Scaling limits of SMC genealogies

Jere Koskela, Paul Jenkins[1], **Adam M. Johansen**[2], and Dario Spanò, all at Department of Statistics, University of Warwick.
arxiv: 1804.01811

Imperial College Workshop — 9th May, 2018

[1]Also at Department of Computer Science, University of Warwick
[2]Also at The Alan Turing Institute

Outline

Sequential Monte Carlo

for $i \in \{1, \ldots, N\}$: Sample $\mathbf{X}_0^{(i)} \sim \mu_0$.
for $i \in \{1, \ldots, N\}$: Set

$$w_0^{(i)} \leftarrow \frac{g_0(\mathbf{X}_0^{(i)})}{\sum_{j=1}^{N} g_0(\mathbf{X}_0^{(j)})}.$$

for $t \in \{1, \ldots, T\}$ **do**
    Sample $(a_t^{(1)}, \ldots, a_t^{(N)}) \sim \text{Categorical}(w_{t-1}^{(1)}, \ldots, w_{t-1}^{(N)})$.
    for $i \in \{1, \ldots, N\}$: Sample $\mathbf{X}_t^{(i)} \sim \mu_t(\cdot|\mathbf{X}_{t-1}^{(a_t^{(i)})})$.
    for $i \in \{1, \ldots, N\}$: Set

$$w_t^{(i)} \leftarrow \frac{g_t(\mathbf{X}_{t-1}^{(a_t^{(i)})}, \mathbf{X}_t^{(i)})}{\sum_{j=1}^{N} g_t(\mathbf{X}_{t-1}^{(a_t^{(j)})}, \mathbf{X}_t^{(j)})}.$$

end for

Example: Hidden Markov Model

- Let $\{\mathbf{X}_t\}_{t\geq 0}$ be a discrete time Markov process with transition density $p(\mathbf{x}, \mathbf{x}')$ and initial density $\pi(\mathbf{x})$.
- Suppose a noisy observation $\mathbf{Y}_t$ with density $g(\mathbf{y}|\mathbf{x})$ is made of each state $\mathbf{X}_t$.
- Conventional particle filters target $\mathbb{P}(\mathbf{X}_{0:T} \in d\mathbf{x}_{0:T} | Y_{0:T} = y_{0:T})$.
- E.g. the bootstrap particle filter (Gordon et al., 1993): $\mu_0 = \pi$, $\mu_t(d\mathbf{x}_t|\mathbf{x}_{t-1}) = p(\mathbf{x}_{t-1}, \mathbf{x}_t)dx_t$, and $g_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = g(\mathbf{y}_t|\mathbf{x}_t)$.

Path degeneracy/storage

- Suppose $T \gg 1$.
- Mergers due to resampling mean that the smoothing distributions at times $t \ll T$ are approximated with $m \ll N$ paths.

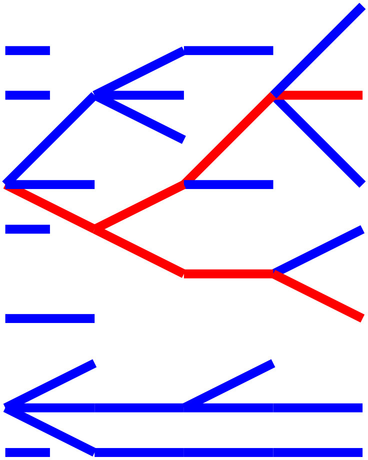
  $\Rightarrow$ High variance estimators.
- Loss of paths also means that fewer than $N \times T$ particles need to be stored, reducing memory cost.
- Aim: characterization of $\mathbb{E}[T_{MRCA}]$, $\text{Var}(T_{MRCA})$, etc.
- Related work: Jacob et al. (2015).

The coalescent process (Kingman, 1982)

- Let $\{R_t^{(n)}\}_{t \geq 0}$ be a partition-valued process.
- $R_0^{(n)} = \{\{1\}, \ldots, \{n\}\}$.
- Each pair of blocks $\{i\}, \{j\}$ merge at rate 1.
- A "death" process of rate $\binom{k}{2}$ where k is the number of blocks.

The genealogical process

- Let $\{G_t^{(n,N)}\}_{t \geq 0}$ be the genealogy of $n \leq N$ particles sampled randomly from an N-particle SMC algorithm of interest.
- $G_0^{(n,N)} = \{\{1\}, \ldots, \{n\}\}$.
- $i \sim j$ in $G_t^{(n,N)} \Rightarrow$ particles i and j have a common ancestor t generations ago.
- $G^{(2,7)}$ illustrated.



