State-dependent speciation and extinction (SSE) models

Alex Beams

Department of Mathematics
Simon Fraser University

PhylogeogRaphy Workshop

October 1, 2025

1. MASTER EQUATIONS FOR BIRTH-DEATH MODELS

2. LIKELIHOOD FOR BISSE

3. SIMULATING AND FITTING MODELS IN R

Table of Contents

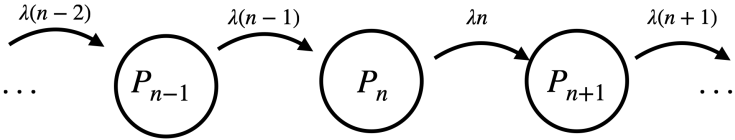
THE BIRTH-DEATH MODEL

We start with the most basic birth-death model – sometimes called the **Yule Model**

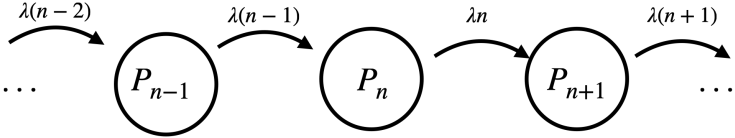
Individuals/species replicate at some rate, but they never die
Specifically, assume the following:

- we start with one individual/species at time $t = 0$
- Each individual/species produces a new individual/species at rate λ
- (So, the time until a lineage gives "birth" is $T \sim \text{Exp}(\lambda)$)

The Yule model



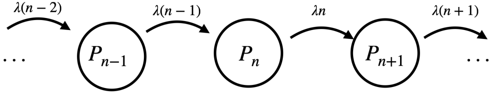
This is a continuous-time Markov chain



Continuous-time Markov chain with a countable state space!

What is the probability the system has n individuals at time t, given that we start with one individual at time $t = 0$?

MASTER EQUATION:



We need to keep track of the ways to move in and out of the state characterized by n individuals alive at time t

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1}$$

A birth moves us out of P_n to P_{n+1}, and the only way to enter into P_n is from P_{n-1}

MASTER EQUATION:

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1}$$

The solution to this Master equation is $P_n(t)$, the probability of n individuals at time t

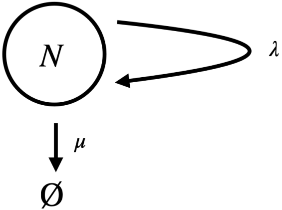
Can solve this using (i) generating functions, (ii) guess and check (the solution is a time-dependent geometric distribution), or (iii) mathematical induction[1]

---

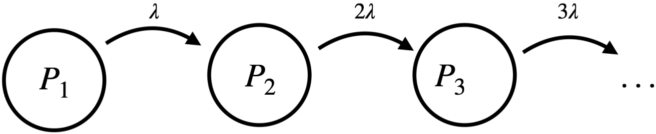
[1] Tanja Stadler et al. *Decoding genomes: from sequences to phylodynamics.* ETH Zurich, 2024.

ADDING DEATH

What if individuals/lineages can also die at rate μ?

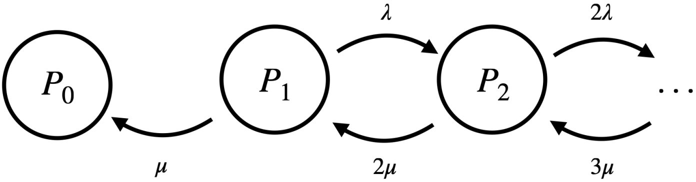


WITHOUT DEATH



In the Yule model, extinction is impossible

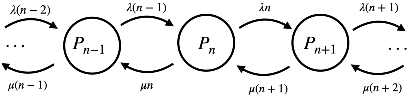
ADDING DEATH



In the birth-death model, extinction is possible: once we enter the state with 0 lineages, we are stuck – can't have birth from nothing

The state $n = 0$ is an **absorbing state** of this continuous-time Markov chain

Adding death[2]



$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1} \quad \text{for } n \geq 1,$$

$$\ldots$$

$$\frac{dP_1}{dt} = -\lambda P_1 - \mu P_1 + 2\mu P_2,$$

$$\frac{dP_0}{dt} = \mu P_1.$$

---

[2] Sean Nee, Robert Mccredie May, and Paul H Harvey. “The reconstructed evolutionary process”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 344.1309 (1994), pp. 305–311.

