try and achieve this by first moving terms outside the integral

$$= n!\lambda^n (\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\}$$

$$\times \int_{x_1}^{\infty} \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{n-1}}{(1-e^{-(\lambda-\mu)t})^{n-1}} \frac{(1-e^{-(\lambda-\mu)t})^{n-1}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} e^{-(\lambda-\mu)t} dt$$

Then we match common terms in the integral

$$= n!\lambda^n (\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\}$$

$$\times \int_{x_1}^{\infty} \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{n-1-n-1}}{(1-e^{-(\lambda-\mu)t})^{n-1-n+1}} e^{-(\lambda-\mu)t} dt$$

$$= n!\lambda^n (\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\}$$

$$\times \int_{x_1}^{\infty} \frac{(\lambda-\mu e^{-(\lambda-\mu)t})^{-2}}{(1-e^{-(\lambda-\mu)t})^{0}} e^{-(\lambda-\mu)t} dt$$

Cancellation gives the following result

$$= n!\lambda^n (\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \int_{x_1}^{\infty} \frac{e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2} dt$$

Then assuming $\mu \neq 0$ we can solve by substitution, let $z = \lambda - \mu e^{-(\lambda-\mu)t}$

$$\implies \frac{dt}{dz} = \mu(\lambda-\mu)e^{-(\lambda-\mu)t} \implies dt = \frac{dz}{\mu(\lambda-\mu)e^{-(\lambda-\mu)t}}$$

Then finding the end points $z_1 = \lambda - \mu e^{-(\lambda-\mu)x_1}$; $z_2 = \lambda - \mu e^{-(\lambda-\mu)\infty} = \lambda$

Assuming $0 \leq \mu \leq \lambda < \infty$

Step by step we then solve the integral

$$\implies f(x|n) = n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\}$$

$$\times \int_{(\lambda-\mu e^{-(\lambda-\mu)x_1})^2}^{\lambda^2} \frac{e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2} \times \frac{dz}{\mu(\lambda-\mu)e^{-(\lambda-\mu)t}}$$

$$= n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \int_{\lambda-\mu e^{-(\lambda-\mu)x_1}}^{\lambda} \frac{1}{z^2} \times \frac{dz}{\mu(\lambda-\mu)}$$

Next we plug in the bounds of the integral to get the following

$$= n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \left[\frac{-1}{z\mu(\lambda-\mu)} \right]_{\lambda-\mu e^{-(\lambda-\mu)x_1}}^{\lambda}$$

$$= n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \left[\frac{1}{\mu(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)x_1})} - \frac{1}{\lambda\mu(\lambda-\mu)} \right]$$

Further simplification leads to

$$= n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\}$$

$$\times \left[\frac{\lambda}{\lambda\mu(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)x_1})} - \frac{(\lambda-\mu e^{-(\lambda-\mu)x_1})}{\lambda\mu(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)x_1})} \right]$$

$$= n!\lambda^n(\lambda-\mu)^2 \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \left[\frac{\lambda-\lambda+\mu e^{-(\lambda-\mu)x_1}}{\lambda\mu(\lambda-\mu)(\lambda-\mu e^{-(\lambda-\mu)x_1})} \right]$$

$$= n!\lambda^{n-1}(\lambda-\mu) \left\{ \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \left[\frac{e^{-(\lambda-\mu)x_1}}{\lambda-\mu e^{-(\lambda-\mu)x_1}} \right]$$

$$= n!\lambda^{n-1}(\lambda-\mu) \frac{e^{-(\lambda-\mu)x_1}}{\lambda-\mu e^{-(\lambda-\mu)x_1}} \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2}$$

If the joint probability of ordered speciation times were independent we would expect the result to be some factorial times a product, for example

$$f(x|n) = (n-1)! \prod_{i=1}^{n-1} f(x_i|n)$$

However this is not the case, therefore we can not prove there is independence between speciation times when conditioning on t_{or} and assuming a uniform prior. This result is true in all cases, general λ, μ, Yule case and the critical case. However, we know from Section 2.2 that the speciation times are identically distributed. So while we don't have iid speciation times, we do have identical speciation times.

Given we don't not have independent speciation times, we can not write the process as a Poisson point process, because points have to be independent. This completes the proof. □

Note that if we were to again condition that $x_1 = t_{or}$, the first speciation time is the origin time, we would have independent speciation times. This is how we proved independence in Section 4.

We add a more detailed proof to the following corollary from Gernhard [5] Section 6.

Corollary 8.1.1 (+)**.** *Alternatively, if we set $x_0 := t_{or}$, that is the 0th order statistic is equal to the origin time, then we have iid points*

(See Gernhard [5] Section 6)

Proof.

$$f(t, x_1, \ldots, x_{n-1}|n) = f(x_1, \ldots, x_{n-1}|t, n) q_{or}(t|n)$$

by Conditional Probability $\quad \mathbb{P}(A; B) = \mathbb{P}(A|B)\mathbb{P}(B)$

$$\iff f(x_0, x_1, \ldots, x_{n-1}|n) = f(x_1, \ldots, x_{n-1}|t, n) q_{or}(t|n)$$

Then we plug in the corresponding values from Equation 5.1 and Equation 8.1 to get

$$\begin{aligned} f(t, x_1, \ldots, x_{n-1}|n) = (n-1)! \frac{(1-e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} n\lambda^n (\lambda-\mu)^2 \\ \times \prod_{i=1}^{n-1} \left\{ \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \cdot \frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right\} \end{aligned}$$

Then we isolate terms in product

$$\begin{aligned} = n!\lambda^n(\lambda-\mu)^2 \frac{(1-e^{-(\lambda-\mu)t})^{n-1} e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}} \left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)^{n-1} \\ \times \prod_{i=1}^{n-1} \left\{ \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2} \right\} \end{aligned}$$

$$= n!\lambda^n\lambda^{n-1}(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)t})^{n-1-n+1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1-n+1}}$$
$$\times\prod_{i=1}^{n-1}\left\{\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2}\right\}$$

Then we can simplify further

$$= n!\lambda\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^2}\prod_{i=1}^{n-1}\left\{\lambda\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2}\right\}$$
$$= n!\lambda\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)x_0}}{(\lambda-\mu e^{-(\lambda-\mu)x_0})^2}\prod_{i=1}^{n-1}\left\{\lambda\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2}\right\}\quad\text{given } t := x_0$$
$$= n!\prod_{i=0}^{n-1}\left\{\lambda\frac{(\lambda-\mu)^2e^{-(\lambda-\mu)x_i}}{(\lambda-\mu e^{-(\lambda-\mu)x_i})^2}\right\}$$

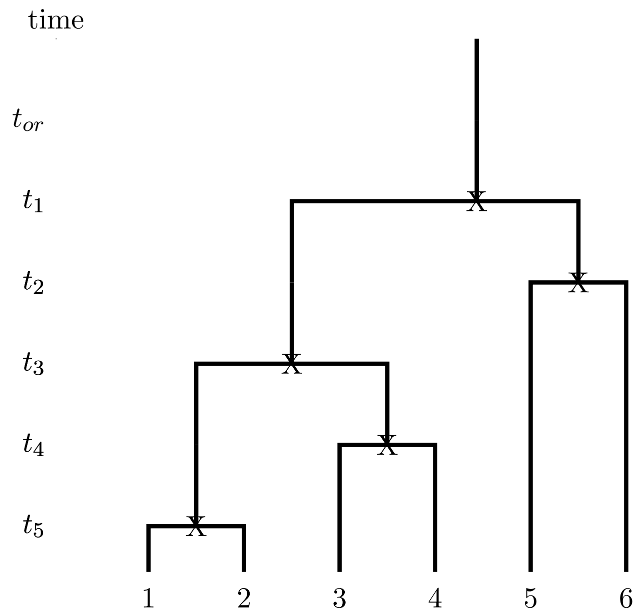
Therefore we have the ordered speciation times, $x_0, x_1, \ldots, x_{n-1}$ are n iid random variables, if we set $x_0 := t$ and assume a uniform prior for t.

They are iid because we have an $n!$ on the outside corresponding to the number of combinations we can order the times, the product indicates independence, and they all have the same density. □

8.2 Point Process of Coalescent

In the original birth death process, species split (w.r. λ) and go extinct (w.r. μ). In the reconstructed process, which we have been examining in this paper, we condition on the fact there are n extant species, or n species at the end of the process. In the following section, we generate a new process. We set the beginning of the process to be the present, and have species coalesce. This is a coalescent model, which is the standard neutral model for population genetics [5].

Review the following graph of a point process, where $0 < t_5 < \ldots < t_1 < t_{or}$.



Instead of starting at t_{or} and having speciation events, we start at the present time, 0, and have coalescent events. For example, in this new process, species 1 and 2 exist and then coalesce at random time t_5, and at exact point $(1.5, t_5)$. Then the next set of species coalesce, and so on.

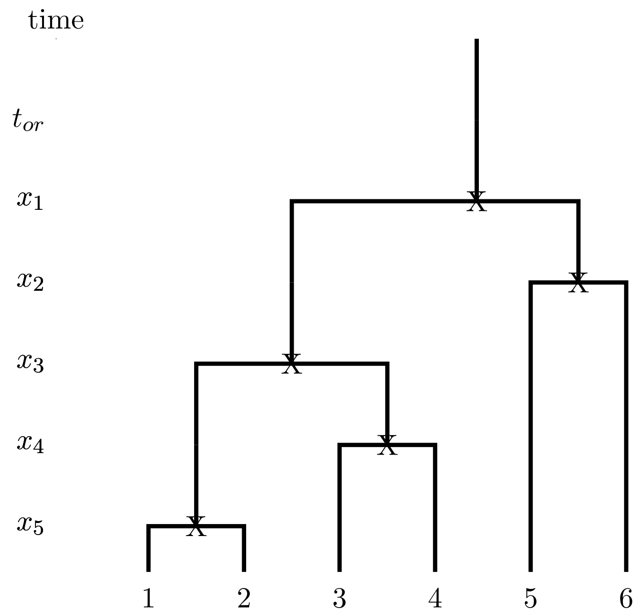
Firstly, the rate at which the first two species coalesce is Poisson distributed with rate $\binom{n}{2}\lambda$. λ, because that is the rate for speciation/coalescence. $\binom{n}{2}$, because we are choosing two species from our n extant species to coalesce. Hence the time taken from the present to the first coalescent event has distribution exponential$\left(\binom{n}{2}\lambda\right)$. We intend to show that this process does not have a Poisson process representation with iid coalescent points [5].

