Contents

1 Motivation

2 Diffusion Approximations
   - Wright-Fisher Model
   - Diffusion Processes
   - Stationary Distributions
   - Characterization of Markov Processes
   - Applications of Generators
   - Time Reversal

3 Structured Coalescents
   - Derivation
   - Applications
   - Linked Loci

4 Fecundity Variance Polymorphism
   - Fecundity Variance in Neutral Populations
   - Diploid Model
   - Haploid Model
   - Fluctuating Selection

5 Common Ancestor Process
   - Substitution Rates and Divergence
Contents II

- Derivation
- Genic Selection
- Diploid Models
Motivation

Coalescents, Exchangeability and Selection

**Question:** Why isn’t the genealogy at a selected locus described by Kingman’s coalescent?

Consider the following generalization of the Wright-Fisher model.

**Assumptions:**

- $N$ haploid individuals (constant population size).
- Alleles $A_1$, $A_2$ with frequencies $p$ and $q$.
- Each individual alive in generation $t + 1$ chooses its parent from the preceding generation independently and with replacement with the following probabilities:

  - each $A_1$-type individual: \[ \frac{1}{N} \frac{1 + s/N}{1 + ps/N} \]
  - each $A_2$-type individual: \[ \frac{1}{N} \frac{1}{1 + ps/N} \]

Here $s/N$ can be interpreted as the selection coefficient of $A_1$ relative to $A_2$. 

Suppose that we have sampled two individuals at random from the population.
If $p$ and $q$ denote the frequencies of $A_1$ and $A_2$ in the preceding generation, then the
probability that the sampled individuals have a common ancestor in that generation is:

$$
P\{\text{C.A.} \} = P\{\text{C.A. is an } A_1 \text{ parent} \} + P\{\text{C.A. is an } A_2 \text{ parent} \}$$
Suppose that we have sampled two individuals at random from the population. If \( p \) and \( q \) denote the frequencies of \( A_1 \) and \( A_2 \) in the preceding generation, then the probability that the sampled individuals have a common ancestor in that generation is:

\[
P\{\text{C.A.}\} = P\{\text{C.A. is an } A_1 \text{ parent}\} + P\{\text{C.A. is an } A_2 \text{ parent}\}
\]

\[
= Np \left( \frac{1}{N} \frac{1 + s/N}{1 + ps/N} \right)^2 + Nq \left( \frac{1}{N} \frac{1}{1 + ps/N} \right)^2
\]

where the remainder term, \( O(N^{-3}) \), is uniformly bounded in \( p \in [0,1] \).
Suppose that we have sampled two individuals at random from the population.

If \( p \) and \( q \) denote the frequencies of \( A_1 \) and \( A_2 \) in the preceding generation, then the probability that the sampled individuals have a common ancestor in that generation is:

\[
\begin{align*}
P\{\text{C.A.}\} & = P\{\text{C.A. is an } A_1 \text{ parent}\} + P\{\text{C.A. is an } A_2 \text{ parent}\} \\
& = Np \left( \frac{1}{N} \left( 1 + \frac{s}{N} \right) \right)^2 + Nq \left( \frac{1}{N} \left( 1 + \frac{ps}{N} \right) \right)^2 \\
& = \frac{p}{N} \left( 1 + \frac{qs}{N} + O(N^{-2}) \right) + \frac{q}{N} \left( 1 - \frac{ps}{N} + O(N^{-2}) \right) \\
& = \frac{1}{N} + O(N^{-3}),
\end{align*}
\]

where the remainder term, \( O(N^{-3}) \), is uniformly bounded in \( p \in [0, 1] \).
An Erroneous Calculation: If $\tau^{(N)}$ is the time to the most recent common ancestor for the sample, then seemingly

$$P\{\tau^{(N)} > t + 1\} = P\{\tau^{(N)} > t + 1|\tau^{(N)} > t\}P\{\tau^{(N)} > t\}$$

If true, this would imply that

$$\lim_{N \to \infty} P\{\tau^{(N)} > Nt\} = e^{-t},$$

which appears to show that $\tau^{(N)}$ converges in distribution to an exponentially distributed random variable with mean $1$. 
An Erroneous Calculation: If $\tau^{(N)}$ is the time to the most recent common ancestor for the sample, then seemingly

$$\mathbb{P}\{\tau^{(N)} > t + 1\} = \mathbb{P}\{\tau^{(N)} > t + 1 | \tau^{(N)} > t\} \mathbb{P}\{\tau^{(N)} > t\}$$

$$= \left(1 - \frac{1}{N} + O(N^{-3})\right) \mathbb{P}\{\tau^{(N)} > t\} \quad \text{(by the Markov property)}$$
An Erroneous Calculation: If $\tau^{(N)}$ is the time to the most recent common ancestor for the sample, then seemingly

$$
\mathbb{P}\{\tau^{(N)} > t + 1\} = \mathbb{P}\{\tau^{(N)} > t + 1|\tau^{(N)} > t\}\mathbb{P}\{\tau^{(N)} > t\}
$$

$$
= \left(1 - \frac{1}{N} + O(N^{-3})\right)\mathbb{P}\{\tau^{(N)} > t\} \quad \text{(by the Markov property)}
$$

$$
= \left(1 - \frac{1}{N} + O(N^{-3})\right)^{t+1} \quad \text{(by induction)}.
$$
**An Erroneous Calculation:** If $\tau^{(N)}$ is the time to the most recent common ancestor for the sample, then seemingly

\[
P\{\tau^{(N)} > t + 1\} = P\{\tau^{(N)} > t + 1 | \tau^{(N)} > t\}P\{\tau^{(N)} > t\}
\]

\[
= \left(1 - \frac{1}{N} + O(N^{-3})\right) P\{\tau^{(N)} > t\} \quad \text{(by the Markov property)}
\]

\[
= \left(1 - \frac{1}{N} + O(N^{-3})\right)^{t+1} \quad \text{(by induction)}.
\]

If true, this would imply that

\[
\lim_{N \to \infty} P\{\tau^{(N)} > Nt\} = \lim_{N \to \infty} \left(1 - \frac{1}{N} + O(N^{-3})\right)^{Nt}
\]

\[
= e^{-t},
\]

which appears to show that $\tau^{(N)}$ converges in distribution to an exponentially distributed random variable with mean 1.
To see where this reasoning goes wrong, let $n'_1$ and $n'_2$ denote the number of ancestors with genotype $A_1$ or $A_2$ in the preceding generation, and let $n' = n'_1 + n'_2$.

Then

$$
\mathbb{P}\left\{ n'_1 = 2, n'_2 = 0 \mid n' = 2 \right\} = \mathbb{P}\left\{ \text{there are two distinct } A_1 \text{ ancestors} \mid n' = 2 \right\}
$$
To see where this reasoning goes wrong, let $n'_1$ and $n'_2$ denote the number of ancestors with genotype $A_1$ or $A_2$ in the preceding generation, and let $n' = n'_1 + n'_2$.

Then

$$
\mathbb{P}\{n'_1 = 2, n'_2 = 0 \mid n' = 2\} = \mathbb{P}\{\text{there are two distinct } A_1 \text{ ancestors} \mid n' = 2\}
$$

$$
= \frac{\mathbb{P}\{\text{the sample has two distinct } A_1 \text{ ancestors}\}}{\mathbb{P}\{\text{the sample has two distinct ancestors}\}}
$$
To see where this reasoning goes wrong, let \( n'_1 \) and \( n'_2 \) denote the number of ancestors with genotype \( A_1 \) or \( A_2 \) in the preceding generation, and let \( n' = n'_1 + n'_2 \).

Then

\[
\mathbb{P}\{n'_1 = 2, n'_2 = 0 \mid n' = 2\} = \mathbb{P}\{\text{there are two distinct } A_1 \text{ ancestors} \mid n' = 2\}
\]

\[
= \frac{\mathbb{P}\{\text{the sample has two distinct } A_1 \text{ ancestors}\}}{\mathbb{P}\{\text{the sample has two distinct ancestors}\}}
\]

\[
= \frac{1}{1 - \frac{1}{N} + O(N^{-3})} Np(Np - 1) \left( \frac{1}{N} \left( 1 + \frac{s}{N} \right) \right)^2
\]

\[
= p^2 - \frac{pq}{N} + \frac{2p^2qs}{N} + O(sN^{-2}).
\]
Similar calculations show that:

\[
\mathbb{P}\{n_1' = 2, n_2' = 0 \mid n' = 2\} = p^2 - \frac{pq}{N} + \frac{2p^2qs}{N} + O(sN^{-2})
\]

\[
\mathbb{P}\{n_1' = 1, n_2' = 1 \mid n' = 2\} = 2pq + \frac{2pq}{N} + \frac{2pq(q - p)s}{N} + O(sN^{-2})
\]

\[
\mathbb{P}\{n_1' = 0, n_2' = 2 \mid n' = 2\} = q^2 - \frac{pq}{N} - \frac{2pq^2s}{N} + O(sN^{-2}).
\]
Similar calculations show that:

\[
P\{n'_1 = 2, n'_2 = 0 \mid n' = 2\} = p^2 - \frac{pq}{N} + \frac{2p^2qs}{N} + O(sN^{-2})
\]

\[
P\{n'_1 = 1, n'_2 = 1 \mid n' = 2\} = 2pq + \frac{2pq}{N} + \frac{2pq(q - p)s}{N} + O(sN^{-2})
\]

\[
P\{n'_1 = 0, n'_2 = 2 \mid n' = 2\} = q^2 - \frac{pq}{N} - \frac{2pq^2s}{N} + O(sN^{-2}).
\]

In contrast, the distribution of the composition of a random sample containing two individuals is:

\[
P\{A_1 + A_1\} = p \frac{Np - 1}{N - 1} = p^2 - \frac{pq}{N}
\]

\[
P\{A_1 + A_2\} = 2pq + \frac{2pq}{N}
\]

\[
P\{A_2 + A_2\} = q^2 - \frac{pq}{N}.
\]
Conclusions:

- In a neutral model ($s = 0$), the distribution of the types of the ancestors of a random sample is the same as the distribution of the types of a random sample from the ancestral population.

- However, with selection, these two distributions differ. In particular, selectively favorable types are over-represented in the ancestors relative to the population frequencies.

- This is where the induction argument on the preceding slide breaks down, i.e.,

  $\mathbb{P}\{\tau^{(N)} > t + 1 | \tau^{(N)} > t\} \neq \mathbb{P}\{\tau^{(N)} > 1\}$

- These discrepancies are of order $O(N^{-1})$ and so cannot be neglected.
**Problem:** Without exchangeable dynamics, the genealogy of a random sample is no longer a Markov process.

- Lineages conditioned not to coalesce are no longer a random sample from the population.
- Coalescent rates of a non-random sample are frequency-dependent.
**Problem:** Without exchangeable dynamics, the genealogy of a random sample is no longer a Markov process.

- Lineages conditioned not to coalesce are no longer a random sample from the population.
- Coalescent rates of a non-random sample are frequency-dependent.

**Remedy:** Augment the process with sufficient information to restore the Markov property. For models with selection, this has been done in two ways:

- In the **Ancestral Selection Graph**, the additional information is carried by a set of virtual particles that account for selective deaths (Krone & Neuhauser 1997; Donnelly & Kurtz 1999).
- Alternatively, it suffices to record the types of the lineages and the genetic composition of the ancestral population (Hudson et al. 1988).
The second approach leads to a **structured coalescent**:

- We can think of the population as being subdivided into different genetic backgrounds.
- Even when there are selective differences between genotypes, individuals that have the same genotype are **exchangeable** - the demography does not depend on how these individuals are ordered.
- Because of this exchangeability, coalescence within backgrounds can be described by a simple modification of Kingman’s coalescent.

The structured coalescent can be represented as a Markov process

\[ \Gamma(t) = (n_1(t), n_2(t), p(t)), \]

where \( n_i(t) \) is the number of ancestral lineages in the \( A_i \) background and \( p(t) \) is the ancestral frequency of \( A_1 \) at time \( t \) in the past.
However, to describe this process, we will need to understand how the allele frequencies evolve when:

- $N$ is very large and time is measured in units of $N$ generations.
- Time is followed from the present into the past.

