

Class 5: Markov Genealogy Processes

Kunyang He

2025-10-23

Table of contents

1	What the paper does and why it matters	1
1.1	From population dynamics to trees	1
1.2	Two likelihood formulas (conditioning then integrating)	2
1.3	Examples (how the filter specializes)	3
2	Answers to the slide questions (slides 14–17)	3
2.1	Provenance of questions	3
2.2	(Slide 14) Kolmogorov equations	3
2.3	(Slide 15) Why is the pruned-tree process Markovian?	4
2.4	(Slide 16.1) What changes if there is population structure and migration?	4
2.5	(Slide 16.2) How is Eq.~(16) derived?	5
2.6	(Slide 17.1) Compare the linear birth–death–sampling filter to Stadler (2010) . .	5
2.7	(Slide 17.2) Compare the SIR filter to Volz (2009)	6
3	How to compute the likelihood in practice (algorithm sketch)	6
4	Notation crib (quick reference)	6

1 What the paper does and why it matters

1.1 From population dynamics to trees

Population process.

Let $X_t \in \mathbb{Z}^d$ be a non-explosive continuous-time Markov jump process with initial law p_0 and hazards $\alpha_u(t, x)$ for jumps $x \mapsto x + u$.¹

The Kolmogorov backward and forward equations are, respectively,

$$\partial_s F(s, x) = - \sum_u \alpha_u(s, x) [F(s, x + u) - F(s, x)], \quad (2)$$

$$\partial_t w(t, x) = \sum_u [\alpha_u(t, x - u) w(t, x - u) - \alpha_u(t, x) w(t, x)]. \quad (3)$$

Master and history processes.

A *master process* $A_t = (t, \omega|_t)$ stores the entire point-process realization up to t , and the *history* H_t records the ordered list of event times and types (t_k, u_k) .²

¹See definitions and Eqs. (1)–(3), pp. 2–3 of the paper.

²Master/history constructions: pp. 3–4.

Inventory and genealogy.

Define I_t to be the *inventory* of live individuals (with globally unique names). The paper encodes genealogies as *node sequences* whose “pockets” contain colored balls (green = internal node, black = extant leaf, blue = sample, red = sampled lineage that has since died), yielding a *unique* diagrammatic representation of a tree/forest (Fig.~2, p.~6).³ Birth, death, and sampling events update the node sequence by local operations (Fig.~3 for births/samples; Fig.~4 for deaths).⁴

Pruning and the visible genealogy.

Prune away unsampled leaves (drop every black ball) to obtain the *visible genealogy* V_t relating only samples (Fig.~5, p.~9). The embedded chain $W_i := V_{S_i}$, at successive sample times S_i , adds one lineage each step and is trivially Markov (Fig.~6, p.~9).⁵ Let $\ell(t, V)$ be the lineage count at time t in V (right panel of Fig.~5, p.~9).

1.2 Two likelihood formulas (conditioning then integrating)

Theorem 1 (conditioning on the history).

Given the population *history* H , the likelihood of W_i factors into a product over history events via combinatorial terms q_{jk} that encode whether an event (birth/sample) is compatible with the way the j -th sampled lineage attaches to W_{j-1} . The result is

$$\mathbb{P}(W_i | H) = \prod_{j=1}^i \prod_{k=1}^K q_{jk},$$

with explicit q_{jk} in Eq.~(16), obtained by elementary counting arguments.⁶

Theorem 2 (integrating out the history).

Re-ordering products yields a history-integrated likelihood for the visible genealogy V_t :

$$\mathbb{P}(V_t | H_t) = \prod_{e \in U(h)} \left(1 - \frac{\binom{\ell(e, V_t)}{2}}{\binom{I(x_e)}{2}} \right) \cdot \prod_{e \in L(V_t)} \left(1 - \frac{\ell(e, V_t)}{I(x_e)} \right) \cdot \prod_{e \in C(V_t)} \frac{1}{\binom{I(x_e)}{2}} \cdot \prod_{e \in D(V_t)} \frac{1}{I(x_e)},$$

where $U(h)$ are unobserved births in the history. Intuitively: at a coalescence time $e \in C(V)$ the probability that the population birth responsible links two visible lineages is $1/\binom{I}{2}$, etc.⁷

Unnormalized nonlinear filter (DMZ analogue).

