

~~What good is it?~~

The Coalescent

①

What good is it?

Haplotype test -

~~What good is it?~~

Objectives

WF forward in time \rightarrow Fixation

Backward - there is a genealogy
 n lineages $\rightarrow n-1 \rightarrow n-2 \dots 1$ (entire sample coalesces)

never $n \rightarrow n-2$ (assume $n \ll N$) Sample size is very small compared to pop. size

Topology

What else is needed to draw a tree?

Branch lengths

SIDEBAR 1 exp exponential distribution
Waiting time to failure

$$\text{Pr}(\text{fail each day}) = x$$

$$\text{Pr}(\text{not fail}) = 1-x$$

$$\text{Pr}(\text{4 days to fail}) = (1-x)(1-x)(1-x)(1-x) = (1-x)^4 \quad \text{geometric}$$

$$(1-x)^t \approx e^{-xt} \quad \leftarrow \text{an important approximation - true for small } x$$

$$\text{Pr}(\text{fail at time } t) = x e^{-xt}$$

$$\text{Mean} = \frac{1}{x}$$

$x \leftarrow \text{rexp}(1000, .5)$ \leftarrow to generate ¹⁰⁰⁰ random draws from an exponential distribution with parameter .5

Topology - pick pairs of lineages until only 1 lineage

(2)

Branch lengths

$n=2$ prob. draw same allele = $\frac{1}{2N}$

prob draw different allele = $1 - \frac{1}{2N}$

prob 2 alleles ~~do not~~ coalesce = $1 - \frac{1}{2N}$

$Pr(\text{coal at } t) = (1 - \frac{1}{2N})^{t-1} \cdot \frac{1}{2N} \approx \frac{1}{2N} e^{-\frac{t}{2N}}$

with n lineages, any pair could coalesce

$\binom{n}{2} = \frac{n(n-1)}{2}$ so $Pr(\text{coal}) = \frac{n(n-1)}{2} \cdot \frac{1}{2N}$

= prob of exactly 1 coal.

$= \frac{\binom{n}{2}}{2N} e^{-\frac{\binom{n}{2}}{2N} t}$

SIDE BAR 2

Poisson distribution

Sitting on a pier, trying to catch fish.
very early. sun just rising.

Throw line in water, reel in.

$Pr(\text{catch is small, say } \mu)$

Cast again are caught each hour

How many fish in 1 hr.?

$Pr(k \text{ fish} | \mu) = \frac{e^{-\mu} \mu^k}{k!}$

(Poisson mean μ)

$x \leftarrow rpois(10000, 2)$ $\leftarrow R$ expression to generate 10000 draws from Poisson.

Poisson dist because it grows fish caught

Expected time to ^{next} coalesce

$= \frac{2N}{\binom{n}{2}} = \frac{4N}{n(n-1)}$

↓
1st time we f
↓
then slower

(3)

How many ~~and~~ ^{negating} seg site (μ) expected between 2 alleles ($n=2$)

$$E(\text{TMRCA}) = E(\text{exp. on } 2N) = E\left(\frac{1}{2N} e^{-\frac{t}{2N}}\right) = 2N$$

$\swarrow \searrow \mu \rightarrow 2\mu$ $E(S) = E(T) \times \mu \times 2 = 4N\mu$

2 Stochastic things - time to coal + time to μ
These can be simulated independently.

~~or lineages~~

~~Pr. Each pair of lineages has prob $\frac{1}{2N}$ to coalesce each gen.~~

~~There are $\binom{n}{2} = \frac{n(n-1)}{2}$ pairs that could coal.~~

~~So $\text{Pr}(\text{coalescence}) = \frac{n(n-1)}{2} / 2N$~~

~~So dist of times~~

To page 9

(4) (5)

Total tree length $E(T_{tot}) = \sum_{i=2}^n i T(i) = \sum_{i=2}^n i \frac{4N}{i(i-1)}$
 $= \sum_{i=2}^n \frac{4N}{i-1} = 4N \sum_{i=1}^{n-1} \frac{1}{i}$

What fraction of tree is the last coal? ($\sim \frac{1}{2}$)

Total # μ ? $E(S) = \mu \times \text{Total branch length}$
 $= \mu \cdot 4N \sum_{i=1}^{n-1} \frac{1}{i} = 4N\mu \sum_{i=1}^{n-1} \frac{1}{i}$

Fraction on external branches (SFS) # singletons = 0

How many branches with n tips?
 $2(n-1)$

How many are external?
 n



Drawing a tree (fun)

2 task topology & branch lengths

3rd task - mutation - of any sort

topo

Pick pairs of nodes \rightarrow Calc. \rightarrow Draw

choose(2, 5)

node indices

Branch lengths

Draw fm exp. r_{exp}

Calc exp. mean / # nodes

Draw it

\rightarrow Node times

Now we tree

Add μ

~~Sitting on a pier. Throw line with reel in.
 Pr catch is small size fish
 Fish cast again
 How many fish in 1 hr~~