STATS 700-002 Class 5.
Markov genealogy processes

Aaron King and Edward Ionides

September 25, 2025

Outline

King AA, Lin Q, Ionides EL. (2022) Markov genealogy processes. *Theoretical Population Biology* **143**:77–91. (doi:10.1016/j.tpb.2021.11.003).

1. A general framework for building phylogenies resulting from dynamic models of compartmentalized populations.
2. Theorems on finding the likelihood.
3. Computational strategies to implement the resulting formulas.

Various relevant processes

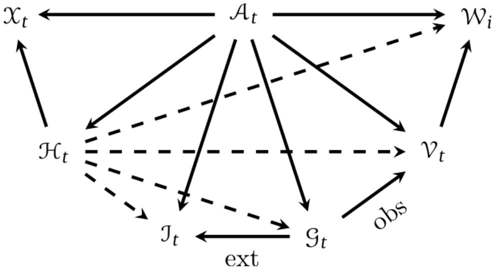
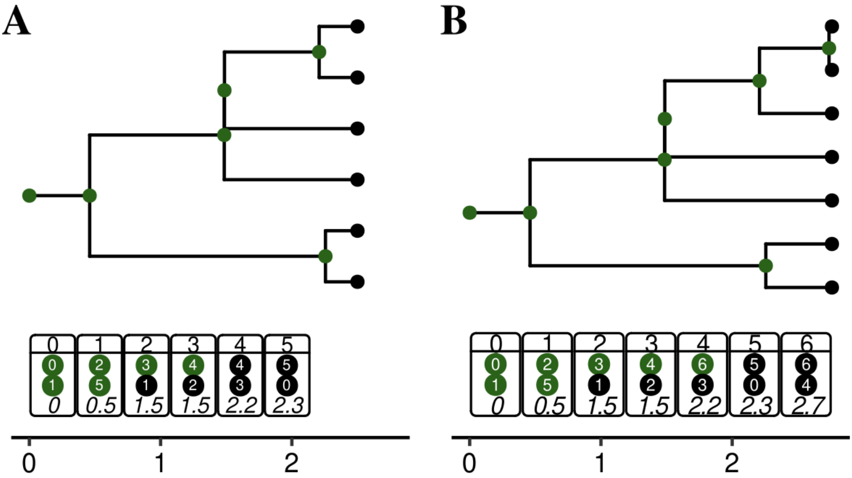
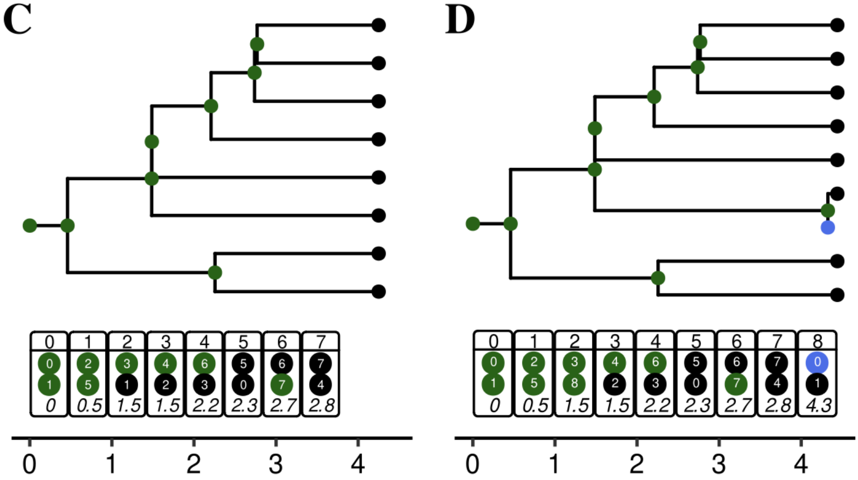


Fig. 1. Relations among the various Markov processes discussed in the paper. Deterministic maps are indicated with solid arrows; random maps are shown as dashed arrows. All the maps shown commute. $\mathcal{X}_t$ is the *population process*, a model of the dynamics of some system, which we take as a starting point. $\mathcal{H}_t$ is the *history process*, which records the full history of $\mathcal{X}_t$. $\mathcal{I}_t$ is the *inventory process*: at each time t, $\mathcal{I}_t$ is an inventory of all extant individuals in the population, each of which has a globally unique name. $\mathcal{G}_t$ is the *genealogy process*, which captures the precise genealogical relationships among all individuals in $\mathcal{I}_t$, as well as among any samples that have been taken from the population. $\mathcal{V}_t$ is the *visible genealogy process*, which is $\mathcal{G}_t$ pruned so that only relationships among samples remain. Finally $\mathcal{W}_i$ is the *embedded chain of the visible genealogy process*, which is $\mathcal{V}_{s_i}$, s_i being the time of the ith sample. All of these processes can be obtained via deterministic procedures applied to the *master process* $\mathcal{A}_t$, as described in the text.

