Scaling limits of sequential Monte Carlo genealogies

Jere Koskela[1], Paul Jenkins[1,2], Adam Johansen[1,3], and Dario Spanò[1]
Departments of [1]Statistics and [2]Computer Science, University of Warwick, and the [3]Alan Turing Institute

Warwick Statistics

Introduction

- Sequential Monte Carlo (SMC): a class of numerical schemes for non-linear filtering with a wide range of applications [Del04].
- $(X_t, Y_t)_{t \geq 1}$: a Markov chain in which Y_t is a noisy observation of X_t.
- $X_t | X_{t-1}$ has transition density $p(x'|x)$ and $Y_t | X_t$ has density $g(y|x)$.
- Interested in functionals of the *smoothing density* of $X_{1:T} | X_0, Y_{1:T}$:

$$P(x_1, \ldots, x_T) \propto \prod_{t=1}^{T} p(x_t | x_{t-1}) g(y_t | x_t).$$

- Particle filter with system size $N \in \mathbb{N}$ and proposal density $q(x'|x)$:
  1. Set $x_0^{(i)} \leftarrow x_0$ and $w_i \leftarrow 1/N$ for $i \in \{1, \ldots, N\}$.
  2. For each $t \in \{1, \ldots T\}$:
     1. Sample $a_t^{(i)} \sim \text{Categorical}(w_1, \ldots, w_N)$ for $i \in \{1, \ldots, N\}$.
     2. Sample $x_t^{(i)} \sim q(\cdot | x_{t-1}^{(a_t^{(i)})})$ for $i \in \{1, \ldots, N\}$.
     3. Set $\tilde{w}_i \leftarrow p(x_t^{(i)} | x_{t-1}^{(a_t^{(i)})}) g(y_t | x_t^{(i)}) / q(x_t^{(i)} | x_{t-1}^{(a_t^{(i)})})$ for $i \in \{1, \ldots, N\}$.
     4. Set $w_i \leftarrow \tilde{w}_i / \sum_{j=1}^{N} \tilde{w}_j$ for $i \in \{1, \ldots, N\}$.
- A P-functional $f(x_1, \ldots, x_T)$ can be approximated as

$$\mathbb{E}_P[f(X_1, \ldots, X_T)] \approx \sum_{i=1}^{N} w_i f(x_1^{(a_2^{\cdot^{\cdot^{(i)}}})}, \ldots, x_{T-1}^{(a_T^{(i)})}, x_T^{(i)}). \quad (1)$$

Path storage and degeneracy

- A naive implementation requires $O(N \times T)$ storage.
- But common ancestry induced by resampling (step 2.1 in the algorithm) means that many of these states are not evaluated in (1).
- Hence the storage cost can be reduced.
- Common ancestry also means that times $t \ll T$ will be estimated using fewer than N realisations, increasing variance.
- Increased variance due to loss of paths is known as *path degeneracy* [LC95].
- This work identifies the asymptotic distribution of genealogical trees and hence provides the first tool for quantifying path degeneracy *before* running the algorithm.
- The limit is most conveniently expressed in terms of the n-coalescent [Kin82].

The n-coalescent and the genealogical process

- n-coalescent: a continuous-time process initialised from $\{\{1\}, \ldots, \{n\}\}$ in which each pair of blocks merges at rate 1.
- It terminates once one block remains, resulting in a random tree.
- A realisation is depicted in Figure 1.

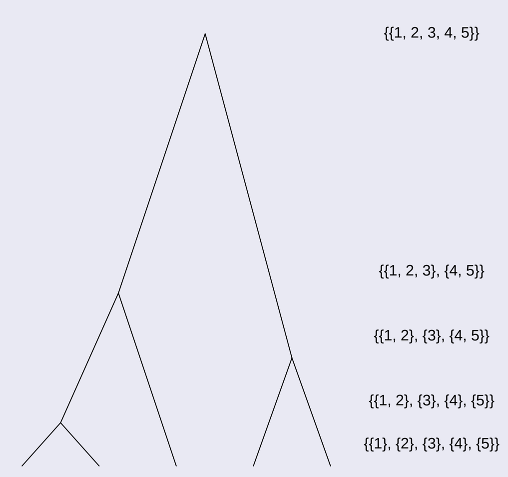

Figure 1: A realisation of the 5-coalescent.