Define partial weights $w(t, x, V)$ that (i) evolve between genealogy-event times by a modified forward equation, and (ii) jump at visible events:

$$\begin{aligned} \partial_t w(t, x, V) = & \sum_u [\alpha_u(t, x - u) w(t, x - u, V) - \alpha_u(t, x) w(t, x, V)] \\ & - \sum_{u \in B} \alpha_u(t, x - u) \frac{\binom{\ell(t, V)}{2}}{\binom{I(x)}{2}} w(t, x - u, V) \\ & - \sum_{u \in G} \alpha_u(t, x - u) w(t, x - u, V) \\ & + \text{Dirac-}\delta \text{ updates at } t \in C(V), D(V), L(V), \end{aligned} \quad (21)$$

with $w(0, x, V) = p_0(x)$ and $L_V = \sum_x w(t(V), x, V)$.⁸ This recasts full-information phylo-

³Genealogy and colored-balls representation: pp. 5–7; Fig. 2 (p. 6).

⁴Local update rules and illustrations: Fig. 3 (p. 7) and Fig. 4 (p. 8).

⁵Pruning and W_i : pp. 8–9; Lemma 1 on p. 9.

⁶Theorem 1 and case-by-case q_{jk} in Eq. (16): pp. 9–10.

⁷Statement and proof sketch of Theorem 2: p. 10.

⁸Filter (21) and event-time updates (20); initial condition (22); likelihood (23). pp. 11–12.

dynamics as a plug-and-play Monte Carlo Feynman–Kac problem (particle filtering between genealogy events, with weight *and state* updates at those events; see Figs.~7–9 for examples and validation).⁹

1.3 Examples (how the filter specializes)

For the linear birth–death–sampling model (n alive, per-capita rates λ, δ, ψ), the between-event evolution is

$$\partial_t w(t, n) = \lambda(n-1) \left(1 - \frac{\binom{\ell}{2}}{\binom{n}{2}}\right) w(t, n-1) + \delta(n+1) w(t, n+1) - (\lambda + \delta + \psi) n w(t, n),$$

and the coalescence/direct-descent/leaf updates are $w \mapsto \frac{\lambda(n-1)}{\mu} \frac{1}{\binom{n}{2}} w$, $w \mapsto \frac{\psi}{\mu} w$, and $w \mapsto \frac{\psi}{\mu} (n-\ell) w$, respectively (Eq.~(24)–(25)).¹⁰ Analogous specializations are written out for SIR and SIRS on pp.~12–13 and illustrated in Figs.~8–9.¹¹

2 Answers to the slide questions (slides 14–17)

2.1 Provenance of questions

The questions appear on slides 14–17 of the class deck titled “*STATS 700-002 Class 5. Markov genealogy processes*”.¹²

2.2 (Slide 14) Kolmogorov equations

14.1 “For any one example in §3, what is the corresponding Kolmogorov forward equation (Eq.~3)?”

Take the linear birth–death–sampling model with state $X_t = (n_t, g_t)$ and hazards

$$\alpha_{(+1,0)}(t, (n, g)) = \lambda(t) n, \quad \alpha_{(-1,0)}(t, (n, g)) = \delta(t) n, \quad \alpha_{(0,+1)}(t, (n, g)) = \psi(t) n.$$

Let $p_{n,g}(t) = \mathbb{P}\{X_t = (n, g)\}$. Plugging these α into Eq.~(3) gives the forward (master) equation

$$\begin{aligned} \frac{\partial}{\partial t} p_{n,g}(t) &= \lambda(t)(n-1) p_{n-1,g}(t) + \delta(t)(n+1) p_{n+1,g}(t) + \psi(t) n p_{n,g-1}(t) \\ &\quad - [\lambda(t) + \delta(t) + \psi(t)] n p_{n,g}(t). \end{aligned} \quad (\star)$$

Remarks. (i) If one marginalizes out g the sampling term vanishes from the n -marginal; (ii) analogous forward equations follow for SIR/SIRS by substituting the α_u listed on p.~5.¹³

14.2 “Why are the forward and backward equations adjoint?”