Rescaling time

- For $i \in \{1, \ldots, N\}$ and $t \in \mathbb{N}$, let $\nu_t^{(i)}$ be the number of children of particle i, t generations *ago*.
- Standing assumption: $(\nu_t^{(1)}, \ldots, \nu_t^{(N)})$ is exchangeable for every t.
- Define

$$c_N(t) := \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}[(\nu_t^{(i)})_2] \sim \frac{1}{N} \mathbb{E}[(\nu_t^{(1)})_2],$$

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

- For multinomial resampling:

$$c_N(t) = \sum_{i=1}^{N} \mathbb{E}[(w_t^{(i)})^2] \approx 1/\mathrm{ESS}(t)$$

Convergence theorem

Main Result (Koskela et al., 2018)

Suppose (for all sufficiently large N):

$$\mathbb{E}\left[\prod_{d=1}^{k}\prod_{a_d=1}^{\alpha_d}\prod_{b_d=1}^{\beta_d}\prod_{c_d=1}^{\gamma_d}(\nu^{(1)}_{r^{(d)}_{a_d}})_2(\nu^{(2)}_{s^{(d)}_{b_d}})^n(\nu^{(3)}_{t^{(d)}_{c_d}})_2(\nu^{(4)}_{t^{(d)}_{c_d}})^2\right] \leq C^{\sum_{d=1}^{k}\alpha_d+\beta_d+\gamma_d},$$

and an analogous lower bound with constant C_*, with $0 < C_* \leq C < \infty$ (and the standing exchangeability assumption).
Then: $(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ converges (in the sense of fdds) to a time-inhomogeneous Kingman coalescent with merger rate varying between C_*/C and C/C_* as $N \to \infty$ for fixed n.

Extends an argument of Möhle (1998) to a *non-Markovian* setting.

Sketch proof

- Let ξ and η be partitions of $\{1, \ldots, n\}$, with the block sizes of η in terms of the blocks of ξ being $b_1, \ldots, b_{|\eta|}$,
i.e. $b_1 + \ldots + b_{|\eta|} = |\xi|$.
- The conditional one-step transition probability of $G_t^{(n,N)}$ given family sizes is

$$p_{\xi\eta}(t) := \frac{1}{(N)_{|\xi|}} \sum_{i_1 \neq \ldots \neq i_{|\eta|}=1}^{N} (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_{|\eta|})})_{b_{|\eta|}}.$$

- FDDs:

$$\mathbb{P}(G_{\tau_N(t_1)}^{(n,N)} = \eta_1, \ldots, G_{\tau_N(t_k)}^{(n,N)} = \eta_k | G_{\tau_N(t_0)}^{(n,N)} = \eta_0)$$
$$= \mathbb{E}\left[\prod_{d=1}^{k} \left\{ \prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} P_N(r) \right\}_{\eta_{d-1}\eta_d}\right].$$

Sketch proof II

- For a single time interval

$$\left\{ \prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} P_N(r) \right\}_{\eta_{d-1}\eta_d} = \sum_{\boldsymbol{\xi}} \prod_{s=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} p_{\xi_{s-1}\xi_s}(s),$$

  where $\boldsymbol{\xi} = (\xi_{\tau_N(t_{d-1})}, \xi_{\tau_N(t_{d-1})+1}, \dots, \xi_{\tau_N(t_d)})$.
- Each partition in $\boldsymbol{\xi}$ is either equal to its predecessor, or obtained from its predecessor by either
  1. exactly one binary merger (α_d),
  2. a set of mergers at least one of which is larger than binary (β_d),
  3. more than one binary merger, but no larger than binary mergers (γ_d).

Sketch proof III

- The assumed bounds imply $\tau_N(t) \sim Nt$.
- More precisely the number of $\boldsymbol{\xi}$'s is bounded above by

$$\binom{\tau_N(t_d) - \tau_N(t_{d-1})}{\alpha_d + \beta_d + \gamma_d} 2^{|\eta_{d-1}|(\alpha_d+\beta_d+\gamma_d)} \leq \frac{(2^{|\eta_{d-1}|}(t_d - t_{d-1})N)^{\alpha_d+\beta_d+\gamma_d}}{C_*^{\alpha_d+\beta_d+\gamma_d}(\alpha_d + \beta_d + \gamma_d)!}.$$

- An elementary but **tortuous** counting argument shows that the probability of each fixed $\boldsymbol{\xi}$ is bounded above by

$$\frac{C^{\alpha_d+\beta_d+\gamma_d}}{N^{\alpha_d+2\beta_d+2\gamma_d}}.$$

- Thus, $\beta_d > 0$ or $\gamma_d > 0 \Rightarrow$ probability $\to 0$.

