Kingman's n-coalescent

From **?** via **?**, Chapter 3 and **?**.

In the notation of **?**, T_i; $i = 2, \ldots, n$ is the ith coalescence time, that is, the length of time for which there are exactly i branches in the sample genealogy. The n-coalescent is the process in which these times are i.i.d. Exponentials:

$$T_i \sim \mathrm{Exp}\left(\binom{i}{2}\right)$$

having

$$\mathbb{E}(T_i) = \frac{2}{i(i-1)}, \qquad \mathrm{Var}(T_i) = \left(\frac{2}{i(i-1)}\right)^2$$

By noting that the number of possible pairs from i lineages is $\frac{i(i-1)}{2}$, the process can equivalently be formulated as a Poisson process where pairs of lineages coalesce independently at rate 1, with the pair to coalesce being chosen uniformly at random (**?**, Section 3.2).

? writes the same process in terms of the infinitesimal generator Q of a Markov process $\{R_r\}$ on the set of equivalence relations on n elements, having entries

$$q_{\xi\eta} = \begin{cases} -\frac{1}{2}b(b-1) & \text{if } \xi = \eta \\ 1 & \text{if } \xi \prec \eta \\ 0 & \text{otherwise} \end{cases}$$

where b is the number of equivalence classes of ξ, and $\xi \prec \eta$ means that η is a state with exactly one more pair of lineages coalesced compared to ξ. [check this is equivalent to Wakeley's formulation]

Notes from ?

Framework

- Current generation is generation 0; generations enumerated backward in time.

- Generations are non-overlapping.

Notation

M_r population size at generation r

$N := M_0$ initial population size

$v_i^{(r)}$ (random) number of offspring of individual i of generation r

$n \leq N$ size of the sample of individuals in generation 0 to be considered

T_n the number of generations since the most recent common ancestor (MRCA) of the sample

$\mathcal{R}_r$ the equivalence relation that contains the pair (i, j) iff individuals i and j in the sample have a common ancestor in generation r

$\{\mathcal{R}_r\}_{r \in \mathbb{N}}$ will be referred to as the *ancestral process*

Δ the minimal relation $\{(i, i); i = 1, \ldots, n\}$

Θ the maximal relation $\{(i, j); i, j = 1, \ldots, n\}$

$p_{\xi\eta}(r)$ the transition probability $\mathbb{P}(\mathcal{R}_r = \eta \mid \mathcal{R}_{r-1} = \xi)$ of the ancestral process

c_r the probability that a random pair of distinct individuals from generation r have a common ancestor in generation $r - 1$, called the *coalescence probability*

$\sigma^2(r)$ the expected *mean crowding* of the offspring variables of generation r

X_r the number of descendants in generation r in the forward genealogical process

$\pi_{ij}(r)$ the transition probability $\mathbb{P}(X_{r-1} = j \mid X_r = i)$ of the forward genealogical process

$(x)_y$ denotes the descending factorial $x(x-1) \cdots (x - y + 1)$

Calculating the coalescence rate

A combinatorial argument allows us to derive an expression for the transition probability $p_{\xi\eta}(r)$ of the ancestral process:

$$p_{\xi\eta}(r) = \frac{1}{(M_{r-1})_b} \sum_{\substack{i_1,\dots,i_a=1 \\ \text{distinct}}}^{M_r} \mathbb{E}\left[(v_{i_1}^{(r)})_{b_1} \cdots (v_{i_a}^{(r)})_{b_a}\right] \tag{2}$$

The sum is over all the possible ordered choices of the a parents from generation r. Inside the sum is the expected number of ways to have at least the required number of offspring from each parent, given this choice of parents. Thus the whole sum represents the probability of finding a set of parents that produce the required number of offspring. Then dividing by the number of ordered ways to choose b offspring from generation $r-1$ ensures that the parents produce the *correct* ordered offspring. Overall then we have the probability that exactly the right subsets of offspring from generation $r-1$ coalesce in generation r, counting all the different parents to which they could coalesce.

In the special case where $v_1^{(r)}, \dots, v_{M_r}^{(r)}$ are exchangeable, it is not necessary to have separate terms for different combinations of individuals $i_1, \dots, i_a$, so (2) simplifies to

$$p_{\xi\eta}(r) = \frac{(M_r)_a}{(M_{r-1})_b} \mathbb{E}\left[(v_1^{(r)})_{b_1} \cdots (v_a^{(r)})_{b_a}\right] \tag{3}$$