Consider $\sum_x f(x) w(t, x) = \mathbb{E}[f(X_t)]$. Differentiating and using the forward equation (3) yields

$$\frac{d}{dt} \sum_x f(x) w(t, x) = \sum_{x,u} \left[f(x+u) - f(x) \right] \alpha_u(t, x) w(t, x).$$

Equivalently, holding t fixed and differentiating $F(s, x) = \mathbb{E}[f(X_t) \mid X_s = x]$ in s gives the backward equation (2):

$$\partial_s F(s, x) = - \sum_u \alpha_u(s, x) [F(s, x+u) - F(s, x)].$$

⁹Computation and examples: pp. 12–13; Fig. 7 (linear birth–death–sampling), Fig. 8 (SIR), Fig. 9 (SIRS).

¹⁰Specialization for the linear model and comparison to closed-form likelihoods: p. 12; Fig. 7.

¹¹SIR/SIRS specializations and illustrations: pp. 12–13; Figs. 8–9.

¹²See slides 14–17 in the deck.

¹³Examples and rate functions: pp. 5–6.

Hence $\frac{d}{dt}\langle f, w \rangle = \langle \mathcal{L}f, w \rangle = \langle f, \mathcal{L}^*w \rangle$ with \mathcal{L} the backward generator and \mathcal{L}^* the forward operator.¹⁴

2.3 (Slide 15) Why is the pruned-tree process Markovian?

Although a function of a Markov process need not be Markov, the pruned process $V_t = \text{obs}(G_t)$ is a pure-jump Markov process under the paper's assumptions (exchangeable individuals; births, deaths, samplings occur singly and do not coincide). Intuition and sketch:

- The full genealogy G_t is Markov by construction (it is a deterministic update of the Markov history H_t at each jump).
- Pruning drops exactly those black balls corresponding to extant, unsampled tips (Fig.~5, p.~9). The only future events that can change V_t are: a birth that coalesces two *visible* lineages, a direct-descent sample that attaches to a visible lineage, or a leaf sample that starts a new visible lineage. Each such transition depends on the present through $(\ell(t, V_t), I(X_t))$ and current hazards $\alpha_u(t, X_t)$, but not on the earlier path once V_t and X_t at time t are given.
- The embedded chain $W_i = V_{S_i}$ is *trivially* Markov because W_i contains W_{i-1} as an induced sub-tree (Lemma~1), and the attachment distribution for the new lineage depends only on W_{i-1} via $\ell(\cdot, W_{i-1})$.¹⁵

Formally, one verifies the Markov property from the small- Δ transition probabilities obtained by combining Eq.~(21) (between-event evolution) with the three event-time updates in Eq.~(20), whose factors depend only on the *current* V_t and X_t .¹⁶

2.4 (Slide 16.1) What changes if there is population structure and migration?

The “absence of structure” assumption ensures exchangeability: at any event a uniformly chosen live individual is equally likely to be the parent, die, or be sampled. With population structure (say, C demes or risk groups) and migration, three aspects change:

State and hazards.

Augment X_t to $X_t = (\mathbf{s}_t, \mathbf{i}_t, \dots)$ with $\mathbf{i}_t = (i_t^{(1)}, \dots, i_t^{(C)})$ etc. Birth (infection), death (recovery), sampling, and migration rates become *typed* hazards $\alpha_u^{(c \rightarrow c')}(t, x)$ that specify parent and child compartments.

Attachment probabilities.