This motivates our consideration of diffusion approximations, which we turn to next.
**Discrete-time Markov chains:**

Let $E = \{e_1, \cdots, e_n\}$ be a finite set. We say that a sequence of $E$-valued random variables, $(X_n : n \in \mathbb{N})$, is a **discrete-time Markov chain** if for every $n, k \geq 0$ and every $A \subset E$:

$$
\mathbb{P}\{X_{n+k} \in A | (X_0, \cdots, X_n)\} = \mathbb{P}\{X_{n+k} \in A | X_n\}.
$$
Discrete-time Markov chains:

Let $E = \{e_1, \cdots, e_n\}$ be a finite set. We say that a sequence of $E$-valued random variables, $(X_n : n \in \mathbb{N})$, is a discrete-time Markov chain if for every $n, k \geq 0$ and every $A \subset E$:

$$
\mathbb{P}\{X_{n+k} \in A | (X_0, \cdots, X_n)\} = \mathbb{P}\{X_{n+k} \in A | X_n\}.
$$

Any discrete-time Markov chain can be characterized by its transition matrix $P = (p_{ij})$, where

$$
p_{ij} = \mathbb{P}\{X_{n+1} = e_j | X_n = e_i\}.
$$

The multi-step transition probabilities can be found by raising $P$ to the appropriate power:

$$
\mathbb{P}\{X_{n+k} = e_j | X_n = e_i\} = (P^k)_{ij}.
$$
Example: The neutral Wright-Fisher model

If \( p(t) \) denotes the frequency of \( A_1 \) in generation \( t \), then conditional on \( p(t) = p \), the distribution of \( p(t+1) \) is

\[
Np(t + 1) \sim \text{Binomial}(N, p).
\]

Consequently, the transition probabilities for this chain are:

\[
P\{ p(1) = p' \mid p(0) = p \} = \binom{N}{Np'} p^{Np'} (1 - p)^{N(1-p')}.\]

Remark: This model is easy to describe and simulate, but is mathematically intractable.
Using the properties of the binomial distribution, we can show that:

\[ \mathbb{E}_p[p(1) - p] = 0 \]
\[ \mathbb{E}_p[(p(1) - p)^2] = \frac{p(1 - p)}{N} \]
\[ \mathbb{E}_p[(p(1) - p)^e] = O(N^{-2}) \quad \text{if } e \geq 3. \]

In other words, when \( N \) is large, the fluctuations of the Wright-Fisher process tend to be small, with jumps typically of order \( N^{-1/2} \).
For example, notice that the expected ‘heterozygosity’ decays geometrically:

\[
\mathbb{E}_p \left[ p(1)(1 - p(1)) \right] = \mathbb{E}_p \left[ p(1) - p(1)^2 \right] \\
= p - p^2 - \frac{p(1-p)}{N} \\
= \left(1 - \frac{1}{N}\right)p(1-p).
\]
For example, notice that the expected ‘heterozygosity’ decays geometrically:

$$\mathbb{E}_p \left[ p(1)(1 - p(1)) \right] = \mathbb{E}_p \left[ p(1) - p(1)^2 \right]$$

$$= p - p^2 - \frac{p(1 - p)}{N}$$

$$= \left(1 - \frac{1}{N}\right)p(1 - p).$$

Similarly, the Markov property can be used to show that:

$$\mathbb{E}_p \left[ p(t)(1 - p(t)) \right] = \left(1 - \frac{1}{N}\right)^{t-1} p(1 - p) \approx e^{-t/N} p(1 - p).$$

In other words, genetic drift tends to remove variation from the population and the expected heterozygosity decays geometrically at rate \((1 - 1/N)\).
These results suggest that allele frequency fluctuations occur on a time scale that is naturally measured in units of $N$ generations. This can also be deduced from the behavior of the transition probabilities:

$$P^{(N)}(\tau, p, A) \equiv \mathbb{P}_p\{p(N\tau) \in A\}$$

where $A \subset [0, 1]$.

In principle, the transition probabilities can be calculated exactly by raising the transition matrix to an appropriate power. However, this is cumbersome unless $N$ is very small.
These results suggest that **allele frequency fluctuations occur on a time scale that is naturally measured in units of $N$ generations.** This can also be deduced from the behavior of the transition probabilities:

$$P^{(N)}(\tau, p, A) \equiv \mathbb{P}_p\{p(N\tau) \in A\}$$

where $A \subset [0, 1]$.

In principle, the transition probabilities can be calculated exactly by raising the transition matrix to an appropriate power. However, this is cumbersome unless $N$ is very small.

Alternatively, suppose that the limit

$$P(\tau, p, A) = \lim_{N \to \infty} P^{(N)}(\tau, p, A) = \lim_{N \to \infty} \mathbb{P}_p\{p(N\tau) \in A\}$$

is well-defined and is a smooth function of $p$ and $\tau$. 
Then,

\[ P(\tau, p, A) \approx \mathbb{P}_p \{ p(N\tau) \in A \} \]
Then,

\[ P(\tau, p, A) \approx \mathbb{P}_p \{ p(N\tau) \in A \} \]

\[ = \mathbb{E}_p \left[ \mathbb{P} \{ p(N\tau) \in A | p(1) \} \right] \quad \text{(a standard trick)} \]
Then,

\[ P(\tau, p, A) \approx \mathbb{P}_p \{ p(N\tau) \in A \} \]

\[ = \mathbb{E}_p \left[ \mathbb{P} \{ p(N\tau) \in A | p(1) \} \right] \quad \text{(a standard trick)} \]

\[ = \mathbb{E}_p \left[ P^{(N)}(\tau - 1/N, p(1), A) \right] \quad \text{(by the Markov property)} \]
Then,

\[ P(\tau, p, A) \approx P_p\{p(N\tau) \in A\} \]

\[ = \mathbb{E}_p\left[ \mathbb{P}\{p(N\tau) \in A|p(1)\} \right] \quad \text{(a standard trick)} \]

\[ = \mathbb{E}_p\left[ P^{(N)}(\tau - 1/N, p(1), A) \right] \quad \text{(by the Markov property)} \]

\[ \approx \mathbb{E}_p\left[ P(\tau - 1/N, p(1), A) \right] \]
Then,

\[ P(\tau, p, A) \approx \mathbb{P}_p\{p(N\tau) \in A\} \]

\[ = \mathbb{E}_p\left[ \mathbb{P}\{p(N\tau) \in A|p(1)\} \right] \quad \text{(a standard trick)} \]

\[ = \mathbb{E}_p\left[ p(N)(\tau - 1/N, p(1), A) \right] \quad \text{(by the Markov property)} \]

\[ \approx \mathbb{E}_p\left[ p(\tau - 1/N, p(1), A) \right] \]

\[ = \mathbb{E}_p\left[ P(\tau - 1/N, p, A) + \partial_p P(\tau - 1/N, p, A) \delta p \right. \]

\[ + \frac{1}{2} \partial_{pp} P(\tau - 1/N, p, A)(\delta p)^2 + O((\delta p)^3) \]

\[ \text{(with } \delta p = p(1) - p \text{ typically small)} \]
Then,

\[
P(\tau, p, A) \approx \mathbb{P}_p\{p(N\tau) \in A\}
\]

\[
= \mathbb{E}_p\left[\mathbb{P}\{p(N\tau) \in A|p(1)\}\right] \quad \text{(a standard trick)}
\]

\[
= \mathbb{E}_p\left[P^{(N)}(\tau - 1/N, p(1), A)\right] \quad \text{(by the Markov property)}
\]

\[
\approx \mathbb{E}_p\left[P(\tau - 1/N, p(1), A)\right]
\]

\[
= \mathbb{E}_p\left[P(\tau - 1/N, p, A) + \partial_p P(\tau - 1/N, p, A)\delta p + \frac{1}{2} \partial_{pp} P(\tau - 1/N, p, A)(\delta p)^2 + O((\delta p)^3)\right]
\]

(with \(\delta p = p(1) - p\) typically small)

\[
= P(\tau - 1/N, p, A) + \frac{p(1 - p)}{2N} \partial_{pp} P(\tau - 1/N, p, A) + O(N^{-2}).
\]
This can be rearranged to:

\[
\frac{P(\tau, p, A) - P(\tau - 1/N, p, A)}{1/N} = \frac{1}{2} p(1 - p) \partial_{pp} P(\tau - 1/N, p, A) + O(N^{-1}).
\]
This can be rearranged to:

\[
P(\tau, p, A) - P(\tau - 1/N, p, A) = \frac{1}{2} p(1 - p) \partial_{pp} P(\tau - 1/N, p, A) + O(N^{-1}).
\]

Then, letting \( N \to \infty \), we obtain the following initial value problem

\[
\begin{aligned}
\partial_t P(\tau, p, A) &= \frac{1}{2} p(1 - p) \partial_{pp} P(\tau, p, A) \\
P(0, p, A) &= 1 \text{ if } p \in A.
\end{aligned}
\]

Thus, if we can find a solution to (*), then we can approximate the transition probabilities of the Wright-Fisher process by undoing the scaling:

\[
P_p\{p(N\tau) \in A\} \approx P(\tau, p, A).
\]
The following figure shows a series of simulations of the Wright-Fisher model for 100 generations for \( N = 10 \) (blue), 100 (red), 1000 (orange), and 10,000 (green).

Notice that both the size of the fluctuations and the total change in \( p \) over 100 generations decrease as the population size is increased.
A different picture emerges if we plot each sample path against time measured in units of \( N \) generations. Here the total rescaled time is \( \tau = 1 \).

In this case, the typical jump sizes shrink as \( N \) increases, but the ‘roughness’ of the paths increases. This suggests that the processes \((p^N(t) : t \geq 0)\) themselves converge to a stochastic process with continuous sample paths.
One-dimensional Diffusion Processes

Informal Definition: A Markov process \( X = (X(t) : t \geq 0) \) is a one-dimensional diffusion process on the set \( E = [l, r] \subset \mathbb{R} \) if:

- \( X \) has continuous sample paths with values in \( E \);
- The transition probabilities \( P(t, x, A) = \mathbb{P}_x\{X(t) \in A\} \) satisfy the initial value problem

\[
\frac{\partial_t P(t, x, A)}{2} = a(x) \frac{\partial_{xx} P(t, x, A)}{2} + b(x) \frac{\partial_x P(t, x, A)}{2} \quad \text{for } t > 0, \quad x \in E
\]

\[
P(0, x, A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}
\]

Remark: The function \( a(x) \) must be non-negative in \( E \).
Interpretation of the coefficients $a(x)$ and $b(x)$

- The **infinitesimal drift** $b(x)$ determines the expected change in a small increment of $X$ starting at $x$:

$$b(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x [X(t) - x]$$

- The **infinitesimal variance** $a(x)$ determines the variance of a small increment of $X$ starting at $x$:

$$a(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x [(X(t) - x)^2]$$

**Caveat:** There is an unfortunate overlap of terminology. The **infinitesimal drift** of a diffusion describes its expected change. In contrast, **genetic drift** refers to stochastic fluctuations in allele frequencies that are described by the **infinitesimal variance** $p(1-p)$ of the Wright-Fisher diffusion.
**Example: Brownian motion and random walks**

Suppose that \((Z_n : n \geq 0)\) is a collection of IID random variables with distribution

\[
P\{Z_n = 1\} = P\{Z_n = -1\} = \frac{1}{2},
\]

and let

\[
S_N = \sum_{n=1}^{N} Z_n
\]

be the \(n'\)th partial sum. The process \((S_n : n \geq 0)\) is a Markov chain and is called a simple random walk.
Example: Brownian motion and random walks

Suppose that \((Z_n : n \geq 0)\) is a collection of IID random variables with distribution

\[
P\{Z_n = 1\} = P\{Z_n = -1\} = \frac{1}{2},
\]

and let

\[
S_N = \sum_{n=1}^{N} Z_n
\]

be the \(n\)'th partial sum. The process \((S_n : n \geq 0)\) is a Markov chain and is called a simple random walk.

Recall that the Central Limit Theorem tells us that when \(N\) is large, the distribution of

\[
\frac{1}{\sqrt{N}} S_N
\]

is approximately normal with mean 0 and variance 1.
Now, suppose that we define a related sequence of continuous-time stochastic processes, $W^{(N)} = (W_t^{(N)} : t \geq 0)$ by setting

$$W_t^{(N)} = N^{-1/2} S_{\lfloor Nt \rfloor};$$

here $\lfloor Nt \rfloor$ denotes the greatest integer less than $Nt$ (a real number).
Now, suppose that we define a related sequence of continuous-time stochastic processes, $W^{(N)} = (W_t^{(N)} : t \geq 0)$ by setting

$$W_t^{(N)} = N^{-1/2} S_{\lfloor Nt \rfloor} ;$$

here $\lfloor Nt \rfloor$ denotes the greatest integer less than $Nt$ (a real number).

Notice that the increments are also sums of IID random variables,

$$W_{t+s}^{(N)} - W_t^{(N)} = N^{-1/2} \left( S_{\lfloor N(t+s) \rfloor} - S_{\lfloor Nt \rfloor} \right)$$

$$= N^{-1/2} \left( \sum_{n=1}^{\lfloor N(t+s) \rfloor} Z_n - \sum_{n=1}^{\lfloor Nt \rfloor} Z_n \right)$$

$$= s^{1/2} (Ns)^{-1/2} \sum_{n=\lfloor Nt \rfloor+1}^{\lfloor N(t+s) \rfloor} Z_n$$

$$\sim \mathcal{N}(0; s).$$
From this decomposition, we can deduce that when \( N \) is large the process \( W^{(N)} \) has the following properties:

- If \( 0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \), then the increments
  \[ W^{(N)}(t_2) - W^{(N)}(t_1) \quad \text{and} \quad W^{(N)}(t_4) - W^{(N)}(t_3) \]
  are independent;

- The distribution of each increment \( W^{(N)}(t + s) - W^{(N)}(t) \) is approximately normal with mean 0 and variance \( s \);

- The sample paths of \( W^{(N)} \) are ‘nearly’ continuous, i.e., we are unlikely to observe any large jumps by the process over short time intervals.
In fact, as $N \to \infty$, it can be shown that the sequence of processes $W^{(N)}$ converges to a continuous-time Markov process $W = (W(t); t \geq 0)$ with the following properties:

- If $0 \leq t_1 \leq t_2 \leq t_3$, then the increments
  
  $$W(t_2) - W(t_1) \quad \text{and} \quad W(t_4) - W(t_3)$$

  are independent;

- The distribution of each increment $W(t + s) - W(t)$ is exactly normal with mean 0 and variance $s$;

- $W$ has continuous sample paths.

The process $W$ is called a one-dimensional Brownian motion.
Because the increments are normally-distributed, we know that the transition probabilities of the Brownian motion can be written as:

\[
P(t, x, A) = \mathbb{P}_x\{W(t) \in A\} = \int_A \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/(2t)} dy.
\]

**Exercise:** Show that the transition probabilities satisfy the following equation:

\[
\partial_t P(t, x, A) = \frac{1}{2} \partial_{xx} P(t, x, A).
\]

In other words, Brownian motion is a one-dimensional diffusion with infinitesimal variance \(a(x) = 1\) and infinitesimal drift \(b(x) = 0\). In fact, it can be regarded as the canonical example of a diffusion process.
Stochastic Differential Equations

Suppose that \( X \) is a one-dimensional diffusion with infinitesimal variance \( a(x) \) and infinitesimal drift \( b(x) \). Then \( X \) is a solution to the following stochastic differential equation:

\[
dX_t = b(X_t)dt + \sqrt{a(X_t)}dB_t,
\]

where \( dB_t \) is the stochastic differential of a Brownian motion \( B_t \).