We also add to the following proof of the theorem described in Gernhard [5].

Theorem 8.2 (+)**.** *The coalescent process does not have a Poisson process representation with iid coalescent points.*

(See Gernhard [5] Section 6)

Proof. Let $x = (x_1, x_2, \ldots, x_{n-1})$ be the order statistic of the coalescent times. Where $x_1 > x_2 > \ldots > x_{n-1}$. Here we have the time between two coalescent events, $x_i - x_{i+1}$, distributed exponentially with rate $\binom{i+1}{2}\lambda$. See the graph below for more detail, note the time axis (y-axis).



Hence we can find the distribution for the joint order statistic by rewriting them as differences.

$f(x|n) = f(x_1, x_2, \ldots, x_{n-1}|n)$

Then we can rewrite as the following, because they are equivalent

$$= f(x_1 - x_2, \ldots, x_{n-2} - x_{n-1}, x_{n-1} - 0|n)$$

$$= f(x_{n-1}|n) \prod_{i=1}^{n-2} f(x_i - x_{i+1}|n) \quad \text{by independence of exponential random times}$$

$$\iff f(x|n) = \left(\lambda \binom{n}{2} e^{-\lambda\binom{n}{2}x_{n-1}}\right) \prod_{i=1}^{n-2} \lambda \binom{i+1}{2} e^{-\lambda\binom{i+1}{2}(x_i - x_{i+1})}$$

Then we expand the binomial coefficients to simplify

$$= \lambda \frac{n!}{2!(n-2)!} exp\left\{-\lambda \frac{n!}{2!(n-2)!}\right\} \prod_{i=1}^{n-2} \lambda \frac{(i+1)!}{2!(i-1)!} exp\left\{-\lambda \frac{(i+1)}{2!(i-1)!}(x_i - x_{i+1})\right\}$$

$$= \lambda \frac{n(n-1)}{2} e^{-\lambda\frac{n(n-1)}{2}x_{n-1}} \prod_{i=1}^{n-2} \lambda \frac{(i+1)i}{2} \frac{e^{-\lambda\frac{(i+1)i}{2}x_i}}{e^{-\lambda\frac{(i+1)i}{2}x_{i+1}}}$$

$$= \frac{n!(n-1)!}{2^{n-1}} e^{-\lambda \frac{n(n-1)}{2} x_{n-1}} \prod_{i=1}^{n-2} \frac{e^{-\lambda \frac{(i+1)i}{2} x_i}}{e^{-\lambda \frac{(i+1)i}{2} x_{i+1}}}$$

Then note that in the product the denominator of term i will cancel most of the numerator of $i+1$ leaving only the value i

For example take $i = 1, 2, 3, 4 \implies \frac{e^{-\lambda \frac{2 \cdot 1}{2} x_1}}{e^{-\lambda \frac{2 \cdot 1}{2} x_2}} \times \frac{e^{-\lambda \frac{3 \cdot 2}{2} x_2}}{e^{-\lambda \frac{3 \cdot 2}{2} x_3}} \times \frac{e^{-\lambda \frac{4 \cdot 3}{2} x_3}}{e^{-\lambda \frac{4 \cdot 3}{2} x_4}} \times \frac{e^{-\lambda \frac{5 \cdot 4}{2} x_4}}{e^{-\lambda \frac{5 \cdot 4}{2} x_5}} \ldots$

$$= e^{-\lambda x_1} e^{-\lambda (3-1) x_2} e^{-\lambda (6-3) x_3} e^{-\lambda (10-6) x_4} \ldots$$

$$= e^{-\lambda x_1} e^{-\lambda 2 x_2} e^{-\lambda 3 x_3} e^{-\lambda 4 x_4} \ldots$$

$$\implies f(x|n) = \frac{n!(n-1)!}{2^{n-1}} \prod_{i=1}^{n-1} \lambda e^{-\lambda i x_i}$$

Then suppose we condition on the most recent common ancestor, x_1, then we get

$$f(x|n, x_1) = \frac{f(x|n)}{f(x_1|n)} \quad \text{By Conditional Probability}$$

$$= \frac{n!(n-1)!}{f(x_1|n) 2^{n-1}} \prod_{i=1}^{n-2} \lambda e^{-\lambda i x_i}$$

$$= h(x_1, n) \prod_{i=1}^{n-2} \lambda e^{-\lambda i x_i}$$

Where h is some function of x_1 and n . If the $n-2$ coalescent points were iid with some desnity function g, we would have $f(x|n, x_1) = (n-2)! \prod_{i=1}^{n-2} g(x_i, x_1, n)$. But, given there is a value i present in $e^{-\lambda i x_i}$, this property is not satisfied, hence the $n-2$ points are not iid. However, we still have that each rank orientated tree shape is equally likely, see Chapter 2.3. This means that each permutation of s_i has the same probability. Therefore we have that each s_i is identically distributed, but not independent. This completes the proof.

□

8.3 Backwards Conditioned Birth Death Process

In the forwards process, we have extant species speciate and die with exponential waiting times. We also condition on the n extant species, that exist in the present. In the backwards, which we will examine now, species coalesce with exponential waiting times.

8.3.1 Under the cBDP, the Probability Density Function for Time Two Species Coalesce is $f(s|t)$ from Chapter 4

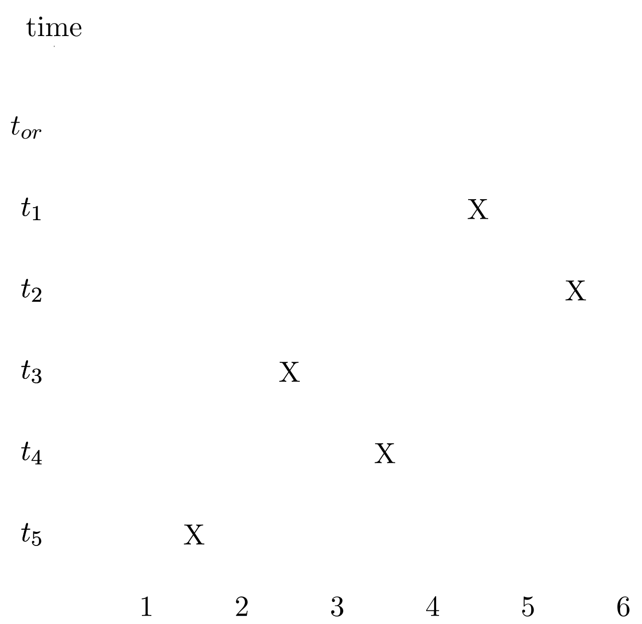
Theorem 8.3 (+). *Under the cBDP, the Probability Density Function for Time Two Species Coalesce is $f(s|t)$ from Chapter 4. Coalescent times are also Independent.*

(See Gernhard [5] Theorem 6.2)

$$f(s|t_{or}=t) = \begin{cases} \left(\frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)s}}{(\lambda-\mu e^{-(\lambda-\mu)s})^2}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) & \text{if } s \leq t \\ 0 & \text{otherwise} \end{cases} \tag{8.2}$$

$$F(s|t_{or}=t) = \begin{cases} \left(\frac{1-e^{-(\lambda-\mu)s}}{\lambda-\mu e^{-(\lambda-\mu)s}}\right)\left(\frac{\lambda-\mu e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}}\right) & \text{if } s \leq t \\ 1 & \text{otherwise} \end{cases} \tag{8.3}$$

Proof. This proof is intuitive. Think of the process graphically, and consider the Poisson point process below. Species coalesce at points $(i+1/2, s_i)$. For example the first coalescent point is $(1.5, t_5)$. These coalescent points on the Poisson point process are the same points as the forward process. We derived and proved the probability density for the forward speciation times were $f(s|t)$, in Equation 4.1. We also proved that speciation times are identically distributed and independent. Therefore the backwards coalesce points also has the density function $f(s|t)$ from Equation 4.1, given they are the same points. This completes the proof. □



8.3.2 Calculating Time Between kth and $k+1$th Speciation Event

Our goal in this section is to find the density for the time between the kth and $k+1$th speciation event in the backwards process, not that $k+1$th event is closer to 0, the present.

The theorem below is displayed on Pg. 775, Section 6 of Gernhard [5], there is a minimal proof provided, so we prove this result in full.

Theorem 8.4 (+)**.** *The probability density of the time difference between the kth speciation event and the $k+1$th speciation event is:*

(See Gernhard [5] Section 6)

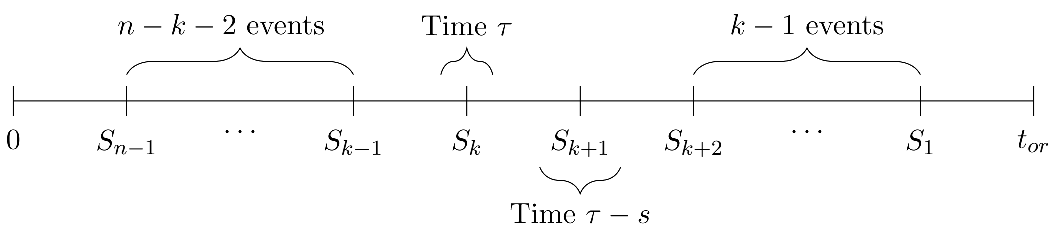
$$f_{\mathscr{A}^{k}_{n,t}-\mathscr{A}^{k+1}_{n,t}} = \int_{s}^{t} (n-1)(n-2)\binom{n-3}{k-1}(1-F(\tau))^{k-1}F(\tau-s)^{n-k-2}f(\tau)f(\tau-s)d\tau \qquad (8.4)$$

Proof. First note that since the $n-1$ points in the point process are iid, with a density $f(s|t)$, as explained in the theorem prior. The joint density of two time points, j_1 and j_2 is as follows

$$g(s_{j_1}, s_{j_2}|t) = f(s_{j_1}|t)f(s_{j_2}|t)$$

This is by independence.

Next we need to examine the timeline of events so it clear what we have to account for in determining our density function:



So we have four separate set of events we need to account for. Those before the k, the kth event, the $k+1$th event and those after the $k+1$th event.

