Q1) M haplotypes
\[ z = (z_1, \ldots, z_M) : \text{number of offsprings of the } M \text{ haplotypes in generation } k \]

Show that
\[ P(\bar{z} = z) = \frac{M!}{z_1! z_2! \cdots z_M!} \frac{M^M}{z_1 + \cdots + z_M} \]

v1) In a W-F model, each child independently chooses a parent (with replacement) from 1, \ldots, M haplotypes, each with probability \( \frac{1}{M} \).

\[ \Rightarrow \text{Total number of times each parent is chosen} \]
\[ z \sim \text{Multinomial}(n, p=(\frac{1}{M}, \ldots, \frac{1}{M})) \]
\[ \Rightarrow P(\bar{z} = z) = \binom{M}{z_1} \binom{M-z_1}{z_2} \cdots \binom{M-z_{M-1}}{z_M} \frac{M!}{z_1! \cdots z_M!} \left( \frac{1}{M} \right)^M \]

v2) Each child chooses a parent with probability \( \frac{1}{M} \), independently of other children.

\[ \Rightarrow \text{Any specific choice of parent by } M \text{ children has probability } \left( \frac{1}{M} \right)^M \]

There are \( \binom{M}{z_1} \binom{M-z_1}{z_2} \cdots \binom{M-z_{M-1}}{z_M} \) assignments of \( M \) children to \( M \) parents, such that each parent has \( z = z \) offsprings.

\[ \Rightarrow P(\bar{z} = z) = \binom{M}{z_1} \binom{M-z_1}{z_2} \cdots \binom{M-z_{M-1}}{z_M} \cdot \left( \frac{1}{M} \right)^M \]

\[ = \frac{M!}{z_1! \cdots z_M!} \left( \frac{1}{M} \right)^M \]
\[ P(T = t) = P(T_1 = z_1) P(T_2 = z_2 | T_1 = z_1) \ldots P(T_M = z_M | T_1 = z_1, T_2 = z_2, \ldots, T_{M-1} = z_{M-1}) \]

because \( N \) offspring choose parents independently with probability \( \frac{1}{M} \),

\[ P(T_2 = z_2 | T_1 = z_1) = \left( \frac{M - z_1}{z_2} \right) \left( \frac{1}{M-1} \right)^{z_2 - 1} \left( \frac{1}{M-1} \right)^{M-z_1-z_2} \]

because \( M-z_1 \) offspring choose from \( M-1 \) parents independently with probability \( \frac{1}{M-1} \). \( z_1 \) cannot be chosen.

\[ = \left( \frac{M}{M} \right)^{z_1} \left( \frac{1}{M} \right)^{z_2} \left( \frac{1}{M-1} \right)^{M-z_1-z_2} \left( \frac{1}{M-1} \right)^{M-z_1-z_2} \ldots \]

\[ = \frac{M!}{z_1! \ldots z_M!} \frac{(M-1)^{M-z_1}}{M^{M-z_1}} \frac{(M-2)^{M-z_1-z_2}}{(M-1)^{M-z_1-z_2}} \ldots \]

\[ = \frac{M!}{z_1! \ldots z_M!} \frac{M^{M-z}}{M^{M-z}} \frac{(M-1)^{M-z_1-z_2}}{(M-1)^{M-z_1-z_2}} \ldots \]

\[ = \frac{M!}{z_1! \ldots z_M!} \frac{M^{M-z}}{M^{M-z}} \]
a) Each of the \( M \) children choose parent \( i \) independently with probability \( \frac{1}{M} \).

\[
\Rightarrow Z_i \sim \text{Binomial}(M, \frac{1}{M})
\]

i.e.,

\[
P(Z_i = z_i) = \binom{M}{z_i} \left( \frac{1}{M} \right)^{z_i} \left( 1 - \frac{1}{M} \right)^{M-z_i}
\]

b) From a), we know

\[
P(Z_i = 0) = \binom{M}{0} \left( \frac{1}{M} \right)^{0} \left( 1 - \frac{1}{M} \right)^{M} = (1 - \frac{1}{M})^{M}
\]

Note that \( Z_i \)'s are not independent, so the number of parents with no offspring is not binomial.

Define \( X = \sum_{i=1}^{M} \mathbb{1}(Z_i = 0) \), the \# of parents with no offspring.

Then,

\[
E[X] = E[\sum_{i=1}^{M} \mathbb{1}(Z_i = 0)]
\]

Linearity of expectation.

\[
= \sum_{i=1}^{M} E[\mathbb{1}(Z_i = 0)]
\]

\[
= \sum_{i=1}^{M} P(Z_i = 0) = \sum_{i=1}^{M} \left( 1 - \frac{1}{M} \right)^{M}
\]

\[
= M \left( 1 - \frac{1}{M} \right)^{M} \quad \text{as required.}
\]

Notice that \( \frac{E[X]}{\frac{M}{2}} \rightarrow 1 \) as \( M \rightarrow \infty \).

so expected \# of parents with no offspring is linear in \( M \) with slope \( \frac{1}{2} \).
From lecture notes, \( T_j \sim \text{Exponential}(\frac{1}{2}) \)

and \( T_j \)'s are independent.

\[
\Rightarrow T_3 \sim \text{Exp}(3), \quad T_2 \sim \text{Exp}(2)
\]