Now let us consider the coalescence time T_n. In general we have $\mathbb{P}(T_n > l) = \mathbb{P}(\mathcal{R}_l \neq \Theta)$. In the case $n = 2$ (i.e. time to coalescence for a pair of individuals), we can say something further:

$$\mathbb{P}(T_2 > l) = \mathbb{P}(\mathcal{R}_l = \Delta)$$

$$= \prod_{r=1}^{l} p_{\Delta\Delta}(r)$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \sum_{j \neq i} \mathbb{E}\left[v_i^{(r)} v_j^{(r)}\right]$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}\left[v_i^{(r)} \sum_{j \neq i} v_j^{(r)}\right]$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}\left[v_i^{(r)}(M_{r-1} - v_i^{(r)})\right] \tag{4}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ \mathbb{E}\left[M_{r-1} \sum_{i=1}^{M_r} v_i^{(r)}\right] - \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})^2\right] \right\}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ \mathbb{E}\left[M_{r-1}^2\right] - \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})^2\right] \right\}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ M_{r-1}(M_{r-1} - 1) + \sum_{i=1}^{M_r} \mathbb{E}\left[v_i^{(r)}\right] - \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})^2\right] \right\}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ M_{r-1}(M_{r-1} - 1) + \sum_{i=1}^{M_r} \mathbb{E}\left[v_i^{(r)} - (v_i^{(r)})^2\right] \right\}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ M_{r-1}(M_{r-1} - 1) - \sum_{i=1}^{M_r} \mathbb{E}\left[v_i^{(r)}(v_i^{(r)} - 1)\right] \right\}$$

$$= \prod_{r=1}^{l} \frac{1}{(M_{r-1})_2} \left\{ (M_{r-1})_2 - \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})_2\right] \right\}$$

$$= \prod_{r=1}^{l} \left\{ 1 - \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})_2\right] \right\} \tag{5}$$

Now we have $\mathbb{P}(T_2 > l)$ in the form $\prod(1 - c_r)$, where c_r represents the *coalescence probability*, i.e. the probability that a randomly chosen pair of (distinct) individuals from generation $r - 1$ have a common ancestor in generation r. We can derive an equivalent expression for the coalescence probability, $c_r = \sigma^2(r)/(M_{r-1} - 1)$, where $\sigma^2(r)$ is the expected *mean crowding* of the offspring variables $\{v_1^{(r)}, \ldots, v_{M_r}^{(r)}\}$.

$$c_r = \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}\left[(v_i^{(r)})_2\right] \tag{6}$$

$$= \frac{1}{M_{r-1} - 1} \frac{1}{M_{r-1}} \mathbb{E}\left[\sum_{i=1}^{M_r}(v_i^{(r)})_2\right]$$

$$= \frac{1}{M_{r-1} - 1} \frac{1}{\mathbb{E}[M_{r-1}]} \mathbb{E}\left[\sum_{i=1}^{M_r}(v_i^{(r)})_2\right]$$

$$= \frac{1}{M_{r-1} - 1} \frac{1}{\mathbb{E}[\sum_{i=1}^{M_r} v_i^{(r)}]} \mathbb{E}\left[\sum_{i=1}^{M_r}(v_i^{(r)})_2\right]$$

$$= \frac{1}{M_{r-1} - 1} \mathbb{E}\left[\frac{\sum_{i=1}^{M_r}(v_i^{(r)})_2}{\sum_{i=1}^{M_r} v_i^{(r)}}\right] \tag{7}$$

Example: generalised Wright-Fisher model

Model as Wright-Fisher, except we don't assume exchangeability of offspring distributions. Instead of choosing parents multinomially with equal weights, we choose them multinomially with unequal weights. Thus the marginal offspring distributions are $v_i^{(r)} \sim \text{Bin}(M_{r-1}, p_i)$. We therefore have $\mathbb{E}(v_i^{(r)}) = M_{r-1}p_i$ and $\mathbb{E}(v_i^{(r)}2) = M_{r-1}p_i(1 - p_i + M_{r-1}p_i)$. Hence

$$c_r = \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}[(v_i^{(r)})_2]$$

$$= \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r} \mathbb{E}[v_i^{(r)}2 - v_i^{(r)}]$$

$$= \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r}[M_{r-1}p_i(1 - p_i + M_{r-1}p_i - 1)]$$

$$= \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r}[M_{r-1}p_i^2(M_{r-1} - 1)]$$

