Reading notes on **Exact phylodynamic likelihood via structured Markov genealogy processes** (KLI25)

Kunyang He

2025-10-23

Table of contents

Source: King, Lin & Ionides (2025), *Exact phylodynamic likelihood via structured Markov genealogy processes*, arXiv:2405.17032.

1 Executive summary

- Starting from a discretely structured, time-inhomogeneous **Markov jump process** (MJP) on the population, the paper constructs a **stochastic process on genealogies** and derives **exact** likelihoods for (i) *pruned* genealogies (colored by demes) and (ii) *obscured* genealogies (colors removed). These likelihoods are expressed as solutions of **forward filter equations** whose structure is determined entirely by the population model.

- **Key idea.** Condition on the population history and factor the genealogy likelihood **forward in time** across jump times. A local multiplier φ_u (one per jump mark u) is a product of binomial ratios built from: deme occupancy $n(x)$, event **production** r_u, pruned-lineage count ℓ, and **saturation** s (how many emergent lineages belong to the pruned genealogy).

- **Main results.** (i) Conditional-on-history factorization (Theorem 1). (ii) Integrating over histories yields a **filter equation** (Theorem 2) with an adjoint/backward form. (iii) For obscured genealogies, introduce an auxiliary process over colorings with importance weights; this gives another filter/adjoint pair (Theorems 4–6).

- **Algorithms.** Appendix B shows the filters correspond to an event-driven **SMC** (particle filter): simulate the MJP forward, multiply weights at genealogical events, optionally apply between-event decays. Algorithm B1 yields unbiased likelihood estimates.

- **Sanity checks.** The framework specializes to (a) the **Kingman coalescent** (via the Moran model) and (b) the **linear birth–death** model with sampling-through-time (Stadler), reproducing known closed forms.

2 Set-up: from a population MJP to a genealogy process

2.1 Population process

Let $(X_t)_{t \geq 0}$ be a (possibly time-inhomogeneous) MJP on a complete metric space X with transition rate kernel $\alpha(t, x, x') = \sum_{u \in U} \alpha_u(t, x, x')$, where U is a countable **mark set** (event types). Deterministic-time jumps (e.g., conditioning on sample times) are accommodated via a singular part of the Kolmogorov forward equation.

2.2 Demes and occupancy

A **deme** index set $\mathcal{D}$ partitions exchangeable subpopulations (e.g., incubation vs. infectious, strains, severity, behavior). The **deme occupancy** $n : X \to \mathbb{Z}_{\geq 0}^{\mathcal{D}}$ maps a population state x to deme-wise counts (e.g., $n(x) = (E, I)$ in SEIRS). *Figure 1 (p. 4)* illustrates SEIRS, two-strain competition, severity progression, and behavioral heterogeneity, highlighting which compartments harbor lineages.

2.3 Genealogy objects

A genealogy is $G = (T, Z, Y)$: time T, a c'adl'ag partition-valued tree Z_t, and coloring Y assigning a deme and a node-count along branches. **Pruning** removes extant tips and recursively prunes childless internal nodes; **obscuring** further erases colors. Two local statistics drive the likelihoods: the **lineage count** $\ell(Y_t)$ and **saturation** $s(Y_{t^-}, Y_t)$. *(Figures 5–6)*.

2.4 Event types and production

Pure event types are birth, death, migration, sample, and neutral (compound events allowed under mild restrictions). Define **production** $r_u = (r_{u,i})_{i \in \mathcal{D}}$: the number of emergent lineages by deme at a jump with mark u *(Figure 4)*.

3 Likelihood for pruned genealogies

3.1 Binomial ratio and local multiplier

For vectors $n, r, \ell, s \in \mathbb{Z}_{\geq 0}^{\mathcal{D}}$, define the deme-wise **binomial ratio**

$$\prod_{i \in \mathcal{D}} \frac{\binom{n_i - \ell_i}{r_i - s_i}}{\binom{n_i}{r_i}} \in [0, 1],$$

with the natural zero convention when constraints fail. For a jump of mark u producing state x (from some x'), the local factor is

$$\varphi_u(x, y^-, y) = \left[\prod_{i \in \mathcal{D}} \frac{\binom{n_i(x) - \ell_i(y)}{r_{u,i} - s_i(y^-, y)}}{\binom{n_i(x)}{r_{u,i}}} \right] Q_u(y^-, y),$$

where $Q_u \in \{0, 1\}$ is a **compatibility indicator** determined by the local genealogy.