Sketch proof IV

- More careful counting with $\beta_d = \gamma_d = 0$ shows that the FDDs of $G^{(n,N)}_{\tau_N(t)}$ are asymptotically bounded above by those of the Kingman coalescent with clock speed C.
- A completely analogous programme using lower bounds shows that the FDDs are bounded below by those of a Kingman coalescent with clock speed C_*.
- Exchangabiliy $\Rightarrow$ merger rates between all pairs are equal at all times $\Rightarrow$ Kingman coalescent.

Corollary 1

The genealogy of n particles sampled uniformly at random from an N-particle filter with multinomial resampling converges to a time-inhomogeneous Kingman coalescent under the time-scaling $\tau_N(t)$ whenever

$$a \leq g_t(x_{t-1}, x_t) \leq A$$

for some $0 < a \leq A < \infty$, uniformly in time and (almost) both arguments.

Jacob et al. (2015) employ essentially the same assumption.

Sketch proof

- Conditional on weights, the offspring counts have multinomial distributions with parameters $(N; w_t^{(1)}, \ldots, w_t^{(N)})$.
- Upper and lower bounds on potentials imply

$$\left(\frac{a}{AN}\right)^j \leq (w_t^{(i)})^j \leq \left(\frac{A}{aN}\right)^j .$$

- The required upper and lower bounds follow from these two observations, and standard but tedious moment-calculations for multinomial distributions.

Corollary 2

Let $T_N^{(n)}$ be the total tree height of $n \leq N$ particles, and $L_N^{(n)}$ be the total branch length. Under the preceding assumptions,

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

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

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

$$\mathrm{Var}(L_N^{(n)}) \leq \left(\frac{2\pi^2}{3}\right)\left(\frac{C}{C_*^2}\right)^2$$

and similar lower bounds are also obtained (with the roles of C and C_* are exchanged and $O(N^{-1})$ correction).

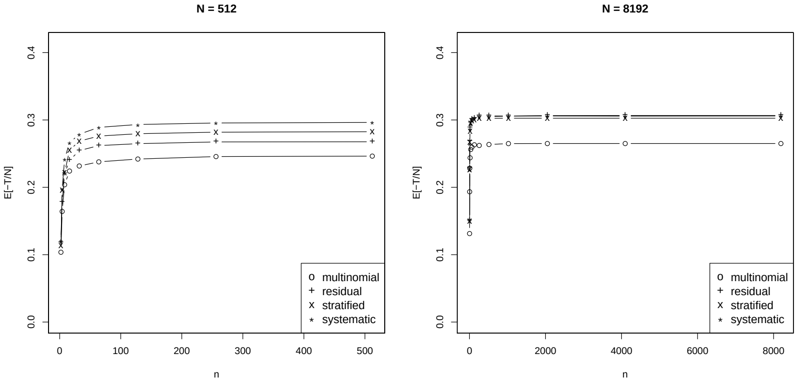
A numerical example

- Take the earlier HMM to be

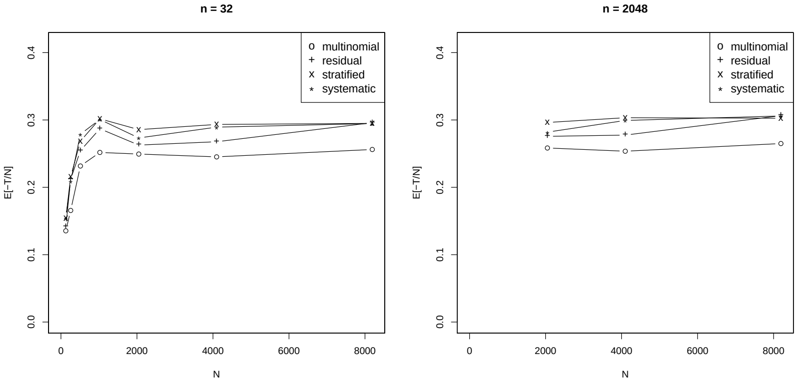
$$X_{t+1} = (1-\Delta)X_t + \sqrt{\Delta}\xi_t$$
$$X_0 \sim N(0,1),$$
$$Y_t | X_t \sim N(X_t, \sigma^2),$$

  where $\xi_t \sim N(0,1)$.
- $\Delta = \sigma = 0.1$.
- Simulations using a boostrap particle filter suggest that the Kingman scalings hold, even when $n = N$, for various resampling schemes.
- This example violates the assumed lower bounds on g_t.

