The hydrodynamic limit of multiple merger coalescent processes that come down from infinity

Helmut Pitters

joint work with Luke Miller
Supervisor: Alison Etheridge

Department of Statistics
University of Oxford

June 19, 2015

Outline

What is a coalescent process?

- Markov process
- encodes dynamics of particles grouped into so-called blocks
- as time passes, only mergers of (some or all) blocks may occur

Origins: Kingman's coalescent models the genealogy of individuals in population genetics (Kingman, Tavaré, Griffiths, Watterson).

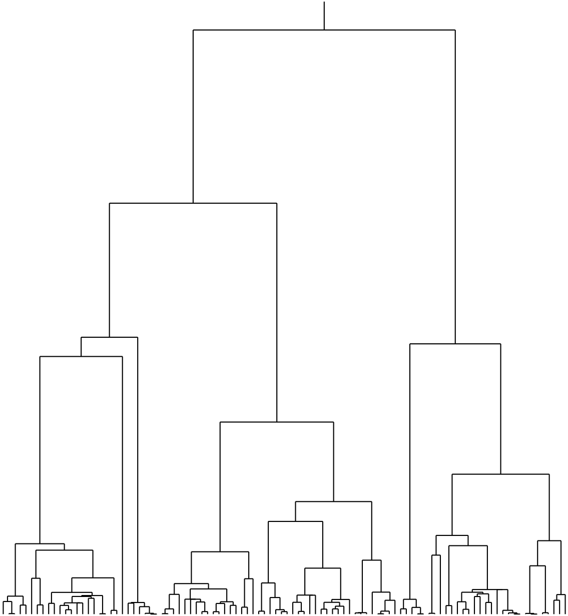


Figure : Simulation of Kingman coalescent tree with sample size $n = 100$.

Why study multiple merger coalescents?

- allow for multiple mergers instead of just binary mergers
- better null models than Kingman's coalescent for the genealogy of highly fecund populations

Figure : Examples of highly fecund populations. Left: Atlantic cod (gadus morhua); right: pacific oyster (crassostrea gigas)

- model the genealogy of populations subject to selection
- occur as rescaling limits in the theory of spin glasses/statistical physics
- rich mathematical structure

Definition of multiple merger coalescents I

Fix a sample size $n \geq 2$. The Λ n*-coalescent* $\{\Pi^n(t), t \geq 0\}$

- is a Markov process of jump-hold type,
- has state space the partitions of $[n] = \{1, \ldots, n\}$,
- and rates

$$\lambda_{m,k} = \text{rate at which any specific } k \text{ out of } m \text{ blocks merge}$$
$$= \int_0^1 x^k (1-x)^{m-k} \frac{\Lambda(dx)}{x^2},$$

where Λ is a finite measure on $[0, 1]$.

[see Donnelly, Kurtz 1999, Pitman 1999, Sagitov 1999]

Definition of multiple merger coalescents II

There exists a Markov process

$$\{\Pi(t), t \geq 0\}$$

on the partitions of $\mathbb{N}$ such that for any n

$$\text{restriction of } \{\Pi(t), t \geq 0\} \text{ to } [n] \quad =_d \quad \{\Pi^n(t), t \geq 0\}.$$

Π is referred to as the Λ *coalescent*.

Examples of Λ coalescents

$\Lambda(dx)$	name
$\delta_1(dx)$	star-shaped coalescent
$\delta_0(dx)$	Kingman coalescent
xdx	Bolthausen-Sznitman coalescent
$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}dx$	beta coalescents

Beta function: $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx \qquad (a,b>0).$

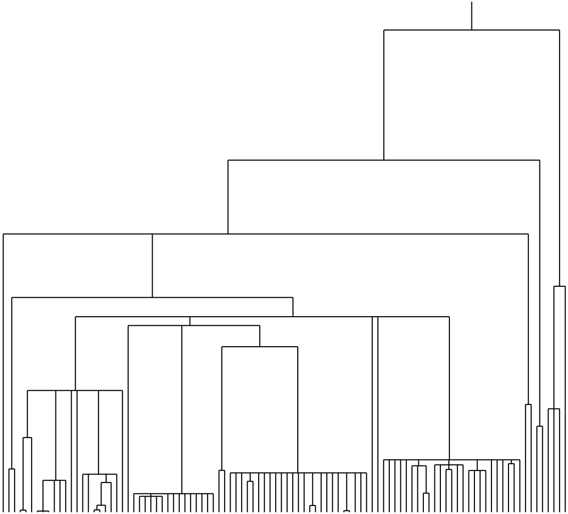


Figure : beta$(0.9,\ 1.1)$ coalescent tree with sample size $n = 100$.