3.2 Theorem 1 (conditional-on-history factorization)

Let $P = (T, Z, Y)$ be a pruned genealogy and $H = (X, U)$ a realized population history. Then

$$\mathbb{P}\{P_T = P \mid H_T = H\} = \mathbf{1}\{\operatorname{ev}(H) \supseteq \operatorname{ev}(P)\} \prod_{t \in \operatorname{ev}(H)} \varphi_{U_t}(X_t, Y_{t^-}, Y_t),$$

i.e., the likelihood **factors forward** over jump times using only local data through φ_u.

3.3 Theorem 2 (filter equation)

Marginalizing over histories yields a forward-time **filter** $w(t, x)$ with $w(0, x) = p_0(x)$:

$$\frac{\partial w}{\partial t}(t, x) = \sum_u \int w(t, x') \, \alpha_u(t, x', x) \, \varphi_u(x, Y_{t^-}, Y_t) \, dx' - \sum_u \int w(t, x) \, \alpha_u(t, x, x') \, dx', \quad t \notin \operatorname{ev}(P),$$

and at $t \in \operatorname{ev}(P)$,

$$w(t, x) = \sum_u \int w(t^-, x') \, \alpha_u(t, x', x) \, \varphi_u(x, Y_{t^-}, Y_t) \, dx'.$$

The **likelihood** is $L = \int w(T, x) \, dx$. An adjoint/backward form evolves $F(s, x)$ from $F(T, x) = 1$ and yields $L = \int F(0, x) \, p_0(x) \, dx$.

4 Likelihood for obscured genealogies

When colors are unobserved, introduce a finite-state auxiliary process y_t that proposes **paintings** of the fixed tree Z with kernels q (initial) and π (transition). With

$$\beta_u(t, x, x', y, y') = \alpha_u(t, x, x') \, \pi_u(t, x, x', y, y'), \quad \Psi_u = \frac{\varphi_u(x', y, y')}{\pi_u(t, x, x', y, y')},$$

a filter on (x, y) analogous to the pruned case yields the likelihood of the obscured genealogy $V = (T, Z)$; an adjoint counterpart is also available (Theorems 4–6). This is **importance sampling over colorings** with freedom to design efficient proposals π.

5 Filter equations & SMC

Appendix B formalizes **filter equations** with **boosts** at jumps and optional **decay** between jumps. Lemma B3 shows that augmenting the driver with a weight process V_t (multiplying by boosts at jumps and decaying exponentially) gives $\int w(t,x)\,dx = \mathbb{E}[V_t]$ and leads directly to an **event-driven SMC** implementation (Algorithm B1): simulate the MJP forward; at each genealogical event time do a singular update (reweight and optionally resample); between events accumulate deterministic decay; average weights for an unbiased likelihood estimator.

6 Special cases

6.1 Moran model $\Rightarrow$ Kingman coalescent

With a constant-size Moran population (Poisson event rate μ), the filter reduces to the classical Kingman coalescent likelihood. For m contemporaneous samples at T, the log-likelihood takes the familiar form

$$\log L = |B| \log\left(\mu \binom{n}{2}\right) - \mu \binom{n}{2} \sum_{i=1}^{m} \binom{i}{2} s_i,$$

where B are branch times and s_i are coalescent-interval lengths; inline/Poisson sampling factors appear as additional singular updates.

6.2 Linear birth–death with sampling-through-time

With per-capita birth λ, death μ, through-time sampling ψ, and bulk sampling probability ρ at T, the adjoint filter admits closed-form solutions coinciding with Stadler (2010), and generalizes to time-varying rates.

7 References (for convenience)

- King, A. A., Lin, Q., & Ionides, E. L. (2025). *Exact phylodynamic likelihood via structured Markov genealogy processes.* arXiv:2405.17032. DOI: 10.48550/arXiv.2405.17032.
- Vaughan, T. G. & Stadler, T. (2025). Bayesian phylodynamic inference of multitype population trajectories using genomic data. *Molecular Biology and Evolution*, **42**: msaf130.