1 Wright Fisher (WF) or Coalescent model

We have a fixed population of $N$ individuals going from generation to generation. The individuals have allele $A$ or $a$. In each generation, individual chooses a single parent from the previous generation, and adopts their allele.

**Discrete time modeling.** Let $N_t$ denote the number of copies of allele $A$ at time $t$, so that $N_0 = 1$. Under selection with parameter $s$,

$$\begin{align*}
\Pr(A) &= c(1 + s) \frac{N_{t-1}}{N} \\
\Pr(a) &= c \left(1 - \frac{N_{t-1}}{N}\right)
\end{align*}$$

Where the normalizing constant

$$c = \frac{1}{(1 + s) \frac{N_{t-1}}{N} + \left(1 - \frac{N_{t-1}}{N}\right)} = \frac{1}{1 + \frac{s N_{t-1}}{N}}$$

Therefore,

$$E(N_t) = N \Pr(A) = \frac{(1 + s)}{1 + \frac{s N_{t-1}}{N}} N_{t-1}$$

**Continuous time.** The growth follows a logistic curve. Our goal is to get the rate of growth. In continuous time, suppose that we are sampling individuals at time $t + \delta t$

$$\begin{align*}
\Pr(A) &= c(1 + s)^{\delta t} \frac{N_t}{N} \\
\Pr(a) &= c \left(1 - \frac{N_t}{N}\right)
\end{align*}$$

Where the normalizing constant

$$c = \frac{1}{(1 + s)^{\delta t} \frac{N_t}{N} + \left(1 - \frac{N_t}{N}\right)} \approx \frac{1}{1 + \frac{s \delta t N_t}{N}}$$

Therefore,

$$E(N_{t+\delta t}) = N \Pr(A) = \frac{(1 + s \delta t)}{1 + \frac{s \delta t N_t}{N}} E(N_t)$$

$$E(N_{t+\delta t}) + \frac{s \delta t N_t}{N} E(N_{t+\delta t}) = E(N_t) + s \delta t E(N_t)$$

$$E(N_{t+\delta t}) - E(N_t) = s \delta t \left( E(N_t) - \frac{N_t}{N} E(N_{t+\delta t}) \right)$$

Or,

$$\frac{E(N_{t+\delta t}) - E(N_t)}{\delta t} = s \left( N_t - \frac{N_t N_{t+\delta t}}{N} \right)$$
Note that
\[ \lim_{\delta t \to 0} N_{t+\delta t} = N_t \]

Therefore,
\[ \frac{dE(N_t)}{dt} = s \frac{N_t(N - N_t)}{N} \]

Suppose the population goes from \( N_i \) to \( N_j \) in time \( \tau \) Then,
\[
\int_{t=0}^{\tau} s dt = \int_{N_t=N_i}^{N_j} dE(N_t) \left( \frac{1}{N_t} + \frac{1}{N-N_t} \right)
\]
\[ s \tau = \ln \frac{N_t}{N-N_t} \bigg|_{N_i}^{N_j} \]

This gives us the basic logistic regression equation for time as:
\[ \tau = \frac{1}{s} \ln \left( \frac{N_j(N-N_i)}{N_i(N-N_j)} \right) \tag{1} \]

and for population growth as
\[ \tau s = \ln \left( \frac{N_j(N-N_i)}{N_i(N-N_j)} \right) \tag{2} \]
\[ e^{\tau s} = \frac{(\frac{N}{N_j} - 1)}{(\frac{N}{N_i} - 1)} \tag{3} \]
\[ N_t \simeq \frac{N}{1 + Ne^{-\tau s}} \tag{4} \]

where Eqn. 4 is obtained by choosing \( N_i = 1 \) (population at time 0), and \( N_j = N_t \). We can use Eqn. 1 to compute times to fixation, as well as the time to reach different population milestones.