Type of a partition

Consider a partition π of $[n]$. For fixed $i \in [n]$ let

$$\mathfrak{c}_i \pi = \#\{B \in \pi \colon \#B = i\}$$

denote the number of blocks of π of size i. We call

$$\mathfrak{c}\pi = (\mathfrak{c}_1 \pi, \ldots, \mathfrak{c}_n \pi)$$

the *type* of π.

Hydrodynamic limit

Goal: We would like to understand the evolution of the relative block size frequencies of $\Pi^n(t)$

$$\{n^{-1}(\mathfrak{c}_1\Pi^n(t\tau_n), \ldots, \mathfrak{c}_n\Pi^n(t\tau_n)), t \geq 0\},$$

as $n \to \infty$, with a suitable time-scaling τ_n.

Provided this limit exists, it yields information about the distribution of the marginal

$$\Pi(t)$$

due to exchangeability.
Fact: $\Pi^n(t)$ is an exchangeable random partition.

Exchangeable random partitions

A (random) partition Π of $\mathbb{N}$ is called *exchangeable* if its distribution is invariant under the action of any finite permutation, i.e. iff for all $n \in \mathbb{N}$

$$\sigma\Pi =_d \Pi \qquad \text{for any permutation } \sigma \text{ of } [n].$$

An exchangeable random partition of $[n]$ is defined in complete analogy.

Asymptotic frequencies

Given a partition $\pi = (B_1, B_2, \ldots)$, and a block B of π, let

$$|B| := \lim_n \frac{\#(B \cap [n])}{n}$$

denote the *asymptotic frequency* of B, if this limit exists.

Exchangeable random partitions: a simple example I

Fix $(c_1, \ldots, c_n) \in \mathbb{N}_0^n$ such that $\sum_i c_i = b$, and $\sum_i i c_i = n$. Define a random partition Π of $[n]$ with fixed block sizes $(c_1, \ldots, c_n)$ by

$$\mathbb{P}\{\Pi = \pi\} = \begin{cases} \left(\frac{n!}{\prod_{j=1}^n j!^{c_j} c_j!} \right)^{-1} & \text{if } \mathfrak{c}\pi = (c_1, \ldots, c_n), \\ 0 & \text{otherwise.} \end{cases}$$

Notice the *Faá di Bruno coefficients*

$$\#\{\pi \colon \mathfrak{c}\pi = (c_1, \ldots, c_n)\} = \frac{n!}{\prod_{j=1}^n j!^{c_j} c_j!}.$$

Exchangeable random partitions: a simple example II

Explicit construction of Π

1. Partition $1, 2, \ldots, n$ into some partition, π say, with $b = \sum_i c_i$ blocks, the first c_1 blocks being singletons, the next c_2 blocks being doubletons, etc., i.e.

$$\pi = 1|2|\cdots|c_1|c_1+1, c_1+2|\cdots|c_1+2c_2-1, c_1+2c_2|\cdots \ .$$

2. Let

$$\Pi = \Sigma\pi$$

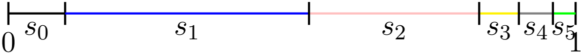
   be a relabelling of the elements of π by a permutation Σ of $[n]$ drawn uniformly at random.

Kingman's paintbox I

We construct an exchangeable random partition of $\mathbb{N}$ as follows. Fix a *tiling*

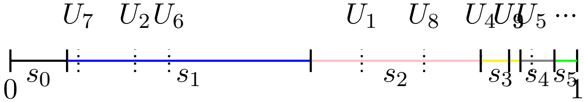
$$s \in \mathcal{S}_0 := \left\{ s = (s_0, s_1, \ldots) \colon s_i \geq 0, s_1 \geq s_2 \geq \cdots, \sum_i s_i = 1 \right\}$$

of the unit interval. Think of the s_is as boxes of different colours.



