STATS 700-002 Class 4.
Sampling through time in birth-death trees

Edward Ionides and Aaron King

September 18, 2025

Stadler (2010)

Stadler, T. (2010). Sampling-through-time in birth–death trees. *Journal of Theoretical Biology*, 267(3), 396-404. https://doi.org/10.1016/j.jtbi.2010.09.010.

- Here, we start developing models for heterochronous genealogical trees (not all leaves are at the same time).

Birth-death trees vs the coalescent

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1), 58-66. https://doi.org/10.1016/j.jtbi.2009.07.018

- Deals with homochronous sampling (i.e., all samples at the end time, T)

Birth-death trees vs the coalescent

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1), 58-66. https://doi.org/10.1016/j.jtbi.2009.07.018

- Deals with homochronous sampling (i.e., all samples at the end time, T)
- Similar to Kingman's coalescent, but not identical

Birth-death trees vs the coalescent

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, 261(1), 58-66. https://doi.org/10.1016/j.jtbi.2009.07.018

- Deals with homochronous sampling (i.e., all samples at the end time, T)
- Similar to Kingman's coalescent, but not identical
- What are the advantages and disadvantages?

Definitions

- Notice the deletion of ancestors of unsampled lineages. We will call this *pruning*.

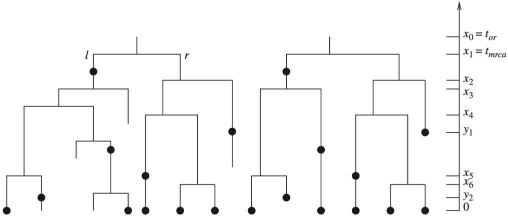

Fig. 1. An example of an oriented tree with root edge induced by the birth–death process displaying all sampled and non-sampled individuals is shown on the left. The left descendant of a bifurcation is l, the right descendant is r. Note that we only label one pair of descendants with l and r for easier readability. The right tree is the corresponding sampled tree with $n=5$ extant individuals, $m=2$ extinct individuals without sampled descendants, and $k=3$ extinct individuals with sampled descendants. In the sampled tree, the bifurcation times are at $x=(x_1,\ldots,x_{n+m-1})$ and the sampling times of extinct individuals without sampled descendants are at $y=(y_1,y_2)$. The time of origin is $t_{or}=x_0$ and the time of the most recent common ancestor of the extant species is $t_{mrca}=x_1$.

Lemma 2.3

- How do you prove the first identity?

$$\sum_{N=n}^{\infty} \binom{N}{n} a^N = \frac{a^n}{(1-a)^{n+1}},$$

Lemma 2.3

- How do you prove the first identity?

$$\sum_{N=n}^{\infty} \binom{N}{n} a^N = \frac{a^n}{(1-a)^{n+1}},$$

- You can always ask AI

Lemma 2.3

- How do you prove the first identity?

$$\sum_{N=n}^{\infty} \binom{N}{n} a^N = \frac{a^n}{(1-a)^{n+1}},$$

- You can always ask AI
- How do you know if AI got it right?

Master equations

$$\dot{p}_0(t) = \mu - (\lambda + \mu + \psi)p_0(t) + \lambda p_0(t)^2, \qquad p_0(0) = 1 - \rho,$$
$$\dot{p}_1(t) = -(\lambda + \mu + \psi)p_1(t) + 2\lambda p_0(t)p_1(t), \qquad p_1(0) = \rho.$$

- What are these, and where do they come from?

An informal derivation

let $A_0(t)$, $A_1(t)$ be the events corresponding to $p_0(t)$, $p_1(t)$.

$$P[A_0(t+h)] = E\Big[P[A_0(t+h)\,\Big|\,\text{outcomes in } [t,t+h]]\Big] =$$

$$(1-[\lambda+\mu+\phi]h)P[A_0(t)] \tag{1}$$

$$+\lambda h(P[A_0])^2 \tag{2}$$

$$+\mu h \tag{3}$$

$$+o(h) \tag{4}$$

Now subtract $P[A_0(t)]$ from both sides, divide by h, and take a limit.

Note: we've passed a limit through expectation without justification.

Theorem 3.1.

How did Stadler guess this solution?

Theorem 3.5.

- This is a key result, since it gives the likelihood of a given genealogy under the linear birth-death-sampling model.

Theorem 3.5.

- This is a key result, since it gives the likelihood of a given genealogy under the linear birth-death-sampling model.
- How is this derived?

References I