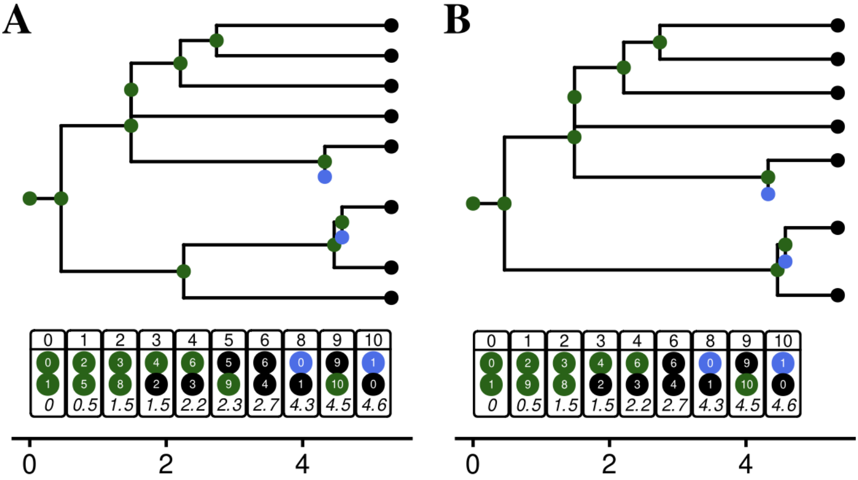
Birth


A
B
0 1 2 3 4 5
0 1 2 3 4 5
1 5 1 2 3 0
0 0.5 1.5 1.5 2.2 2.3
0 1 2
0 1 2 3 4 5 6
0 2 3 4 6 5 6
1 5 1 2 3 0 4
0 0.5 1.5 1.5 2.2 2.3 2.7
0 1 2

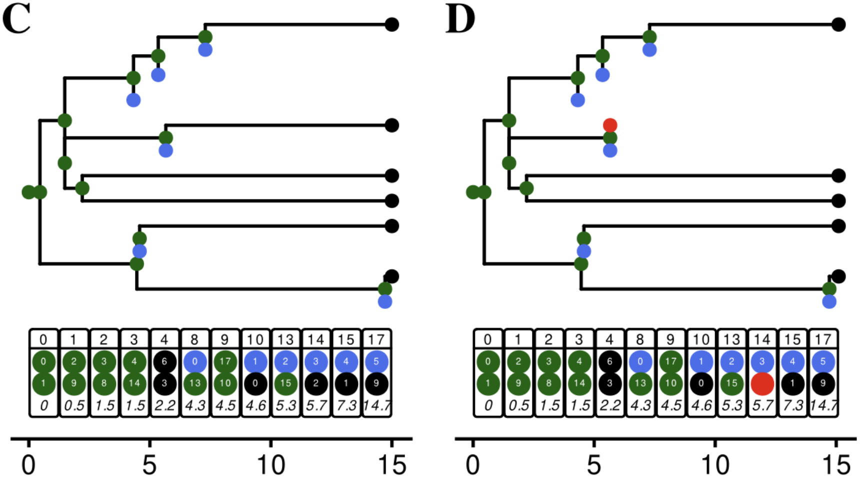
Sampling


C
D
0 1 2 3 4 5 6 7
0 2 3 4 6 5 6 7
1 5 1 2 3 0 7 4
0 0.5 1.5 1.5 2.2 2.3 2.7 2.8
0 1 2 3 4 5 6 7 8
0 2 3 4 6 5 6 7 0
1 5 8 2 3 0 7 4 1
0 0.5 1.5 1.5 2.2 2.3 2.7 2.8 4.3
0 1 2 3 4
0 1 2 3 4