Kingman's paintbox II

Let (U_i) be a sequence of i.i.d. uniform $[0,1]$ random variables.



Define the partition Π of $\mathbb{N}$ via

i, j are in the same block in $\Pi \iff$
U_i, U_j fell into the same paintbox, not s_0

Then Π is an exchangeable random partition.
In fact, any exchangeable random partition of $\mathbb{N}$ can be constructed from a (possibly random) tiling of $[0,1]$.

Kingman's paintbox III

Now start with a random partition Π of $\mathbb{N}$. For any block B of Π the law of large numbers yields that its asymptotic frequency

$$|B| = \lim_n \frac{\#(B \cap [n])}{n} \in [0,1]$$

exists. If $|B| > 0$, we have recovered a fragment in the tiling $S = S(\Pi) \in \mathcal{S}_0$ corresponding to Π. Moreover,

$$s_0 = 1 - \sum_{B \in \Pi} |B|$$

is the proportion of singletons in Π.

Kingman's correspondence

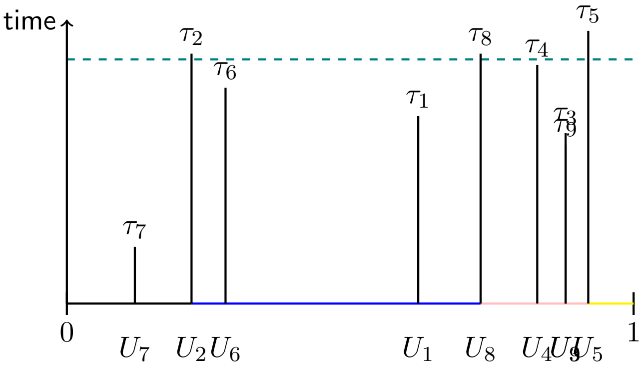
Theorem (Kingman's correspondence)

There is a bijection between the set of exchangeable random partitions Π *and the set of probability distributions on* $\mathcal{S}_0$.

$$\Pi \text{ exchangeable random partition of } \mathbb{N}$$
$$\longleftrightarrow s \text{ (random) tiling of } [0,1]$$

Aldous' construction of Kingman's coalescent I

Let (U_i) be i.i.d. uniform $[0, 1]$,
let (E_i) be independent exponentials, where E_i has parameter $\binom{i}{2}$,

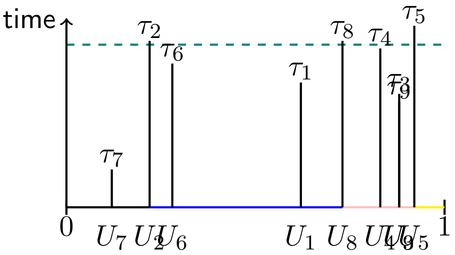


$$\tau_i := \sum_{k \geq i+1} E_k < \infty.$$

Attach a stick of length τ_i to U_i.

Aldous' construction of Kingman's coalescent II

Define $f\colon [0,1] \to [0,\infty)$ by $f(u) \coloneqq \tau_j$ if $u = U_j$ and $f(u) \coloneqq 0$ otherwise.



Then $\{S(t), t \geq 0\}$ defined by

$$S(t) \coloneqq \text{open connected components of } \{u \in (0,1)\colon f(u) \leq t\}$$

is equal in law to the asymptotic frequencies of Kingman's coalescent.

Hydrodynamic limit of Kingman's coalescent

Goal: Quantify behaviour of $\#\Pi(t)$ for small times t.
Idea: "Approximate" $\Pi(t)$ by $\Pi^n(t)$ for large n.