The uniform $\beta_{u,x}$ over the first $I(x)$ individuals (used in Eq.~(10)) is replaced by compartment-weighted choices. In particular, the case-by-case combinatorics in Eq.~(16) change by replacing population-wide counts with compartmental counts. For example, at a sampling time in compartment c ,

$$q_{jk} = \begin{cases} 1 - \frac{1}{I_k^{(c)} - \ell_{jk}^{(c)}}, & \text{sample not direct-descent in } c, \\ \frac{1}{I_k^{(c)} - \ell_{jk}^{(c)}}, & \text{direct descent in } c, \end{cases}$$

where $I_k^{(c)} = I^{(c)}(x_k)$ and $\ell_{jk}^{(c)}$ is the number of visible lineages in compartment c just prior to t_k (generalizing Lemma~1 to keep track of per-compartment lineage counts).

Birth (coalescence) terms.

For within-compartment births ($c \rightarrow c$), the non-coalescence and coalescence factors in

¹⁴Eqs. (2)–(3) and discussion: pp. 2–3.

¹⁵Lemma 1 and the embedded chain: p. 9; Fig. 6.

¹⁶Event-time updates and filter equation: pp. 11–12.

Eq.~(16e–f) become

$$1 - \frac{\binom{\ell_{jk}^{(c)}}{2}}{\binom{I_k^{(c)}}{2}} \quad \text{and} \quad \frac{1}{\binom{I_k^{(c)}}{2}},$$

respectively. For cross-compartment births ($c \rightarrow c'$), a visible coalescence can only occur if the parent is in the compartment containing the pre-existing visible lineage; this introduces cross-terms of the form $\ell_{jk}^{(c)} / ((\binom{I_k^{(c)}}{1})(\binom{I_k^{(c')}}{1}))$ when the new lineage lies in c' and attaches to a visible lineage in c . The DMZ filter (21) then inherits these typed combinatorial weights by replacing ℓ and I with their compartmental analogues inside the birth/sampling factors.

Bottom line.

All constructions (master/history process, genealogy updates, pruning, W_i) remain intact if event types carry compartment labels; the counting in Theorems~1–2 and the “measurement” factors in Eq.~(21) simply become block-structured by compartment. The plug-and-play particle filter carries through unchanged, now on the higher-dimensional state X_t .¹⁷

2.5 (Slide 16.2) How is Eq.~(16) derived?

Eq.~(16) lists the q_{jk} by a complete partition of cases. Each term is a *counting argument* under exchangeability. Let $I_k = I(x_k)$ be population size just before t_k and $\ell_{jk} = \ell(t_k, W_{j-1})$. Then:

1. If $t_k \notin [a_j, s_j]$ or $u_k \notin B \cup G$, the event is irrelevant: $q_{jk} = 1$.
2. If $t_k \in A(W_{j-1})$ (an earlier attachment time), again $q_{jk} = 1$.
3. If $t_k \in (a_j, s_j) \setminus A(W_{j-1})$ and $u_k \in G$ (a sample occurred but not the j -th lineage), there are I_k equally likely individuals to be sampled, but ℓ_{jk} of them would have made t_k an attachment time; exclude these and also the one individual of the j -lineage: $q_{jk} = 1 - \frac{1}{I_k - \ell_{jk}}$.
4. If $t_k = a_j$ and $u_k \in G$ (direct-descent attachment), the unique sampled individual must be the single individual on the j -lineage among $I_k - \ell_{jk}$ admissible individuals: $q_{jk} = \frac{1}{I_k - \ell_{jk}}$.
5. If $t_k \in (a_j, s_j)$ and $u_k \in B$ (a birth that does *not* cause attachment), potential parent–child *unordered* pairs number $\binom{I_k}{2}$. Of these, $\binom{\ell_{jk}}{2}$ involve two visible lineages (which would have coalesced); among the remaining pairs, exactly ℓ_{jk} pairs involve the single j -lineage individual and one visible lineage in W_{j-1} (which would also have caused attachment). Excluding all attachment-causing pairs gives

$$q_{jk} = 1 - \frac{\ell_{jk}}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}.$$