Forward vs. Backward equations

$$\begin{aligned}
\frac{dP_n}{dt} &= -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1} \quad \text{for } n \geq 1, \\
&\ldots \\
\frac{dP_1}{dt} &= -\lambda P_1 - \mu P_1 + 2\mu P_2, \\
\frac{dP_0}{dt} &= \mu P_1.
\end{aligned}$$

This equation (and the others up to now) is a **Forward equation**. Given an initial condition at time $t = 0$, the $P_n(t)$ characterize the system at some time, t, in the future

FORWARD VS. BACKWARD EQUATIONS

However, we can also pose a different question: given a measurement of the system at the present time, what was the likely configuration of the system some time in the past?

Solutions to the **Forward Kolmogorov equations** tell us how the system propagates forward in time from an initial condition

Solutions to the **Backward Kolmogorov equation** will characterize the system's earlier configuration

Which of these equations seems more relevant for reconstructing evolutionary scenarios?

BACKWARD KOLMOGOROV EQUATION FOR A BIRTH-DEATH MODEL

Let's derive a Backward Kolmogorov equation. The logic is slightly different.

Let's still use the convention that time is measured in the forward sense, that is, we observe $P_n(t)$ but want to understand $P_n(s)$ for $s < t$

Let's start by reexamining where this comes from:

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1}$$

FORWARD KOLMOGOROV EQUATION DERIVATION FOR A BIRTH-DEATH MODEL

This ODE

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1},$$

arises from the following:

$$\begin{aligned} P_n(t+\Delta t) = &(1 - n(\lambda+\mu)\Delta t)P_n(t) + \\ &(n-1)\lambda \Delta t P_{n-1}(t) + \\ &(n+1)\mu \Delta t P_{n+1} \end{aligned}$$

Start in P_n and nothing happens; or, start in P_{n-1} and there is a birth; or, start in P_{n+1} and then there is a death

FORWARD VS. BACKWARD KOLMOGOROV EQUATION DERIVATIONS FOR A BIRTH-DEATH MODEL

This derivation relates the system at time $t + \Delta t$ to the system at earlier time t:

$$\begin{aligned} P_n(t+\Delta t) = &(1 - n(\lambda+\mu)\Delta t)P_n(t) + \\ &((n-1)\lambda\Delta t)\, P_{n-1}(t) + \\ &((n+1)\mu\Delta t)\, P_{n+1}(t). \end{aligned}$$

We can use similar logic to relate the system at time $t - \Delta t$ to the system at a later time t:

$$\begin{aligned} P_n(t-\Delta t) = &(1 - n(\lambda+\mu)\Delta t)P_n(t) + \\ &(n\lambda\Delta t)\, P_{n+1}(t) + \\ &(n\mu\Delta t)\, P_{n-1}(t). \end{aligned}$$

FORWARD VS. BACKWARD KOLMOGOROV EQUATION DERIVATIONS FOR A BIRTH-DEATH MODEL

Forward Equation in English: we can end up in P_n at $t + \Delta t$ if we start from $P_n(t)$ and nothing happens, if we start in $P_{n-1}(t)$ and there is a birth, or we start in $P_{n+1}(t)$ and there is a death:

$$\begin{aligned} P_n(t+\Delta t) =& (1 - n(\lambda+\mu)\Delta t) P_n(t) + \\ & ((n-1)\lambda\Delta t)\, P_{n-1}(t) + \\ & ((n+1)\mu\Delta t)\, P_{n+1}(t). \end{aligned}$$