Thus, the increments of \( X \) over short time intervals are approximately equal to:

\[
X_{t+\delta} \approx X_t + b(X_t)\delta t + \sqrt{a(X_t)}(B_{t+\delta} - B_t)
\]
Multidimensional Diffusion Processes

Informal Definition: A Markov process $X = (X(t) : t \geq 0)$ is a multi-dimensional diffusion process on a set $E \subset \mathbb{R}^n$ if:

- $X$ has continuous sample paths with values in $E$;
- The transition probabilities

$$P(t, x, A) = \mathbb{P}_x\{X(t) \in A\}$$

satisfy the initial value problem

$$\partial_t P(t, x, A) = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{x_i x_j} P(t, x, A) + \sum_{i=1}^n b_i(x) \partial_{x_i} P(t, x, A)$$

$$P(0, x, A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$
Interpretation of the coefficients $a(x)$ and $b(x)$

- The infinitesimal drift $b(x) = (b_1(x), \cdots, b_n(x))$ satisfies:
  \[
  b_i(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[X_i(t) - x_i]
  \]

- The infinitesimal variance-covariance matrix $a(x) = (a_{ij}(x))$ satisfies:
  \[
  a_{ij}(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[(X_i(t) - x_i)(X_j(t) - x_j)]
  \]

**Remark:** The variance-covariance matrix must be non-negative definite, i.e., for every $x \in E$ and any vector $v = (v_1, \cdots, v_n)$,

\[
\sum_{i,j=1}^{n} a_{ij}(x)v_i v_j \geq 0.
\]
**Diffusion Approximations for Discrete-time Markov Chains**

**Theorem:** Suppose that $X^{(N)} = (X^{(N)}(n) : n \geq 0)$ is a sequence of discrete-time Markov chains and let $\epsilon_N$ be a sequence of positive numbers tending to 0 such that the following limits exist:

\[
\lim_{N \to \infty} \epsilon_N^{-1} \mathbb{E}_x [X^{(N)}(1) - x] = b(x)
\]

\[
\lim_{N \to \infty} \epsilon_N^{-1} \mathbb{E}_x [(X^{(N)}(1) - x)^2] = a(x)
\]

\[
\lim_{N \to \infty} \epsilon_N^{-1} \mathbb{E}_x [(X^{(N)}(1) - x)^n] = 0 \text{ if } n \geq 3.
\]

Then, the interpolated processes $(X^{(N)}(\lfloor \epsilon_N^{-1} t \rfloor) : t \geq 0)$ converge to the diffusion process $X$ with infinitesimal drift $b(x)$ and infinitesimal variance $a(x)$.

**Intuition:** When $N$ is large, $X(N)$ makes many small jumps between times $\epsilon_N^{-1} t$ and $\epsilon_N^{-1}(t + s)$. 
Example: A Wright-Fisher model with selection and mutation

Assumptions:

- \( N \) haploid individuals (constant population size).
- Alleles \( A_1, A_2 \) with frequencies \( p \) and \( q \).
- Each individual alive in generation \( t + 1 \) chooses its parent from the preceding generation independently and with replacement with the following probabilities:
  
  \[
  \begin{align*}
  \text{each } A_1\text{-type individual:} & \quad \frac{1}{N} \frac{1 + s(p)/N}{1 + ps(p)/N} \\
  \text{each } A_2\text{-type individual:} & \quad \frac{1}{N} \frac{1}{1 + ps(p)/N}.
  \end{align*}
  \]

- Note that the selection coefficient \( s(p)/N \) may be frequency-dependent.
- Following selection, each \( A_1 \)-type individual mutates to \( A_2 \) with probability \( \mu_2 = \theta_2/N \), while each \( A_2 \)-type individual mutates to \( A_1 \) with probability \( \mu_1 = \theta_1/N \).
Let $p^{(N)} = (p^{(N)}(n) : n \geq 0)$ be the Wright-Fisher process and set $\epsilon_N = N^{-1}$. Then,

$$\lim_{N \to \infty} N \mathbb{E}_p[p^N(1) - p] = (1 - p)\theta_1 - p\theta_2 + s(p)p(1 - p)$$

$$\lim_{N \to \infty} N \mathbb{E}_p[(p^N(1) - p)^2] = p(1 - p)$$

$$\lim_{N \to \infty} N \mathbb{E}_p[(p^N(1) - p)^n] = 0 \text{ if } n \geq 3.$$ 

**Exercise:** Verify these limits using the properties of the binomial distribution.

Using the theorem on the previous slide, we know that the interpolated processes $(p^N([Nt]) : t \geq 0)$ converge to the diffusion process with

$$a(p) = p(1 - p) \quad \text{(infinitesimal variance)}$$

$$b(p) = (1 - p)\theta_1 - p\theta_2 + s(p)p(1 - p) \quad \text{(infinitesimal drift)}.$$
The Kolmogorov Forward Equation

As we saw with Brownian motion, it is often the case that the transition probabilities of a diffusion process have densities:

\[ \mathbb{P}_x \{ X(t) \in A \} = \int_A p(t, x, y) dy. \]
The Kolmogorov Forward Equation

As we saw with Brownian motion, it is often the case that the transition probabilities of a diffusion process have densities:

\[ \mathbb{P}_x \{ X(t) \in A \} = \int_A p(t, x, y) dy. \]

Then, provided that \( a(x) \) and \( b(x) \) are smooth, it can be shown that the transition densities satisfy a partial differential equation involving the variable \( y \) which is called the Kolmogorov forward equation:

\[
\partial_t p(t, x, y) = \frac{1}{2} \partial_{yy} (a(y) p(t, x, y)) - \partial_y (b(y) p(t, x, y))
\]

\[ p(0, x, y) = \delta_x(dy) \]

where the identity \( p(0, x, y) = \delta_x(dy) \) means:

\[
\lim_{t \to 0} \int_E f(y) p(t, x, y) dy = f(x).
\]
Stationary Distributions

**Definition:** Let $X = (X_t : t \geq 0)$ be a Markov process on $E$. A **stationary distribution** for $X$ is a probability distribution $\pi$ on $E$ such that if $X_0$ has distribution $\pi$, then $X_t$ has distribution $\pi$ for all $t \geq 0$. 

This property can also be expressed in terms of expectations. If $f$ is continuous and $\pi$ is a stationary distribution for $X$, then 

$$E_\pi f(X_0) = E_\pi f(X_t),$$

i.e.,

$$\int f(x) \pi(dx) = \int E_x f(X_t) \pi(dx)$$

holds for all $t \geq 0$. 

Jay Taylor ()
Diffusions and Structured Coalescents
2009 36 / 160
Stationary Distributions

Definition: Let $X = (X_t : t \geq 0)$ be a Markov process on $E$. A stationary distribution for $X$ is a probability distribution $\pi$ on $E$ such that if $X_0$ has distribution $\pi$, then $X_t$ has distribution $\pi$ for all $t \geq 0$.

This property can also be expressed in terms of expectations. If $f$ is continuous and $\pi$ is a stationary distribution for $X$, then

$$\mathbb{E}_\pi[f(X_0)] = \mathbb{E}_\pi[f(X_t)]$$

i.e.

$$\int_E f(x) \pi(dx) = \int_E \mathbb{E}_x[f(X_t)] \pi(dx)$$

holds for all $t \geq 0$. 
Stationary Distributions of Diffusion Processes

Let $X = (X_t : t \geq 0)$ be a diffusion process on $[l, r]$ with infinitesimal variance $a(x)$ and infinitesimal drift $b(x)$, and suppose that:

- $X$ has transition density $p(t, x, y)$.
- $X$ has a unique stationary distribution $\pi(x)dx$.
- For every $x \in [l, r]$, 
  \[ \lim_{t \to \infty} p(t, x, y) = \pi(y). \]

Remark: The third property is an ergodic condition. It says that if we wait long enough, the distribution of the process will approach the stationary distribution.
Under these conditions, we can use the forward equation to calculate:

\[ 0 = \partial_t \pi(y) = \partial_t \lim_{t \to \infty} p(t, x, y) \]
Under these conditions, we can use the forward equation to calculate:

\[
0 = \partial_t \pi(y) \\
= \partial_t \lim_{t \to \infty} p(t, x, y) \\
= \lim_{t \to \infty} \partial_t p(t, x, y) \quad \text{(if we can interchange the limit and differentiation)}
\]
Under these conditions, we can use the forward equation to calculate:

\[
0 = \partial_t \pi(y)
\]

\[
= \partial_t \lim_{t \to \infty} p(t, x, y)
\]

\[
= \lim_{t \to \infty} \partial_t p(t, x, y) \quad \text{(if we can interchange the limit and differentiation)}
\]

\[
= \lim_{t \to \infty} \left( \frac{1}{2} \partial_{yy} (a(y)p(t, x, y)) - \partial_y (b(y)p(t, x, y)) \right)
\]
Under these conditions, we can use the forward equation to calculate:

\[
0 = \partial_t \pi(y) \\
= \partial_t \lim_{t \to \infty} p(t, x, y) \\
= \lim_{t \to \infty} \partial_t p(t, x, y) \quad \text{(if we can interchange the limit and differentiation)} \\
= \lim_{t \to \infty} \left( \frac{1}{2} \partial_{yy}(a(y)p(t, x, y)) - \partial_y(b(y)p(t, x, y)) \right) \\
= \frac{1}{2} \partial_{yy}(a(y)\pi(y)) - \partial_y(b(y)\pi(y))
\]

(again assuming that we can interchange the limit and differentiation).

In other words, \( \pi(x) \) is a stationary solution to the forward equation

\[
\frac{1}{2} (a(x)\pi(x))'' - (b(x)\pi(x))' = 0.
\]
This can be integrated to give
\[ \frac{1}{2} (a(x)\pi(x))' - (b(x)\pi(x)) = C, \]
where $C$ is a constant.
This can be integrated to give

\[
\frac{1}{2} (a(x)\pi(x))' - (b(x)\pi(x)) = C,
\]

where \(C\) is a constant.

It can be shown that \(\pi(x)\) is integrable \((\int \pi(x)dx < \infty)\) only if \(C = 0\) (zero flux condition).

Consequently, \(\pi(x)\) satisfies the first-order linear equation

\[
\frac{1}{2} a(x)\pi'(x) + \left(\frac{1}{2} a'(x) - b(x)\right) \pi(x) = 0,
\]

which can be rewritten as

\[
\frac{\pi'(x)}{\pi(x)} = \frac{2b(x)}{a(x)} - \frac{a'(x)}{a(x)}.
\]
Both sides of this equation can be integrated to obtain

\[ \ln(\pi(x)) = 2 \int_c^x \frac{b(y)}{a(y)} dy - \ln(a(x)) + C, \]

where \( C \) is a (new) constant of integration and \( c \) is an arbitrary point in \((l, r)\).
Both sides of this equation can be integrated to obtain
\[
\ln(\pi(x)) = 2 \int_c^x \frac{b(y)}{a(y)} dy - \ln(a(x)) + C,
\]
where \( C \) is a (new) constant of integration and \( c \) is an arbitrary point in \((l, r)\).

**Key Result:** The density of the stationary distribution (if it exists) is
\[
\pi(x) = \frac{1}{C} \frac{1}{a(x)} \exp \left( 2 \int_c^x \frac{b(y)}{a(y)} dy \right),
\]
where the normalizing constant \( C < \infty \) must be chosen (if possible) so that
\[
\int_l^r \pi(x) dx = 1.
\]

**Caveat:** If \( C \) cannot be so chosen, then any stationary distribution of the diffusion process does not have a density on \([l, r]\).
Example: Since the variance and drift coefficients of the (neutral) Wright-Fisher diffusion are $a(p) = p(1-p)$ and $b(p) = \theta_1(1-p) - \theta_2 p$, the density of the stationary distribution of this process is

$$
\pi(p) = \frac{1}{C} \frac{1}{p(1-p)} \exp \left\{ 2 \int_c^p \frac{\left( \theta_1(1-q) - \theta_2 q \right)}{q(1-q)} dq \right\} 
$$

$$
= \frac{1}{C} p^{2\theta_1-1}(1-p)^{2\theta_2-1}.
$$

To complete the calculation, notice that

$$
C = \int_0^1 p^{2\theta_1-1}(1-p)^{2\theta_2-1} dp = \beta(2\theta_1, 2\theta_2) < \infty,
$$

is just the Beta function with arguments $2\theta_1$ and $2\theta_2$. This is finite as long as $\theta_1$ and $\theta_2$ are both positive.
Thus, the stationary distribution of the Wright-Fisher diffusion with mutation is just the Beta distribution with parameters $2\theta_1$ and $2\theta_2$:

$$
\pi(p) = \frac{1}{\beta(2\theta_1, 2\theta_2)} p^{2\theta_1-1}(1 - p)^{2\theta_2-1}
$$

$$
= \frac{1}{\beta(2N\mu_1, 2N\mu_2)} p^{2N\mu_1-1}(1 - p)^{2N\mu_2-1}
$$

The shape of the distribution reflects the competing effects of genetic drift, which tends to remove variation from the population, and mutation, which creates variation.
Characterization of Markov Processes

Suppose that $X = (X_t : t \geq 0)$ is a continuous-time Markov process with values in a set $E$. $X$ can be described in several ways.