\text{v1) } W = T_3 + T_2, \quad \text{then}

\[
P(W \leq w) = \int_0^w P(T_3 \leq w-t_2 \mid T_2 = t_2) f_{T_2}(t_2) \, dt_2
\]

\[
= P(T_3 \leq w-t_2) \quad \text{by independence.}
\]

\[
= \int_0^w (1 - e^{-3(w-t_2)}) e^{-t_2} \, dt_2
\]

\[
= \left[ e^{-t_2} - e^{-3w-2t_2} \right]_0^w
\]

\[
= 1 - e^{-w} - e^{-3w} (e^{2w} - 1) \cdot \frac{1}{2}
\]

\[
\Rightarrow P(W \leq w) = 1 - \frac{3}{2} e^{-w} + \frac{e^{-3w}}{2}
\]

differentiate to get density function:

\[
f_W(w) = \frac{d}{dw} P(W \leq w) = \frac{3}{2} (e^{-w} - e^{-3w})
\]

\text{v2) use formula for convolution of two independent random variables:}

\[
f_W(w) = \int_0^w f_{T_3}(w-t_2) f_{T_2}(t_2) \, dt_2
\]

\[
= \int_0^w 3 e^{-3(w-t_2)} e^{-t_2} \, dt_2 = \int_0^w 3 e^{-3w} e^{2t_2} \, dt_2
\]

\[
= \frac{3}{2} (e^{-w} - e^{-3w}).
\]
At $k=4$, the first event is a coalescence of any two haplotypes.

At $k=3$, depending on which lineages coalesce, we obtain a different tree shape:

- either A: coalesce same lineage that coalesced at $k=4$.
- B: coalesce the other two lineages.

At $k=2$, the remaining lineages must coalesce.

At $k=3$, we had a choice of coalescing with 2 lineages to obtain tree A. For B, we could only coalesce one pair.

\[ P(A) = \frac{2}{3}, \quad P(B) = \frac{1}{3} \]

i.e., expect an asymmetric tree shape!
Q4) Independently simulate two things:

1) times: \( T_j \sim \text{Exp}(\frac{1}{2}) \)

2) shape: which lineages to coalesce.

For times, use that if \( U \sim \text{Uniform}(0,1) \), then

\[ P(\ U \leq F(t)) = F(t), \text{ with } F(t) = P(T_j \leq t). \]

Therefore

\[ P(\ F^{-1}(U) \leq t) = F(t) \]

\[ \Rightarrow F^{-1}(U) \sim \text{Exp}(\frac{1}{2}). \]

We have that

\[ F(t) = 1 - e^{-\left(\frac{1}{2}\right)t} \]

\[ \Rightarrow -F(t) + 1 = e^{-\left(\frac{1}{2}\right)t} \]

\[ \Rightarrow t = -\left(\frac{1}{2}\right) \log \left[ 1 - F(t) \right] \]

So

\[ F^{-1}(U) = -\left(\frac{1}{2}\right) \log \left[ 1 - U \right] \leq -\left(\frac{1}{2}\right) \log U \]

For shape, choose a pair uniformly at random out of all remaining lineages.

If you want to see how these trees should look like, you can use the "rcoal" function in the "ape" package in R.

```r
> library (ape)
> rt <- rcoal(10)
> plot.phylo (rt)
```
Q5) From the notes, the probability that any lineage has \( z \) descendants while \( k \) lineages remain is given by

\[
P_k(z = z) = \frac{(r-2-1)}{(k-2)} \frac{(k-2)}{(r-1)}
\]

At \( k = 2 \), \( P_2(z = z) = \frac{(r-2-1)}{(2-2)} = \frac{1}{r-1} \)

i.e. uniform.

Q6) We take a sub-sample of \( n \) haplotypes from \( r \) in haplotypes.

\( (z, r-z) \): number of descendants of the two lineages at \( k = 2 \).

\[
z = 5, \quad r-z = 3
\]

a) The MRCA of the sub-sample is not the same as that of the \( r \) samples, if they coalesce before the final coalescence happens.

This happens if we sample \( n \) haplotypes from \( z \) or \( n \) haplotypes from \( r-z \).

\[
\binom{z}{n}: \text{# ways of choosing } n \text{ from } z
\]
\[
\binom{r-z}{n}: \text{# ways of choosing } n \text{ from } r-z
\]
\[
\binom{r}{n}: \text{# ways of choosing } n \text{ from } r
\]

\[
P(\text{MRCA of } n \neq \text{MRCA of } r) = \frac{\binom{z}{n} + \binom{r-z}{n}}{\binom{r}{n}}
\]
\[ p(\text{same MRCA}) = 1 - p(\text{MRCA of } u \oplus \text{MRCA of } r) \]
\[ = 1 - \frac{\left(\frac{2}{n}\right) + \left(\frac{r-2}{n}\right)}{\binom{n}{r-1}} \]
\[ (\text{all conditional on } t) \]

b) unconditional probability of event \( X = \{ \text{same MRCA}\} \)

\[ p(X) = \sum_{t=1}^{r-1} p(X \mid t) p(t) \]
\[ = \sum_{t=1}^{r-1} \left[ 1 - \frac{\left(\frac{2}{n}\right) + \left(\frac{r-2}{n}\right)}{\binom{n}{r-1}} \right] \]
\[ = \frac{1}{r-1} \]
\[ \text{from a) from Q5} \]

\[ = 1 - \frac{2}{\binom{n}{r-1}(r-1)} \sum_{t=1}^{r-1} \left(\frac{t}{n}\right) \]
\[ = \binom{r}{n+1} \text{ (e.g. using Pascal's triangle)} \]

\[ = 1 - \frac{2 \cdot r! \cdot n! \cdot (r-n)!}{(n+1)! \cdot (r-n-1)! \cdot r! \cdot (r-1)!} = 1 - \frac{2(r-n)}{n+1}(r-1) \]

In the limit \( r \to \infty \), we get \( 1 - \frac{2}{n+1} = \frac{n-1}{n+1} \) as required.