

Kingman limit for non-neutral populations with applications to sequential Monte Carlo

Suzie Brown

University of Warwick

with Paul Jenkins, Adam Johansen & Jere Koskela

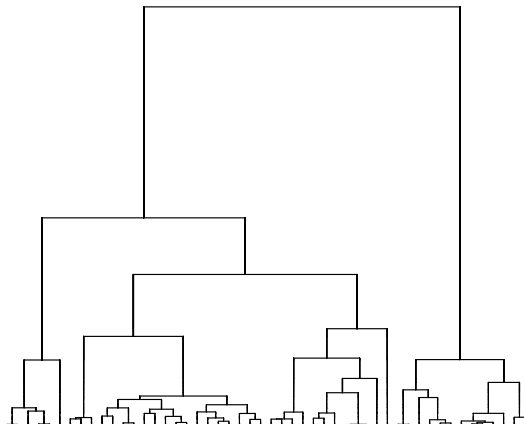
22 July 2021

Outline

1. Kingman's n -coalescent & population models
2. A history of convergence to the coalescent
3. Application to sequential Monte Carlo

Kingman's n -coalescent¹

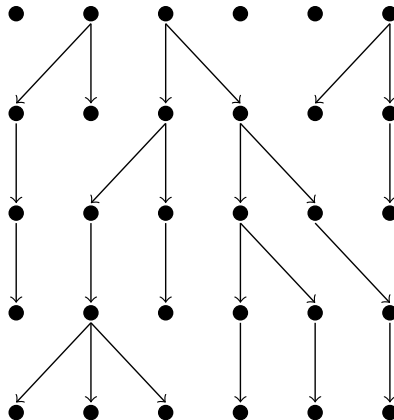
- ▶ Continuous-time Markov chain on the space of partitions of $\{1, \dots, n\}$
- ▶ Single pair mergers only
- ▶ Each pair merges independently at rate 1 (total merge rate $\binom{k}{2}$ while there are k distinct lineages)
- ▶ Exchangeable



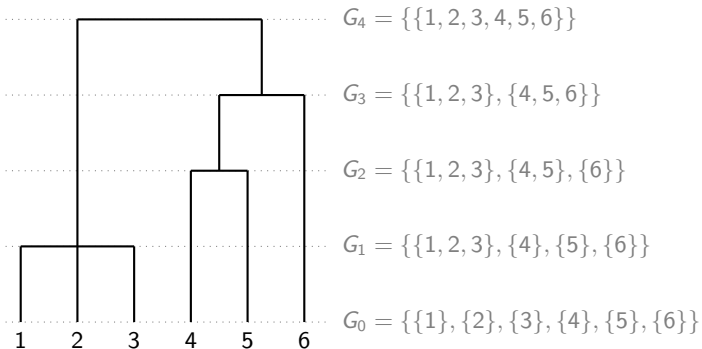
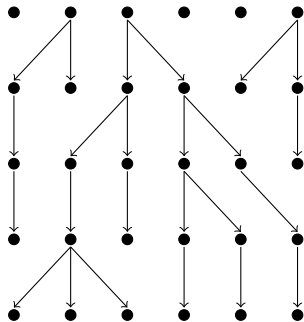
¹JFC Kingman, *Stochastic Processes & their Applications*, 1982.

Question

Under what conditions does a population have genealogies that are asymptotically distributed as n -coalescents?



Encoding genealogies



Common assumptions on population

- ▶ discrete generations
- ▶ population size $N(t)$ at generation t ; define $N := N(0)$
- ▶ offspring counts $(\nu_1(t), \dots, \nu_N(t))$ at generation t
- ▶ define the coalescence rate $c_N(t) := \frac{1}{(N)_2} \sum_{i=1}^N (\nu_i(t))_2$
- ▶ consider a random sample of n individuals from the terminal generation
- ▶ scale time to obtain a non-trivial limiting process
- ▶ take $N \rightarrow \infty$

Kingman's sufficient conditions²

Population model:

- ▶ fixed population size $N(t) \equiv N$
- ▶ offspring counts $(\nu_1(t), \dots, \nu_N(t))$ i.i.d. over t
- ▶ (ν_1, \dots, ν_N) exchangeable

Time scale: $N\sigma^{-2}$

Conditions:

- ▶ $\lim_{N \rightarrow \infty} \text{Var}[\nu_1] = \sigma^2 \in (0, \infty)$
- ▶ $\mathbb{E}[\nu_1^k] \leq M_k$ for each $k \in \mathbb{N}$

Then the FDDs of the rescaled sample genealogies converge to those of the n -coalescent as $N \rightarrow \infty$.

²JFC Kingman, *Stochastic Processes & their Applications*, 1982.

Möhle's necessary & sufficient conditions³

Population model:

- ▶ fixed population size $N(t) \equiv N$
- ▶ offspring counts $(\nu_1(t), \dots, \nu_N(t))$ i.i.d. over t
- ▶ (ν_1, \dots, ν_N) exchangeable

Time scale:

$$\frac{1}{\mathbb{E}[c_N]} = \frac{N-1}{\text{Var}[\nu_1]}$$

³M Möhle, *Advances in Applied Probability*, 2000.

Möhle's necessary & sufficient conditions³

Conditions:

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[(\nu_1)_3]}{N\mathbb{E}[(\nu_1)_2]} = 0$$

if and only if the FDDs of the rescaled sample genealogies converge to those of the n -coalescent as $N \rightarrow \infty$.

³M Möhle, *Advances in Applied Probability*, 2000.