- The same encoding defines a genealogical process $(G_t^{(n,N)})_{t \geq T}$ of $n \leq N$ particles sampled uniformly from a particle filter.
- The initial state is $\{\{1\}, \ldots, \{n\}\}$.
- Indices i and j belong to the same block in $G_t^{(n,N)}$ if particles i and j have a common ancestor t generations ago.

Assumptions and the time scale

- $\nu_t^{(i)}$: random number of offspring of particle i at time t.
- **Assumption 1**: Conditional on $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$, all valid assignments of offspring to parents are equally likely.
- **Assumption 2**: The densities $p(x'|x)$, $q(x'|x)$, and $g(y|x)$ are bounded from above and away from 0.
- Numerical experiments suggest that neither is necessary.
- Assumption 1 $\Rightarrow$ the marginal probability of two blocks merging in one generation at time t is

$$c_N(t) := \frac{1}{N(N-1)} \sum_{i=1}^{N} \mathbb{E}[\nu_t^{(i)}(\nu_t^{(i)} - 1)].$$

- The generalised inverse defines a time change

$$\tau_N(t) := \max \left\{ s \geq 0 : \sum_{r=1}^{s} c_N(r) \geq t \right\},$$

which is the correct rescaling for a non-trivial limiting genealogy.

The convergence theorem

- Suppose Assumptions 1 and 2 hold, and that multinomial resampling is done at every time step.
- Then $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ converges to a Markov process $(G_t^{(n)})_{t \geq 0}$ as $N \to \infty$ when n is fixed.
- $(G_t^{(n)})_{t \geq 0}$ admits only binary mergers.
- The rate at which blocks merge is bounded between two constants

$$0 < \frac{C_*}{C} \leq 1 \leq \frac{C}{C_*} < \infty.$$

- C_* and C are determined by cumbersome bounds on joint moments of family sizes which hold under Assumption 2.
- If the vector of family sizes

$$(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$$

is exchangeable at each time, then $(G_t^{(n)})_{t \geq 0}$ is an n-coalescent with variable merger rate between C_*/C and C/C_*.

Consequences of the convergence theorem

- $T_N^{(n)}$: number of generations separating n leaves of the genealogical tree from their most recent common ancestor.
- Under Assumptions 1 and 2, the following bounds hold for any sufficiently large N:

$$\mathbb{E}\left[\frac{T_N^{(n)}}{N}\right] \leq \frac{2C}{C_*^2}\left(1 - \frac{1}{n}\right),$$

$$\mathbb{E}\left[\frac{T_N^{(n)}}{N}\right] \geq \frac{2C_*}{C^2}\left(1 - \frac{1}{n}\right) + O(N^{-1}),$$

$$\text{Var}\left(\frac{T_N^{(n)}}{N}\right) \leq \left(\frac{4\pi^2}{3} - 12 + O(n^{-1})\right)\left(\frac{C}{C_*^2}\right)^2,$$

$$\text{Var}\left(\frac{T_N^{(n)}}{N}\right) \geq \left(\frac{4\pi^2}{3} - 12 + O(n^{-1})\right)\left(\frac{C_*}{C^2}\right)^2 + O(N^{-1}).$$