$$= \frac{1}{(M_{r-1})_2} \sum_{i=1}^{M_r}[(M_{r-1})_2 p_i^2]$$

$$= \sum_{i=1}^{M_r} p_i^2$$

and

$$\sigma^2(r) = \frac{1}{M_{r-1}} \sum_{i=1}^{M_r} \mathbb{E}[(v_i^{(r)})_2]$$
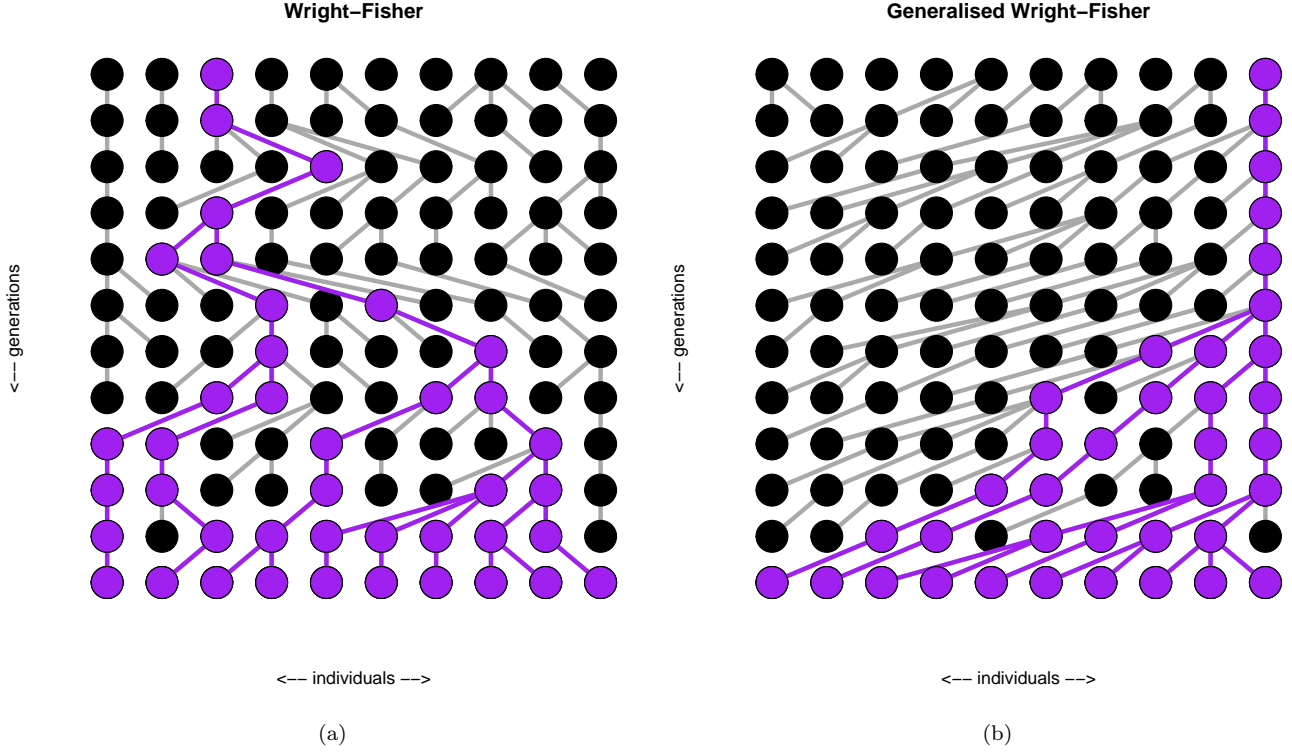
$$= (M_{r-1} - 1) \sum_{i=1}^{M_r} p_i^2$$

Figure 1: Realisations from a Wright-Fisher model with $N = 10$ individuals and 12 generations. (a) is from the standard model with uniform weights ("fitnesses"). (b) is from the generalised model with weighting on individual i proportional to i.

Special case: uniform weights

Suppose $p_i = 1/M_r$ for all i (the standard Wright-Fisher model). Then we have

$$c_r = \sum_{i=1}^{M_r} p_i^2 = \sum_{i=1}^{M_r} \frac{1}{M_r^2} = \frac{1}{M_r}$$

and if the population size doesn't depend on r (i.e. $M_r \equiv N$), we can find the full distribution of T_2:

$$\mathbb{P}(T_2 > l) = \prod_{r=1}^{l}(1 - c_r) = \prod_{r=1}^{l}\left(1 - \frac{1}{M_r}\right) = \left(1 - \frac{1}{N}\right)^l$$

and deduce

$$\mathbb{E}[T_2] = \sum_{l=0}^{\infty}\mathbb{P}(T_2 > l) = \sum_{l=0}^{\infty}(1 - c_r)^l = c^{-1} = M_r = N$$

A realisation of 12 generations of this model with $N = 10$ is shown in Figure 1a.

Special case: linearly increasing weights

Suppose $p_i = 2i/(M_r(M_r - 1))$ for each i. Then

$$c_r = \sum_{i=1}^{M_r} p_i^2 = \sum_{i=1}^{M_r} \frac{4i^2}{M_r^2(M_r + 1)^2} = \frac{4}{M_r^2(M_r + 1)^2}\sum_{i=1}^{M_r} i^2 = \frac{2(2M_r + 1)}{3M_r(M_r + 1)}$$

and for constant population size

$$\mathbb{E}[T_2] = \frac{3N(N + 1)}{2(2N + 1)}$$

A realisation of 12 generations of this model with $N = 10$ is shown in Figure 1b.