1. Transition Function

The transition function of a Markov process is defined by the equation

$$P(t, x, A) = \mathbb{P}_x\{X_t \in A\} = \mathbb{P}\{X_t \in A | X_0 = x\},$$

where $t \geq 0$, $x \in E$ and $A \subset E$.

Remarks:

- A Markov process is fully determined by its transition function.
- Unfortunately, the transition function is usually not known explicitly.
2. Transition Semigroup

The transition semigroup of a Markov process is the collection of operators \((T_t : t \geq 0)\) defined by the equation

\[ T_t f(x) = \mathbb{E}_x[f(X_t)] = \int_E f(y) P(t, x, dy) \]

where \(x \in E\) and \(f : E \to \mathbb{R}\) is a bounded, continuous function.

Remarks:

- \(T_t\) is called an operator because it maps functions to functions: \(T_t f\) is a function from \(E\) into \(\mathbb{R}\).
- Notice that \(T_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)\).
- A Markov process is uniquely determined by its transition semigroup.
- Sadly, the transition semigroup also is usually not known explicitly.
3. **Infinitesimal Generator**

The **infinitesimal generator** of a Markov process $X$ is the operator defined by the equation

$$Gf(x) = \frac{d}{dt} T_t f(x) \bigg|_{t=0} = \lim_{t \to 0} \frac{\mathbb{E}_x [f(X_t)] - f(x)}{t},$$

provided the limit exists for all $x \in E$.

**Remarks:**

- The generator specifies the rate of change of the expected value of a function evaluated along sample paths of $X$ started at $x$.
- A Markov process is uniquely determined by its generator.
- Mercifully, the generator often can be found explicitly.
Continuous-time Markov chains

Let $E = \{e_1, \cdots, e_n\}$. We say that a sequence of $E$-valued random variables, $(X_t : t \geq 0)$, is a continuous-time Markov chain if for every $t, s \geq 0$ and every subset $A \subset E$:

$$
\mathbb{P}\{X_{t+s} \in A | X_u, 0 \leq u \leq t\} = \mathbb{P}\{X_{t+s} \in A | X_t\}.
$$

A continuous-time Markov chain can be characterized by its rate matrix, $Q = (q_{ij})$, where

$$
q_{ij} = \lim_{s \downarrow 0} \frac{1}{s} \mathbb{P}\{X_{t+s} = e_j | X_t = e_i\} \text{ if } j \neq i
$$

$$
q_{ii} = -\sum_{j \neq i} q_{ij}.
$$
**Example:** Suppose that $X$ is a continuous-time Markov chain on a finite set $E = \{e_1, \cdots, e_n\}$ with rate matrix $Q$. Then, for any function $f : E \to \mathbb{R}$,

$$Gf(e_i) = \lim_{t \to 0} \frac{1}{t} \left( E_X[f(X_t)] - f(e_i) \right)$$
Example: Suppose that $X$ is a continuous-time Markov chain on a finite set $E = \{e_1, \cdots, e_n\}$ with rate matrix $Q$. Then, for any function $f : E \to \mathbb{R}$,

$$Gf(e_i) = \lim_{t \to 0} \frac{1}{t} \left( E_X[f(X_t)] - f(e_i) \right)$$

$$= \lim_{t \to 0} \frac{1}{t} \sum_j P_{e_i} \{X_t = e_j\} (f(e_j) - f(e_i))$$
Example: Suppose that $X$ is a continuous-time Markov chain on a finite set $E = \{e_1, \cdots, e_n\}$ with rate matrix $Q$. Then, for any function $f : E \to \mathbb{R}$,

\[
Gf(e_i) = \lim_{t \to 0} \frac{1}{t} \left( E_x [f(X_t)] - f(e_i) \right) = \lim_{t \to 0} \frac{1}{t} \sum_j \mathbb{P}_{e_i} \{X_t = e_j\} (f(e_j) - f(e_i)) = \lim_{t \to 0} \frac{1}{t} \sum_{j \neq i} (q_{ij} t + o(t)) (f(e_j) - f(e_i))
\]
**Example:** Suppose that $X$ is a continuous-time Markov chain on a finite set $E = \{e_1, \cdots, e_n\}$ with rate matrix $Q$. Then, for any function $f : E \rightarrow \mathbb{R}$,

$$Gf(e_i) = \lim_{t \to 0} \frac{1}{t} \left( E_X[f(X_t)] - f(e_i) \right)$$

$$= \lim_{t \to 0} \frac{1}{t} \sum_j \mathbb{P}_{e_i} \{ X_t = e_j \} (f(e_j) - f(e_i))$$

$$= \lim_{t \to 0} \frac{1}{t} \sum_{j \neq i} (q_{ij} t + o(t)) (f(e_j) - f(e_i))$$

$$= \sum_{j \neq i} q_{ij} (f(e_j) - f(e_i)).$$

Thus, for a continuous-time Markov chain, the generator is just a difference operator.
Example: Kingman’s Coalescent

Recall that in Kingman’s coalescent, pairs of lineages coalesce at rate $\binom{n}{2}$ when there are $n$ lineages.

Consequently, if we let $n(t)$ denote the number of lineages at time $t$ (in the past), then $N = (n(t): t \geq 0)$ is a continuous-time Markov chain (a pure death process) with generator

$$Gf(n) = \binom{n}{2} \left[ f(n-1) - f(n) \right].$$

Exercise: Write down the generator for the continuous-time Markov chain that describes the number of branches in the Ancestral Selection Graph.
Example: Moran model

Assumptions:

- Constant population size: $N$ haploid adults.
- Overlapping generations.
- At rate 1 each individual gives birth to a single offspring.
- Each birth is accompanied by the death of a single adult individual chosen uniformly at random from the population (including the parent, but excluding the offspring).
- Two alleles $A_1$ and $A_2$.

Remark: An alternative version of the Moran model is sometimes studied in which offspring are not permitted to replace their own parent.
Let $X(t)$ denotes the number of copies of $A_1$ in the population at time $t$. Then $X(t)$ only changes during reproductive events, according to the following rules:

<table>
<thead>
<tr>
<th>parental genotype</th>
<th>deceased genotype</th>
<th>probability</th>
<th>change in $X(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$A_1$</td>
<td>$p^2$</td>
<td>0</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$A_2$</td>
<td>$p(1-p)$</td>
<td>+1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$A_1$</td>
<td>$p(1-p)$</td>
<td>-1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$A_2$</td>
<td>$(1-p)^2$</td>
<td>0</td>
</tr>
</tbody>
</table>

where $p = X(t)/N$ is the frequency of $A_1$.

**Remark:** Notice that $X(t)$ is a birth-death process, i.e., the number of copies of $A_1$ can only increase or decrease by one at each jump.
Likewise, \((p(t) : t \geq 0)\) is a continuous-time Markov chain, and the rate matrix can be found by multiplying each transition probability by the total rate, \(N\), at which reproductive events occur:

\[
\begin{align*}
Q_{p,p+1/N} &= Np(1 - p) \\
Q_{p,p-1/N} &= Np(1 - p) \\
Q_{p,p} &= -2Np(1 - p) \\
Q_{p,q} &= 0 \text{ if } q \neq p - 1/N, p, p + 1/N
\end{align*}
\]

Consequently, the generator of the Moran model is the difference operator:

\[
G_N f(p) = Np(1 - p)\left[ f(p + 1/N) - f(p) \right] + Np(1 - p)\left[ f(p - 1/N) - f(p) \right].
\]
**Example:** Suppose that $X = (X(t) : t \geq 0)$ is a one-dimensional diffusion process. If $f(x)$ is twice-continuously differentiable, then

$$Gf(x) = \lim_{t \to 0} \frac{1}{t} \left( \mathbb{E}_x[f(X_t)] - f(x) \right)$$
Example: Suppose that $X = (X(t) : t \geq 0)$ is a one-dimensional diffusion process. If $f(x)$ is twice-continuously differentiable, then

$$Gf(x) = \lim_{t \to 0} \frac{1}{t} \left( \mathbb{E}_x[f(X_t)] - f(x) \right)$$

$$= \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x \left[ f'(x)(X(t) - x) + \frac{1}{2} f''(x)(X(t) - x)^2 + O \left( (X(t) - x)^3 \right) \right]$$
Example: Suppose that $X = (X(t) : t \geq 0)$ is a one-dimensional diffusion process. If $f(x)$ is twice-continuously differentiable, then

$$Gf(x) = \lim_{t \to 0} \frac{1}{t} \left( \mathbb{E}_x [f(X_t)] - f(x) \right)$$

$$= \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x \left[ f'(x)(X(t) - x) + \frac{1}{2} f''(x)(X(t) - x)^2 + O\left((X(t) - x)^3\right) \right]$$

$$= \frac{1}{2} a(x)f''(x) + b(x)f'(x).$$

Thus the generator of a diffusion process is a second-order differential operator.
Application: Generators and Time Changes

It is sometimes useful to consider a Markov process $X$ run on a different time scale:

$$\hat{X}(t) \equiv X(\lambda t),$$

where $\lambda > 0$ is constant. This is an example of a time change.

Suppose that $G$ is the generator of $X$. Then the generator of the rescaled process is:

$$\hat{G}f(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)]$$

$$= \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]$$

$$= \lambda \mathbb{E}_x[f(X(\lambda t)) - f(x)],$$

i.e., rescaling a Markov process corresponds to multiplying its generator by a constant.
Application: Generators and Time Changes

It is sometimes useful to consider a Markov process $X$ run on a different time scale:

$$\hat{X}(t) \equiv X(\lambda t),$$

where $\lambda > 0$ is constant. This is an example of a time change.

Suppose that $G$ is the generator of $X$. Then the generator of the rescaled process is:

$$\hat{G}f(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)]$$

$$= \lambda \lim_{t \to 0} \frac{1}{\lambda t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]$$

$$= \lambda \lim_{t \to 0} \frac{1}{\lambda t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]$$
Application: Generators and Time Changes

It is sometimes useful to consider a Markov process $X$ run on a different time scale:

$$\hat{X}(t) \equiv X(\lambda t),$$

where $\lambda > 0$ is constant. This is an example of a time change.

Suppose that $G$ is the generator of $X$. Then the generator of the rescaled process is:

$$\hat{G}f(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)]$$

$$= \lim_{t \to 0} \frac{1}{t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]$$

$$= \lambda \lim_{t \to 0} \frac{1}{\lambda t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]$$

$$= \lambda Gf(x),$$

i.e., rescaling a Markov process corresponds to multiplying its generator by a constant.
Application: Generators and Convergence of Markov Processes

Suppose that

- $E$ is a set.
- For each $N \geq 1$, $X^N$ is an $E$-valued Markov process with generator $G^N$.
- $X$ is an $E$-valued Markov process with generator $G$.

We can sometimes show that the sequence of processes $\{X^N : N \geq 1\}$ converges to $X$ by showing that the generators of these processes converge:

$$\lim_{N \to \infty} G^N f(x) = Gf(x)$$

whenever $Gf$ is defined.
Example: Diffusion Approximation for the Moran Model

Recall that the generator of the Moran model $p_N(t)$ is:

$$G_N f(p) = Np(1 - p)(f(p + 1/N) - f(p)) + Np(1 - p)(f(p - 1/N) - f(p)).$$

If $N$ is large and $f$ is smooth, then we can approximate $G_N f$ by the first few terms in its Taylor series expansion:

$$G_N f(p) = Np(1 - p) \left( \frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) + Np(1 - p) \left( -\frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) = \frac{1}{N} p(1 - p)f''(p) + O(N^{-2}).$$
This limit vanishes as $N$ tends to infinity, essentially because the frequency is constant in an infinite population measured in units of generations.

However, if we consider the process, $\hat{p}^N(t) = p^N(Nt)$, with time measured in units of $N$ generations, then

$$\lim_{N \to \infty} \hat{G}^N f(p) = \lim_{N \to \infty} NG^N f(p) = p(1 - p)f''(p).$$

Thus, the diffusion approximation for the Moran model is a Wright-Fisher diffusion with

$$a(p) = 2p(1 - p) \quad \text{(infinitesimal variance)}$$
$$b(p) = 0 \quad \text{(infinitesimal drift)}.$$
Comparison of the Wright-Fisher and Moran models:

Diffusion approximations:

- **Wright-Fisher:** \( G_{WF} f(p) = \frac{1}{2} p(1 - p) f''(p) \)
- **Moran:** \( G_M f(p) = p(1 - p) f''(p) \).