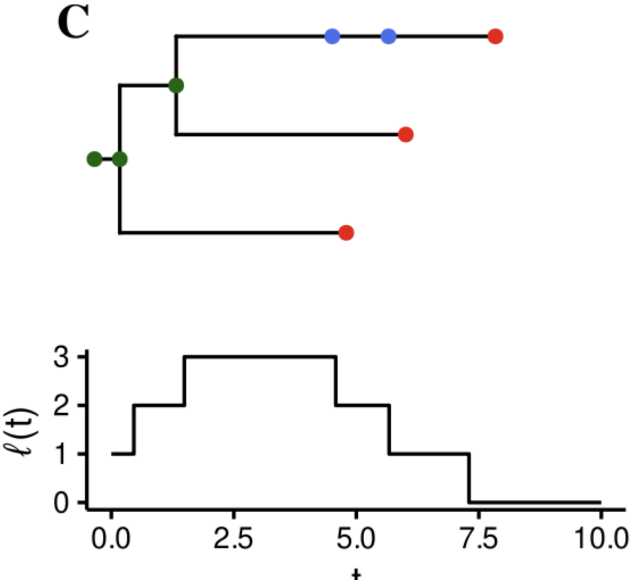
Death on an unobserved branch



Death on an observed branch



The lineage function, $\ell(t)$



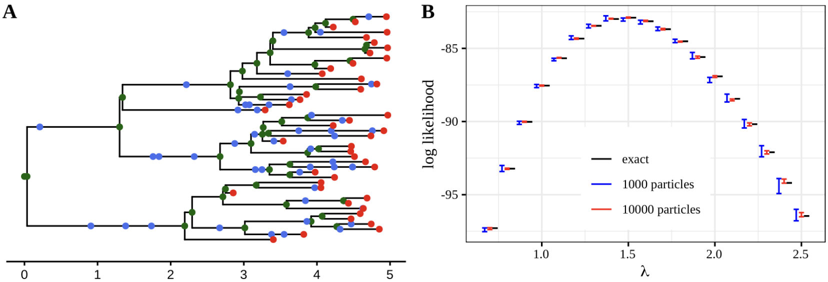
Theorem 2

$$P_{\mathcal{V}_t|\mathcal{H}_t}(\mathcal{V}_t|h) = \frac{\prod_{e \in \mathrm{U}(h)} \left(1 - \frac{\binom{\ell(e,\mathcal{V}_t)}{2}}{\binom{I(x_e)}{2}}\right) \prod_{e \in \mathrm{L}(\mathcal{V}_t)} \left(1 - \frac{\ell(e,\mathcal{V}_t)}{I(x_e)}\right)}{\prod_{e \in \mathrm{C}(\mathcal{V}_t)} \binom{I(x_e)}{2} \prod_{e \in \mathrm{D}(\mathcal{V}_t)} I(x_e)}.$$

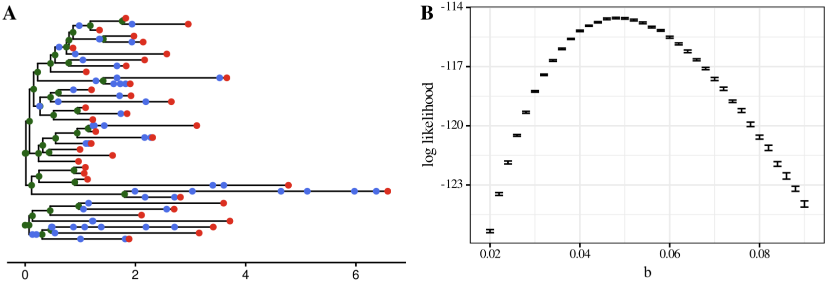
Integrating over the population dynamics via sequential Monte Carlo (SMC), also known as the particle filter.

- Each term in Theorem 2 can be treated as a "measurement"
- This can be mapped on to the computational task of a hidden Markov model, also known as a state space model or a partially observed Markov process (POMP) model
- A standard method for likelihood calculation in is the **particle filter**

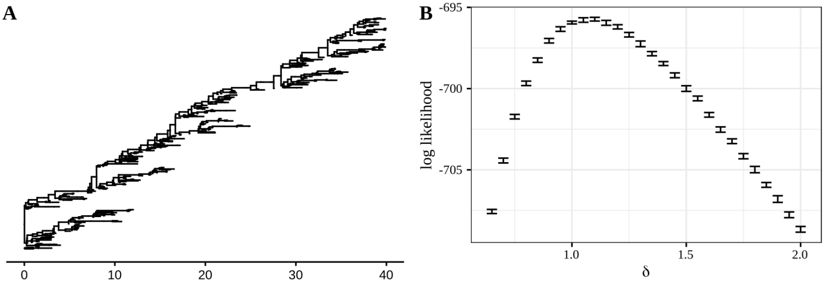
Example: Stadler's birth-death model



Example: SIR model



Example: SIRS model



Kolmogorov equations

- For any one of the examples in Section 3, what is the corresponding Kolmogorov forward equation (Eq. 3)?
- Why are the forward and backward equations adjoint? Solution

Why is the pruned-tree process Markovian?

- This may be unexpected.
- The full genealogy process is Markovian.
- The pruned tree is a function of the full tree, and functions of Markov processes are not generally Markov.

Questions

1. Regarding the absence of structure assumption, if there is population structure and migration events are occurring, how would the theory change?
2. How is Eq. 16 derived?

Comparing with previous results

1. Compare the form of the filter equation for the linear birth-death-sampling model (§6) to that obtained by Stadler (2010).
2. Compare the form of the filter equation for the SIR model (§6) to that obtained by Volz (2009).