We assume the kth speciation event occurs at some time τ and the $k+1$th speciation event occurs at some time $\tau - s$. Given there are $n-1$ speciation events, we have $n-1$ possibilities choosing the points for the kth speciation event. We also have $n-2$ possibilities to choose the point for the $k+1$th speciation event. Therefore we can write the joint density function as

$$g(\tau, \tau - s|t) = (n-1)(n-2)f(\tau|t)f(\tau - s|t)$$

The probability we have some $k-1$ speciation points of the remaining $n-3$ points being earlier than τ is

$$\binom{n-3}{k-1}(1-F(\tau))^{k-1}$$

Because:

First note that we are choosing $k-1$ objects from $n-3$ total objects,
So we include the binomial coefficient to account for all possible orderings

$$\mathbb{P}(k-1 \text{ points of remaining } n-3 > \tau) = \binom{n-3}{k-1}\mathbb{P}(s_1, \ldots, s_{k-1} > \tau)$$

Next we account for independence between speciation events, see Corollary 4.3.1

$$= \binom{n-3}{k-1}\mathbb{P}(s_1 > \tau) \cdots \mathbb{P}(s_{k-1} > \tau)$$

Then we account for the fact speciation times are identically distributed

$$= \binom{n-3}{k-1}\prod_{i=1}^{k-1}(1-\mathbb{P}(s_i \leq \tau))$$
$$= \binom{n-3}{k-1}(1-F(\tau))^{k-1}$$

Next the probability that the remaining $n-k-2$ speciation points happened after $\tau - s$ is

$$F(\tau - s)^{n-k-2}$$

By a similar process above:

$$\mathbb{P}(s_{k+2}, \ldots, s_{n-1} \leq \tau - s) = \prod_{i=k+2}^{n-3}\mathbb{P}(s_i \leq \tau - s) = F(\tau - s)^{n-k-2}$$

Note we don't have a binomial coefficient because we have already *chosen* the points to be less than τ. We are now selecting all the remaining points, rather than choosing them.

Hence we need to combine the results to find the density: $f_{\mathscr{A}_{n,t}^{k} - \mathscr{A}_{n,t}^{k+1}}(s)$. Our initial result is intuitively, the probability $s_1, \ldots, s_{k+2}$ occur before the kth event (τ), the probability that the kth event occurs at τ, the $k+1$th event occurs at $\tau - s$ and the probability the $s_{k-2}, \ldots, s_1$ occur after $\tau - s$. We combine results below and separate probabilities given independence. Then we integrate out τ, so our difference is a function of s.

$$
\begin{aligned}
f_{\mathscr{A}^k_{n,t}-\mathscr{A}^{k+1}_{n,t}}(s) &= \int_s^t (n-1)(n-2)\binom{n-3}{k-1} \\
&\quad \times \mathbb{P}(s_{n-1},\ldots,s_{k+1} \leq \tau, s_k = \tau, s_{k-1} = \tau - s, s_1,\ldots,s_{k-1} > \tau - s)d\tau \\
&= \int_s^t (n-1)(n-2)\binom{n-3}{k-1}\mathbb{P}(s_{n-1},\ldots,s_{k+1} \leq \tau)\mathbb{P}(s_k = \tau) \\
&\quad \times \mathbb{P}(s_{k-1} = \tau - s)\mathbb{P}(s_1,\ldots,s_{k-1} > \tau - s)d\tau \\
&= \int_s^t (n-1)(n-2)\binom{n-3}{k-1}(1-F(\tau))^{k-1}F(\tau-s)^{n-k-2}f(\tau)f(\tau-s)d\tau
\end{aligned}
$$

This completes the proof

□

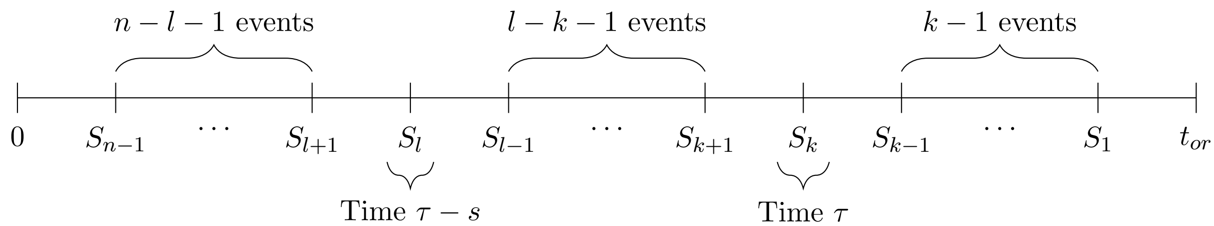
The previous theorem assumes the $k+1$th event comes after the kth event. We can also assume that some speciation event k comes before some general speciation event l. Then we get a different result, the takes into account the speciation events the occur between event k and event l. Note that, if the kth event comes before lth event $k < l$. But the speciation times have the quality $S_k > S_l$, because *before* means further from the present, 0, and closer to the origin, t_{or}.

The theorem below is also present in Gernhard [5], but only a sketch proof is provided, we provide the full proof below.

Theorem 8.5 (+). *The probability density function for the time difference between the kth speciation event and the lth speciation event ($l > k$) is*

$$
\begin{aligned}
f_{\mathscr{A}^k_{n,t}-\mathscr{A}^l_{n,t}}(s) &= \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1} \\
&\quad \times (1-F(\tau|t))^{k-1}(F(\tau|t)-F(\tau-s|t))^{l-k-1}F(\tau-s|t)^{n-l-1} \\
&\quad \times f(\tau|t)f(\tau-s|t)d\tau
\end{aligned}
\tag{8.5}
$$

Proof. In this scenario, we need to account for a few things. All events that happened before l, the events between l and k, the events after k, then also the event k and l. Therefore inside our density function, we have five clear probabilities to account for, see the timeline below.



All these events and joint events are independent. So we take the product of the probability for each of these events and integrate out τ because we want to have a function of s.

First we also have to account for the total combinations of events, s_k can be any number of $n-1$ events, S_l can be any number of $n-2$ events (removing the kth event). Then we choose $k-1$ events to occur after k from a remaining $n-3$ events, which gives $\binom{n-3}{k-1}$. Then we also account for all the events between l and k. We choose $l-k-1$ events from a remaining $n-k-2$ events, giving $\binom{n-k-2}{l-k-1}$. This leaves $n-l-1$ events left to order from a total remaining $n-l-1$ events, which has one total combination.

Hence we multiply our density by

$$(n-1)(n-2)\binom{n-3}{k-1}\binom{n-k-2}{l-k-1}$$

to account for all possible orderings of speciation times.

- The $n-l-1$ events that occur after l, closer to the present, are independent of one another and have a common distribution, $\mathbb{P}(S_i \leq S_l) = \mathbb{P}(S_i \leq \tau - s) = F(\tau - s|t)$. Then given independence the joint distribution they occur before S_l is $F(\tau|t)^{n-l-1}$
- Next, the $k-1$ events that occur before k and closer to the origin are also iid and have a distribution of $\mathbb{P}(S_i > S_k) = 1 - \mathbb{P}(S_i \leq S_k) = \mathbb{P}(S_i \leq \tau - s) = 1 - F(\tau)$. This implies the joint distribution they occur after S_k is$(1 - F(\tau|t))^{k-1}$
- The density that the kth speciation event occurs at τ is just the density function $f(\tau|t)$
- The density that the lth speciation event occurs at $\tau - s$ is the density function $f(\tau - s|t)$
- Finally, the distribution of events that occurs between S_k and S_l are iid with common distribution, $\mathbb{P}(S_l \leq S_i \leq S_k) = \mathbb{P}(S_i \leq S_k) - \mathbb{P}(S_i \leq S_l) = \mathbb{P}(S_i \leq \tau) - \mathbb{P}(S_i \leq \tau - s) = F(\tau|t) - F(\tau - s|t)$. Then the joint distribution they all occur between the kth and lth events is $(F(\tau|t) - F(\tau - s|t))^{l-k-1}$

Now we can combine these results to get a general formula for $f_{\mathscr{A}^k_{n,t} - \mathscr{A}^l_{n,t}}(s)$

$$f_{\mathscr{A}_{n,t}^{k}-\mathscr{A}_{n,t}^{l}}(s) = \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1}$$
$$\times \mathbb{P}(S_{n-1} \leq \ldots \leq S_{l+1} \leq S_l \leq S_{l-1} \leq \ldots \leq S_{k+1} \leq S_k \leq S_{k-1} \leq \ldots \leq S_1 | n, t) d\tau$$

Then by independence of speciation times (see Corollary 4.3.1), we have the following

$$= \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1}\mathbb{P}(S_{n-1}, \ldots, S_{l+1} \leq S_l | n, t)$$
$$\times \mathbb{P}(S_l \leq S_{l-1}, \ldots, S_{k+1} \leq S_k | n, t)\mathbb{P}(S_{k+1}, \ldots, S_1 > S_k | n, t) d\tau$$
$$= \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1}\mathbb{P}(S_{n-1}, \ldots, S_{l+1} \leq \tau - s | n, t)$$
$$\times \mathbb{P}(\tau - s \leq S_{l-1}, \ldots, S_{k+1} \leq \tau | n, t)\mathbb{P}(S_{k+1}, \ldots, S_1 > \tau | n, t)$$
$$\times \mathbb{P}(S_k = \tau)\mathbb{P}(S_l = \tau - s) d\tau$$

then by independence and identical distribution we have

$$= \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1}F(\tau - s|t)^{n-l-1}$$
$$\times (F(\tau|t) - F(\tau - s|t))^{l-k-1}(1 - F(\tau|t))^{k-1}f(\tau|t)f(\tau - s|t)d\tau$$
$$= \int_s^t (n-1)(n-2)\binom{n-2}{k-1}\binom{n-k-2}{l-k-1}$$
$$\times (1 - F(\tau|t))^{k-1}(F(\tau|t) - F(\tau - s|t))^{l-k-1}F(\tau - s|t)^{n-l-1}$$
$$\times f(\tau|t)f(\tau - s|t)d\tau \quad \text{as desired}$$

This completes the proof

□

8.4 Backwards Process of the Yule Model

8.4.1 Known Tree Age

Under the Yule Model, we can calculate the time between any two speciation events in the reconstructed tree analytically. The time between the kth event and the lth event $l > k$ given the time between the nth speciation event (today) and the origin of the tree.

The following theorem is not proved by Gernhard, but rather just printed, we give the full proof below.