The theorem of Möhle

Algebraic tools for proof

Here are a few identities and inequalities that will be referred to when proving the theorem. (8) is obtained using a multinomial expansion, (9) using Bernoulli's inequality, and (10)–(13) by expanding factorials.

$$\sum_{i_1\ldots i_m=1}^{n}\prod_{j=1}^{m} x_{i_j} = \prod_{j=1}^{m}\sum_{i=1}^{n} x_i = \left(\sum_{i=1}^{n} x_i\right)^m \tag{8}$$

$$(k-x)^n = k^n\left(1-\frac{x}{k}\right)^n \leq k^n - nxk^{n-1} \tag{9}$$

$$n^a \geq (n)_a \tag{10}$$

$$(n)_a \leq (n)_b\, n^{a-b}, \text{ if } 0 \leq b \leq a \tag{11}$$

$$\frac{n^{a-b}}{(n)_a} = \frac{1}{(n)_b} + O(n^{-b-1}) \tag{12}$$

$$\frac{1}{(n)_b} = \frac{1}{n^b} + O(n^{-b-1}) \tag{13}$$

Now for simplicity we assume a constant population size $M_r \equiv N$, which for the purposes of SMC will generally be satisfied.

Theorem 1. *Let $T \subset \mathbb{R}$ and suppose there is a function $\tau : T \to \mathbb{N}_0$ satisfying:*

(A) correct limiting coalescence rate

$$\forall t \in T,\ \lim_{N\to\infty}\sum_{r=1}^{\tau(t)} c_r = t$$

(B) variance of coalescence rate goes to zero

$$\forall t \in T,\ \lim_{N\to\infty}\sum_{r=1}^{\tau(t)} c_r^2 = 0$$

(C) no triple coalescences

$$\forall t \in T, \forall k \in \mathbb{N}_0,\ \lim_{N\to\infty}\sup_{r\leq\tau(t)}\frac{1}{N^3 c_r}\sum_{i=1}^{N}\mathbb{E}\left[(v_i^{(r)})_2(v_i^{(r)})^k\right] = 0$$

(D) only one coalescence at a time

$$\forall t \in T,\ \lim_{N\to\infty}\sup_{r\leq\tau(t)}\frac{1}{N^4 c_r}\sum_{i,j=1}^{N}\mathbb{E}\left[(v_i^{(r)})_2(v_j^{(r)})^2\right] = 0$$

Then the finite-dimensional distributions of $\{\mathcal{R}_{\tau(t)}\}_{t\in T}$ converge to those of the n-coalescent (with time restricted to T) in the limit $N \to \infty$.

Note that, since (being a probability) $c_r \geq 0$ for all r; under (A), (B) is equivalent to

$$\forall t \in T,\ \lim_{N\to\infty}\sup_{r\leq\tau(t)} c_r = 0 \tag{14}$$

Proof. We first bound the transition probability $p_{\xi\eta}(r)$ as given by (2), in each of four possible cases. This will show that the only type of coalescence event to occur at any one time in the limit $N \to \infty$ is a merger of exactly one pair of lineages (**Case 1.** below). Assume the offspring numbers $v_i^{(r)}$ are known (so we drop the expectations).

Case 1. η is obtained from ξ by merging exactly one pair of lineages, i.e. $b_1 = 2, b_2 = \cdots = b_a = 1$, and $b = a + 1$. We derive an upper bound:

$$
\begin{aligned}
p_{\xi\eta}(r) &= \frac{1}{(N)_b} \sum_{\substack{i_1,\ldots,i_a=1 \\ \text{distinct}}}^{N} (v_{i_1}^{(r)})_2 (v_{i_2}^{(r)})_1 \cdots (v_{i_a}^{(r)})_1 \\
&\leq \frac{1}{(N)_b} \sum_{i_1,\ldots,i_a=1}^{N} (v_{i_1}^{(r)})_2 (v_{i_2}^{(r)})_1 \cdots (v_{i_a}^{(r)})_1 && \text{dropping distinctness} \\
&= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \sum_{i_2,\ldots,i_a=1}^{N} v_{i_2}^{(r)} \cdots v_{i_a}^{(r)} \\
&= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 (v_1^{(r)} + \cdots + v_N^{(r)})^{a-1} && \text{using (8)} \\
&= \frac{N^{b-2}}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 && \text{using (1)} \\
&= \left(\frac{1}{(N)_2} + O(N^{-3}) \right) \sum_{i=1}^{N} (v_i^{(r)})_2 && \text{using (12)} && (15) \\
&= c_r + o(c_r) && \text{by (6) and (C)}
\end{aligned}
$$

and a lower bound:

$$
p_{\xi\eta}(r) = \frac{1}{(N)_b} \sum_{\substack{i_1,\ldots,i_a=1 \\ \text{distinct}}}^{N} (v_{i_1}^{(r)})_2 (v_{i_2}^{(r)})_1 \cdots (v_{i_a}^{(r)})_1
$$

