

Λ -coalescents arising in a population with dormancy

by Jason Schweinsberg
University of California San Diego

With Fernando Cordero, Adrián González Casanova, and Maite Wilke-Berenguer

Outline of Talk

1. The Wright-Fisher model and Kingman's coalescent
2. Coalescents with multiple mergers
3. A model involving dormancy
4. Results for the genealogy of the population

The Wright-Fisher Model

One of the earliest models in population genetics, goes back to Fisher (1921) and Wright (1930).

- ▶ The population has fixed size N .
- ▶ Generations do not overlap.
- ▶ Each individual has one parent, chosen at random from the individuals in the previous generation.

Sample two individuals in generation 0. The probability that they have the same parent is $1/N$.

Let T be the number of generations we have to go back before they have the same ancestor. Then

$$P(T > Nx) = \left(1 - \frac{1}{N}\right)^{\lfloor Nx \rfloor} \approx e^{-x}.$$

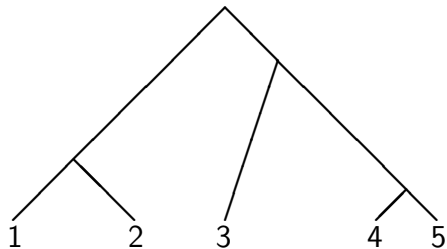
T/N has approximately an exponential distribution with rate 1.

Kingman's Coalescent (Kingman, 1982)

Continuous-time Markov chain $(\Pi(t), t \geq 0)$ taking values in the set \mathcal{P}_n of partitions of $\{1, \dots, n\}$.

Here $\Pi(0)$ consists of n singletons. Any transition that involves merging two blocks into one happens at rate 1. No other transitions are possible.

When there are k blocks, the distribution of the time until the next merger is exponential with rate $k(k-1)/2$. Then two randomly chosen blocks merge.



One time unit in Kingman's coalescent corresponds to N generations in the Wright-Fisher model.

A Limit Theorem

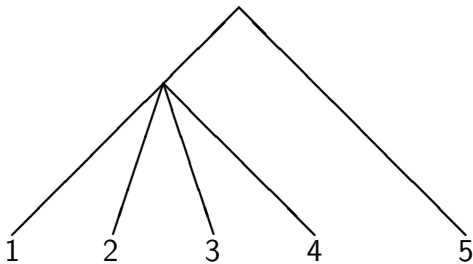
Theorem (Kingman, 1982): Suppose a population evolves according to the Wright-Fisher model with population size N . Sample n individuals at random from the population in generation zero. Let $\Psi_N(k)$ be the partition of $\{1, \dots, n\}$ such that i and j are in the same block if and only if the i th and j th sampled individuals have the same ancestor in generation $-k$. Let $(\Pi(t), t \geq 0)$ be Kingman's coalescent. Then, as $N \rightarrow \infty$,

$$(\Psi_N(\lfloor Nt \rfloor), t \geq 0) \Rightarrow (\Pi(t), t \geq 0).$$

Here \Rightarrow denotes weak convergence of stochastic processes with respect to the Skorohod topology.

Coalescents with multiple mergers (Λ -coalescents)

Introduced by Pitman (1999), Sagitov (1999), and Donnelly and Kurtz (1999).
More than two ancestral lines can merge at a time.



Applications of coalescents with multiple mergers:

- ▶ Large family sizes (many lineages trace back to an individual with large number of offspring).
- ▶ Natural selection (many lineages trace back to an individual who got a beneficial mutation).

Definition and construction of the Λ -coalescent

Let Λ be a finite measure on $[0, 1]$. The Λ -coalescent is a continuous-time \mathcal{P}_n -valued Markov chain such that when there are b blocks, each k -tuple ($2 \leq k \leq b$) of blocks merges at rate

$$\lambda_{b,k} = \int_0^1 p^{k-2}(1-p)^{b-k} \Lambda(dp).$$

Put $\Lambda = a\delta_0 + \Lambda_0$, with $\Lambda_0(\{0\}) = 0$.

- ▶ Each pair of lineages merges at rate a .
- ▶ Construct a Poisson point process on $[0, \infty) \times (0, 1]$ with intensity

$$dt \times p^{-2}\Lambda_0(dp).$$

If (t, p) is a point of this Poisson process, then a p -merger occurs at time t . Each lineage independently participates with probability p .