- [JMR15] showed that $\mathbb{E}[T_N^{(N)}] = O(N \log N)$, also under Assumption 2.
- Our result provides an $O(N)$ lower bound, showing their bound is tight up to a $\log N$ factor.
- $L_N^{(n)}$: total branch length of the tree connecting n leaves to their most recent common ancestor.
- Under Assumptions 1 and 2, the following bounds hold for any sufficiently large N:

$$\mathbb{E}\left[\frac{L_N^{(n)}}{N}\right] \leq \frac{2C}{C_*^2}(\log n + \gamma_{EM} + O(n^{-1})),$$

$$\mathbb{E}\left[\frac{L_N^{(n)}}{N}\right] \geq \frac{2C_*}{C^2}(\log n + \gamma_{EM} + O(n^{-1})) + O(N^{-1}),$$

$$\text{Var}\left(\frac{L_N^{(n)}}{N}\right) \leq \left(\frac{2\pi^2}{3} + O(n^{-1})\right)\left(\frac{C}{C_*^2}\right)^2,$$

$$\text{Var}\left(\frac{L_N^{(n)}}{N}\right) \geq \left(\frac{2\pi^2}{3} + O(n^{-1})\right)\left(\frac{C_*}{C^2}\right)^2 + O(N^{-1}),$$

where $\gamma_{EM} \approx 0.577$ is the Euler-Mascheroni constant.

- All of these follow from elementary calculations for the n-coalescent on e.g. page 76 of [Wak09].
- In addition to contributing to SMC, this also extends the domain of attraction of the n-coalescent.
- Previous work has focused on *neutral* systems [Möh98], where family sizes do not depend on particle locations.
- Our theorem identifies the n-coalescent as the genealogical process of *selective* particle systems as well.

Simulation study: set up

- We conducted a simulation study to verify these bounds for finite N.
- The model was the discretised Ornstein-Uhlenbeck process

$$X_{t+1} = (1 - \Delta)X_t + \sqrt{\Delta}\xi_t,$$
$$X_0 \sim N(0, 1),$$
$$Y_t | X_t \sim N(X_t, \sigma^2),$$

where $\xi_t \sim N(0, 1)$ is white noise, and with $\Delta = \sigma = 0.1$ as well as a time horizon $T = 40\,960$.

- We ran 1 000 realisations of a boostrap particle filter with $q(x'|x) = p(x'|x)$ and stored the tree heights.
- Figures 2 and 3 show the corresponding estimates of the mean and variance of $T_N^{(n)}$.
- Assumption 2 fails for this experiment.
- Assumption 1 also fails for the three resampling schemes other than multinomial (see [DCM05] for details of these schemes).