$$
= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \sum_{\substack{i_2,\ldots,i_a=1 \\ \text{distinct}\neq i}}^{N} (v_{i_2}^{(r)})_1 \cdots (v_{i_a}^{(r)})_1
$$

$$
\geq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \sum_{j\neq i} \sum_{\substack{i_3,\ldots,i_a \\ \neq i}} \left[v_j \cdot v_{i_3} \cdots v_{i_a} - \binom{b-2}{2} (v_j^{(r)})^2 \cdot v_{i_4} \cdots v_{i_a} \right] \tag{16}
$$

$$
= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \left[\sum_{i_2,\ldots,i_a \neq i} v_{i_2} \cdots v_{i_a} - \binom{b-2}{2} \sum_{i_3,\ldots,i_a \neq i} v_{i_3} \cdots v_{i_a} \sum_{j \neq i} (v_j^{(r)})^2 \right]
$$

$$
= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \left[(N-v_i)^{b-2} - \binom{b-2}{2} \sum_{j\neq i} (v_j^{(r)})^2 N^{b-4} \right] \qquad \text{using (8)}
$$

$$
= \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 (N-v_i)^{b-2} - \frac{1}{(N)_b} \binom{b-2}{2} \sum_{i\neq j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2 N^{b-4}
$$

$$
\geq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 (N)^{b-2} - \frac{1}{(N)_b} (b-2) \sum_{i=1}^{N} (v_i^{(r)})_2 v_i^{(r)} N^{b-3}
$$

$$
- \frac{1}{(N)_b} \binom{b-2}{2} \sum_{i\neq j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2 N^{b-4} \qquad \text{using (9)}
$$

$$
\geq \frac{1}{(N)_2} \sum_{i=1}^{N} (v_i^{(r)})_2 - \frac{(b-2)N^{b-3}}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 v_i^{(r)}
$$

$$
- \frac{N^{b-4}}{(N)_b} \binom{b-2}{2} \sum_{i\neq j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2
$$

$$
= \frac{1}{(N)_2} \sum_{i=1}^{N} (v_i^{(r)})_2 - \left[\frac{(b-2)}{N^3} + O(N^{-4}) \right] \sum_{i=1}^{N} (v_i^{(r)})_2 v_i^{(r)}
$$

$$
- \left[\frac{1}{N^4} \binom{b-2}{2} + O(N^{-5}) \right] \sum_{i\neq j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2 \qquad \text{using (12), (13) (17)}
$$

$$
= c_r + o(c_r) \qquad \text{by (6),(C),(D)}
$$

Hence in this case $p_{\xi\eta}(r) = c_r + o(c_r)$. The inequality (16) is obtained by bounding the number of configurations with distinct parents by the the number of configurations with not necessarily distinct parents minus the number with at least one pair of parents chosen indistinctly. This leaves us with only the distinct-parents configurations since all indistinct choices must necessarily have a pair of parents chosen indistinctly, and the inequality arises from double-counting.

Case 2. η is obtained from ξ by merging three or more lineages into one, possibly as well as other simultaneous mergers,

i.e. $b_1 \geq 3$.

$$p_{\xi\eta}(r) = \frac{1}{(N)_b} \sum_{\substack{i_1,\ldots,i_a=1 \\ \text{distinct}}}^{N} (v_{i_1}^{(r)})_{b_1} (v_{i_2}^{(r)})_{b_2} \cdots (v_{i_a}^{(r)})_{b_a}$$

$$\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_{b_1} \sum_{i_2,\ldots,i_a=1}^{N} (v_{i_2}^{(r)})_{b_2} \cdots (v_{i_a}^{(r)})_{b_a}$$

$$\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_{b_1} (v_1^{(r)} + \cdots + v_a^{(r)})^{b_2+\cdots+b_a} \qquad\qquad \text{using (8)}$$

$$\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_{b_1} N^{b-3} \qquad\qquad \text{since } b_2 + \cdots + b_a \leq b - 3 \qquad (18)$$

$$\leq \frac{N^{b-3}}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 (v_i^{(r)})^{b_1-2} \qquad\qquad \text{using (11)} \qquad (19)$$

$$= o(c_r) \qquad\qquad \text{using (C)}$$

Case 3. η is obtained from ξ via two or more pair mergers, with no merges of more than two lineages, i.e. $2 = b_1 = b_2 \geq b_3 \geq \cdots \geq b_a \geq 1$.