Theorem 8.6 (+)**.** *Under the Yule model, where $\mu = 0$, if we assume the tree age is known, the probability density function for the time difference between the kth speciation event and the lth* *speciation event* $(l > k)$ *is:*

$$f_{\mathscr{A}^k_{n,t}-\mathscr{A}^l_{n,t}}(s) = \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} B_{i,j} e^{\lambda(n-l)s} \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda(n-k+i)(t-s)} - e^{\lambda j(t-s)}\right)$$

$$\textit{Where } B_{i,j} = k(k+1)\binom{l}{k+1}\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j} \tag{8.6}$$

(See Gernhard [5] Section 6)

Proof. Firstly we can simplify

$$(n-1)(n-2)\binom{n-3}{k-1}\binom{n-k-2}{l-k-1}F(\tau-s|t)^{n-l-1}$$

to match Gernhards

$$\begin{aligned}
(n-1)(n-2)\binom{n-3}{k-1}\binom{n-k-2}{l-k-1} &= \frac{(n-1)(n-2)(n-3)!(n-k-2)!}{(k-1)!(n-3-k+1)!(l-k-1)!(n-k-2-l+k+1)!} \\
&= \frac{(n-1)!(n-k-2)!}{(k-1)!(n-k-2)!(l-k-1)!(n-l-1)!} \\
&= \frac{(n-1)!}{(k-1)!(l-k-1)!(n-l-1)!} \times \frac{l!}{l!} \times \frac{k(k+1)}{k(k+1)} \\
&= k(k+1) \times \frac{l!}{(k+1)!(l-k-1)!} \times \frac{(n-1)!}{l!(n-l-1)!} \\
&= k(k+1)\binom{l}{k+1}\binom{n-1}{l}
\end{aligned}$$

$$\begin{aligned}
f_{\mathscr{A}^k_{n,t}-\mathscr{A}^l_{n,t}}(s) = \int_s^t & (n-1)(n-2)\binom{n-3}{k-1}\binom{n-k-2}{l-k-1} \\
& \times (1-F(\tau|t))^{k-1}(F(\tau|t)-F(\tau-s|t))^{l-k-1}F(\tau-s|t)^{n-l-1} \\
& \times f(\tau|t)f(\tau-s|t)d\tau
\end{aligned}$$

Then we plug in values for $F(s|t)$ from Equations 4.6 and 4.7 and the above simplification

$$\implies f_{\mathscr{A}^k_{n,t}-\mathscr{A}^l_{n,t}}(s) = \int_s^t k(k+1)\binom{l}{k+1}\binom{n-1}{l}$$

$$\times \left(1 - \frac{1-e^{-\lambda\tau}}{1-e^{-\lambda t}}\right)^{k-1} \left(\frac{1-e^{-\lambda\tau}}{1-e^{-\lambda t}} - \frac{1-e^{-\lambda(\tau-s)}}{1-e^{-\lambda t}}\right)^{l-k-1} \left(\frac{1-e^{-\lambda(\tau-s)}}{1-e^{-\lambda t}}\right)^{n-l-1}$$

$$\times \left(\frac{\lambda e^{-\lambda\tau}}{1-e^{-\lambda t}}\right)\left(\frac{\lambda e^{-\lambda(\tau-s)}}{1-e^{-\lambda t}}\right) d\tau$$

After plugging the formulas, we now have to simplify everything

$$= \int_s^t k(k+1)\binom{l}{k+1}\binom{n-1}{l}$$

$$\times \left(\frac{1-e^{-\lambda t}-1+e^{-\lambda\tau}}{1-e^{-\lambda t}}\right)^{k-1} \left(\frac{e^{-\lambda(\tau-s)}-e^{-\lambda\tau}}{1-e^{-\lambda t}}\right)^{l-k-1} \left(\frac{1-e^{-\lambda(\tau-s)}}{1-e^{-\lambda t}}\right)^{n-l-1}$$

$$\times \left(\frac{\lambda e^{-\lambda\tau}}{1-e^{-\lambda t}}\right)\left(\frac{\lambda e^{-\lambda(\tau-s)}}{1-e^{\lambda t}}\right) d\tau$$

Here we multiply brackets and extract common terms

$$= \int_s^t k(k+1)\binom{l}{k+1}\binom{n-1}{l}$$

$$\times \left(\frac{e^{-\lambda\tau}-e^{-\lambda t}}{1-e^{-\lambda t}}\right)^{k-1} \left(\frac{e^{-\lambda\tau}(e^{\lambda s}-1)}{1-e^{-\lambda t}}\right)^{l-k-1} \left(\frac{1-e^{-\lambda(\tau-s)}}{1-e^{-\lambda t}}\right)^{n-l-1}$$

$$\times \left(\frac{\lambda^2}{(1-e^{-\lambda t})^2}\right)\left(e^{-\lambda(2\tau-s)}\right) d\tau$$

Next we isolate the relevant integral parts inside the integral

$$= \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\left(\frac{(e^{\lambda s}-1)^{l-k-1}}{(1-e^{-\lambda t})^{n-l-1+k-1+l-k-1+2}}\right)$$

$$\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k-1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1} d\tau$$

$$= \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\frac{(e^{\lambda s}-1)^{l-k-1}}{(1-e^{-\lambda t})^{n-1}}$$

$$\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k-1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}\left(e^{-\lambda\tau}\right)^2\left(e^{\lambda s}\right) d\tau$$

Then we apply the binomial theorem around: $\left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}$

$$= \sum_{i=0}^{k-1}\binom{k-1}{i}\left(e^{-\lambda\tau}\right)^i(-1)^{k-1-i}\left(e^{-\lambda t}\right)^{k-1-i}$$

We can also apply the binomial theorem around: $\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}$

$$= \sum_{j=0}^{n-l-1} \binom{n-l-1}{j} (-1)^{n-l-1-j} \left(e^{-\lambda(\tau-s)}\right)^{n-l-1-j}$$

We then add these into our equation, so we have

$$f_{\mathscr{A}^k_{n,t}-\mathscr{A}^l_{n,t}}(s) = \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\frac{(e^{\lambda s}-1)^{l-k-1}}{(1-e^{-\lambda t})^{n-1}} \times \frac{\left(e^{\lambda t}\right)^{n-1}}{\left(e^{\lambda t}\right)^{n-1}}$$
$$\times \int_s^t \sum_{i=0}^{k-1} \binom{k-1}{i} \left(e^{-\lambda\tau}\right)^i (-1)^{k-1-i} \left(e^{-\lambda t}\right)^{k-1-i}$$
$$\times \left(e^{-\lambda\tau}\right)^{l-k-1} \sum_{j=0}^{n-l-1} \binom{n-l-1}{j} (-1)^{n-l-1-j} \left(e^{-\lambda(\tau-s)}\right)^{n-l-1-j} \left(e^{-\lambda\tau}\right)^2 \left(e^{\lambda s}\right) d\tau$$

Next we combine like terms as seen below

$$= \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} \binom{k-1}{i}\binom{n-l-1}{j} (-1)^{k-1-i}(-1)^{n-l-1-j} \left(e^{-\lambda t}\right)^{k-1-i}$$
$$\times \int_s^t \left(e^{-\lambda\tau}\right)^i \left(e^{-\lambda\tau}\right)^{l-k-1} \left(e^{-\lambda(\tau-s)}\right)^{n-l-1-j} \left(e^{-\lambda\tau}\right)^2 \left(e^{\lambda s}\right) d\tau$$

Then we try to remove everything except τ terms from the integral

$$= \lambda^2 \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} k(k+1)\binom{l}{k+1}\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}$$
$$\times (-1)^{n+k-l-i-j}(-1)^{-2}\frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-1} \left(e^{-\lambda t}\right)^{k-1-i}$$
$$\times \int_s^t \left(e^{-\lambda\tau}\right)^{l-k-1+i+2} \left(e^{-\lambda(\tau-s)}\right)^{n-l-1-j} e^{\lambda s} d\tau$$

Next we combine like terms with powers

$$= \lambda^2 \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} k(k+1)\binom{l}{k+1}\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}(-1)^{n+k-l-i-j}$$
$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-1-k+1+i}$$
$$\times \int_s^t \left(e^{-\lambda\tau}\right)^{l-k-1+i+2} \left(e^{-\lambda\tau}\right)^{n-l-1-j} \left(e^{\lambda s}\right)^{n-l-1-j+1} d\tau$$

$$= \lambda^2 \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-k+i} \left(e^{\lambda s}\right)^{n-l-j} \times \int_s^t \left(e^{-\lambda \tau}\right)^{l-k-1+i+n-l-1-j+2} d\tau$$

Now we completely isolate the integral term, so we can solve it

$$= \lambda^2 \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-k+i} \left(e^{\lambda s}\right)^{n-l-j} \times \int_s^t \left(e^{-\lambda \tau}\right)^{n-k+i-j} d\tau$$

Now we solve the integral to get

$$= \lambda^2 \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-k+i} \left(e^{\lambda s}\right)^{n-l-j} \times \left[-\frac{\left(e^{-\lambda \tau}\right)^{n-k+i-j}}{\lambda(n-k+i-j)} \right]_s^t$$

Solving for the integral bounds gives us

$$= \lambda^2 \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda t}\right)^{n-k+i} \left(e^{\lambda s}\right)^{n-l-j} \times \left[\frac{\left(e^{-\lambda s}\right)^{n-k+i-j} - \left(e^{-\lambda t}\right)^{n-k+i-j}}{\lambda(n-k+i-j)} \right]$$

In the next steps we try to isolate like terms to simplify the equation

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda s}\right)^{n-l} \left(e^{\lambda t}\right)^{n-k+i} \left(e^{-\lambda s}\right)^{j} \times \left[\frac{\left(e^{-\lambda s}\right)^{n-k+i-j} - \left(e^{-\lambda t}\right)^{n-k+i-j}}{n-k+i-j} \right]$$

Here we attempt to simplify so more multiplying variables into the square brackets as shown

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda s}\right)^{n-l}$$

$$\times \left[\frac{\left(e^{\lambda t}\right)^{n-k+i} \left(e^{-\lambda s}\right)^{j} \left(e^{-\lambda s}\right)^{n-k+i-j} - \left(e^{\lambda t}\right)^{n-k+i} \left(e^{-\lambda s}\right)^{j} \left(e^{-\lambda t}\right)^{n-k+i-j}}{n-k+i-j} \right]$$

We then simplify more by finding like terms inside the bracket

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda s}\right)^{n-l} \left[\frac{\left(e^{\lambda t}\right)^{n-k+i} \left(e^{-\lambda s}\right)^{n-k+i} - \left(e^{-\lambda s}\right)^{j} \left(e^{-\lambda t}\right)^{-j}}{n-k+i-j} \right]$$

Combine like terms in the bracket

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda s}\right)^{n-l} \left[\frac{\left(e^{\lambda (t-s)}\right)^{n-k+i} - \left(e^{\lambda (t-s)}\right)^{j}}{n-k+i-j} \right]$$