Somewhat related studies of asymptotic properties of beta coalescents:

- Berestycki, Berestycki, Schweinsberg 2007, 2008
- Berestycki, Berestycki, Limic 2010
- Limic, Talarczyk-Noble 2013, 2015

Hydrodynamic limit of Kingman's coalescent

Heuristics

- waiting time in state $N^n(t) := \#\Pi^n(t) \approx \mathrm{Exp}(\binom{N^n(t)}{2})$
- for small t and large n
  - (rate at which $N^n(t)$ decreases) $= \frac{-1}{\mathbb{E}\left[\mathrm{Exp}\left(\binom{N^n(t)}{2}\right)\right]} \approx -\frac{1}{2}N^n(t)^2$,
  - hence $N^n(t)/n$ should be approximated by the ODE

$$c'(t) = -\frac{1}{2}c(t)^2, \qquad c(0) = 1$$

with solution

$$c(t) = \frac{2}{2+t}.$$

Hydrodynamic limit of Kingman's coalescent

Theorem

As $n \to \infty$

$$\{n^{-1}\#\Pi^n(t/n), t \geq 0\} \to \left\{\frac{2}{2+t}, t \geq 0\right\},$$

in the Skorohod topology.

Cf. Aldous 1999 and Wattis 2008.

Hydrodynamic limit of Kingman's coalescent

Theorem

For fixed $d \in \mathbb{N}$ *as* $n \to \infty$

$$\{n^{-1}(\mathfrak{c}_1\Pi^n(t/n), \ldots, \mathfrak{c}_d\Pi^n(t/n)), t \geq 0\} \\ \to \{(c_1(t), \ldots, c_d(t)), t \geq 0\}$$

in the Skorohod topology, where

$$c_j(t) = c(t)^2(1 - c(t))^{j-1}, \quad c(t) = \frac{2}{2+t} \qquad (t \geq 0, j \in \mathbb{N}).$$

Cf. Aldous 1999 and Wattis 2008.

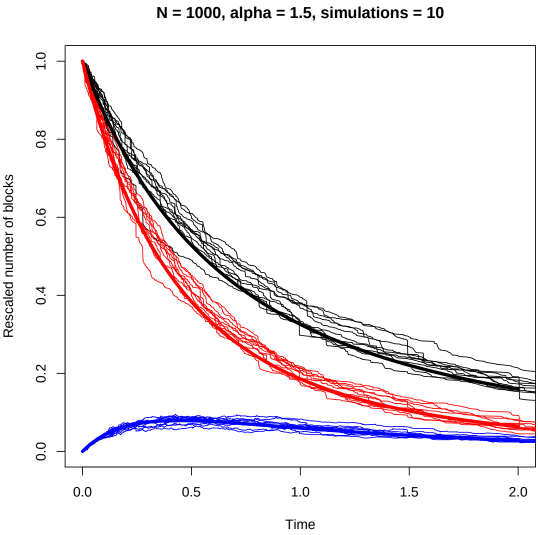


Figure : Simulated relative sizes of blocks in a beta$(0.5, 1.5)$ coalescent and analytic solution (thick line). black: total number of blocks, red: singletons, blue: doubletons.

Hydrodynamic limit of beta coalescents

Theorem (Miller, P. 2014)

Consider $beta(a,b)$ *coalescents with* $a < 1$*. Then as* $n \to \infty$

$$\{n^{-1} \# \Pi^n(n^{a-1}t), t \geq 0\} \to \{c(t), t \geq 0\},$$

in the Skorohod topology, where

$$c(t) = \left(\frac{(2-a)\Gamma(b)}{(2-a)\Gamma(b) + \Gamma(a+b)t} \right)^{\frac{1}{1-a}}.$$

Bell polynomials I

For each finite set F_n with n elements we are given a construction V that associates with F_n a set of V-*structures*, $V(F_n)$, so

$$F_n \mapsto V(F_n),$$

such that $\#V(F_n) = v_n$, for some fixed sequence $v_\bullet = (v_n)$. V is called a *species of combinatorial structures*.