$$p_{\xi\eta}(r) = \frac{1}{(N)_b} \sum_{\substack{i_1,\ldots,i_a=1 \\ \text{distinct}}}^{N} (v_{i_1}^{(r)})_2 (v_{i_2}^{(r)})_2 (v_{i_3}^{(r)})_{b_3} \cdots (v_{i_a}^{(r)})_{b_a}$$

$$\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \sum_{j=1}^{N} (v_j^{(r)})_2 \sum_{i_3,\ldots,i_a=1}^{N} (v_{i_3}^{(r)})_{b_3} \cdots (v_{i_a}^{(r)})_{b_a} \qquad \text{dropping distinctness}$$

$$\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_2 \sum_{j=1}^{N} (v_j^{(r)})^2 \sum_{i_3,\ldots,i_a=1}^{N} (v_{i_3}^{(r)})_{b_3} \cdots (v_{i_a}^{(r)})_{b_a} \qquad \text{using (10)}$$

$$\leq \frac{1}{(N)_b} \sum_{i,j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2 (v_1^{(r)} + \cdots + v_N^{(r)})^{b_3+\cdots+b_a}$$

$$= \frac{N^{b-4}}{(N)_b} \sum_{i,j=1}^{N} (v_i^{(r)})_2 (v_j^{(r)})^2 \qquad\qquad\qquad (20)$$

$$= o(c_r) \qquad\qquad \text{using (D)}$$

Case 4. $\eta = \xi$, i.e. $b_1 = \cdots = b_a = 1$, and $a = b$.

$$
\begin{aligned}
p_{\xi\xi}(r) &= \frac{1}{(N)_a} \sum_{\substack{i_1,\dots,i_a=1 \\ \text{distinct}}}^{N} (v_{i_1}^{(r)})_1 \cdots (v_{i_a}^{(r)})_1 \\
&= \frac{1}{(N)_a} \sum_{\substack{i_1,\dots,i_a=1 \\ \text{distinct}}}^{N} v_{i_1}^{(r)} \cdots v_{i_a}^{(r)} \\
&= \frac{1}{(N)_a} \left[\sum_{i_1,\dots,i_a=1}^{N} v_{i_1}^{(r)} \cdots v_{i_a}^{(r)} - \binom{a}{2} \sum_{j=1}^{N} (v_j^{(r)})^2 \sum_{i_3,\dots,i_a=1}^{N} v_{i_3}^{(r)} \cdots v_{i_a}^{(r)} \right] \quad (21) \\
&= \frac{1}{(N)_a} \left[(v_1^{(r)} + \cdots + v_N^{(r)})^a - \binom{a}{2} \sum_{i=1}^{N} (v_i^{(r)})^2 (v_1^{(r)} + \cdots + v_N^{(r)})^{a-2} \right] \quad \text{using (8)} \\
&= \frac{1}{(N)_a} \left[N^a - \binom{a}{2} N^{a-2} \sum_{i=1}^{N} (v_i^{(r)})^2 \right] \\
&\geq 1 - \binom{a}{2} \frac{N^{a-2}}{(N)_a} \sum_{i=1}^{N} (v_i^{(r)})^2 \quad \text{using (10)} \\
&= 1 - \binom{a}{2} \left[\frac{1}{(N)_2} + O(N^{-3}) \right] \sum_{i=1}^{N} (v_i^{(r)})^2 \quad \text{using (12)} \quad (22) \\
&= 1 - c_r + o(c_r)
\end{aligned}
$$

The equality (21) is obtained in the same way as (16), with no double-counting. We have now shown that the only coalescence events having positive probability in the limit $N \to \infty$ are staying the same (**Case 4.**) or merging a single pair of lineages (**Case 1.**). All other possibilities have asymptotic probability $o(c_r)$.

It remains to show that the finite-dimensional distributions converge to those of the Kingman coalescent. Because the processes considered are Markov even when viewed as coalescing backwards in time, it suffices to show that the generators of the process converge to the generators of the Kingman coalescent (this will no longer be the case for the processes considered in **?**). $\qquad\square$

For the argument of **?**, a different form is needed for the upper bound on triple mergers (**Case 2.**). Starting from (18), we obtain:

$$
\begin{aligned}
p_{\xi\eta}(r) &\leq \frac{1}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})_{b_1} N^{b-3} \\
&\leq \frac{N^{b-3}}{(N)_b} \sum_{i=1}^{N} (v_i^{(r)})^{b_1} \\
&= \left[\frac{1}{N^3} + O(N^{-4}) \right] \sum_{i=1}^{N} (v_i^{(r)})^{b_1} \quad (23)
\end{aligned}
$$

Coalescence in the Moran and Wright-Fisher models

The results in this section come from my own reasoning and investigation, though I'm sure that they can (if correct) be found somewhere else.

Moran model

We would like to know about the distribution of the *full coalescence time* τ, the first generation backwards in time at which all N individuals in generation 0 have coalesced. Let us consider the process by which coalescence occurs in the Moran model.

The appearance of an absorbing Markov chain

In any one generation, at most one pair can coalesce (alternatively, the same individual dies and gives birth so nothing changes). However, a pair coalescence does not always represent a coalescence in the ancestry of interest, i.e. the ancestral line of the N generation-0 individuals. We denote by c_r the number of "useful" pair coalescences that have happened up to and including generation r. Thus $c_0 = 0$. Note that the maximum value of c_r is $N - 1$ (indicating that all N lineages have coalesced and the MRCA has been reached). It is also evident that $N - c_r$ gives the number of individuals in generation r that are part of a lineage continuing to generation 0.