Möhle's sufficient conditions for weak convergence⁴

Population model:

- ▶ population size $N(t)$ any deterministic function of t ; $N := N(0)$
- ▶ offspring counts $(\nu_1(t), \dots, \nu_N(t))$ independent over t
- ▶ *random assignment condition*: given $(\nu_1(t), \dots, \nu_N(t))$, assignment of offspring to parents is uniform over all valid assignments

Time scale: some function $\tau_N(t)$ such that for all t

- ▶ $\lim_{N \rightarrow \infty} \sum_{r=1}^{\tau_N(t)} \mathbb{E}[c_N(r)] = t$
- ▶ $\lim_{N \rightarrow \infty} \sum_{r=1}^{\tau_N(t)} \mathbb{E}[c_N(r)]^2 = 0$

⁴M Möhle, *Journal of Applied Probability*, 1998, 1999.

Möhle's sufficient conditions for weak convergence⁴

Conditions: for all $t > 0, k \geq 0$

$$\limsup \frac{1}{N(t-1)^3 c_N(t)} \sum_{i=1}^{N(t)} \mathbb{E}[(\nu_i(t))_2 (\nu_i(t))^k] = 0$$

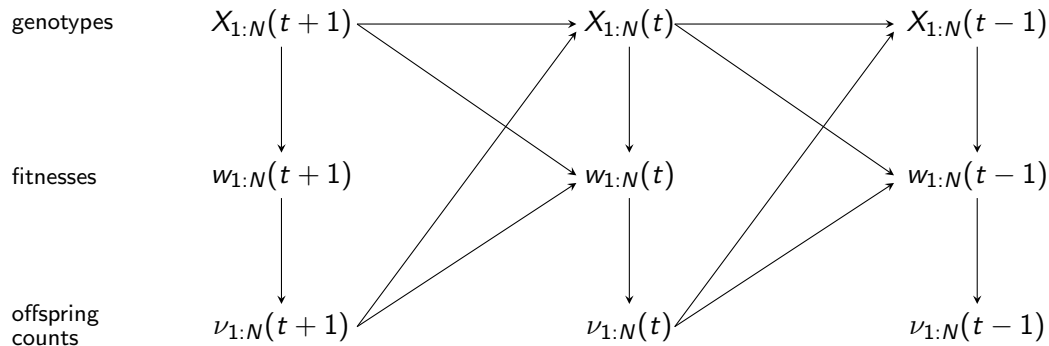
and

$$\limsup \frac{1}{N(t-1)^4 c_N(t)} \sum_{i=1}^{N(t)} \sum_{j=1}^{N(t)} \mathbb{E}[(\nu_i(t))_2 (\nu_j(t))^2] = 0.$$

Then the rescaled sample genealogies converge weakly to the n -coalescent as $N \rightarrow \infty$.

⁴M Möhle, *Journal of Applied Probability*, 1998, 1999.

A non-neutral population model



Sufficient conditions for weak convergence⁵

Population model:

- ▶ fixed population size $N(t) \equiv N$
- ▶ offspring counts $(\nu_1(t), \dots, \nu_N(t))$ conditionally dependent over t as in previous slide
- ▶ random assignment condition

Time scale:

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

and assume that $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t

⁵S Brown, PA Jenkins, AM Johansen, J Koskela, *Electronic Journal of Probability*, 2021.

Sufficient conditions for weak convergence⁵

Conditions: there exists a deterministic sequence $b_N \rightarrow 0$ such that for all N, t

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E}_t[(\nu_i(t))_3] \leq b_N \sum_{i=1}^N \mathbb{E}_t[(\nu_i(t))_2]$$

Then the rescaled sample genealogies converge weakly to the n -coalescent as $N \rightarrow \infty$.

⁵S Brown, PA Jenkins, AM Johansen, J Koskela, *Electronic Journal of Probability*, 2021.

Sequential Monte Carlo

Aim: simulate a particle system that approximates a given sequence of probability distributions.

Initialise N particles by sampling their genotypes $X_{1:N}$ from some distribution $\mu(\cdot)$

Then iterate these steps:

1. **(mutation)** update the genotypes via Markov kernel M_t
 2. **(fitness)** calculate fitness scores by applying function g_t to the genotypes
 3. **(selection)** resample particles according to their fitnesses
-
- ▶ μ , (M_t) and (g_t) are chosen such that the particle system approximates the desired distributions
 - ▶ The selection step induces a genealogy, which affects performance of the algorithm

Sequential Monte Carlo genealogies

- ▶ Most of the popular sequential Monte Carlo algorithms have asymptotically Kingman genealogies⁶
- ▶ The time scale differs depending on the particular algorithm, and is an indicator of performance
- ▶ Explicitly characterising the time scale would allow better tuning and comparisons between algorithms

⁶S Brown, PA Jenkins, AM Johansen, J Koskela, *Electronic Journal of Probability*, 2021.

In conclusion...

- ▶ Many neutral population models are known to have asymptotically Kingman genealogies
- ▶ We add to these a large class of non-neutral models, having a particular conditional independence structure rather than independent generations
- ▶ For these models, weak convergence to the n -coalescent is proved under simple sufficient conditions
- ▶ This result is applied to several popular sequential Monte Carlo algorithms
- ▶ The next step is to describe explicitly the time scale function in these cases

References

- 1,2 JFC Kingman (1982) *On the genealogy of large populations*, Stochastic Processes and their Applications.
- 3 M Möhle (2000) *Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models*, Advances in Applied Probability.
- 4a M Möhle (1998) *Robustness results for the coalescent*, Journal of Applied Probability.
- 4b M Möhle (1999) *Weak convergence to the coalescent in neutral population models*, Journal of Applied Probability.
- 5,6 S Brown, PA Jenkins, AM Johansen, J Koskela (2021) *Simple conditions for convergence of sequential Monte Carlo genealogies with applications*, Electronic Journal of Probability.