Then we turn exponents into multiplication to match Gernhard's [5] notation

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} (-1)^{n+k-l-i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda s}\right)^{n-l} \left[\frac{\left(e^{\lambda (n-k+i)(t-s)}\right) - \left(e^{\lambda j(t-s)}\right)}{n-k+i-j} \right]$$

Reorganise equation to get the final result

$$= \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} \frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$

$$\times \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda (n-l) s}\right) \left(e^{\lambda (n-k+i)(t-s)} - e^{\lambda j(t-s)}\right)$$

Then let $B_{i,j} = k(k+1) \binom{l}{k+1} \binom{n-1}{l} \binom{k-1}{i} \binom{n-l-1}{j} \frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$

$$\implies f_{\mathscr{A}_{n,t}^{k} - \mathscr{A}_{n,t}^{l}}(s) = \lambda \sum_{i=0}^{k-1} \sum_{j=0}^{n-l-1} B_{i,j} e^{\lambda(n-l)s} \frac{(e^{\lambda s}-1)^{l-k-1}}{(e^{\lambda t}-1)^{n-1}} \left(e^{\lambda (n-k+i)(t-s)} - e^{\lambda j(t-s)}\right) \quad \text{as desired}$$

This completes the proof the $f_{\mathscr{A}_{n,t}^{k} - \mathscr{A}_{n,t}^{l}}(s)$, the probability density function of the difference between the kth and lth speciation time, given $\mu = 0$, $l > k$, n extant species and origin time t. Note that when we set $k = k-1$ and $l = k$ we have $\mathscr{A}_{n,t}^{k-1} - \mathscr{A}_{n,t}^{k}$ is the the time until the coalescent event of k extant species.

□

8.4.2 Unknown Tree Age

In assuming we don't know the origin time of the tree, we assign a uniform prior for the time of origin from 0 to ∞. Indicating there is equal probability of when the first ancestor was created across any point in time. Therefore, we only condition on the n extant species. In the pure birth process this would be the same as starting the tree, then waiting until the tree has n species and stopping [5].

Following the $k-1$th speciation event, we have k species and therefore a waiting time that is exponentially distributed with rate λk. This is because we are in a pure birth markov process where speciation events are occurs at a Poisson distribution with rate λk.

Therefore, the coalescent process will also have an exponentially distributed waiting time of rate λk, as it is essentially the backward process. This is not the case in the process with extinction events because information about extinction events is lost when examining a backwards process [5].

Note that in Gernhard's *Conditioned Reconstructed Process* [5], she writes this rate as simply rate k, with λ omitted.

This theorem is present in Gernhard's [5], but we provide additional details in the proof.

Theorem 8.7 (+)**.** *Examining the tree at the nth speciation event is the same as examining the tree today [5]*

(See Gernhard [5] Section 6)

Proof. We can find the time between the $n-1$th and today (time 0), by reexamining the equation for the kth speciation time in the Yule Process, for an unknown tree age, see Equation 6.4.

$$\begin{aligned}
f_{\mathscr{A}_n^k}(s) &= (k+1)\binom{n}{k+1}\lambda\frac{(e^{\lambda s}-1)^{n-k-1}}{e^{\lambda sn}} \\
\implies f_{\mathscr{A}_n^{n-1}}(s) &= ((n-1)+1)\binom{n}{(n-1)+1}\lambda\frac{(e^{\lambda s}-1)^{n-(n-1)-1}}{e^{\lambda sn}} \\
&= (n-1+1)\binom{n}{n-1+1}\lambda\frac{(e^{\lambda s}-1)^{n-n+1-1}}{e^{\lambda sn}} \\
&= n\binom{n}{n}\lambda\frac{(e^{\lambda s}-1)^{0}}{e^{\lambda sn}} \\
&= n\lambda\frac{1}{e^{\lambda sn}} \\
&= n\lambda e^{-\lambda sn}
\end{aligned}$$

This is just an exponential distribution with rate λn. This implies examining the tree today, is equivalent to examining the tree at the nth speciation event [5]. This completes the proof. $\square$

We can also establish the time between events l and k assuming $l > k$, that is l is a speciation events that occurs first.

For the theorem below, Gernhard [5] provides only the result. We prove the following theorem from scratch.

Theorem 8.8 (+). *The time difference between the kth and lth speciation event ($l > k$), given an unknown origin is*

$$f_{\mathscr{A}_n^k - \mathscr{A}_n^l} = \lambda(k+1)\binom{l}{k+1}e^{-l\lambda s}(e^{\lambda s}-1)^{l-k-1} \tag{8.7}$$

(See Gernhard [5] Pg. 776)

Proof. We will use Equation 8.5 for $f_{\mathscr{A}_n^k - \mathscr{A}_{n,t}^l}$ and the law of total probability to integrate out t. We use the derivation for the probability density of the origin time given n extant species, see Chapter: 5.

$$\begin{aligned}
f_{\mathscr{A}_n^k - \mathscr{A}_n^l} &= \int_s^\infty f_{\mathscr{A}_{n,t}^k - \mathscr{A}_{n,t}^l} q_{or}(t|n) \\
&= \int_s^\infty \int_s^t (n-1)(n-2)\binom{n-3}{k-1}\binom{n-k-2}{l-k-1} \\
&\times (1-F(\tau|t))^{k-1}(F(\tau|t)-F(\tau-s|t))^{l-k-1}F(\tau-s|t)^{n-l-1} \\
&\times f(\tau|t)f(\tau-s|t)d\tau \\
&\times n\lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)t})^{n-1}e^{-(\lambda-\mu)t}}{(\lambda-\mu e^{-(\lambda-\mu)t})^{n+1}}dt
\end{aligned}$$

Then using simplifications from the previous proof for Equation 8.5 and letting $\mu = 0$

$$\begin{aligned}
\implies f_{\mathscr{A}_n^k - \mathscr{A}_n^l} &= \int_s^\infty \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\frac{(e^{\lambda s}-1)^{l-k-1}}{(1-e^{-\lambda t})^{n-1}} \\
&\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k-1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}\left(e^{-\lambda\tau}\right)^2\left(e^{\lambda s}\right)d\tau \\
&\times n\lambda^n(\lambda)^2\frac{(1-e^{-\lambda t})^{n-1}e^{-\lambda t}}{(\lambda)^{n+1}}dt
\end{aligned}$$

Then we can simplify moving terms outside of the integral

$$= \int_s^\infty \lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}\frac{(e^{\lambda s}-1)^{l-k-1}}{(1-e^{-\lambda t})^{n-1}}$$
$$\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k-1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}\left(e^{-\lambda\tau}\right)^2\left(e^{\lambda s}\right)d\tau$$
$$\times n\lambda(1-e^{-\lambda t})^{n-1}e^{-\lambda t}dt$$

We can simplify by moving more terms to outside

$$= \int_s^\infty n\lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}e^{-\lambda t}$$
$$\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k-1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}\left(e^{-\lambda\tau}\right)^2\left(e^{\lambda s}\right)d\tau dt$$

Next we match common terms

$$= \int_s^\infty \lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}e^{-\lambda(t-s)}$$
$$\times \int_s^t \left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}\left(e^{-\lambda\tau}\right)^{l-k+1}\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}d\tau dt$$

Then we apply the binomial theorem around: $\left(e^{-\lambda\tau}-e^{-\lambda t}\right)^{k-1}$

$$= \sum_{i=0}^{k-1}\binom{k-1}{i}\left(e^{-\lambda\tau}\right)^i(-1)^{k-1-i}\left(e^{-\lambda t}\right)^{k-1-i}$$

We can also apply the binomial theorem around: $\left(1-e^{-\lambda(\tau-s)}\right)^{n-l-1}$

$$= \sum_{j=0}^{n-l-1}\binom{n-l-1}{j}(-1)^{n-l-1-j}\left(e^{-\lambda(\tau-s)}\right)^{n-l-1-j}$$

Then we add in those binomial theorem results

$$= \int_s^\infty n\lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}e^{-\lambda(t-s)}$$
$$\times \int_s^t \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\left(e^{-\lambda\tau}\right)^i(-1)^{k-1-i}\left(e^{-\lambda t}\right)^{k-1-i}\left(e^{-\lambda\tau}\right)^{l-k+1}$$
$$\times (-1)^{n-l-1-j}\left(e^{-\lambda\tau}e^{\lambda s}\right)^{n-l-1-j}d\tau dt$$

Next we can simplify the integrals more

$$= \int_s^\infty n\lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}e^{-\lambda t}e^{\lambda s}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}(-1)^{k-1-i+n-l-1-j}\left(e^{-\lambda t}\right)^{k-1-i}\left(e^{\lambda s}\right)^{n-l-1-j}$$

$$\times \int_s^t \left(e^{-\lambda\tau}\right)^{n-l-1-j+i+l-k+1} d\tau dt$$

Here we continue to simplify the exponents on the integrals, so its in its simplest form

$$= \int_s^\infty n\lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}(-1)^{n+k-l-i-j}\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j}$$
$$\times \int_s^t \left(e^{-\lambda\tau}\right)^{n-k+i-j} d\tau dt$$

Then we solve the integral

$$= \int_s^\infty n\lambda^3 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}(-1)^{n+k-l-i-j}\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j}$$
$$\times \left[\frac{\left(e^{-\lambda\tau}\right)^{n-k+i-j}}{-\lambda(n-k+i-j)}\right]_s^t dt$$

Then we can take the resulting bounds to get

$$= \int_s^\infty n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j}$$
$$\times \left[\left(e^{-\lambda s}\right)^{n-k+i-j} - \left(e^{-\lambda t}\right)^{n-k+i-j}\right] dt$$

Next we simplify more by multiplying into the square brackets

$$= \int_s^\infty n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$
$$\times \left[\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j}\left(e^{-\lambda s}\right)^{n-k+i-j} - \left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j}\left(e^{-\lambda t}\right)^{n-k+i-j}\right] dt$$

We combine like terms in the square brackets to get the following

$$= \int_s^\infty n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$
$$\times \left[\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{n-l-j-n+k-i+j} - \left(e^{-\lambda t}\right)^{k-i+n-k+i-j}\left(e^{\lambda s}\right)^{n-l-j}\right] dt$$

Then we simplify those like terms to get

$$= \int_s^\infty n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$
$$\times \left[\left(e^{-\lambda t}\right)^{k-i}\left(e^{\lambda s}\right)^{k-i-l} - \left(e^{-\lambda t}\right)^{n-j}\left(e^{\lambda s}\right)^{n-j-l}\right] dt$$