The crucial observation is that a pair coalescence corresponds to a coalescence in the N lineages of interest (increasing the value of c_r by 1) if and only if the pair consist of two distinct "useful" individuals out of these $N - c_r$. Since c_{r+1} depends only on N and c_r, $\{c_r\}_{r \in \mathbb{N}}$ is a Markov chain, with state space $\{0, \ldots, N-1\}$, absorbing state $N - 1$ and transition probabilities derived below.

The first time (backwards in time) that a pair coalesces will always be useful, since all N individuals are useful. Thus the probability $\mathbb{P}(c_{r+1} = 1 \mid c_r = 0)$ is simply the probability of choosing two *distinct* individuals from generation $r - 1$.

$$p_{0,1} := \mathbb{P}(c_{r+1} = 1 \mid c_r = 0) = \frac{N}{N} \frac{N-1}{N} = 1 - \frac{1}{N}$$

Now consider the transition probabilities from a general state i. Given $c_r = i$, c_{r+1} can take only the values i or $i + 1$. To increase by 1, we must have a useful pair coalescence. Thus the probability $\mathbb{P}(c_{r+1} = i + 1 \mid c_r = i)$ is the probability of choosing two distinct individuals from the $N - i$ useful individuals.

$$p_{i,i+1} := \mathbb{P}(c_{r+1} = i + 1 \mid c_r = i) = \frac{N-i}{N} \frac{N-i-1}{N} \qquad i = 1, \ldots, N-2$$

and we see that the expression for $p_{0,1}$ was a special case of this. The probability of staying in state i will be $p_{i,i} = 1 - p_{i,i+1}$ for $i = 1, \ldots, N-2$. The Markov chain is now fully specified:

$$p_{i,j} := \mathbb{P}(c_{r+1} = j \mid c_r = i) = \begin{cases} 1 & \text{if } i = j = N-1 \\ 1 - \frac{(N-i)(N-i-1)}{N^2} & \text{if } j = i; \ i = 0, \ldots, N-2 \\ \frac{(N-i)(N-i-1)}{N^2} & \text{if } j = i+1; \ i = 0, \ldots, N-2 \\ 0 & \text{otherwise} \end{cases} \qquad (24)$$

Now let us return to the quantity of interest τ, which is the time to absorption of this Markov chain. we use without proof the following results about absorbing Markov chains. Let Q be the part of the transition matrix P containing only transition probabilities between transient states (in this case the first $N - 2$ rows and columns of P). Define $R := (I - Q)^{-1}$. Then $\mathbb{E}(\tau \mid c_0 = i)$ is given by the ith entry of $t := R\mathbf{1}$, and $\mathrm{Var}(\tau \mid c_0 = i)$ is given by the ith entry of $(2R - I)t - t^2$, where t^2 is the element-wise square of t. Our chain starts at 0, so we take the corresponding entry to find $\mathbb{E}(\tau)$ and $\mathrm{Var}(\tau)$.

We can easily calculate these quantities numerically for any fixed N. Figure 2 shows the behaviour of $\mathbb{E}(\tau)$ over a range of population sizes N. It seems that when τ is measured in rescaled generations, its expectation is linear in N. Can we prove this analytically? We see a similar story in the standard deviation of τ under rescaling, although the relationship appears to be linear only asymptotically. Can we prove this analytically?

A loose bound

Let us consider another quantity, $\mathbb{P}(\tau \le N)$. This is the probability that N generations back, all N lineages have coalesced. Since at most one pair of lineages can coalesce in one step, we have $\mathbb{P}(\tau \le N) = \mathbb{P}(\tau = N)$. Now we can put a (seemingly loose, at least for larger N) bound on this probability. The event $\{\tau = N\}$ can only happen if there is some pair coalescence in every generation $1, \ldots, N$ (regardless of whether it is a useful coalescence). The probability of having no coalescence in a given generation is $1/N$, and is independent between generations. Therefore