Since \( G_M f(p) = 2G_{WF} f(p) \), we know that the diffusion approximation, \( p_M \), for the Moran model is a time change of the diffusion approximation, \( p_{WF} \), for the Wright-Fisher model:

\[
p_M(t) = p_{WF}(2t).
\]

**Conclusion:** Apart from a rescaling of time, the Moran model and the Wright-Fisher model are very similar when \( N \) is large. In fact, a very large class of neutral population genetical models share this property, i.e., the diffusion approximation is robust.
Time Reversal of Markov Chains

Suppose that $X$ is a discrete-time Markov chain on a set $E = \{e_1, \cdots, e_N\}$ with transition matrix $P = (p_{ij})$ and stationary distribution $\pi$. We will assume that $X = (X(n) : n \in \mathbb{Z})$ is defined at all times in the future and the past, and that $X$ is stationary, i.e., for every $n$,

$$\mathbb{P}\{X(n) = e_k\} = \pi_k.$$

When these conditions are satisfied, a related process, $\tilde{X} = (\tilde{X}(n) : n \in \mathbb{Z})$, called the stationary time reversal of $X$, can be defined by setting

$$\tilde{X}(n) = X(-n).$$

In other words, $\tilde{X}$ is simply the process $X$ run backwards in time.
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

$$\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j | \tilde{X}(0) = e_i\}$$
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

$$\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j | \tilde{X}(0) = e_i\}$$

$$= \mathbb{P}\{X(-1) = e_j | X(0) = e_i\}$$
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

$$\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j | \tilde{X}(0) = e_i\}$$

$$= \mathbb{P}\{X(-1) = e_j | X(0) = e_i\}$$

$$= \frac{\mathbb{P}\{X(-1) = e_j, X(0) = e_i\}}{\mathbb{P}\{X(0) = e_j\}}$$
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

\[
\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j | \tilde{X}(0) = e_i\} \\
= \mathbb{P}\{X(-1) = e_j | X(0) = e_i\} \\
= \frac{\mathbb{P}\{X(-1) = e_j, X(0) = e_i\}}{\mathbb{P}\{X(0) = e_j\}} \\
= \mathbb{P}\{X(0) = e_i | X(-1) = e_j\} \left( \frac{\mathbb{P}\{X(0) = e_j\}}{\mathbb{P}\{X(-1) = e_i\}} \right) \quad \text{(Bayes’ formula)}
\]
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

$$
\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j|\tilde{X}(0) = e_i\}
$$

$$
= \mathbb{P}\{X(-1) = e_j|X(0) = e_i\}
$$

$$
= \mathbb{P}\{X(-1) = e_j, X(0) = e_i\} \frac{\mathbb{P}\{X(0) = e_j\}}{\mathbb{P}\{X(0) = e_i\}}
$$

$$
= \mathbb{P}\{X(0) = e_i|X(-1) = e_j\} \left( \frac{\mathbb{P}\{X(0) = e_j\}}{\mathbb{P}\{X(-1) = e_i\}} \right) \quad \text{(Bayes’ formula)}
$$

$$
= \mathbb{P}\{X(0) = e_i|X(-1) = e_j\} \left( \frac{\pi_j}{\pi_i} \right) \quad \text{($X$ is stationary)}
$$
It can be shown that $\tilde{X}$ is itself a Markov chain. The transition probabilities can be calculated in the following way:

$$
\tilde{p}_{ij} = \mathbb{P}\{\tilde{X}(1) = e_j | \tilde{X}(0) = e_i\} \\
= \mathbb{P}\{X(-1) = e_j | X(0) = e_i\} \\
= \frac{\mathbb{P}\{X(-1) = e_j, X(0) = e_i\}}{\mathbb{P}\{X(0) = e_j\}} \\
= \mathbb{P}\{X(0) = e_i | X(-1) = e_j\} \left( \frac{\mathbb{P}\{X(0) = e_j\}}{\mathbb{P}\{X(-1) = e_i\}} \right) \quad \text{(Bayes’ formula)} \\
= \mathbb{P}\{X(0) = e_i | X(-1) = e_j\} \left( \frac{\pi_j}{\pi_i} \right) \quad \text{(X is stationary)} \\
= p_{ji} \left( \frac{\pi_j}{\pi_i} \right) .
$$
It follows that the transition matrices, $P$ and $\tilde{P}$, of $X$ and $\tilde{X}$ are related by the following simple expression:

$$
\tilde{P} = \Pi^{-1} P^T \Pi,
$$

where

- $\Pi = diag(\pi_i)$ is the diagonal matrix with the elements of the stationary distribution $\pi$ along the diagonal;
- $\Pi^{-1} = diag(\pi^{-1})$ is the matrix inverse of $\Pi$;
- $P^T = (p_{ji})$ is the transpose of $P$.

In particular,

$$
\mathbb{E}_\pi \left[ \mathbb{E}_{\tilde{X}(0)}[f(X(n))] \cdot g(\tilde{X}(0)) \right] = \mathbb{E}_\pi \left[ f(X(0)) \cdot \mathbb{E}_{\tilde{X}(0)}[g(\tilde{X}(n))] \right]. \quad (*)
$$

**Exercise:** Verify (*).
Time Reversal of Continuous-time Markov Processes

Suppose that \( X = (X(t) : t \in \mathbb{R}) \) is a stationary Markov process on \( E \) with generator \( G \) and stationary distribution \( \pi(x)dx \). Then \( \tilde{X} = (\tilde{X}(t) : t \geq 0) \) is defined by setting:

\[
\tilde{X}(t) = X(-t).
\]
Time Reversal of Continuous-time Markov Processes

Suppose that \( X = (X(t) : t \in \mathbb{R}) \) is a stationary Markov process on \( E \) with generator \( G \) and stationary distribution \( \pi(x)dx \). Then \( \tilde{X} = (\tilde{X}(t) : t \geq 0) \) is defined by setting:

\[
\tilde{X}(t) = X(-t).
\]

If \( (T_t : t \geq 0) \) and \( (\tilde{T}_t : t \geq 0) \) are the transition semigroups for \( X \) and \( \tilde{X} \) respectively, it can be shown that a continuous-time version of (\( * \)) is satisfied:

\[
E_\pi \left[ E_{\tilde{X}(0)}[f(X(t))] \cdot g(\tilde{X}(0))] \right] = E_\pi \left[ f(X(0)) \cdot E_X[\tilde{g}(\tilde{X}(t))] \right], \quad \text{i.e.,}
\]

\[
\int_E T_t f(x) g(x) \pi(x) dx = \int_E f(x) \tilde{T}_t g(x) \pi(x) dx.
\]
A similar identity can be derived for the generators $G$ and $\tilde{G}$ of the two processes.

Notice that

$$\int_E \left( T_t f(x) - f(x) \right) g(x) \pi(x) dx = \int_E f(x) \left( \tilde{T}_t g(x) - g(x) \right) \pi(x) dx,$$

holds for every $t \geq 0$. 
A similar identity can be derived for the generators $G$ and $\tilde{G}$ of the two processes.

Notice that

$$\int_{E} \left( T_{t} f(x) - f(x) \right) g(x) \pi(x) dx = \int_{E} f(x) \left( \tilde{T}_{t} g(x) - g(x) \right) \pi(x) dx,$$

holds for every $t \geq 0$.

Consequently, dividing both sides by $t$ and letting $t \to 0$ shows that

$$\int_{E} Gf(x)g(x)\pi(x)dx = \int_{E} f(x)\tilde{G}g(x)\pi(x)dx.$$

Sometimes this can be used to identify $\tilde{G}$.
Suppose that $X = (X_t : t \geq 0)$ is a stationary diffusion process on $[l, r]$ with

- generator $Gf(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x)$
- stationary distribution $\pi(x) dx$.

Recall that the density $\pi(x)$ satisfies the following identities:

\[
\frac{1}{2} (a(x)\pi(x))'' - (b(x)\pi(x))' = 0
\]
\[
\frac{1}{2} (a(x)\pi(x))' - b(x)\pi(x) = 0.
\]
**Reminder: Integration-by-Parts**

Using the product rule for differentiation, we have:

\[
\begin{align*}
  u(r)v(r) - u(l)v(l) &= \int_{l}^{r} (u(x)v(x))' \, dx \\
  &= \int_{l}^{r} (u'(x)v(x) + u(x)v'(x)) \, dx.
\end{align*}
\]

Therefore,

\[
\int_{l}^{r} u'(x)v(x) \, dx = u(r)v(r) - u(l)v(l) - \int_{l}^{r} u(x)v'(x) \, dx.
\]
Suppose that $f, g$ are twice differentiable, with $f(x) = 0$ when $x \notin [\epsilon, 1 - \epsilon]$ for some $\epsilon > 0$. Then

$$
\int_{l}^{r} Gf(x)g(x)\pi(x)dx
= \int_{l}^{r} \left( \frac{1}{2} a(x)f''(x) + b(x)f'(x) \right) g(x)\pi(x)dx
$$
Suppose that $f, g$ are twice differentiable, with $f(x) = 0$ when $x \not\in [\epsilon, 1 - \epsilon]$ for some $\epsilon > 0$. Then

$$\int_{l}^{r} Gf(x)g(x)\pi(x)dx$$

$$= \int_{l}^{r} \left( \frac{1}{2} a(x)f''(x) + b(x)f'(x) \right) g(x)\pi(x)dx$$

$$= - \int_{l}^{r} \left[ f'(x) \frac{1}{2} (g(x)a(x)\pi(x))' + f(x)(g(x)b(x)\pi(x))' \right] dx$$
Suppose that \( f, g \) are twice differentiable, with \( f(x) = 0 \) when \( x \notin [\epsilon, 1 - \epsilon] \) for some \( \epsilon > 0 \). Then

\[
\int_{l}^{r} Gf(x)g(x)\pi(x)dx = \int_{l}^{r} \left( \frac{1}{2} a(x)f''(x) + b(x)f'(x) \right) g(x)\pi(x)dx
\]

\[
= -\int_{l}^{r} \left[ f'(x) \frac{1}{2} (g(x)a(x)\pi(x))' + f(x)(g(x)b(x)\pi(x))' \right] dx
\]

\[
= \int_{l}^{r} f(x) \left[ \frac{1}{2} (g(x)a(x)\pi(x))'' - (g(x)b(x)\pi(x))' \right] dx
\]
Suppose that \( f, g \) are twice differentiable, with \( f(x) = 0 \) when \( x \notin [\epsilon, 1 - \epsilon] \) for some \( \epsilon > 0 \). Then

\[
\int_{l}^{r} Gf(x)g(x)\pi(x)dx
\]

\[
= \int_{l}^{r} \left( \frac{1}{2} a(x)f''(x) + b(x)f'(x) \right) g(x)\pi(x)dx
\]

\[
= - \int_{l}^{r} \left[ f'(x) \frac{1}{2} (g(x)a(x)\pi(x))' + f(x)(g(x)b(x)\pi(x))' \right] dx
\]

\[
= \int_{l}^{r} f(x) \left[ \frac{1}{2} (g(x)a(x)\pi(x))'' - (g(x)b(x)\pi(x))' \right] dx
\]

\[
= \int_{l}^{r} f(x) \left[ \frac{1}{2} a(x)\pi(x)g''(x) + (a(x)\pi(x))'g'(x) + \frac{1}{2} g(x)(a(x)\pi(x))''
\right.
\]

\[
- b(x)\pi(x)g'(x) - g(x)(b(x)\pi(x))'
\]

\[
] dx
\]
= \int_{l}^{r} f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + (a(x) \pi(x))' g'(x) \right. \\
\left. + g(x) \left( \frac{1}{2} (a(x) \pi(x))'' - (b(x) \pi(x))' \right) \right] dx
\[
\begin{align*}
\int f(x) & \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + (a(x) \pi(x))' g'(x) \\
+ g(x) \left( \frac{1}{2} (a(x) \pi(x))'' - (b(x) \pi(x))' \right) \right] dx \\
& = \int f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + 2 \pi(x) b(x) g'(x) \right] dx
\end{align*}
\]
\[
\int_{l}^{r} f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + (a(x) \pi(x))' g'(x) \right. \\
\left. + g(x) \left( \frac{1}{2} (a(x) \pi(x))'' - (b(x) \pi(x))' \right) \right] dx \\
= \int_{l}^{r} f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + 2 \pi(x) b(x) g'(x) \right] dx \\
= \int_{l}^{r} f(x) \left( \frac{1}{2} a(x) g''(x) + b(x) g'(x) \right) \pi(x) dx
\]
\[
\begin{align*}
&= \int_\ell^r f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + (a(x) \pi(x))' g'(x) \
&\quad + g(x) \left( \frac{1}{2} (a(x) \pi(x))'' - (b(x) \pi(x))' \right) \right] \, dx \\
&= \int_\ell^r f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + 2\pi(x) b(x) g'(x) \right] \, dx \\
&= \int_\ell^r f(x) \left( \frac{1}{2} a(x) g''(x) + b(x) g'(x) \right) \pi(x) \, dx \\
&= \int_\ell^r f(x) Gg(x) \pi(x) \, dx
\end{align*}
\]

It follows that \( \tilde{G} g = G g \) for all \( g \) for which these expressions are defined.
\[ = \int_I^r f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + (a(x) \pi(x))' g'(x) \right. \\
+ g(x) \left( \frac{1}{2} (a(x) \pi(x))'' - (b(x) \pi(x))' \right) \right] dx \\
= \int_I^r f(x) \left[ \pi(x) \left( \frac{1}{2} a(x) g''(x) - b(x) g'(x) \right) + 2 \pi(x) b(x) g'(x) \right] dx \\
= \int_I^r f(x) \left( \frac{1}{2} a(x) g''(x) + b(x) g'(x) \right) \pi(x) dx \\
= \int_I^r f(x) Gg(x) \pi(x) dx \\
= \int_I^r f(x) \tilde{G}g(x) \pi(x) dx. \]

It follows that \( \tilde{G}g = Gg \) for all \( g \) for which these expressions are defined.
Main Result: The stationary time reversal of a one-dimensional diffusion process has the same distribution as the forwards-in-time process.

Remarks:

- Invariance under stationary time reversal also holds true for any birth-death process with a stationary distribution, in particular, for the Moran model with mutation.

- Invariance does not hold (in general) for multi-dimensional diffusions. Instead,

\[
Gf(x) = \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x) \partial_{ij} f(x) + \sum_{i=1}^{n} b_{i}(x) \partial_{i} f(x)
\]

\[
\tilde{G}f(x) = \frac{1}{2} \sum_{i,j=1}^{n} a_{ij}(x) \partial_{ij} f(x) + \sum_{i=1}^{n} \left( \frac{1}{\pi(x)} \sum_{j=1}^{n} \partial_{j}(a_{ij}(x)\pi(x)) - b_{i}(x) \right) \partial_{i} f(x)
\]
Coalescent Processes in Populations with Genetic Structure

The Structured Coalescent for a Wright-Fisher Diffusion

Key ideas:

- Think of the population as being subdivided into different genetic backgrounds.
- Individuals that have the same genotype are exchangeable.
- Consequently, coalescence within backgrounds can be described by a modification of Kingman’s coalescent.