Simulation study: results

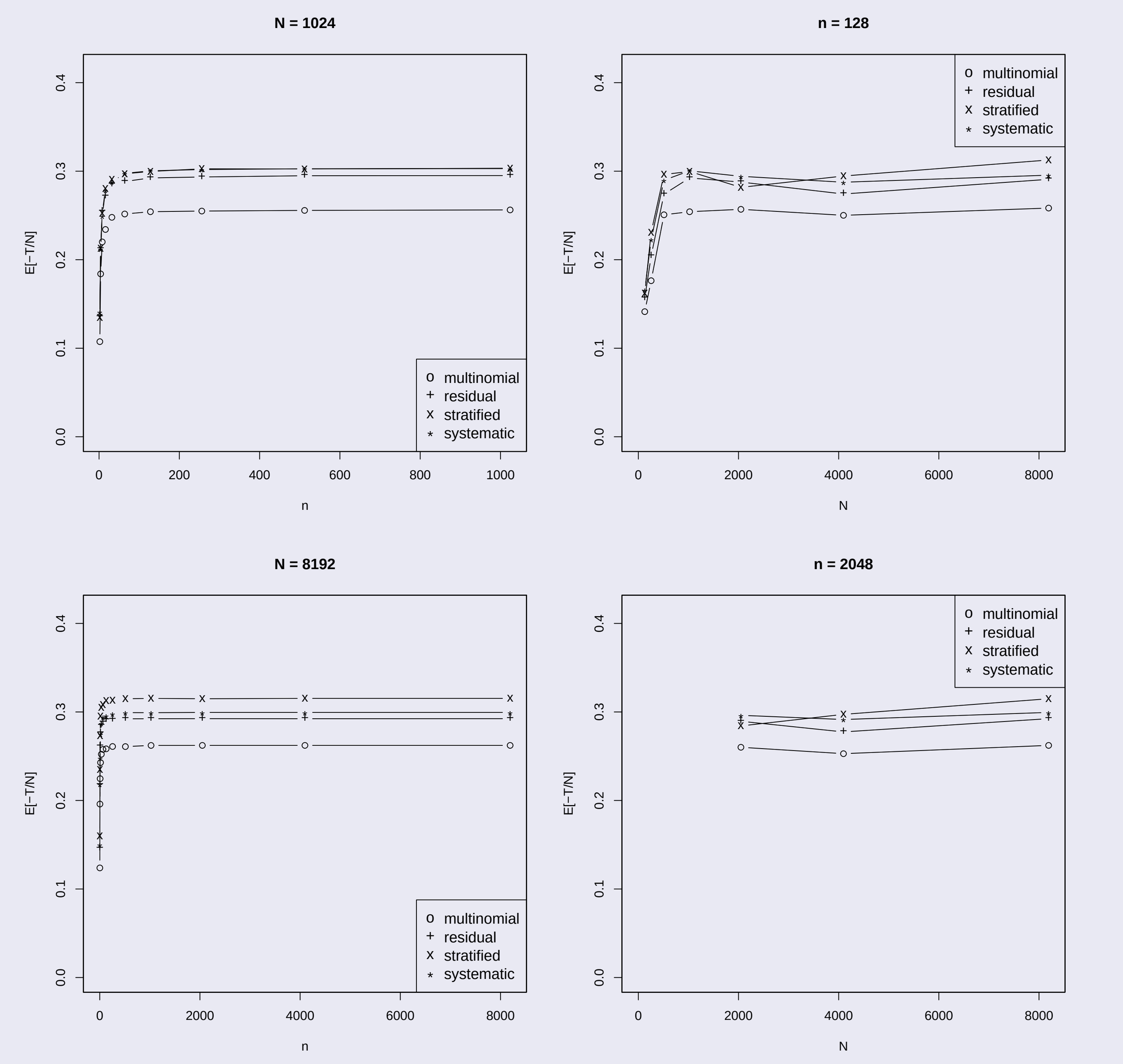

Figure 2: Averaged tree heights from 1 000 realisations for fixed N as a function of n on the left, and vice versa on the right.