6. If $t_k = a_j$ and $u_k \in B$ (the attachment was a coalescing birth), among the same $\binom{I_k}{2} - \binom{\ell_{jk}}{2}$ admissible pairs, there is *exactly one* that coalesces the j -lineage with the specific earlier visible lineage it attaches to; hence

$$q_{jk} = \frac{1}{\binom{I_k}{2} - \binom{\ell_{jk}}{2}}.$$

Collecting the cases yields Eq.~(16) as printed in the paper.¹⁸

2.6 (Slide 17.1) Compare the linear birth–death–sampling filter to Stadler (2010)

For the linear model, the MGP specialization (Eq.~(24)–(25)) yields an exact likelihood when integrated across genealogy events; Fig.~7B (p.~12) shows particle-filter estimates converging to

¹⁷The paper notes that generalizations (including overdispersion and simultaneous events) are feasible with more intricate combinatorics; see Discussion, p. 13.

¹⁸See bullets (a)–(f) and Eq. (16) on p. 10; Figs. 7–9 make the cases visually clear.

the *closed-form* log-likelihood of Stadler’s sampling-through-time birth–death model. The mapping of notation is direct: (λ, δ, ψ) here correspond to Stadler’s per-capita speciation, extinction, and sampling rates; the MGP likelihood factorizes into the same birth/sampling/branch contributions as Stadler’s product over event times, but MGP arrives there via a forward-in-time DMZ filter that *also* extends seamlessly to time/state-dependent and nonlinear hazards.¹⁹

2.7 (Slide 17.2) Compare the SIR filter to Volz (2009)

Specializing Eq.~(21) to SIR (p.~12) shows that the *instantaneous* intensity with which two *visible* lineages coalesce at time t is

$$\lambda_{\text{coal}}(t) = b(t) s(t) i(t) \cdot \frac{\binom{\ell(t)}{2}}{\binom{i(t)}{2}} = \binom{\ell(t)}{2} \cdot \frac{2 b(t) s(t)}{i(t) - 1} \approx \binom{\ell(t)}{2} \cdot \frac{2 b(t) s(t)}{i(t)}.$$

In the large-population, small-sample limit ($\ell \ll i$), this reduces to a standard coalescent with *per-pair* rate $\approx 2 b(t) s(t)/i(t)$, which is the Volz-type SIR coalescent rate; MGP additionally treats direct-descent sampling (blue nodes) explicitly and does *not* rely on independence of sample vs. population branching events, so it remains valid outside the asymptotic regime where coalescent approximations are accurate.²⁰

3 How to compute the likelihood in practice (algorithm sketch)

Treat the terms in Theorem~2 (or the Dirac- δ factors in Eq.~(21)) as *measurements* in a partially observed Markov process:

1. Simulate many particles of the population process X_t between consecutive genealogy-event times $e \in E(V)$.
2. At each e , *update both the weight and state* of every particle using the appropriate factor from Eq.~(20) (coalescence, direct-descent, leaf).
3. Normalize/resample as in standard particle filtering; continue to the next event. Sum weights at $t(V)$ to obtain L_V .

This is the Monte Carlo Feynman–Kac scheme advertised on p.~13 and validated in Figs.~7–9.²¹

4 Notation crib (quick reference)

- X_t : population Markov jump process; H_t : history; A_t : master process.
- I_t : inventory (live named individuals); G_t : full genealogy; V_t : pruned (visible) genealogy.
- $W_i = V_{S_i}$: embedded chain at sample times; $\ell(t, V)$: lineage count at time t .
- B, D, G : sets of event types (birth, death, sampling); $C(V), D(V), L(V)$: coalescence, direct-descent, and leaf times in V .

¹⁹See the linear example and Fig. 7: p. 12.

²⁰SIR specialization and Fig. 8: pp. 12–13; the framework unifies coalescent and birth–death viewpoints (Discussion, p. 13).

²¹Algorithmic remarks and “plug-and-play” property: p. 13.