Assumptions of WF model

panmictic (random mating)

$N =$ effective population size constant

Any 2 lineages equally probable to coalesce.

time = number of gens before the present
^ non-overlapping

Simultaneous coalescent events do NOT occur

Exchangeability of lineages

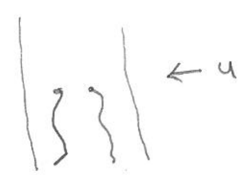
Kingman's Coalescent Model will be a limit of this or a continuous time approximation

Assume panmictic population, time measured in "coalescent units" so that

- the rate of coalesce for 2 lineages is 1

Using $u > 0$ for time in coalescent units and $h(u) =$ probability 2 lineages have not coalesced at time u

then the pertinent diff. eq is



$$\frac{d}{du} h(u) = -h(u)$$

with solution

$$h(u) = e^{-u}$$

Let $P(u) =$ probability 2 lineages did coalesce before time u

$$P(u) = 1 - e^{-u}$$

compared to

$$P(C_2 \le uN) = 1 - \left(1 - \frac{1}{N}\right)^{uN}$$

↑ WF

↓ unit change

See MATLAB plot.

Under the Kingman Coalescent Model, the expected time to coalesce is

$$\mathbb{E}(2 \text{ lineages coalesce}) = \int_0^\infty u\, e^{-u}\, du \qquad u \text{ in coalescence units!}$$

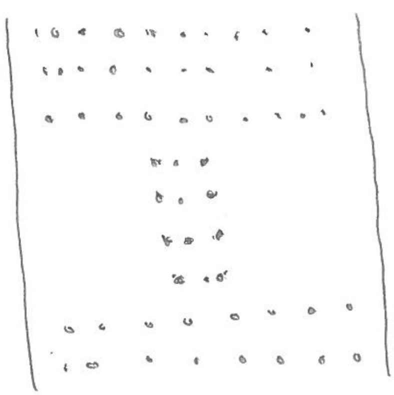
$$= (-1-u)e^{-u}\Big|_0^\infty = 0 - (-1-0)\,e^{-0} = 1 \ ! \qquad \text{as preordained}$$

Some important comments/reiterations/etc.

1) $\mathbb{E}(C_2) = 1$ is a consequence of using COALESCENCE UNITS

$$\Delta u \approx \frac{\Delta t}{N_{WF}}$$

2) Bottlenecks:



Slower $\mathbb{E}_{WF}(C_2) > 3 = N_e$

Speed up in time to coalescence

$\mathbb{E}_{WF}(C_2) = N_e = 3$

The time to coalescence of 2 lineages depends on $N_e(t)$, the current population size. However, in coalescent units (i.e. Kingman coalescent model) $\mathbb{E}(C_2) = 1$ always. This is because if, say $N = N_e$ doubles

original   double N, $1/2$ generation time

$$u = 1 = \frac{\Delta t}{N} = \frac{N \text{ gens}}{N \text{ indiv.}} = \frac{2N \text{ generations}}{2N \text{ individuals}}$$

i.e. you can rescale generation time and population size and still have $u=1$

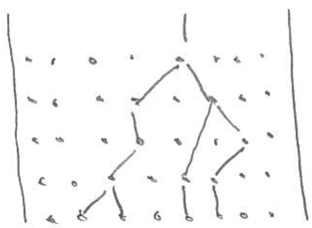
Coalescent units essentially hide any estimation/assumption/knowledge of generation time and population size (Both a plus and a minus)

For a constant population size $N(t) = N$ = constant

$1\ u = t/N$

Coalescent unit

Under the Kingman coalescent we can also compute the expected time that n lineages coalesce into n-1



The crucial assumption is the iid. which makes the rate $\binom{n}{2}$

↑ all pairs equally likely to coalesce

n = 4

In coalescence units,

if $K(u)$ = probability that n lineages remain distinct at time u, then

$$\frac{d}{du} K(u) = -\binom{n}{2} K(u) \qquad K(0) = 1$$

and $K(u) = e^{-\binom{n}{2} u} \qquad u \geq 0$

There are lots of ways to determine

$\mathbb{E}$(time to coalescence for n lineages to n-1)