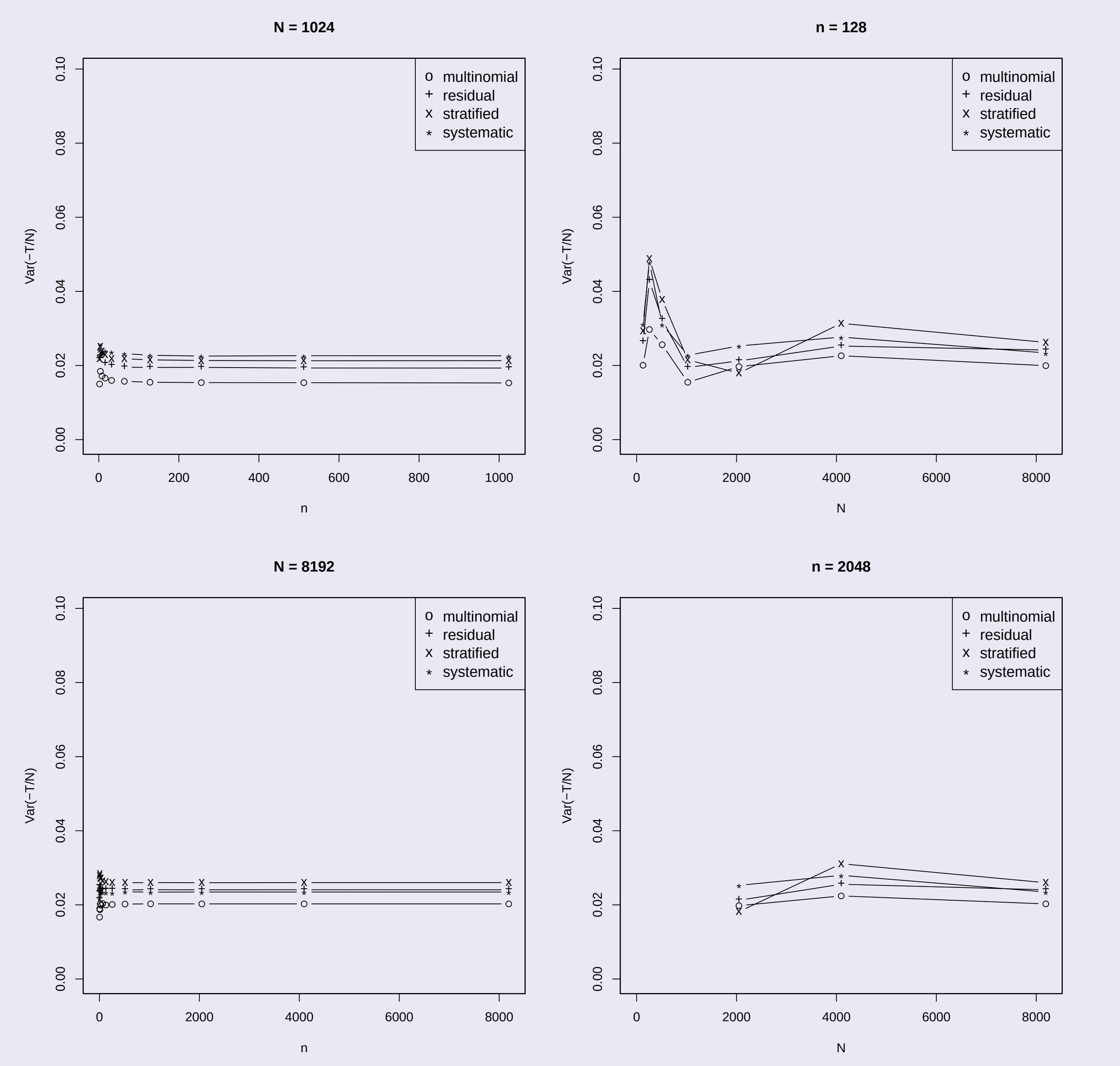

Figure 3: Variance of tree heights from 1 000 realisations for fixed N as a function of n on the left, and vice versa on the right.

Conclusions

- We have related genealogical trees of sequential Monte Carlo algorithms to the tractable n-coalescent.
- This enables *a priori* estimation of functionals of the tree.
- It also extends the domain of attraction of the n-coalescent.
- Simulation studies confirm that asymptotic results accurately describe algorithms with finite N.
- The strong assumptions do not appear to be necessary in practice.
- Our method only works when $n \ll N$, but the result seems to hold even when $n \approx N$.

Acknowledgments and References

JK was supported by EPSRC grant EP/HO23364/1 as part of the MASDOC DTC at the University of Warwick, and by DFG grant BL 1105/3-2. PJ was supported in part by EPSRC grant EP/L018497/1. AJ was partially supported by Lloyd's Register Foundation – Alan Turing Institute Programme on Data-Centric Engineering.

[DCM05] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69, 2005.

[Del04] P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications.* Springer, New York, 2004.

[JMR15] P. E. Jacob, L. M. Murray, and S. Rubenthaler. Path storage in the particle filter. *Stat. Comput.*, 25:487–496, 2015.

[Kin82] J. F. C. Kingman. The coalescent. *Stochast. Process. Applic.*, 13(3):235–248, 1982.

[LC95] J. S. Liu and R. Chen. Blind deconvolution via sequential imputations. *J. Am. Statist. Assoc.*, 90(430):567–576, 1995.

[Möh98] M. Möhle. Robustness results for the coalescent. *J. Appl. Probab*, 35:438–447, 1998.

[Wak09] J. Wakeley. *Coalescent theory: an introduction.* Roberts & Co, 2009.