Cannings models (Cannings, 1974)

- ▶ The population has fixed size N .
- ▶ Generations do not overlap.
- ▶ Let $\nu_{1,N}, \dots, \nu_{N,N}$ denote the numbers of offspring of the N individuals. The distribution of $(\nu_{1,N}, \dots, \nu_{N,N})$ is exchangeable.
- ▶ Family sizes in different generations are independent and identically distributed.

Ancestral process: Sample n individuals from generation 0. Let $\Psi_N(k)$ be the partition of $\{1, \dots, n\}$ such that i and j are in the same block if and only if the i th and j th sampled individuals have the same ancestor in generation $-k$.

The probability that two individuals have the same parent is

$$c_N = \frac{E[(\nu_{1,N})_2]}{N-1}, \quad (m)_k = m(m-1)\dots(m-k+1).$$

Convergence of the genealogy in Cannings models

Theorem (Möhle, 2000): Consider a Cannings model in which

$$\lim_{N \rightarrow \infty} \frac{E[(\nu_{1,N})_3]}{N^2 c_N} = 0.$$

Then the ancestral processes $\Psi_N(\lfloor t/c_N \rfloor, t \geq 0)$ converge as $N \rightarrow \infty$ to Kingman's coalescent.

Theorem (Sagitov, 1999; Möhle and Sagitov, 2001): Suppose

1. $\lim_{N \rightarrow \infty} c_N = 0.$

2. $\lim_{N \rightarrow \infty} \frac{E[(\nu_{1,N})_2(\nu_{2,N})_2]}{N^2 c_N} = 0.$

3. For all $x \in (0, 1)$ such that $\Lambda(\{x\}) = 0$, we have

$$\lim_{N \rightarrow \infty} \frac{N}{c_N} P(\nu_{1,N} > Nx) = \int_x^1 y^{-2} \Lambda(dy).$$

Then the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge to the Λ -coalescent.

Heavy-tailed offspring distributions

Consider the following population model:

- ▶ The population has fixed size N .
- ▶ Generations do not overlap.
- ▶ Numbers of offspring ξ_1, \dots, ξ_N of the N individuals are i.i.d. positive integer-valued random variables.
- ▶ We have $P(\xi_i \geq k) \sim Ck^{-\alpha}$, where $\alpha > 0$.
- ▶ Obtain the next generation by sampling N offspring without replacement.

Sample n individuals in generation 0.

Let $\Psi_N(k)$ be the partition of $\{1, \dots, n\}$ such that i and j are in the same block if and only if the i th and j th sampled individuals have the same ancestor in generation $-k$.

Genealogy of the population

Theorem (Schweinsberg, 2003): As $N \rightarrow \infty$,

- ▶ If $\alpha \geq 2$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge to Kingman's coalescent. When $\alpha > 2$, we have $c_N \sim \sigma^2/N$, where σ^2 is the variance of the number of surviving offspring.
- ▶ If $1 < \alpha < 2$, the processes $(\Psi_N(\lfloor AN^{\alpha-1}t \rfloor), t \geq 0)$ converge, for some constant A , to the Λ -coalescent, where Λ is the Beta($2 - \alpha, \alpha$) distribution (beta coalescent).
- ▶ If $\alpha = 1$, the processes $(\Psi_N(\lfloor (\log N)t \rfloor), t \geq 0)$ converge to the Λ -coalescent, where Λ is uniform distribution on $(0, 1)$, called the Bolthausen-Sznitman (1998) coalescent.
- ▶ If $0 < \alpha < 1$, the processes $(\Psi_N(k))_{k=0}^{\infty}$ converge as $N \rightarrow \infty$ to a discrete-time coalescent process.

Idea of the proof ($1 < \alpha < 2$)

Let $\mu = E[\xi_i]$ be the mean of the offspring distribution.