Our goal is to describe the Markov process,

\[ X_t = (n_1(t), n_2(t), p(t)), \]

where \( n_i(t) \) is the number of ancestral lineages in the \( A_i \) background and \( p(t) \) is the ancestral frequency of \( A_1 \) at time \( t \) in the past.
Changes to \( \Gamma(t) \) occur through the following events:

- Two \( A_1 \) lineages can coalesce.
- Two \( A_2 \) lineages can coalesce.
- Each lineage can migrate between backgrounds due to mutation at the selected site.
- The allele frequencies evolve backwards in time.

**Remark:** We now know that the ancestral process of allele frequencies is just a Wright-Fisher diffusion.
Recall the Wright-Fisher model introduced earlier.

Assumptions:

- $N$ haploid individuals (constant population size).
- Alleles $A_1, A_2$ with frequencies $p$ and $q$.
- Each individual alive in generation $t + 1$ chooses its parent from the preceding generation independently and with replacement with the following probabilities:

  \[
  \begin{align*}
  \text{each } A_1\text{-type individual: } & \frac{1}{N} \frac{1 + s(p)/N}{1 + ps(p)/N} \\
  \text{each } A_2\text{-type individual: } & \frac{1}{N} \frac{1}{1 + ps(p)/N}.
  \end{align*}
  \]

- Following selection, each $A_1$-type individual mutates to $A_2$ with probability $\mu_2 = \theta_2/N$, while each $A_2$-type individuals mutates to $A_1$ with probability $\mu_1 = \theta_1/N$. 

The rate of coalescence of two $A_1$ lineages is equal to:

$$\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two } A_1 \text{ individuals have the same parent} \right\} =$$
The rate of coalescence of two $A_1$ lineages is equal to:

$$\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two } A_1 \text{ individuals have the same parent} \right\} =$$

$$= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two individuals have the same parent} \mid \text{both are } A_1 \right\}$$
The rate of coalescence of two $A_1$ lineages is equal to:

$$\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two } A_1 \text{ individuals have the same parent} \right\} =$$

$$= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two individuals have the same parent} \mid \text{both are } A_1 \right\}$$

$$= \lim_{N \to \infty} \mathbb{P}_p \left\{ \text{two individuals have the same parent and both are } A_1 \right\} \mathbb{P}_p \left\{ \text{two sampled individuals are } A_1 \right\}$$
The rate of coalescence of two $A_1$ lineages is equal to:

$$
\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two } A_1 \text{ individuals have the same parent} \right\} = \\
= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two individuals have the same parent} \mid \text{both are } A_1 \right\} \\
= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two individuals have the same parent and both are } A_1 \right\} / \mathbb{P}_p \left\{ \text{two sampled individuals are } A_1 \right\} \\
= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{two non-mutant individuals have the same parent and both are } A_1 \right\} / \mathbb{P}_p \left\{ \text{two sampled individuals are } A_1 \right\}.
$$

**Remark:** The last identity is justified by the fact that the mutation rate is of order $O(1/N)$.
If $p'$ denotes the frequency in the preceding generation, then

$$\mathbb{P}_p \left\{ \text{two non-mutant individuals have the same parent and both are } A_1 \right\}$$

$$= \mathbb{E}_p \left[ \left( \frac{Np'}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( \frac{1}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( 1 - \frac{\mu_1}{N} \right)^2 \right]$$
If $p'$ denotes the frequency in the preceding generation, then

$$
\mathbb{P}_p\{\text{two non-mutant individuals have the same parent and both are } A_1\} = \mathbb{E}_p \left[ \left( \frac{Np'}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( \frac{1}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( 1 - \frac{\mu_1}{N} \right)^2 \right]
$$

$$
= \mathbb{E}_p \left[ \frac{p'}{N} + O(N^{-2}) \right]
$$
If \( p' \) denotes the frequency in the preceding generation, then

\[
\mathbb{P}_p \left\{ \text{two non-mutant individuals have the same parent and both are } A_1 \right\} = \mathbb{E}_p \left[ \left( \frac{Np'}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( \frac{1}{N} \frac{1 + s(p')/N}{1 + p's(p')/N} \right) \left( 1 - \frac{\mu_1}{N} \right)^2 \right] 
\]

\[
= \mathbb{E}_p \left[ \frac{p'}{N} + O(N^{-2}) \right] 
\]

\[
= \frac{p}{N} + o(N^{-1}),
\]

where the remainder term \( o(N^{-1}) \) in the last line has the property

\[
\lim_{N \to \infty} N \cdot o(N^{-1}) = 0.
\]

This can be deduced from the following statement, which holds for any \( \epsilon > 0 \):

\[
\lim_{N \to \infty} \mathbb{P}_p \left\{ |p' - p| > \epsilon \right\} = 0.
\]
Since

\[ \mathbb{P}_p \left\{ \text{two individuals are } A_1 \right\} = p \frac{Np - 1}{N - 1} = p^2 + O(N^{-1}), \]

it follows that the pairwise coalescence rate for two \( A_1 \) lineages is:

\[ \lim_{N \to \infty} N \left( \frac{p}{N} + o(N^{-1}) \right) / \left( p^2 + O(N^{-1}) \right) = \frac{1}{p}. \]
Since
\[ \mathbb{P}_p \left\{ \text{two individuals are } A_1 \right\} = p \frac{Np - 1}{N - 1} = p^2 + O(N^{-1}), \]
it follows that the pairwise coalescence rate for two \( A_1 \) lineages is:
\[ \lim_{N \to \infty} N \left( \frac{p}{N} + o(N^{-1}) \right) \Big/ \left( p^2 + O(N^{-1}) \right) = \frac{1}{p}. \]

**Intuition:**
- In Kingman’s coalescent, the pairwise coalescent rate is inversely proportional to the population size.
- Likewise, in the structured coalescent, the number of \( A_1 \) individuals is \( Np' \approx Np \), so once we have rescaled time by a factor of \( N \), we expect the pairwise coalescent rate of \( A_1 \) lineages to be \( 1/p \).
It can also be shown that if we consider a sample of \( n_1 A_1 \) individuals, then the probability that more than two of them have a common ancestor in the previous generation (either through a multiple merger or through simultaneous mergers) is of order no greater than \((Np)^{-2}\).

Thus, as long as \( p \) is bounded away from 0, these events are negligible when \( N \) is large. Consequently, the pairwise coalescent rate for the sample is just:

\[
\binom{n_1}{2} \frac{1}{p}.
\]

Similar arguments show that if there are \( n_2 A_2 \) lineages, then the pairwise coalescent rate is:

\[
\binom{n_2}{2} \frac{1}{1 - p}.
\]
Next, we calculate the rate at which an $A_1$ lineage migrates into the $A_2$ background via mutation at the selected site. This is equal to:

$$
\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{an } A_1 \text{ individual has an } A_2 \text{ parent} \right\} =
$$

$$
= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{an individual has an } A_2 \text{ parent | it is } A_1 \right\}
$$

$$
= \lim_{N \to \infty} \frac{N \mathbb{P}_p \left\{ \text{an individual has an } A_2 \text{ parent but is } A_1 \right\}}{\mathbb{P}_p \left\{ \text{a sampled individual is } A_1 \right\}}
$$

$$
= \frac{1}{p} \lim_{N \to \infty} N \mathbb{E}_p \left[ \frac{N q'}{N} \left( \frac{1}{1 + p' s(p')/N} \right) \frac{\mu_1}{N} \right]
$$

$$
= \mu_1 \frac{q}{p}.
$$
Once again, more complicated events (e.g., multiple lineages simultaneously migrating into a different background) are unlikely when \( n_1 + n_2 \ll N \), so the total rate at which one of \( n_1 \) \( A_1 \) lineages migrates into the \( A_2 \) background is just:

\[
n_1 \mu_1 \frac{q}{p}.
\]

Likewise, the total rate at which one of \( n_2 \) \( A_2 \) lineages migrates into the \( A_1 \) background is

\[
n_2 \mu_2 \frac{p}{q}.
\]

Remark: Notice that lineages rapidly migrate out of lineages that are rare.
The infinitesimal generator of the Markov process \( X_t = (n_1(t), n_2(t), p(t)) \) is

\[
\Gamma f(n_1, n_2, p) = \left(\begin{array}{c} n_1 \\ 2 \end{array}\right) \frac{1}{p} \left[ f(n_1 - 1, n_2, p) - f(n_1, n_2, p) \right] + \\
\left(\begin{array}{c} n_2 \\ 2 \end{array}\right) \frac{1}{q} \left[ f(n_1, n_2 - 1, p) - f(n_1, n_2, p) \right] + \\
n_1\mu_1 \frac{q}{p} \left[ f(n_1 - 1, n_2 + 1, p) - f(n_1, n_2, p) \right] + \\
n_2\mu_2 \frac{p}{q} \left[ f(n_1 + 1, n_2 - 1, p) - f(n_1, n_2, p) \right] + \\
\frac{1}{2} pq f''(n_1, n_2, p) + \left(\mu_1 q - \mu_2 p + s(p)pq\right) f'(n_1, n_2, p),
\]

where ' denotes differentiation with respect to \( p \).
**Existence:** Because the coalescence and mutation rates diverge when \( p = 0 \) or \( p = 1 \), it is not obvious that the structured coalescent exists. That it does can be deduced from the following

**Lemma** Let \((p_t : t \geq 0)\) be a Wright-Fisher diffusion. Then, for any \( R < \infty \),

\[
\lim_{k \to \infty} \mathbb{P}_p \left\{ \int_0^{\tau_k} \frac{ds}{p_s} > R \right\} = 1
\]

\[
\lim_{k \to \infty} \mathbb{P}_p \left\{ \int_0^{\tau'_k} \frac{ds}{1 - p_s} > R \right\} = 1,
\]

where \( \tau_k = \inf\{ t > 0 : p_t = 1/k \} \) and \( \tau'_k = \inf\{ t > 0 : p_t = 1 - 1/k \} \).

For the proof, see Barton, Etheridge and Sturm (2004) and Taylor (2007).
The structured coalescent can be used to formulate explicit boundary value problems for various genealogical statistics (Hudson et al. 1988; Barton & Etheridge 2004).

For example, if $\tau(n_1, n_2, p)$ denotes the expected TMRCA for a sample with composition $(n_1, n_2)$ from a population in which $A_1$ has frequency $p$, then

$$
-1 = \binom{n_1}{2} \frac{1}{p} [\tau(n_1 - 1, n_2, p) - \tau(n_1, n_2, p)] + \\
\binom{n_2}{2} \frac{1}{q} [\tau(n_1, n_2 - 1, p) - \tau(n_1, n_2, p)] + \\
n_1 \mu_1 \frac{q}{p} [\tau(n_1 - 1, n_2 + 1, p) - \tau(n_1, n_2, p)] + \\
n_2 \mu_2 \frac{p}{q} [\tau(n_1 + 1, n_2 - 1, p) - \tau(n_1, n_2, p)] + \\
\frac{1}{2} pq \tau''(n_1, n_2, p) + \left( \mu_1 q - \mu_2 p + s(p)pq \right) \tau'(n_1, n_2, p).
$$
The **boundary conditions** for this system of equations are:

\[
\begin{align*}
\tau(n_1, n_2, 0) &= \frac{(n_1 - 1)\tau(n_1 - 1, n_2, 0) + 2\mu_1 \tau(n_1 - 1, n_2 + 1, 0)}{n_1 - 1 + 2\mu_1} \\
\tau(0, n_2, 0) &= \frac{n_2(n_2 - 1)\tau(0, n_2 - 1, 0) + 2 + 2\mu_1 \tau'(0, n_2, 0)}{n_2(n_2 - 1)} \\
\tau(n_1, n_2, 1) &= \frac{(n_2 - 1)\tau(n_1, n_2 - 1, 1) + 2\mu_2 \tau(n_1 + 1, n_2 - 1, 1)}{n_2 - 1 + 2\mu_2} \\
\tau(n_1, 0, 1) &= \frac{n_1(n_1 - 1)\tau(n_1 - 1, 0, 1) + 2 - 2\mu_2 \tau'(n_1, 0, 1)}{n_1(n_1 - 1)} \\
\tau(1, 0, p) &= \tau(0, 1, p) = 0.
\end{align*}
\]

**Caveat:** These equations can be solved numerically, but care is needed because of the singularities at the boundaries. See Barton & Etheridge (2004).
Simulation of the Structured Coalescent - Choice of Initial Configuration

The choice of the initial configuration \((n_1, n_2, p)\) depends on how much information we have about the sample.

Random sample of size \(n\):

- Sample \(p\) from the stationary distribution \(\pi(p)dp\) of the diffusion process.
- Then, conditional on \(p\), sample \(n_1\) from the Binomial distribution with parameters \(n\) and \(p\), and set \(n_2 = n - n_1\).
Simulation of the Structured Coalescent - Choice of Initial Configuration

The choice of the initial configuration \((n_1, n_2, p)\) depends on how much information we have about the sample.

Random sample of size \(n\):

- Sample \(p\) from the stationary distribution \(\pi(p)dp\) of the diffusion process.
- Then, conditional on \(p\), sample \(n_1\) from the Binomial distribution with parameters \(n\) and \(p\), and set \(n_2 = n - n_1\).

Sample with known composition \((n_1, n_2)\):

- Sample \(p\) from the weighted distribution with density

\[
\pi(p|n_1, n_2) = \frac{1}{C} p^{n_1} (1 - p)^{n_2} \pi(p).
\]
Suppose that the density of the stationary distribution can be written as:

$$\pi(p) = \frac{1}{\beta(a, b)} p^{a-1}(1 - p)^{b-1} h(p),$$

where $h : [0, 1] \to \mathbb{R}^+$ is a continuous function. Let $h_{\text{max}}$ denote the maximal value achieved by $h$.