$$\mathbb{P}(\tau \le N) = \mathbb{P}(\tau = N) \le \left(1 - \frac{1}{N}\right)^N \underset{N \to \infty}{\nearrow} e^{-1} \simeq 0.37$$
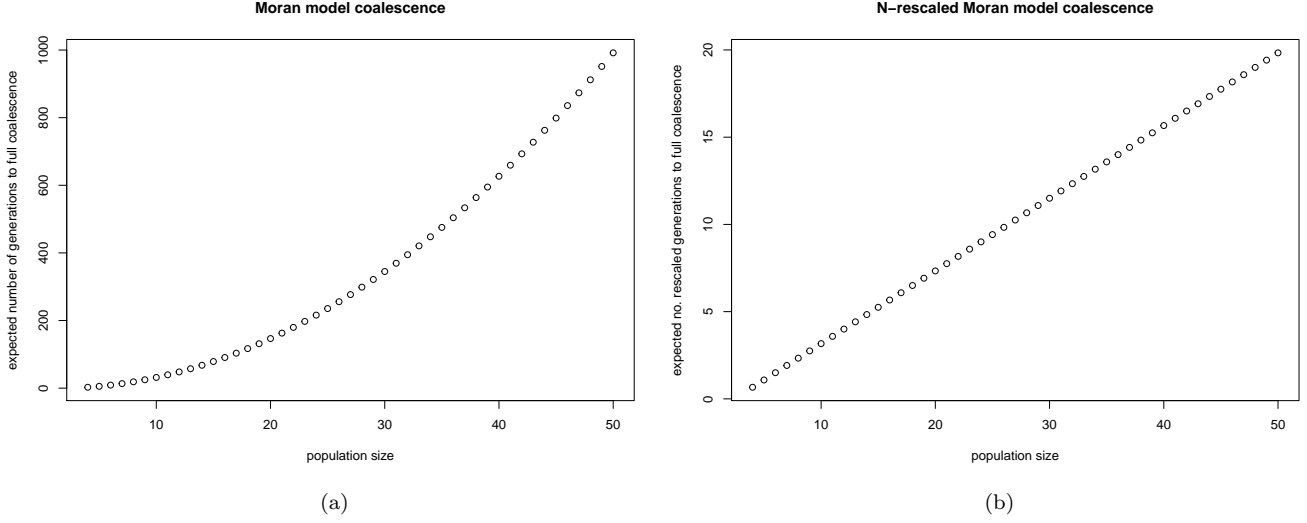
Consider this bound in light of the results in Figure 2.

Figure 2: Expected number of generations to full coalescence $\mathbb{E}(\tau)$ in the Moran model, over a range of population sizes N. In (b) τ is measured in rescaled generations (number of generations divided by N).

Distribution of the full coalescence time

It turns out we can find a closed form (though not a beautiful one) for the full distribution of τ. We divide into three cases for $\mathbb{P}(\tau = k)$:

1. $k < N - 1$. In this case there have not yet been enough steps to transition from state 0 to state N-1, since we can move up at most one state per step.
$$\mathbb{P}(\tau < N - 1) = 0$$

2. $k = N - 1$. The only way to transition from state 0 to state $N - 1$ in $N - 1$ steps is to move up a state at every step.
$$\mathbb{P}(\tau = N - 1) = \prod_{i=0}^{N-2} p_{i,i+1} = \prod_{i=0}^{N-2} \frac{(N-i)(N-i-1)}{N^2}$$

3. $k > N-1$. Now we must make sure we don't arrive in state $N-1$ too early. In order to have $\tau = N-1+M$, we must "rest" (remain in the same state) exactly M times. These rests could happen in any M (not necessarily distinct) states $0, \ldots, N-2$. Additionally we must at some point make all of the $N-1$ transitions up one state, as in the case $k = N - 1$.

$$
\begin{aligned}
\mathbb{P}(\tau = N - 1 + M) &= \prod_{i=0}^{N-2} p_{i,i+1} \sum_{\substack{i_1,\ldots,i_M \\ \in \{0,\ldots,N-2\}}} \prod_{j=1}^{M} p_{i_j,i_j} \\
&= \prod_{i=0}^{N-2} p_{i,i+1} \sum_{\substack{i_1,\ldots,i_M \\ \in \{0,\ldots,N-2\}}} \prod_{j=1}^{M} (1 - p_{i_j,i_j+1}) \\
&= \prod_{i=0}^{N-2} \frac{(N-i)(N-i-1)}{N^2} \sum_{\substack{i_1,\ldots,i_M \\ \in \{0,\ldots,N-2\}}} \prod_{j=1}^{M} \left(1 - \frac{(N-i_j)(N-i_j-1)}{N^2}\right)
\end{aligned}
$$

Wright-Fisher model

For the standard Wright-Fisher model (with constant populations size N and exchangeable offspring distributions i.e. uniform weights), $\{c_r\}_{r \in \mathbb{N}}$ is again a Markov chain, with transition probabilities given by

$$p_{i,j} := \mathbb{P}(c_{r+1} = j \mid c_r = i) = \begin{cases} 1 & \text{if } i = j = N-1 \\ \frac{S_{N-i}^{(N-j)}(N)_{N-j}}{N^{N-i}} & \text{if } j \geq i; \ i,j = 0, \ldots, N-2 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

obtained from **?**, equation 3.12 by noting that the number of lineages "alive" at generation r is $N - c_r$. $S_x^{(y)}$ is the number of ways to partition a set of x elements into y non-empty subsets, called Stirling numbers of the second kind (OEIS A008277).

We can now calculate the expectation and variance of τ numerically as we did for the Moran model.

Notes from ?

Notation