Next we isolate common terms in square brackets

$$= \int_s^\infty n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$
$$\times \left[\left(e^{-\lambda(t-s)}\right)^{k-i}\left(e^{-\lambda s}\right)^{l} - \left(e^{-\lambda(t-s)}\right)^{n-j}\left(e^{-\lambda s}\right)^{l}\right] dt$$

Then we remove those common terms from the brackets and isolate the integral

$$= n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}\left(e^{-\lambda s}\right)^{l}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$
$$\times \int_s^\infty \left[\left(e^{-\lambda(t-s)}\right)^{k-i} - \left(e^{-\lambda(t-s)}\right)^{n-j}\right] dt$$

Next we solve the integral

$$= n\lambda^2 k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}\left(e^{-\lambda s}\right)^{l}$$
$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$

$$\times \left\{ \left[\frac{\left(e^{-\lambda(t-s)}\right)^{k-i}}{-\lambda(k-i)} \right]_0^\infty - \left[\frac{\left(e^{-\lambda(t-s)}\right)^{n-j}}{-\lambda(n-j)} \right]_s^\infty \right\}$$

Then we input the integral bounds to get

$$= n\lambda k(k+1)\binom{l}{k+1}\binom{n-1}{l}(e^{\lambda s}-1)^{l-k-1}\left(e^{-\lambda s}\right)^l$$

$$\times \sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$

$$\times \left\{ \frac{\left(e^{-\lambda(s-s)}\right)^{k-i} - \left(e^{-\lambda(\infty-s)}\right)}{k-i} + \frac{\left(e^{-\lambda(\infty-s)}\right)^{n-j} - \left(e^{-\lambda(s-s)}\right)^{n-j}}{n-j} \right\}$$

Then given $0 < s < t < \infty \implies \infty - s = \infty$ we get the following result

$$\implies f_{\mathscr{A}_n^k - \mathscr{A}_n^l} = \lambda(k+1)\binom{l}{k+1}\left(e^{-\lambda s}\right)^l (e^{\lambda s}-1)^{l-k-1}$$

$$\times n\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} k\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}$$

$$\times \left\{ \frac{1^{k-i}}{k-i} - \frac{1^{n-j}}{n-j} \right\}$$

Then using Proposition 8.8.1 we have our final result

$$= \lambda(k+1)\binom{l}{k+1}\left(e^{-\lambda s}\right)^l (e^{\lambda s}-1)^{l-k-1}$$

$$\times 1$$

$$\implies f_{\mathscr{A}_n^k - \mathscr{A}_n^l} = \lambda(k+1)\binom{l}{k+1}e^{-l\lambda s}(e^{\lambda s}-1)^{l-k-1} \quad \text{as desired}$$

This completes the proof

□

It remains to prove the proposition cited in the theorem

Proposition 8.8.1. *The following equation is equal to a constant* 1

$$\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} nk\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}\left\{\frac{1}{k-i} - \frac{1}{n-j}\right\} = 1$$

Proof.

We can start by inputting the equation

$$\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} nk\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}\left\{\frac{1}{k-i}-\frac{1}{n-j}\right\}$$

We can then simplify by getting a common denominator on the bracketed section

$$=\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} nk\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{n-k+i-j}\left\{\frac{n-j-k+i}{(k-i)(n+j)}\right\}$$

$$=\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1} nk\binom{n-1}{l}\binom{k-1}{i}\binom{n-l-1}{j}\frac{(-1)^{n+k-l-i-j}}{(k-i)(n-j)}$$

Next we expand the binomial coeffcients to get factorials

$$=\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\frac{nk(n-1)!(k-1)!(n-l-1)!}{l!i!j!(n-1-l)!(k-1-i)!(n-l-1-j)!(k-i)(n-j)}(-1)^{n+k-l-i-j}$$

$$=\sum_{i=0}^{k-1}\sum_{j=0}^{n-l-1}\frac{n!k!}{l!i!j!(k-1-i)!(n-l-1-j)!(k-i)(n-j)}(-1)^{n+k-l-i-j}$$

Then we simplify more, to isolate different sums

$$=\frac{n!k!}{l!}\sum_{i=0}^{k-1}\frac{1}{i!(k-1-i)!(k-i)}\sum_{j=0}^{n-l-1}\frac{1}{j!(n-l-1-j)!(n-j)}(-1)^{n+k-l-i-j}$$

$$=\frac{n!k!}{l!}(-1)^{n-k-l}\sum_{i=0}^{k-1}\frac{(-1)^i}{i!(k-i)!}\sum_{j=0}^{n-l-1}\frac{(-1)^j}{j!(n-l-1-j)!(n-j)}$$

We multiply the sums by 1 to get our desired result

$$=\frac{n!k!}{l!}(-1)^{n-k-l}\sum_{i=0}^{k-1}\left\{\frac{(-1)^i}{i!(k-i)!}\times\frac{k!}{k!}\right\}\sum_{j=0}^{n-l-1}\left\{\frac{(-1)^j}{j!(n-l-1-j)!(n-j)}\times\frac{n!l!}{n!l!}\right\}$$

$$=\frac{n!k!}{l!}(-1)^{n-k-l}\left\{\frac{1}{k!}\sum_{i=0}^{k-1}\frac{k!(-1)^i}{i!(k-1)!}\right\}\left\{\frac{l!}{n!}\sum_{j=0}^{n-l-1}\frac{n!(-1)^j}{l!j!(n-l-1-j)!(n-j)}\right\}$$

$$=\frac{n!k!}{l!}(-1)^{n-k-l}\left\{\frac{1}{k!}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\right\}\left\{\frac{l!}{n!}\sum_{j=0}^{n-l-1}\frac{n!(-1)^j}{l!j!(n-l-1-j)!(n-j)}\right\}$$

Then we have the following by the quality of combinations the above two parts both equal 1
Hence we get the final result
$= 1$

$\square$

Chapter 9

Incomplete Sampling

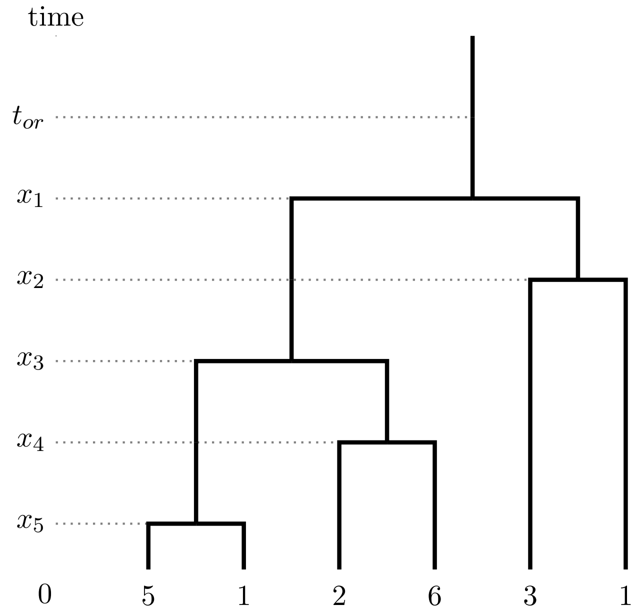
Another paper by Tanja Gernhard on *Incomplete sampling under birth–death models and connections to the sampling-based coalescent* [12] is a sequel to *The Conditioned Reconstructed Process*, the paper this report mainly has explored. Incomplete sampling uses the reconstructed process as a basis for its results. We explain and display some key results, without proving them.

Note that some notation has been changing from these papers to match those of Gernhard's Conditioned Reconstructed Process.

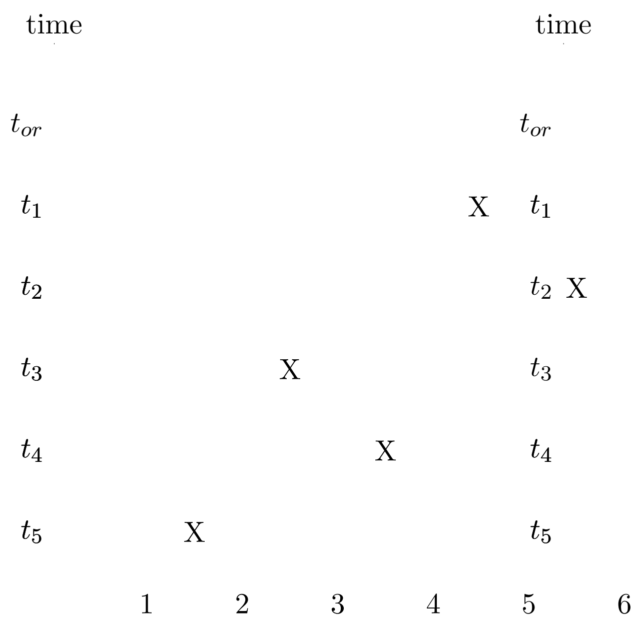
In prior sections we have assumed complete sampling of the tree leaves when deriving results. This is generally unrealistic and in practical use we encounter incomplete sampling. We can derive further results by assigning ρ as the sampling probability for a leaf [12] or m as the number of leaves sampled. For example see Figure 9.1a for the Reconstructed Tree. Then in earlier theorems we proved that there is a Poisson Process representation of the Reconstructed Tree we can use. We use this representation to indicate the difference between complete sampling trees and incomplete sampling of trees.

In Figure 9.1b we have complete sampling with 5 speciation events, when we conduct incomplete sampling, we sample through the leaves. In figure 9.1c, some leaves have been removed, because they were not sampled.

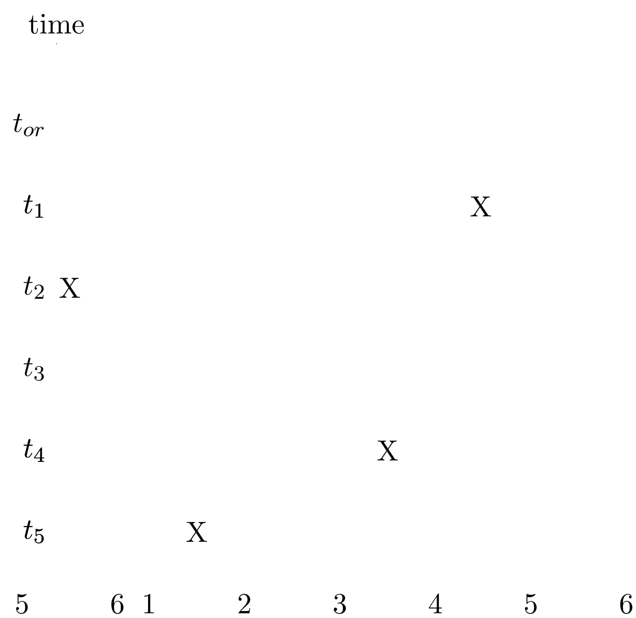
Figure 9.1: Tree Comparison



(a) Reconstructed Tree



(b) Complete Sampling



(c) Incomplete Sampling

9.1 Birth-Death ρ-Sampling Process

A key result we can obtain is the probability density function for joint speciation times $x = x_2, \ldots, x_n$, given incomplete sampling. In the previous paper, *The Conditioned Reconstructed Process*, we also obtained this value. The density function is slightly different, when accounting for ρ, the leaves probability of being sampled.