We get a p -merger with $p \geq x$ if

$$\frac{\xi}{\xi + N\mu} \geq x \quad \iff \quad \xi \geq \frac{x}{1-x} \cdot N\mu$$

The probability of such a family in a given generation is

$$NP\left(\xi \geq \frac{x}{1-x} \cdot N\mu\right) \sim NC\left(\frac{x}{1-x} \cdot N\mu\right)^{-\alpha}.$$

The rate of such mergers in the Beta($2 - \alpha, \alpha$)-coalescent is

$$\frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \int_x^1 p^{-1-\alpha}(1-p)^{\alpha-1} dp = \frac{1}{\alpha\Gamma(\alpha)\Gamma(2-\alpha)} \left(\frac{x}{1-x}\right)^{-\alpha}.$$

Idea of Wright and Vestigian (2019)

In many populations, individuals periodically enter a dormant state, in response to unfavorable environmental conditions.

Wright and Vetsigian postulated that the randomness in the times when individuals emerge from a dormant state could lead to a heavy-tailed offspring distribution.

They found in bacterial experiments that “the heavy-tailed nature of the distribution of descendants can, in our case, be largely explained by phenotypic variability in lag time before exponential growth.”

Our goal: to formulate a population model that incorporates this behavior, and to investigate whether dormancy can give rise to Λ -coalescent genealogies.

A model involving dormancy

Each year begins at time 0 with N dormant individuals and lasts for time T_N .

1. **Activation phase (Spring)**: Each individual wakes up at some time in $(0, t_N)$, then reproduces at rate λ_N , as in a Yule process. The times at which individuals wake up are i.i.d.
2. **Active phase (Summer)**: Between times t_N and T_N , all individuals are awake and reproduce at rate λ_N .
3. **Sampling phase (Winter)**: At time T_N , all individuals enter a dormant state. We choose N individuals at random to survive until the start of the next year. All other individuals die.

Sample n individuals at the beginning of year 0. Let $\Psi_N(k)$ be the partition of $\{1, \dots, n\}$ such that i and j are in the same block if and only if the i th and j th individuals in the sample have the same ancestor at the beginning of year $-k$.

A two-point distribution

Let $\tau_{i,N}$ be the time when the i th individual emerges from dormancy.

Suppose there is no summer, so $T_N = t_N$. Suppose

$$P(\tau_{i,N} = 0) = \omega_N, \quad P(\tau_{i,N} = T_N) = 1 - \omega_N$$

with

$$\lim_{N \rightarrow \infty} N\omega_N = 0.$$

If an individual wakes up at time 0 instead of at time T_N , then it will have approximately $We^{\lambda_N T_N}$ descendants at time T_N , where $W \sim \text{Exponential}(1)$.

If $\lambda_N T_N = \log(\kappa N)$, then the fraction of the population at the end of the year descended from the individual who wakes up early is approximately

$$Y_\kappa = \frac{\kappa W}{\kappa W + 1}.$$

Genealogy of the population

Theorem (Cordero, González Casanova, Wilke-Berenguer, Schweinsberg, 2022):

Assume that $\lambda_N T_N = \beta \log(\kappa N)$. Then:

1. If $\beta > 1$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge as $N \rightarrow \infty$ to the star-shaped coalescent, which is the Λ -coalescent with $\Lambda = \delta_1$.
2. If $\beta = 1$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge as $N \rightarrow \infty$ to the Λ_κ -coalescent, with Λ_κ being the probability measure characterized by

$$y^{-2} \Lambda_\kappa(dy) := \frac{1}{E[Y_\kappa^2]} P(Y_\kappa \in dy).$$

3. If $\beta < 1$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge as $N \rightarrow \infty$ to Kingman's coalescent.

Also,

$$c_N \sim \begin{cases} N\omega_N & \text{if } \beta > 1 \\ N\omega_N E[Y_\kappa^2] & \text{if } \beta = 1 \\ 2\kappa^{-2\beta} \omega_N N^{2\beta-1} & \text{if } \beta < 1 \end{cases}$$

Exponentially increasing rate of exit from dormancy

Suppose there is no summer, so $t_N = T_N$. Suppose

$$\lim_{N \rightarrow \infty} \frac{\log N}{T_N} = 0.$$

Suppose $\lambda_N = \lambda > 0$ for all N .

Suppose $T_N - \tau_{i,N} = \zeta_i \wedge T_N$, where $(\zeta_i)_{i=1}^N$ are i.i.d. We assume there exist constants $\gamma > 0$ and $c > 0$ such that

$$P(\zeta_i > y) \sim ce^{-\gamma y}, \quad \text{as } y \rightarrow \infty.$$

Let $X_{i,N}$ be the number of descendants of the i th individual at the end of spring. The conditional distribution of $X_{i,N}$ given $\zeta_i = u < T_N$ is $\text{Geometric}(e^{-\lambda u})$.