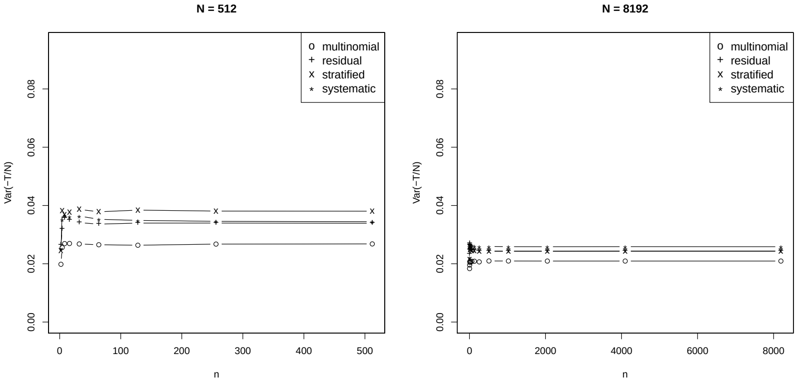
Mean tree height



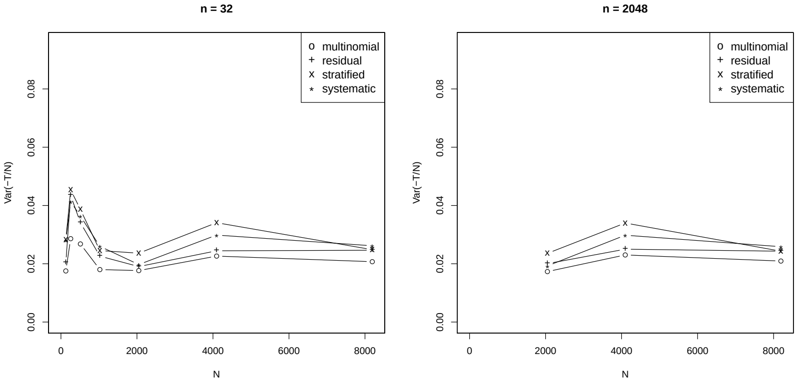
Mean tree height II



Tree height variance



Tree height variance II



Conclusions

- Genealogies of $n \ll N$ particles from N-particle SMC algorithms converge to the Kingman coalescent when time is measured in units of N, as $N \to \infty$.
- Strong technical assumptions (i.e. a compact state space) which do not seem necessary in practice.
- Predicted scalings observed in experiments for finite N, seem to hold even when $n \approx N$.
- Result holds for multinomial resampling, but other schemes agree with predictions empirically.
- This result also demonstrates that the domain of attraction of the Kingman coalescent includes certain non-Markovian genealogies.

Outlook

- Some areas in which genealogical results might be interesting:
  - Variance estimation from SMC output (Lee and Whiteley, 2015).
  - Smoothing and static parameter estimation.
  - Mixing in particle Gibbs/iterated cSMC.
- Room for improvement (selected topics. . .)
  - Relaxing the lower bound on moments.
  - Incorporating other resampling schemes.
  - Obtaining stronger modes of convergence.
  - (Formal analysis of $n \approx N$.)
  - Incorporating conditional SMC.

References

N. J. Gordon, S. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, April 1993.

P. E. Jacob, L. Murray, and S. Rubenthaler. Path storage in the particle filter. *Statistics and Computing*, 25(2):487–496, 2015.

J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248, 1982.

J. Koskela, P. Jenkins, A. M. Johansen, and D. Spanò. Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo. ArXiv mathematics e-print 1804.01811, ArXiv Mathematics e-prints, April 2018.

A. Lee and N. Whiteley. Variance estimation and allocation in the particle filter. Technical Report 1509.00394, Arxiv, 2015.

M. Möhle. Robustness results for the coalescent. *Journal of Applied Probability*, 35:438–447, 1998.

Acknowledgments

- Paul Jenkins was supported in part by EPSRC grant EP/L018497/1.
- Adam Johansen was partially supported by funding from the Lloyd's Register Foundation — Alan Turing Institute Programme on Data-Centric Engineering.
- Jere Koskela was supported by EPSRC as part of the MASDOC DTC at the University of Warwick (Grant No. EP/HO23364/1), and by Deutsche Forschungsgemeinschaft (DFG) grant BL 1105/3-2.