Backward Equation in English: we could have come from P_n at $t - \Delta t$ if nothing happened and we ended up in $P_n(t)$, if there was a birth and we ended up in $P_{n+1}(t)$, or if there was a death and we ended up in $P_{n-1}(t)$

$$\begin{aligned} P_n(t-\Delta t) =& (1 - n(\lambda+\mu)\Delta t) P_n(t) + \\ & (n\lambda\Delta t)\, P_{n+1}(t) + \\ & (n\mu\Delta t)\, P_{n-1}(t). \end{aligned}$$

THE BACKWARD KOLMOGOROV EQUATION DERIVATION FOR A BIRTH-DEATH MODEL

Rearranging terms, dividing by Δt, and taking the limit $\Delta t \to 0$ of this equation,

$$\begin{aligned} P_n(t-\Delta t) =& (1-n(\lambda+\mu)\Delta t)P_n(t) + \\ & (n\lambda\Delta t)\, P_{n+1}(t) + \\ & (n\mu\Delta t)\, P_{n-1}(t), \end{aligned}$$

we obtain

$$\frac{dP_n}{dt} = -\, n(\lambda+\mu)P_n + (n\mu)\, P_{n-1} + (n\lambda)\, P_{n+1}.$$

THE FORWARD VS. BACKWARD KOLMOGOROV EQUATIONS FOR A BIRTH-DEATH MODEL

We had this Forward equation before:

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1},$$

and now we also have this Backward equation:

$$\frac{dP_n}{dt} = -\, n(\lambda + \mu)P_n + (n\mu)\, P_{n-1} + (n\lambda)\, P_{n+1}.$$

Can you see the relationship between these?

THE FORWARD VS. BACKWARD KOLMOGOROV EQUATIONS FOR A BIRTH-DEATH MODEL

We had this Forward equation before:

$$\frac{dP_n}{dt} = -n\lambda P_n + (n-1)\lambda P_{n-1} - \mu n P_n + \mu(n+1)P_{n+1},$$

and now we also have this Backward equation:

$$\frac{dP_n}{dt} = -\, n(\lambda+\mu)P_n + (n\mu)\, P_{n-1} + (n\lambda)\, P_{n+1}.$$

Can you see the relationship between these?

The Forward equation has the form $dP/dt = QP$, and the Backward equation is $dP/dt = Q^*P$, where Q^* is the adjoint of Q.

THE FORWARD VS. BACKWARD KOLMOGOROV EQUATIONS FOR A BIRTH-DEATH MODEL

$$\frac{dP_n}{dt} = - n(\lambda + \mu)P_n + (n\mu)\, P_{n-1} + (n\lambda)\, P_{n+1}.$$

The backward equation is still using the convention that time t increments in forward time. Later, we will transform the time variable so that 0 coincides with the present, and time increases going into the past.

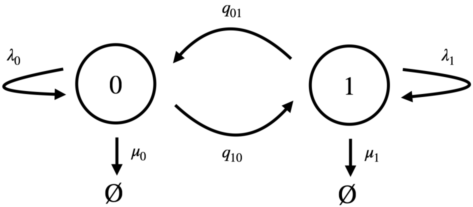
BACKWARD KOLMOGOROV EQUATION FOR A BIRTH-DEATH MODEL

So what? What is the point of doing all of this?

With phylogenetic trees, we observe species at the present day. We want to understand the process that gave rise to an obeservation at the present time. The **likelihood** for the state-dependent speciation and extinction models uses Backward equations to obtain the probability of a given tree.

WHAT IF THERE ARE MULTIPLE STATES?

The **binary state-dependent speciation and extinction (BiSSE)** model extends the birth-death model. Lineages reproduce and go extinct, but possibly at different rates depending on their state. They can migrate/mutate between two states:



We will superimpose this model on a fixed phylogenetic tree next

Table of Contents

BISSE EQUATIONS (ON A SINGLE BRANCH)

Let's first adopt the convention that time, t, increments going backward into the past[3]

We are given a phylogenetic tree, and it is assumed fixed/known.