Theorem 9.1 (+). *The probability density function for speciation times x under a birth death ρ-sampling process is [12] :*

$$f_\rho^{\lambda,\mu}(x|n, x_1) = (n-2)! \left(\prod_{i=2}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{\left(\rho\lambda + (\lambda(1-\rho)-\mu)e^{-(\lambda-\mu)x_i}\right)^2} \right) \times \left(\frac{\rho\lambda + (\lambda(1-\rho)-\mu)\, e^{-(\lambda-\mu)x_1}}{1-e^{-(\lambda-\mu)x_1}} \right)^{n-2} \tag{9.1}$$

(See Gernhard [12] Section 3.1)

This result is conditioning on the time of x_1, the speciation time of the most recent common ancestor of sampled individuals [12]. When we condition on t, the origin time, instead of x_1, the first speciation time, we obtain a similar result. Except notice the product now includes the first speciation event.

Theorem 9.2 (+). *The joint density function for all speciation times conditioning on the time of origin t [12]*

$$f_\rho^{\lambda,\mu}(x|n, t) = (n-1)! \left(\prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{-(\lambda-\mu)x_i}}{\left(\rho\lambda + (\lambda(1-\rho)-\mu)e^{-(\lambda-\mu)x_i}\right)^2} \right) \times \left(\frac{\rho\lambda + (\lambda(1-\rho)-\mu)\, e^{-(\lambda-\mu)t}}{1-e^{-(\lambda-\mu)t}} \right)^{n-1} \tag{9.2}$$

(See Gernhard [12] Equation (2))

The other significant theorem we can obtain is the time of origin, which has its own density function. This allows to further use the law of total probability and find densities for speciation times that do not condition on t. We assume sampling on a tree with n individuals. Under a birth-death incomplete sampling, we take into account ρ, when determining this density. ρ is the probability of sampling a leaf in the tree. Note that, we assume a uniform prior from 0 to ∞ of origin time, when determining the density function for origin time conditional on n extant species.

Theorem 9.3 (+). *The density fore a tree with n individuals to have some time of origin t, given a uniform prior for t and a birth death ρ-sampling process with birth rate λ and death rate μ is the following [12] :*

$$q_\rho^{\lambda,\mu} = n(\lambda-\mu)^2(\lambda\rho)^n \left(\frac{\left(1-e^{-(\lambda-\mu)t}\right)^{n-1} e^{-(\lambda-\mu)t}}{\left(\rho\lambda + (\lambda(1-\rho)-\mu)e^{-(\lambda-\mu)t}\right)^{n+1}} \right) \tag{9.3}$$

(See Gernhard [12] Equation (4))

In this next theorem we make use of the theorem above, with Equation 9.3. The technique used to derive this following formula involves integrating out t from $f_\rho^{\lambda-\mu}(x|n,t)$ using the law of total probability. We get $f_\rho^{\lambda-\mu}(x|n,t)$ from Equation 9.2. This technique of integrating the variable out is something we used in the original paper to derive densities that don't condition on t, time of origin.

Theorem 9.4 (+)**.** *The probability density function for x speciation times, given a uniform from 0 to ∞ on time of origin, and a birth death ρ-sample process is [12] :*

$$f_\rho^{\lambda,\mu} = n!(\lambda-\mu)(\lambda\rho)^{n-1} \frac{e^{-(\lambda-\mu)x_1}}{\rho\lambda_(\lambda(1-\rho)-\mu)e^{-(\lambda-\mu)x_1}} \times \prod_{i=1}^{n-1} \frac{(\lambda-\mu)^2 e^{(\lambda-\mu)x_i}}{\left(\rho\lambda + (\lambda(1-\rho)-\mu)e^{-(\lambda-\mu)x_i}\right)^2} \tag{9.4}$$

(See Gernhard [12] Theorem 3.3)

This formula is only valid for general λ and μ, it does not apply for $\lambda = \mu$, which is the critical case. For the specific scenario we have the theorem below.

Theorem 9.5 (+)**.** *as $\mu \to \lambda$, i.e. the critical case we get the following density for $f_\rho^{\lambda,\lambda}$ [12]*

$$f_\rho^{\lambda,\lambda}(x|n) = n!(\lambda\rho)^{n-1} \frac{1}{1_\rho\lambda x_1} \prod_{i=1}^{n-1} \frac{1}{(1+\rho\lambda x_i)^2} \tag{9.5}$$

(See Gernhard [12] Equation (6))

9.2 Birth Death m-Sampling Process

The next way we could sample leafs in our reconstructed tree, is to use m-sampling. Instead of assigning each leaf a probability of being sampled, we randomly sample m leaves in the tree. This is another way of dealing with incomplete sampling and will have the same desired effect as seen in Figure 9.1c. In this process we derive the same results we did before and in the Conditioned Reconstructed Process. That is we derive density functions, based on different conditions, and examine limits as we our samples approach infinity, that is $m \to \infty$.

We next derive the density that two randomly chosen leaves in a tree of size m have speciation time x, i.e. find $f_m^{\lambda-\mu}(x|2)$

Theorem 9.6 (+)**.** *We have a birth death m-sampling process, with birth rate λ, death rate μ, a uniform prior for origin time t. Suppose we then sample two individuals from the m-size tree. The density of the most recent common ancestor of the two sample individuals is [12]*

$$f_m^{\lambda-\mu}(x|2) = \frac{2\lambda e^{-(\lambda-\mu)x}}{m-1}\left(\frac{1}{a+1}\right)^{m+1} \times \left[(a+1)^{m+1} - \binom{m+1}{2}a^2 - (m+1)a - 1\right]$$
$$\textit{Where } a := \frac{\lambda-\mu}{\lambda} \cdot \frac{e^{-(\lambda-\mu)x}}{(1-e^{-(\lambda-\mu)x}} \tag{9.6}$$

(See Gernhard [12] Theorem 4.1)

Theorem 9.7 (+)**.** *The same density for the limit $\mu \to \lambda$ is the following [12]*

$$f_m^{\lambda,\mu}(x|2) = \frac{2\lambda}{m-1}\left(\frac{\lambda x}{1+\lambda x}\right)^{m+1}\left[\left(1+\frac{1}{\lambda x}\right)^{m+1} - \binom{m+1}{2}\frac{1}{(\lambda x)^2} - \frac{m_1}{\lambda x} - 1\right] \tag{9.7}$$

(See Gernhard [12] Equation (10))

We can examine results when $m \to \infty$.

Theorem 9.8 (+)**.** *The probability density function for the speciation time of the most recent common ancestor in a birth death m-sampling process, as $m \to \infty$ is [12]*

$$f_\infty^{\lambda-\mu}(x|2) = 2ce^{dx} - \frac{d^2}{c}e^{dx}\left(\frac{1}{e^{dx}-1}\right)^2 e^{-(d/c)e^{-dx}/(1-e^{-dx})} - 2d\frac{e^{dx}}{e^{dx}-1}e^{-(d/c)e^{-dx}/(1-e^{-dx})}$$
$$- 2ce^{dx}e^{-(d/c)e^{-dx}/(1-e^{-dx})} \tag{9.8}$$

(See Gernhard [12] Equation (11))

Theorem 9.9 (+)**.** *This result for when $\lambda = \mu$ in the critical case is [12]*

$$f_\infty^{\lambda,\lambda}(x|2) = 2c - \frac{e^{-1/cx}}{cx^2} - \frac{2e^{-1/cx}}{x} - 2ce^{-1/cx} \tag{9.9}$$

(See Gernhard [12] Equation (12))

In the previous theorem we samples with 2 leaves across a tree with m leaves, we can also sample n leaves across a tree with m leaves.

Theorem 9.10 (+). *The probability density function in a birth death m-sampling process, the density function for when we sample n leaves is as follows [12]*

Let s_k be the kth speciation event in the m tree. Let s_k^l be the event that s_k leaves l descendants. Let s_k^{l,l_1} be the event s_k has l descendants, l_1 leaves on the left subtree and $l - l_1$ in the right subtree. Then we have the following:

$$
\begin{aligned}
f_m^{\lambda,\mu}(x|n) &= \sum_{k=1}^{m-1} p_k^n g_k^{\lambda,\mu}(x|m) \\
\text{Where } p_k^n &= \sum_{l=n}^{m-k+1} \frac{\binom{k}{2}\binom{m-k-1}{l-2}}{\binom{l}{2}\binom{m-1}{l}} \sum_{l_1=1}^{l} \frac{\binom{l}{n} - \binom{l_1}{n} - \binom{l-l_1}{n}}{\binom{m}{n}} \\
\text{and } g_k^{\lambda,\mu}(x|m) &= \left(1 - \frac{(k+1)k}{(m+1)m}\right) g_k^{\lambda,\mu}(x|m+1) + \frac{(k+1)k}{(m+1)m} g_m^{\lambda,\mu} g_{k+1}^{\lambda,\mu}(x|m+1) \\
&\text{This is solved by recursion}
\end{aligned}
\tag{9.10}
$$

(See Gernhard [12] Section 4.2)

Chapter 10

Sampling Through Time

This next paper by Tanja Gernhard, *Sampling Through Time in Birth Death Trees* (2010) [13] is a sequel to *The Conditioned Reconstructed Process*, the paper this project has explored. Sampling through time also using the reconstructed process as a basis for its results. We explain and display some key results, without proving them.

Note that some notation has been changing from these papers to match those of Gernhard Reconstructed Process.

The big idea of this paper is in recognising that in some applications we do not have access all the data by time. The paper gives examples of virus sequences or fossil data, both situations where data at certain times is not available [13]. Given this, Gernhard [13] introduced the idea of sampling through time to obtain results given this specific type of data inaccessibility.