Table : Examples of combinatorial species

$V(F_n)$	$\#V(F_n)$
F_n	1
permutations of F_n	$n!$
partitions of F_n	B_n (nth Bell number)

[Further information: Pitman 2006, Combinatorial stochastic processes

For a rich theory of combinatorial species see Bergeron, Labelle, Leroux 2013]

Bell polynomials II

Consider two combinatorial species, V, W, such that for any set F_n with $\#F_n = n$

$$\#V(F_n) = v_n, \qquad \#W(F_n) = w_n.$$

Let

$$(V \circ W)(F_n) = \begin{cases} \text{set of all ways to partition } F_n \text{ into} \\ \text{blocks } \{A_1, \ldots, A_k\} \text{ for some } k, \\ \text{assign each partition a } V\text{-structure} \\ \text{\& assign each block } A_i \text{ a } W\text{-structure.} \end{cases}$$

$(V \circ W)(F_n)$ is called a *composite structure* on F_n.

$$\#(V \circ W)(F_n) = \sum_{\pi \in \mathcal{P}_{[n]}} v_{\#\pi} \prod_{B \in \pi} w_{\#B} = \sum_{k=1}^{n} v_k \sum_{\pi \in \mathcal{P}_{[n],k}} \prod_{B \in \pi} w_{\#B}$$

Bell polynomials III

Recall

$$\#(V \circ W)(F_n) = \sum_{\pi \in \mathcal{P}_{[n]}} v_{\#\pi} \prod_{B \in \pi} w_{\#B} = \sum_{k=1}^{n} v_k \sum_{\pi \in \mathcal{P}_{[n],k}} \prod_{B \in \pi} w_{\#B}.$$

Denote by

$$B_{n,k}(w_\bullet) := \sum_{\pi \in \mathcal{P}_{[n],k}} \prod_{B \in \pi} w_{\#B}$$

the (n, k)*th partial Bell polynomial*, and by

$$B_n(v_\bullet, w_\bullet) := \sum_{k=1}^{n} v_k B_{n,k}(w_\bullet)$$

the n*th complete Bell polynomial*. Then

$$\#(V \circ W)(F_n) = B_n(v_\bullet, w_\bullet).$$

Hydrodynamic limit of beta coalescents

Theorem (Miller, P. 2014)

For fixed $d \in \mathbb{N}$ *we have*

$$\{n^{-1}(\mathfrak{c}_1\Pi^n(n^{a-1}t), \dots, \mathfrak{c}_d\Pi^n(n^{a-1}t)), t \geq 0\} \\ \to \{(c_1(t), \dots, c_d(t)), t \geq 0\}$$

as $n \to \infty$ *in the Skorohod topology, where for each* $i \in \mathbb{N}$

$$c_i(t) = \frac{c(t)^{2-a}}{i!} B_i \left(\left(\frac{1}{1-a} \right)^{\overline{\bullet}} (-c(t)^{1-a})^{\bullet - 1}, (1-a)^{\overline{\bullet}} \right),$$

with $x^{\overline{k}} := x(x+1)\cdots(x+k-1)$ *the ascending factorial power.*

What does this tell us about the asymptotic frequencies of beta coalescents?

Informally,

$$\text{``}S_{\Pi^n}(tn^{a-1}) \to S_\Pi(tn^{a-1})\text{''} \qquad \text{as } n \to \infty.$$

For very large n, at time tn^{a-1} a block of size i has "asymptotic frequency" of order i/n, and there are roughly $nc_i(t)$ of them, hence together occupy a fraction $ic_i(t)$ of the corresponding tiling.

References I

- Aldous 1999, Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists.
- Berestycki 2009, Recent progress in coalescent theory.
- Berestycki, Berestycki, Schweinsberg 2007 Beta-coalescents and continuous stable random trees.
- Berestycki, Berestycki, Schweinsberg 2008 Small-time behaviour of beta coalescents.
- Berestycki, Berestycki, Limic 2010 The Lambda-coalescent speed of coming down from infinity.
- Bergeron, Labelle, Leroux 2013, Introduction to the theory of species of structures.
- Donnelly, Kurtz 1999, Particle representations for measure-valued population models.
- Kingman 1982, The coalescent.

References II

- Limic, Talarczyk-Noble 2013 Second-order asymptotics for the block counting process in a class of regularly varying Lambda-coalescents.
- Limic, Talarczyk-Noble 2015 Diffusion limits at small times for coalescents with a Kingman component.
- Pitman 1999, Coalescents with multiple collisions.
- Pitman 2006, Combinatorial stochastic processes.
- Sagitov 1999, The general coalescent with asynchronous mergers of ancestral lines.
- Wattis 2008, An introduction to mathematical models of coagulation-fragmentation processes: a discrete deterministic mean-field approach.