---

[3] Wayne P Maddison, Peter E Midford, and Sarah P Otto. "Estimating a binary character's effect on speciation and extinction". In: *Systematic biology* 56.5 (2007), pp. 701–710.

BISSE EQUATIONS

Births in this stochastic process will correspond to bifurcations in phylogenetic trees

Deaths/extinctions mean that we will not observe certain lineages

Migrations in the stochastic process will correspond to changes of state on branches of a phylogeny

BISSE EQUATIONS

Introduce some variables:

Let 0 be the time at the present, and t_R be the time that coincides with the root of the tree, so that time t increments going into the past

Let $D_{Ni}(t)$ be the probability a lineage existing at time t in state i evolves into the clade descending from node N that we observe at time 0.

Possible events:

- birth/diversification at rate λ_i
- death/extinction at rate μ_i
- mutation/migration between states at rates q_{10} and q_{01}

Another key variable: let $E_i(t)$ be the probability that a lineage in state i at time t goes extinct before we observe it later at time 0

BISSE EQUATIONS[4]

Now, let's derive a Backward equation for the $D_{Ni}(t)$:

$$\begin{aligned}
D_{N0}(t+\Delta t) =& (1-(\lambda_0+\mu_0+q_{01})\Delta t)D_{N0}(t)+\\
& (q_{01}\Delta t)D_{N1}(t)+(2\lambda_0\Delta t)\,E_0(t)D_{N0}(t),\\
D_{N1}(t+\Delta t) =& (1-(\lambda_1+\mu_1+q_{10})\Delta t)D_{N1}(t)+\\
& (q_{10}\Delta t)D_{N0}(t)+(2\lambda_1\Delta t)\,E_1(t)D_{N1}(t).
\end{aligned}$$

We then apply the standard trick of rearranging terms, dividing by Δt, and taking the limit $\Delta t \to 0$

---

[4] Wayne P Maddison, Peter E Midford, and Sarah P Otto. "Estimating a binary character's effect on speciation and extinction". In: *Systematic biology* 56.5 (2007), pp. 701–710.

BISSE EQUATIONS[5]

$$\frac{dD_{N0}}{dt} = -\left(\lambda_0 + \mu_0 + q_{01})\right)D_{N0}(t) + q_{01}D_{N1}(t) + \\ (2\lambda_0)\,E_0(t)D_{N0}(t),$$

$$\frac{dD_{N1}}{dt} = -\left(\lambda_1 + \mu_1 + q_{10})\right)D_{N1}(t) + q_{10}D_{N0}(t) + \\ (2\lambda_1)\,E_1(t)D_{N1}(t).$$

---

[5] Wayne P Maddison, Peter E Midford, and Sarah P Otto. "Estimating a binary character's effect on speciation and extinction". In: *Systematic biology* 56.5 (2007), pp. 701–710.

BISSE EQUATIONS[6]

$$\begin{aligned}\frac{dD_{N0}}{dt} =& -(\lambda_0 + \mu_0 + q_{01}))D_{N0}(t) + q_{01}D_{N1}(t) + \\ & (2\lambda_0)\,E_0(t)D_{N0}(t), \\ \frac{dD_{N1}}{dt} =& -(\lambda_1 + \mu_1 + q_{10}))D_{N1}(t) + q_{10}D_{N0}(t) + \\ & (2\lambda_1)\,E_1(t)D_{N1}(t).\end{aligned}$$

To solve these equations, we start at the tips of the tree and work up towards nodes, and ultimately, the root of the tree

---

[6] Wayne P Maddison, Peter E Midford, and Sarah P Otto. "Estimating a binary character's effect on speciation and extinction". In: *Systematic biology* 56.5 (2007), pp. 701–710.