**Rejection sampling** with a Beta proposal distribution can be used to sample from $\pi$:

- Sample $x$ from the Beta distribution with parameters $(a, b)$ (e.g., using a uniformly accurate method).
- Sample $u$ from a Uniform $[0, 1]$ distribution.
- If $u < h(x)/h_{\text{max}}$, accept $x$.
- Otherwise, reject $x$ and repeat the first three steps until an $x$ is accepted.

**Remark:** This approach also works with the conditional distribution $\pi(p|n_1, n_2)$. 
Simulation of the Structured Coalescent - Diffusion

To simulate an approximate sample path of the diffusion process:

- Let $V(p)$ and $M(p)$ denote the infinitesimal variance and mean.
- If $\sqrt{V(p)}$ is differentiable, then the diffusion can be simulated by approximating the stochastic differential equation (Euler-Maruyama scheme). If not, then one alternative is:
  - Adaptive time discretization:
    \[ \delta t = \frac{p(1 - p)}{NV(p)}. \]
  - Sample $Np(t + \delta t)$ from $\text{Binom}(N, p + M(p)\delta t)$.
  - Use a normal approximation when $Np'$ is large and a Poisson approximation otherwise.
Simulation of the Structured Coalescent - Jump Events

To simulate the changes in the configuration conditional on \((p(t) : t \geq 0)\),

- Following each jump, simulate four IID mean-one exponentially-distributed RV’s, \(Z_1, \ldots, Z_4\).
- Let \(\tau\) be the time of the last jump. Define a set of stochastic ‘clocks’
  \[
  T_i(t) = \int_{\tau}^{t} R_i(n_1(s), n_2(s), p(s)) \, ds,
  \]
  where \(R_i\) is the rate of the \(i\)'th event.
- The timing and type of the next event is determined by the clock that first exceeds the corresponding exponential RV.
- It is computationally efficient to reuse the same diffusion sample path.
Example: The Effect of Balancing Selection on Genealogies

There are several scenarios in which selection may favor rare alleles:

- Heterozygote advantage
- Negative assortative mating
- Frequency-dependent selection due to predator-prey or host-parasite interactions.

This kind of selection is known as balancing selection and can be modeled by a diffusion approximation with generator:

\[ G_f(p) = \frac{1}{2}pqf''(p) + \left( \mu_1 q - \mu_2 p + \sigma(\bar{p} - p)pq \right)f'(p), \]

where \( \sigma > 0 \) and \( \bar{p} \in [0, 1] \).
Symmetric Balancing Selection and TMRCA

In this case, the stationary distribution has density:

$$\pi(p) = \frac{1}{C} p^{2\mu_1 - 1} q^{2\mu_2 - 1} e^{\sigma p (2\bar{p} - p)}.$$ 

Coalescent times are usually elevated at sites linked to a locus under balancing selection:

- Balancing selection tends to maintain alleles at intermediate frequencies.
- Samples are usually polymorphic for the selected locus.
- Lineages can only coalesce within backgrounds.

(Slatkin 2000; Takahata 2000).
Conditional Genealogies under Balancing Selection

Simulations of the coalescent conditional on the sample composition \((n_1, n_2)\) show that:

- Mean tree depth and tree length typically increase with sample diversity.
- Under balancing selection, the conditional TMRCA is substantially inflated even by singletons.
- Asymmetric balancing selection has a weaker effect than symmetric.
The Genealogy at a Linked Neutral Locus

Kaplan et al. (1988) extended the structured coalescent to models with recombination.

- Suppose that there are two linked loci, the selected locus and a neutral marker locus.
- The aim is to describe the genealogy of a sample at the marker locus.
- We can still think of lineages as belonging to a genetic background defined by the selected locus.
- However, recombination between the two loci can cause lineages to migrate between backgrounds.
Consider the Wright-Fisher model introduced earlier, but modified in the following way:

- Each offspring is recombinant with probability $r/N$.
- A recombinant offspring inherits the marker locus from the chosen parent, but inherits a copy of the selected locus from a randomly sampled individual in the population.

Notice that incorporating recombination into this model has no effect on the marginal dynamics of the allele frequencies at the selected locus. In particular, the diffusion approximation is unchanged.
The rate at which an \( A_1 \) lineage recombines into the \( A_2 \) background is:

\[
\lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{an } A_1 \text{ individual is recombinant} \right\} = \\
= \lim_{N \to \infty} N \mathbb{P}_p \left\{ \text{an individual is recombinant} \mid \text{it is } A_1 \right\} \\
= \lim_{N \to \infty} \frac{N \mathbb{P}_p \left\{ \text{an individual is recombinant and has type } A_1 \right\}}{\mathbb{P}_p \left\{ \text{an individual is } A_1 \right\}} \\
= \frac{1}{p} \lim_{N \to \infty} N \mathbb{E}_p \left[ \frac{N q'}{N} \left( \frac{1}{1 + p' s(p')/N} \right) \frac{r}{N p'} \right] \\
= rq.
\]

Consequently, the rate at which one of \( n_1 \) \( A_1 \) lineages recombines into the \( A_2 \) background is

\[ n_1 rq. \]
With recombination added, the generator of the structured coalescent is:

\[
\Gamma f(n_1, n_2, p) = \binom{n_1}{2} \frac{1}{p} \left[ f(n_1 - 1, n_2, p) - f(n_1, n_2, p) \right] + \\
\binom{n_2}{2} \frac{1}{q} \left[ f(n_1, n_2 - 1, p) - f(n_1, n_2, p) \right] + \\
n_1 \left( \mu_1 \frac{q}{p} + rq \right) \left[ f(n_1 - 1, n_2 + 1, p) - f(n_1, n_2, p) \right] + \\
n_2 \left( \mu_2 \frac{p}{q} + rp \right) \left[ f(n_1 + 1, n_2 - 1, p) - f(n_1, n_2, p) \right] + \\
\frac{1}{2} pqf''(n_1, n_2, p) + \left( \mu_1 q - \mu_2 p + s(p)pq \right) f'(n_1, n_2, p),
\]

where ’ denotes differentiation with respect to \( p \).
The Effect of Balancing Selection on the Genealogy at a Linked Site

Same model of balancing selection as before.

- Coalescent times are increased at linked sites.
- This effect is large only at sites that are very tightly linked to the selected site.
- Lineages can move between backgrounds by recombination as well as mutation.
Fecundity Variance and Neutral Population Genetical Models

A Galton-Watson Process with Culling

- $N$ haploid adults in each generation.
- Non-overlapping generations.
- IID offspring numbers $\eta_1, \ldots, \eta_N$ with
  \[
  \mathbb{E}[\eta_i] = R, \quad \text{Var}[\eta_i] = \sigma^2
  \]
  \[
  \mathbb{P}\{\eta_i \geq 1\} = 1 \quad (R > 1 \text{ is sufficient}).
  \]
- Each generation is formed by sampling $N$ offspring, uniformly at random and without replacement, from the offspring pool.
- Two neutral alleles, $A_1$ and $A_2$, with frequencies $p$ and $q$.
- Each offspring mutates to $A_i$ with probability $\mu_i/N$. 
If time is measured in units of \( N \) generations, then the diffusion approximation has generator:

\[
G_f(p) = \frac{1}{2} \left(1 - \frac{1}{R} + \frac{\sigma^2}{R^2}\right) pq f''(p) + (\mu_2 q - \mu_1 p)f'(p).
\]

- We again encounter the Wright-Fisher diffusion.
- The infinitesimal variance only depends on the mean and the variance of fecundity.
- Two sources of stochasticity: fecundity variance and random survival of offspring.
- Increasing \( \sigma^2 \) causes the allele frequencies to fluctuate more rapidly and also reduces standing variation.
Ancestral Processes and Coalescents

Suppose that we have sampled two individuals at random from the population. The probability that these share a common ancestor in the preceding generation is:

\[ P\{\text{common ancestor}\} = \sum_{i=1}^{N} P\{\text{both are descended from parent } i\} \]
Ancestral Processes and Coalescents

Suppose that we have sampled two individuals at random from the population. The probability that these share a common ancestor in the preceding generation is:

\[
P\{\text{common ancestor}\} = \sum_{i=1}^{N} P\{\text{both are descended from parent } i\}
\]

\[
= \sum_{i=1}^{N} E \left[ \frac{\eta_i(\eta_i - 1)}{Y(Y - 1)} \right] \quad \left( Y = \sum_{i=1}^{N} \eta_i \right)
\]
Ancestral Processes and Coalescents

Suppose that we have sampled two individuals at random from the population. The probability that these share a common ancestor in the preceding generation is:

\[
P\{\text{common ancestor}\} = \sum_{i=1}^{N} P\{\text{both are descended from parent } i\}
\]

\[
= \sum_{i=1}^{N} \mathbb{E} \left[ \frac{\eta_i(\eta_i - 1)}{Y(Y - 1)} \right]
\]

\[
= \frac{1}{N} \mathbb{E} \left[ \left( \frac{N}{Y} \frac{N}{Y - 1} \right) \eta_1(\eta_1 - 1) \right]
\]
Ancestral Processes and Coalescents

Suppose that we have sampled two individuals at random from the population. The probability that these share a common ancestor in the preceding generation is:

\[
\mathbb{P}\{\text{common ancestor}\} = \sum_{i=1}^{N} \mathbb{P}\{\text{both are descended from parent } i\}
\]

\[
= \sum_{i=1}^{N} \mathbb{E} \left[ \frac{\eta_i(\eta_i - 1)}{Y(Y - 1)} \right]
\]

\[
= \frac{1}{N} \mathbb{E} \left[ \left( \frac{N}{Y} \frac{N}{Y - 1} \right) \eta_1(\eta_1 - 1) \right]
\]

\[
= \frac{1}{N} \left( 1 - \frac{1}{R} + \frac{\sigma^2}{R^2} \right) + o\left( N^{-1} \right) \quad \text{(by the SLLN)}.
\]
If time is measured in units of \( N \) generations, then as \( N \to \infty \), the genealogy of a sample tends to Kingman’s coalescent run at a different rate.

- When there are \( n \) lineages, binary mergers occur at rate
  \[
  \binom{n}{2} \left( 1 - \frac{1}{R} + \frac{\sigma^2}{R^2} \right).
  \]
- Each pair is equally likely to coalesce during a merger.
- Thus, increasing \( \sigma^2 \) also increases the rate at which lineages coalesce down to their MRCA.
Duality between the Wright-Fisher diffusion and Kingman’s coalescent

Suppose that $\mu_1 = \mu_2 = 0$. If $\psi(p, n) = p^n$, then

$$G\psi(p, n) = \frac{1}{2} np(1 - p)n(n - 1)p^{n-2}$$

$$= \binom{n}{2} \nu (p^{n-1} - p^n) = \Gamma \psi(p, n),$$

where $\nu = 1 - 1/R + \sigma^2/R^2$, and $\Gamma$ is the generator of the pure death process $(n(t) : t \geq 0)$ describing the number of ancestral lineages in the coalescent.
Duality between the Wright-Fisher diffusion and Kingman’s coalescent

Suppose that $\mu_1 = \mu_2 = 0$. If $\psi(p, n) = p^n$, then

$$G\psi(p, n) = \frac{1}{2}vp(1-p)n(n-1)p^{n-2}$$

$$= \binom{n}{2} v(p^{n-1} - p^n) = \Gamma \psi(p, n),$$

where $v = 1 - 1/R + \sigma^2/R^2$, and $\Gamma$ is the generator of the pure death process $(n(t): t \geq 0)$ describing the number of ancestral lineages in the coalescent.

This identity shows that the Wright-Fisher diffusion and Kingman’s coalescent are dual, i.e.,

$$\mathbb{E}_p [p(t)^n] = \mathbb{E}_p [\psi(p(t), n)] = \mathbb{E}_n [\psi(p, n(t))] = \mathbb{E}_n [p^{n(t)}].$$
Within-Generation Fecundity Variance Polymorphism

Example: Male reproductive success in the Common Side-blotched Lizard

- Three male morphs differentiated by heritable throat color and mating strategy:
  - orange males usurp large territories containing multiple females;
  - blue males defend small territories containing one female;
  - yellow males are female mimics.

- Polymorphism is maintained by frequency-dependent selection (a rock-paper-scissors game: Sinervo & Lively 1996).

- Calsbeek et al. (2002) estimate that $\sigma^2/R^2$ is 0.42 for orange-throated males and 0.22 for blue-throated males.
A Diploid Model with Fecundity Variance Polymorphism

Assumptions

- $N$ diploid individuals.
- Non-overlapping generations.
- Two sexes or mating types (males and females), with $f:m$ sex ratio $\gamma : 1 - \gamma$.
- Two alleles $A_1$ and $A_2$ at an autosomal locus, called the selected locus.
- The diploid genotypes are $A_1A_1$, $A_1A_2$, and $A_2A_2$.
- The genealogy to be described will be that at a linked neutral locus, called the marker locus.
Life cycle:

- Each individual produces a random number of gametes.
- **Fair meiosis**: each gamete is equally likely to inherit either the maternal or paternal copy of the selected locus.
- Recombination between the marker locus and the selected locus occurs with probability $r/2N$.
- Each gamete mutates to $A_i$ with probability $\mu_i/2N$.
- The next generation is formed by sampling $N$ pairs of gametes uniformly at random and without replacement, such that each pair contains one maternal and one paternal gamete (**random mating**).
- The sex of these individuals is determined by choosing $\lfloor N\gamma \rfloor$ of these individuals to be female and the rest to be male.
Fecundity may be sex- and genotype-dependent subject to the following conditions:

- Let $\eta_{ij}^{(\Delta)}(k)$ denote the number of gametes shed by the $k$’th individual with sex $\Delta \in \{m, f\}$ and genotype $A_iA_j$.