10.1 A Key Result

Let Ψ be the probability an individual which is sampled at some time before the present is an extant individual, i.e. it is alive today. Let ρ be the probability an individual at present is sampled.

The first big theorem is below, it is relating to the number of descendants an a individual species has, given sampling through time.

Theorem 10.1 (+). *The probability an individual alive at time t before today has n sampled* *descendants and an arbitrary number of extinct individuals, $\hat{p}_n(t)$, for $\rho > 0$ is the following*

$$\hat{p}_0(t) = p_0(t|\Psi = 0) \tag{10.1}$$

$$\hat{p}_1(t) = p_1(t|\Psi = 0) \tag{10.2}$$

$$\hat{p}_n(t) = \hat{p}_1(t)\left(\frac{\rho\lambda(1 - e^{-(\lambda-\mu)t}}{\lambda\rho + (\lambda(1-\rho) - \mu)e^{-(\lambda-\mu)t}}\right)^{n-1} \tag{10.3}$$

$$\textit{For } \rho = 0, \textit{we have } \hat{p}_0(t) = 1 \textit{ and } \hat{p}_n(t) = 0 \textit{ for all } n > 0 \tag{10.4}$$

(See Gernhard [13] Theorem 3.3)

10.2 Observing a Sample Tree, τ

Another key theorem Gernhard [13] derives is the density function for a sampled tree, assuming we have sampled through time.

Theorem 10.2 (+)**.** *The probability density of a sampled tree, τ, with $n > 1$ extant sampled leaves, $m \geq 0$ extinct sampled leaves, and $k \geq 0$ sampled individuals with sampled descendants, conditioned on the time of the most recent common ancestor being x_1, is,*

$$f[\tau|t_{mrca} = x_1] = \frac{\lambda^{n+m-2}\Psi^{k+m}}{(1-\hat{p}_0(x_1))^2} p_1(x_1) \prod_{i=1}^{n+m-1} p_1(x_i) \prod_{i=1}^{m} \frac{p_0(y_i)}{p_1(y_i)} \tag{10.5}$$

(See Gernhard [13] Theorem 3.8)

We can also condition on the number of extant sampled individuals n, to obtain a different set of results. We let C be some characteristics of the process.

Theorem 10.3 (+)**.** *The probability density of a sampled tree τ with n extant sampled leaves, m extinct sampled leaves, $n + m > 0$, and $k \geq 0$ sampled individuals with sampled descendants, conditioned on sampling n present day individuals is*

$$f[\tau|n] = \frac{F[\tau]}{\int_0^\infty f[n|t_{or} dt_{or}]} \tag{10.6}$$

$$\textit{Where } F[\tau] = \frac{4\rho\lambda^{n+m-1}\Psi^{k+m}}{c_1(c_2+1)(1-c_2+(1+c_2)e^{c_1 x}} \prod_{i=1}^{n+m-1} \lambda p_1(x_i) \prod_{i=1}^{m} \frac{p_0(y_i)}{p_1(y_i)} \tag{10.7}$$

(See Gernhard [13] Theorem 3.11)

We could lastly also condition on $n = 0$, i.e. there are no extant sampled leaves

Theorem 10.4 (+)**.** *The probability density of a sampled tree τ with no extant sampled leaves, $m > 0$ extinct sampled leaves, and $k \geq 0$ sampled individuals with sampled descendants, conditioned on $n = 0$ and the process surviving to the present, is*

$$f[\tau|0, X] = \frac{\lambda(F[\tau] - F[\tau|\rho = 1])}{ln(1/\rho} \tag{10.8}$$

(See Gernhard [13] Equation (11))

Appendix A

Simulation for Expected Time of k-th Speciation Event Code

```
library(ggplot2)
library(tidyverse)

#Create the main model, for c25, c50, c75, c90
mainModel = function(k,n,l,m){

  #Set Rho
  p = m/l

  #Start the first Sum
  sum2 <- numeric(n-k-1+1)
  for(i in 1:length(sum2)){

    #Create the second sum
    sum3 <- numeric(k+i-1)
    for(j in 1:length(sum3)){
      sum3[j] = choose(k+i-1, j)*(((-1)^j)/j)*(1-(1/(1-p))^j)
    }
    sum3 = sum(sum3)

    #Combine two Sums
    sum2[i] = choose(n-k-1,i-1)*(1/((k+i-1+1)*p))*(((1/p) - 1)^(k+i-1))*(
       log(1/(1-p)) - sum3)

  }

  #Add in Previous Values
```

```
  ((k+1)/l)*choose(n, k+1)*((-1)^k)*sum(sum2)

}

#Create the Yule Model
yule <- function(k,n,l){

  #Add an if condition for the n-th speciation event
  if(k+1 <= n){
  sum2 = numeric(n)
  for(i in (k+1):n){
    sum2[i] = (1/(l*i))
  }
  sum(sum2)
  } else
    0
}

#Create the function for Rho = 1
equalModel = function(k,n,l) (n-k)/(l*k)

#Set number of species
n = 10

#Create Empty dataframe
unscaledData <- data.frame(k = c(1:n),
                       Yule = numeric(n),
                       c25 = numeric(n),
                       c50 = numeric(n),
                       c75 = numeric(n),
                       c90 = numeric(n),
                       Equal = numeric(n))

#Create Data frame for each expectation function for each k-th event
for(i in 1:n){
  unscaledData[i,2] = yule(unscaledData[i,1], n, 1)
  unscaledData[i,3] = mainModel(unscaledData[i,1], n, 1, 0.25)
  unscaledData[i,4] = mainModel(unscaledData[i,1], n, 1, 0.5)
  unscaledData[i,5] = mainModel(unscaledData[i,1], n, 1, 0.75)
  unscaledData[i,6] = mainModel(unscaledData[i,1], n, 1, 0.9)
  unscaledData[i,7] = equalModel(unscaledData[i,1], n, 1)



}
```

```
#Scale the time frame
scaledData <- unscaledData%>%
  mutate(Yule = Yule/max(Yule),
         c25 = c25/max(c25),
         c50 = c50/max(c50),
         c75 = c75/max(c75),
         c90 = c90/max(c90),
         Equal = Equal/max(Equal))


#Rework Data for plotting
scaledData <- scaledData%>%
            pivot_longer(cols = c(Yule, c25, c50, c75, c90, Equal),
                           names_to = "rho",
                           values_to = "time")%>%
              mutate(rho = factor(rho,
                     levels = c("Yule", "c25", "c50",
                                "c75", "c90", "Equal")))

unscaledData <- unscaledData%>%
            pivot_longer(cols = c(Yule, c25, c50, c75, c90, Equal),
                           names_to = "rho",
                           values_to = "time")%>%
              mutate(rho = factor(rho,
                     levels = c("Yule", "c25", "c50",
                                "c75", "c90", "Equal")))


#Create plot for scaled Data
ggplot(data = scaledData,
       mapping = aes(y = log(k), x = time, group = rho, colour = rho))+
  geom_line() +
  geom_point(shape = "x", size = 4)+
  scale_colour_discrete(labels=c('\U03C1␣=␣0␣(Yule)',
                                 '\U03C1␣=␣0.25', '\U03C1␣=␣0.5',
                                '\U03C1␣=␣0.75','\U03C1␣=␣0.9',
                                '\U03C1␣=␣1'))+
  labs(color='\U03C1␣=␣\U03BC/\U03BB␣Value',
       title = "Simulation␣of␣the␣Expectation␣for␣the␣k-th␣Speciation␣Time
          ",
       subtitle = "For␣10␣extant␣species␣(n␣=␣10)",
       x = "Time␣Scaled",
       y = "Log(k-th␣Speciation␣Event)")

#Create plot for unscaled Data
ggplot(data = unscaledData,
```

```
     mapping = aes(y = log(k), x = time, group = rho, colour = rho))+
geom_line() +
geom_point(shape = "x", size = 4)+
scale_colour_discrete(labels=c('\U03C1␣=␣0␣(Yule)',
                                '\U03C1␣=␣0.25', '\U03C1␣=␣0.5',
                               '\U03C1␣=␣0.75','\U03C1␣=␣0.9',
                               '\U03C1␣=␣1'))+
labs(color='\U03C1␣=␣\U03BC/\U03BB␣Value',
     title = "Simulation␣of␣the␣Expectation␣for␣the␣k-th␣Speciation␣Time
        ",
     subtitle = "For␣10␣extant␣species␣(n␣=␣10)␣(time␣unscaled)",
     x = "Time␣Scaled",
     y = "Log(k-th␣Speciation␣Event)")
```

Appendix B

Bibliography

[1] *Human evolutionary trees / E. A. Thompson.* Cambridge University Press, 1975.

[2] David Aldous and Lea Popovic. A critical branching process model for biodiversity. *Advances in Applied Probability*, 37(4):1094–1115, 2005.

[3] Norman T. J. Bailey. *The elements of stochastic processes with applications to the natural sciences.* Wiley publication in applied statistics. Wiley, New York, 1964.

[4] H. Dehling and B. Haupt. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, chapter 9.4. Springer Berlin, Heidelberg, 2003.

[5] Tanja Gernhard. The conditioned reconstructed process. *Journal of Theoretical Biology*, 253(1):769–778, 2008.

[6] Tanja Gernhard. New analytic results for speciation times in neutral models. *Bulletin of Mathematical Biology*, 70:1082–1097, 2008.

[7] Tanja Gernhard, Klaas Hartmann, and Mike Steel. Stochastic properties of generalised yule models, with biodiversity applications. *Journal of mathematical biology*, 57(5):713–735, 2008.

[8] Geoffrey. Grimmett and David. Stirzaker. *Probability and random processes.* Oxford University Press, Oxford ;, 3rd ed. edition, 2001.

[9] P.D. Jarvis, B.R. Holland, and J.G. Sumner. Phylogenetic invariants and markov invariants. In *Reference Module in Life Sciences.* Elsevier, 2017.

[10] NN Lebedev. Special functions and their application. *BI-Wissenschaftsverlag, Bibliografies Institut, Zurich*, 1973.

[11] Sarabjeet Singh. Branching processes in disease epidemics. 2014.

[12] Tanja Stadler. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of theoretical biology*, 261(1):58–66, 2009.

[13] Tanja Stadler. Sampling-through-time in birth–death trees. *Journal of theoretical biology*, 267(3):396–404, 2010.

[14] Ilze Ziedins. Stats 320: Poisson processes. University of Auckland Lecture, February-July 2022. Class Name: Stats 320: Stochastic Processes.