- Exponential dist with rate $\binom{n}{2} = \frac{n(n-1)}{2} \Rightarrow$
- cdf: P(n lineages coalesce to n-1 $\leq$ time u) $= 1 - K(u) = 1 - e^{-\binom{n}{2}u}$

$$= 1 - e^{-\binom{n}{2}u}$$

$\Rightarrow$ density fn is $f(u) = \frac{d}{du}\left(1 - e^{-\binom{n}{2}u}\right) = -e^{-\binom{n}{2}u}\left(-\binom{n}{2}\right) = \binom{n}{2} e^{-\binom{n}{2}u} \qquad u > 0$

Again exponential ...

- More formally,

$$\mathbb{E}\left(\begin{array}{c} n \text{ lineages coalesce} \\ \text{to } n-1 \end{array}\right) = \int_0^\infty u \binom{n}{2} e^{-\binom{n}{2}u}\, du = \frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}$$

exponential

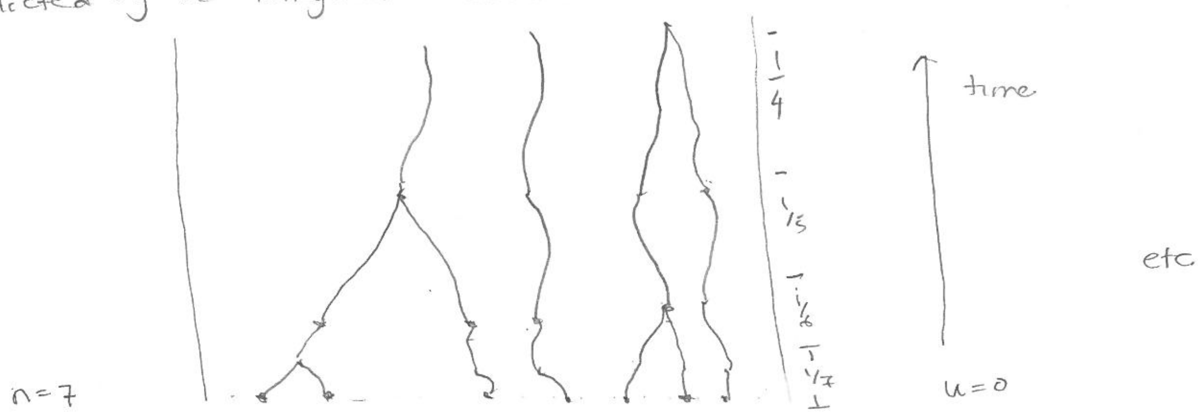
Theorem: Under the Kingman coalescent model, given n lineages in a population, the expected time to the next coalescent event is

$$\mathbb{E}(n \text{ lineages coalesce to } n-1 \text{ lineages}) = \frac{2}{n(n-1)}$$

n	2	3	4	5	6	...	
$\frac{1}{\binom{n}{2}} = \frac{2}{n(n-1)}$	1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{10}$	$\frac{1}{15}$	$1/21$	etc.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1

This has a profound effect on the types of gene trees predicted by the Kingman coalescent model.



i.e. Coalescence events are more numerous near the present.

R example.

Finally, given n lineages then the expected time for them to coalesce down to one lineage is

$$\mathbb{E}(\text{gene tree w/ } n \text{ lineage formed}) = \sum_{i=2}^{n} \frac{1}{\binom{i}{2}} = \sum_{i=2}^{n} \frac{2}{i(i-1)} \overset{\text{partial fractions}}{=} 2\sum_{i=2}^{n}\left[\frac{1}{i-1} - \frac{1}{i}\right] = 2\left(1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \dots \frac{1}{n-1} - \frac{1}{n}\right)$$

$$= 2\left(1 - \frac{1}{n}\right) \quad \text{and} \quad \lim_{n\to\infty} 2\left(1 - \frac{1}{n}\right) = 2 \text{ coalescent units}$$