\[ \begin{align*}
N_0 &= 1 \\
N_1 &= \ln n \\
N_2 &= \frac{N}{\ln N} \\
N_3 &= \frac{N}{2} \\
N_4 &= N - \frac{N}{\ln N} \\
N_5 &= N - \ln N \\
N_f &= N - 1
\end{align*} \]

**Fixation, and other times.** To go from \( N_0 \) to \( N_f \), we have
\[ \tau = \frac{1}{s} \ln(N(N-1)) = \frac{2}{s} \ln N \]
This time can be split into 3 different epochs: \(N_0\) to \(N_2\), \(N_2\) to \(N_4\), and \(N_4\) to \(N_f\). Let the respective times be \(\tau_1, \tau_2, \tau_3\). Then,

\[
\tau_1 = \frac{1}{s} \ln \left( \frac{N}{\ln N} \frac{(N - 1)}{(N - \frac{N}{\ln N})} \right) = \frac{1}{s} \ln \left( \frac{N - 1}{\ln N - 1} \right) \approx \frac{1}{s} (\ln N - \ln \ln N)
\]

\[
\tau_2 = \frac{1}{s} \ln \left( \frac{(N - \frac{N}{\ln N})(N - \frac{N}{\ln N})}{\ln N \ln N} \right) = \frac{2}{s} \ln \ln N
\]

\[
\tau_3 = \frac{1}{s} \ln \left( \frac{(N - 1) \frac{N}{\ln N}}{(N - \frac{N}{\ln N})} \right) = \frac{1}{s} \ln \left( \frac{N - 1}{\ln N - 1} \right) \approx \frac{1}{s} (\ln N - \ln \ln N)
\]

2 Sampling from a population where the beneficial allele has reached fixation

Consider the case when the beneficial allele is just driven to fixation, so that \(N_0 = N\). Sample a pair of individuals. \(Q:\) What is the expected time to coalescence for the pair? Recall from standard coalescent theory that if the population remains constant at \(N\), the time to coalescent is distributed exponentially, with success probability \(\frac{1}{N}\). Thus, probability that we coalesce exactly at \(t\) generations (assuming continuous time) is

\[
P_t = \frac{1}{N} \left( 1 - \frac{1}{N} \right)^t \approx \frac{1}{N} e^{-\frac{t}{N}}
\]

Also, the expected time to coalesce is

\[
\int_{t=1}^{t=\infty} \frac{t}{N} e^{-\frac{t}{N}} dt = N
\]

Similarly, if there are \(k\) individuals, the expected time for a coalescence to occur is \(\frac{N}{\binom{k}{2}}\), and the number of mutations created in the time that \(k\) lineages become \(k-1\) is thus

\[
\mu k \frac{N}{\binom{k}{2}} = \frac{2N\mu}{k-1}
\]

This allow us to calculate the total number of polymorphic sites we expect to see as

\[
\sum_{k=n}^{2} \frac{2N\mu}{k-1} \approx 2N\mu \ln n
\]

We want to do a similar calculation for the case when the individuals are sampled from the process under selection at a time when the beneficial allele has just fixed in the population. In this case, the population is decreasing as we go back in time. Specifically, using an approach similar to Eqn. 4, if a population of \(N_i\) individuals becomes \(N_j\) \(t\) generations later, then

\[
e^{st} = \frac{\frac{N}{N_j} - 1}{\frac{N}{N_j} - 1}
\]

\[
N_i \approx \frac{N}{1 + e^{st}}
\]
The last equation is derived choosing $i = 0$, $N_i = N$, and $N_j = N_t$. Now, the probability that two individuals coalesce at time $\tau$ is

$$P_\tau = \frac{1}{N_\tau} e^{-\int_{t=1}^{\tau} \frac{1}{N_t} dt} \quad (11)$$

Q: Can we solve this expression using Eqn. 10, and then find the expected time to coalesce for (a) 2 individuals, and (b) $k$ individuals?