BiSSE EQUATIONS[7]

$$
\begin{aligned}
\frac{dD_{N0}}{dt} =& -(\lambda_0 + \mu_0 + q_{01}))D_{N0}(t) + q_{01}D_{N1}(t) + \\
& (2\lambda_0)\, E_0(t) D_{N0}(t), \\
\frac{dD_{N1}}{dt} =& -(\lambda_1 + \mu_1 + q_{10}))D_{N1}(t) + q_{10}D_{N0}(t) + \\
& (2\lambda_1)\, E_1(t) D_{N1}(t).
\end{aligned}
$$

So, if A is the ancestor of two sister lineages corresponding to nodes N and M, then at the speciation time t_A we set

$$
\begin{aligned}
D_{A0}(t_A) =& D_{N0}(t_A) D_{M_0}(t_A) \lambda_0, \\
D_{A1}(t_A) =& D_{N1}(t_A) D_{M_1}(t_A) \lambda_1,
\end{aligned}
$$

and then the $D_{Ai}(t_A)$ are the initial conditions for the equations up the branch toward the root

[7] Wayne P Maddison, Peter E Midford, and Sarah P Otto. “Estimating a binary character's effect on speciation and extinction”. In: *Systematic biology*

BISSE EQUATIONS

The $D_{Ni}(T)$ are initialized at the tips, and equal either 0 or 1 because we can observe the tip states

At the root, R, which is at time t_R, we end up two terms

$$\begin{aligned} D_{R0}(T) =& D_{N0}(t_R) D_{M_0}(t_R) \lambda_0, \\ D_{R1}(T) =& D_{N1}(t_R) D_{M_1}(t_R) \lambda_1. \end{aligned}$$

There are different options for averaging these to produce the overall likelihood of the tree, but the default method[8] is take a weighted average

$$\mathcal{L} = D_R(t_R) = pD_{R0}(t_R) + (1-p) D_{R1}(t_R)$$

---

[8] Wayne P Maddison, Peter E Midford, and Sarah P Otto. "Estimating a binary character's effect on speciation and extinction". In: *Systematic biology* 56.5 (2007), pp. 701–710.

BiSSE EQUATIONS

The variable p is defined as the equilibrium probability that a lineage resides in state 0, and is calculated by solving for the equilibrium of the ODEs

$$\begin{aligned}\frac{dn_0}{dt} =& (\lambda_0 - \mu_0)n_0 - q_{01}n_0 + q_{10}n_1, \\ \frac{dn_1}{dt} =& (\lambda_1 - \mu_1)n_1 - q_{10}n_1 + q_{01}n_0,\end{aligned}$$

n_0 and n_1 representing the expected number of lineages in state 0 and 1, respectively.

BISSE LIKELIHOOD

To summarize: the BiSSE likelihood generalizes the Markov models we saw for ancestral character estimation and stochastic character mapping

Where the Markov models calculate transition probabilities across branches of the phylogeny, the BiSSE model also accounts for speciation (λ) and extinction (μ) dynamics that affect the length of the branches

The overall calculation strategy for the BiSSE likelihood is similar:

- Start at the tips and calculate conditional likelihoods of subtrees
- average conditional likelihoods along branches when they meet a node
- upon reaching the root, average the conditional likelihoods to obtain the likelihood of the tree

Table of Contents

Working in R

We will use the R package diversitree to simulate and fit these models to example datasets We will compare our results with

BiSSE to results from ace and simmap

Make sure you have the following R packages installed:

- diversitree
- ape
- phytools

Working in R

These functions simulate trees using these models:

- tree.bd (birth death trees)
- tree.bisse (BiSSE tree)
- tree.mussse (MuSSE tree)
- look at ?tree.bisse for information about others (quasse – quantitative traits, hisse – hidden traits)

WORKING IN R

These functions define likelihood models on trees:

- make.bd (birth death trees)
- make.bisse (BiSSE tree)
- make.mussse (MuSSE tree)
- you get the idea

The **find.mle** function is useful for obtaining Maximum Likelihood Estimates for parameters

The **constrain** function is useful for setting parameters to fixed values in Maximum likelihood estimation