When $\zeta_i \sim \text{Exponential}(\gamma)$, writing $\alpha = \gamma/\lambda$,

$$P(X_{i,N} > k) = \int_0^\infty (1 - e^{-\lambda u})^k \gamma e^{-\gamma u} du = \frac{\alpha \Gamma(k+1) \Gamma(\alpha)}{\Gamma(k+1+\alpha)} \sim \Gamma(\alpha+1) k^{-\alpha}.$$

Genealogy of the population

Theorem (Cordero, González Casanova, Wilke-Berenguer, Schweinsberg, 2022):
Under the conditions described on the previous slide with $\alpha = \gamma/\lambda$:

1. If $\alpha \geq 2$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge as $N \rightarrow \infty$ to Kingman's coalescent.
2. If $1 \leq \alpha < 2$, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge as $N \rightarrow \infty$ to the Λ -coalescent in which Λ is the Beta($2 - \alpha, \alpha$) distribution.
3. If $0 < \alpha < 1$, the processes $(\Psi_N(m))_{m=0}^{\infty}$ converge as $N \rightarrow \infty$ to a discrete-time coalescent process.

Note: We have $c_N \sim AN^{1-\alpha}$ when $1 < \alpha < 2$ and $c_N \sim 1/(\log N)$ when $\alpha = 1$.

Classifying the possible limits

Theorem (Cordero, González Casanova, Wilke-Berenguer, Schweinsberg, 2022):
It is possible for the Λ -coalescent to arise as the limit of the rescaled ancestral processes $(\Psi_N(\lfloor \rho_N t \rfloor), t \geq 0)$ in this population model if and only if

$$\Lambda = b_0 \delta_0 + b_1 \delta_1 + \Lambda',$$

where $b_0, b_1 \geq 0$ and $\Lambda'(dy) = h(y) dy$ with

$$h(y) = \int_0^\infty \frac{1}{\kappa} \left(\frac{y}{1-y} \right)^2 e^{-\frac{y}{\kappa(1-y)}} \eta(d\kappa)$$

for all $y \in (0, 1)$ and $\int_0^\infty (1 \wedge \kappa^2) \eta(d\kappa) < \infty$.

Note: Recall $Y_\kappa = \kappa W / (\kappa W + 1)$ and $y^{-2} \Lambda_\kappa(dy) = E[Y_\kappa^2]^{-1} P(Y_\kappa \in dy)$. Now

$$\Lambda'(dy) = \int_0^\infty y^2 P(Y_\kappa \in dy) \eta(d\kappa) = \int_0^\infty \Lambda_\kappa(dy) E[Y_\kappa^2] \eta(d\kappa).$$

The effect of summer

Theorem (Cordero, González Casanova, Wilke-Berenguer, Schweinsberg, 2022): Suppose, for a model with no summer, the processes $(\Psi_N(\lfloor t/c_N \rfloor), t \geq 0)$ converge to a Λ -coalescent. Let $X_{i,N}$ denote the number of descendants of the i th individual at the beginning of a given year who are alive at time t_N , and suppose

$$\lim_{N \rightarrow \infty} \frac{1}{c_N} E \left[\frac{1}{X_{1,N} + \cdots + X_{N,N}} \right] = 0.$$

Then we get the same limiting genealogy if a summer of any length is added.

Note: Here c_N is the probability that two lineages coalesce during the spring. If there are M individuals alive at the end of the spring, the probability that two lineages coalesce during the summer is $O(1/M)$.

Summary and conclusions

- ▶ In many populations, individuals periodically enter a dormant state.
- ▶ Individuals that emerge unusually early from dormancy may be able to produce more descendants than other individuals.
- ▶ In extreme cases, this could lead to Λ -coalescent genealogies, in which many ancestral lines get traced back to one individual that emerged unusually early from dormancy.
- ▶ Wright and Vestigian (2019): “it is unlikely that the variance diverges with population size for the particular species and conditions we examined”. Therefore, at the moment, we do not have evidence that Λ -coalescent genealogies occur in actual biological populations undergoing dormancy.