- The $\{\eta_{ij}^{(\Delta)}(k)\}$ are independent random variables, which are IID within each sex $\times$ genotype class with

$$
\mathbb{E}[\eta_{ij}^{(\Delta)}(k)] = R^{(\Delta)} \left(1 + \frac{s_{ij}}{2N}\right)
$$

$$
\text{Var}(\eta_{ij}^{(\Delta)}(k)) = \sigma_{ij}^{(\Delta)}
$$

$$
\mathbb{P}\left\{\eta_{ij}^{(f)}(k) > \gamma^{-1}\right\} > 1
$$

$$
\mathbb{P}\left\{\eta_{ij}^{(m)}(k) > (1 - \gamma)^{-1}\right\} > 1.
$$

- The last two conditions could also be replaced by:

$$
\gamma R^{(f)} > 1 \quad \text{and} \quad (1 - \gamma) R^{(m)} > 1.
$$
The corresponding diffusion approximation and coalescent depend on the following sex-averaged quantities:

\[
\begin{align*}
    s_{ij} &= \frac{1}{2} \left( s_{ij}^{(f)} + s_{ij}^{(m)} \right), \\
    V_{ij} &= \left( \frac{\sigma_{ij}^{(f)}}{4\gamma(R^{(f)})^2} + \frac{\sigma_{ij}^{(m)}}{4(1 - \gamma)(R^{(m)})^2} \right), \\
    R &= \left( \frac{1}{4\gamma R^{(f)}} + \frac{1}{4(1 - \gamma)R^{(m)}} \right)^{-1}.
\end{align*}
\]

It will also be useful to introduce

\[
\bar{s} = \frac{1}{2} \left( s_{11} + s_{22} \right) \quad \text{and} \quad \bar{V} = \frac{1}{2} \left( V_{11} + V_{22} \right).
\]
The diffusion approximation for the processes \( p^{(N)}([2Nt]) : t \geq 0 \) has generator

\[
Gf(p) = \frac{1}{2} \left( v_1 + v_2 pq \right) pq f''(p) + \left( \mu_1 q - \mu_2 p + d_1 pq + d_2 (1 - 2p) pq \right) f'(p),
\]

where

- \( v_1 = \frac{1}{4\gamma(1-\gamma)} - \frac{1}{R} + V_{12} = \Psi + V_{12} \) (demographic stochasticity)
- \( v_2 = 4 \left( \bar{V} - V_{12} \right) \) (demographic stochasticity)
- \( d_1 = \frac{1}{2} \left( s_{11} - s_{22} \right) + 2 \left( V_{22} - V_{11} \right) \) (directional selection)
- \( d_2 = \left( s_{12} - \bar{s} \right) + 4 \left( \bar{V} - V_{12} \right) \) (frequency-dependent selection).

Thus, selection favors reduced fecundity variance (Gillespie 1974, 1975, 1977).
The rate of coalescence of two $A_1$ lineages is equal to:

$$\lim_{N \to \infty} 2N \mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled } A_1 \text{ chromosomes are IBD at the marker locus}
\end{array} \right\} =$$

$$= \lim_{N \to \infty} 2N \mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are IBD and both are } A_1
\end{array} \right\}$$

$$= \lim_{N \to \infty} \frac{2N \mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are IBD and both are } A_1
\end{array} \right\}}{\mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are } A_1
\end{array} \right\}}$$

$$= \lim_{N \to \infty} 2N \mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are NRM + IBD, both are } A_1
\end{array} \right\}$$

$$= \lim_{N \to \infty} \frac{2N \mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are NRM + IBD, both are } A_1
\end{array} \right\}}{\mathbb{P}_p \left\{ \begin{array}{l}
\text{two sampled chromosomes are } A_1
\end{array} \right\}}$$

**Notation:** IBD = identical-by-descent (over one generation);
NRM = non-recombinant and non-mutant.
$$\mathbb{P}_p\left\{ \text{two NRM } A_1 \text{ lineages are IBD} \right\} =$$

\[
= \frac{2N - 1}{p(2Np - 1)} \mathbb{E}_p \left[ \frac{N - 1}{2(2N - 1)} \frac{1}{Y^{(f)}(Y^{(f)} - 1)} \left( \sum_{j=1}^{X^{(f)}_{11}} \eta_{11}^{(f,f)}(j) \left( \eta_{11}^{(f,f)}(j) - 1 \right) + \sum_{j=1}^{X^{(f)}_{12}} \eta_{12}^{(f,1)}(j) \left( \eta_{12}^{(f,1)}(j) - 1 \right) \right) \right.
\]

\[
+ \frac{N - 1}{2(2N - 1)} \frac{1}{Y^{(m)}(Y^{(m)} - 1)} \left( \sum_{j=1}^{X^{(m)}_{11}} \eta_{11}^{(m,f)}(j) \left( \eta_{11}^{(m,f)}(j) - 1 \right) \right.
\]

\[
+ \sum_{j=1}^{X^{(m)}_{11}} \eta_{11}^{(m,m)}(j) \left( \eta_{11}^{(m,m)}(j) - 1 \right) + \sum_{j=1}^{X^{(m)}_{12}} \eta_{12}^{(m,1)}(j) \left( \eta_{12}^{(m,1)}(j) - 1 \right) \right]
\]
The limit can be evaluated using the following observations:

- Exchangeability allows each sum to be replaced by the product of the number of terms in the sum and the first term.
- When $N$ is large, $p(t-1) \approx p(t)$, i.e., for any $\epsilon > 0$,
  \[
  \lim_{N \to \infty} \mathbb{P}_p \{ |p(t-1) - p(t)| > \epsilon \} = 0.
  \]
- Similarly, under weak selection and random mating and sex determination,
  \[
  N^{-1}X_{11}^{(f)} \approx \gamma p^2 \\
  N^{-1}X_{12}^{(f)} \approx 2\gamma pq \\
  N^{-1}Y^{(f)} \approx \gamma R^{(f)} \\
  N^{-1}X_{12}^{(m)} \approx (1-\gamma)p^2 \\
  N^{-1}X_{12}^{(m)} \approx 2(1-\gamma)pq \]
  \[
  N^{-1}Y^{(m)} \approx (1-\gamma)R^{(m)}.
  \]

Together, these imply that
\[
\lim_{N \to \infty} 2N \mathbb{P}_p \{ \text{two } A_1 \text{ lineages are IBD} \} = \frac{1}{p} \left( \Psi + V_{11}p + V_{12}q \right).
\]
The other rates can be calculated in a similar fashion and are:

<table>
<thead>
<tr>
<th>Transition</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>two $A_1$ lineages coalesce</td>
<td>$(\begin{pmatrix} n_1 \end{pmatrix}) (\psi + pV_{11} + qV_{12}) / p$</td>
</tr>
<tr>
<td>two $A_2$ lineages coalesce</td>
<td>$(\begin{pmatrix} n_2 \end{pmatrix}) (\psi + qV_{22} + pV_{12}) / q$</td>
</tr>
<tr>
<td>a lineage mutates from $A_1$ to $A_2$</td>
<td>$n_1 \mu_1 (q/p)$</td>
</tr>
<tr>
<td>a lineage mutates from $A_2$ to $A_1$</td>
<td>$n_2 \mu_2 (p/q)$</td>
</tr>
<tr>
<td>a lineage recombines from $A_1$ to $A_2$</td>
<td>$n_1 rq$</td>
</tr>
<tr>
<td>a lineage recombines from $A_2$ to $A_1$</td>
<td>$n_2 rp$</td>
</tr>
</tbody>
</table>
Ancestral Processes and Allele Frequency Dynamics

There are infinitely many combinations of fecundity parameters \( \{V_{ij}, s_{ij}\} \) with \( V_{11} \in [0, 2\bar{V}] \) which have the same diffusion approximation:

\[
\begin{align*}
V_{12} &= v_1 - \Psi, \\
\bar{V} &= \frac{1}{4} v_2 + V_{12} \\
s_{11} - \bar{s} + 4V_{11} &= d_1 - 4\bar{V}, \\
s_{12} - \bar{s} &= d_2 - v_2.
\end{align*}
\]
Ancestral Processes and Allele Frequency Dynamics

There are infinitely many combinations of fecundity parameters \( \{ V_{ij}, s_{ij} \} \) with \( V_{11} \in [0, 2\bar{V}] \) which have the same diffusion approximation:

\[
V_{12} = \nu_1 - \psi, \\
\bar{V} = \frac{1}{4}\nu_2 + V_{12} \\
s_{11} - \bar{s} + 4V_{11} = d_1 - 4\bar{V}, \\
s_{12} - \bar{s} = d_2 - \nu_2.
\]

However, fecundity distributions with different values of \( V_{11} \) have distinct coalescent processes:

\[
\binom{n_1}{2} \frac{1}{p} \left( \psi + pV_{11} + qV_{12} \right) \quad \text{and} \quad \binom{n_2}{2} \frac{1}{q} \left( \psi + qV_{22} + pV_{12} \right).
\]
Neutral Dynamics without Exchangeability

Assume that

- $R = 2$
- $\tilde{V} = V_{12} = 1$
- $\tilde{s} = s_{12} = 0$
- $s_{11} = -s_{22} = -2(V_{22} - V_{11})$

Then the diffusion approximation is

$$Gf(p) = \frac{3}{4} pqf''(p) + (\mu_1 q - \mu_2 p)f'(p),$$

but the genealogical process is Kingman’s coalescent only if

$$V_{11} = V_{22} = V_{12}.$$
Heterozygote Advantage and Balancing Selection

Coalescent times are usually elevated at sites linked to a locus under balancing selection:

- Samples are usually polymorphic for the selected locus.
- Lineages can only coalesce within backgrounds.

However, the pattern observed is sensitive to the fecundity variance of the heterozygote.
Fecundity Variance Polymorphism has a Genome-wide Impact on Relatedness.

The genealogy at an unlinked marker locus can be represented by a stochastic time change of Kingman’s coalescent.

- Unlinked lineages rapidly recombine between backgrounds.
- If \( n(t) = n \) is the number of lineages ancestral to a sample, then binary mergers occur at rate

\[
\binom{n}{2} \left[ \Psi + p^2 V_{11} + 2pq V_{12} + q^2 V_{22} \right],
\]

when \( p(t) = p \) is the frequency of \( A_1 \).

![Graph: Mean TMRCA (n = 10)]
A Haploid Model with Fecundity Variance Polymorphism

Assumptions:

- $N$ asexual adults.
- Genotypes $A_1$ and $A_2$ with random offspring numbers satisfying

$$
\mathbb{E} \left[ \eta_i(k) \right] = R \left( 1 + \frac{s_i}{N} \right)
$$

$$
\text{Var} \left( \eta_i(k) \right) = \sigma_i^2
$$

$$
\gamma_i = 1 - \frac{1}{R} + \frac{\sigma_i^2}{R^2}.
$$

- Mutation, but no recombination.
- $N$ offspring are sampled uniformly at random and without replacement.
The corresponding structured coalescent is determined by the following

**Transition Rates:**

<table>
<thead>
<tr>
<th>Transition</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>two $A_1$ lineages coalesce</td>
<td>$\left(\begin{array}{c} n_1 \nonumber \ 2 \nonumber \end{array}\right) \gamma_1 / p$</td>
</tr>
<tr>
<td>two $A_2$ lineages coalesce</td>
<td>$\left(\begin{array}{c} n_2 \nonumber \ 2 \nonumber \end{array}\right) \gamma_2 / q$</td>
</tr>
<tr>
<td>a lineage mutates from $A_1$ to $A_2$</td>
<td>$n_1 \mu_1 (q/p)$</td>
</tr>
<tr>
<td>a lineage mutates from $A_2$ to $A_1$</td>
<td>$n_2 \mu_2 (p/q)$</td>
</tr>
</tbody>
</table>

**Diffusion Approximation (Gillespie 1975):**

$$Gf(p) = \frac{1}{2} pq \left( p \gamma_2 + q \gamma_1 \right) f''(p) + \left( \mu_1 q - \mu_2 p + (s_1 - s_2 + \gamma_2 - \gamma_1) pq \right) f'(p)$$
An Ancestral Influence Graph

The ancestral process can also be embedded in a graph-valued Markov process. Suppose that $\gamma_1 \geq \gamma_2$ and that the graph contains $n$ lineages. The following events can happen:

- **Coalescence:** Each pair of branches coalesces at rate $\gamma_2$.
- **Potential coalescence:** Each pair of branches is connected by a randomly oriented directed edge at rate $(\gamma_1 - \gamma_2)$.
- **C-splits:** Each branch splits in two and one of the outgoing branches is marked $C$ at rate $\frac{1}{2}(n - 1)(\gamma_1 - \gamma_2)$.
- **S-splits:** Each branch splits in two and one of the outgoing branches is marked $S$ at rate $|s_1 - s_2 + \gamma_2 - \gamma_1|$.

The process is run until there is only one branch left in the graph (the ultimate ancestor).
Resolution of Ancestral Lineages

- Assign mutations working from the UA to the leaves.

- When a directed edge is met, the two branches below the edge coalesce to the originating branch if this branch has type 1. Otherwise, the edge is ignored.

- When two branches meet at a C-split, the branch marked C is retained if it has type 2. Otherwise, the other branch is retained.

- When two branches meet at an S-split, the branch marked S is retained if it has the fitter type. Otherwise, the other branch is retained.
Remarks:

- If $\gamma_1 = \gamma_2$, then this process reduces to the Ancestral Selection Graph.
- If $n(t)$ denotes the number of branches at time $t$, then $(n(t) : t \geq 0)$ is a birth-death process with
  
  \begin{align*}
  \text{birth rate:} & \quad \frac{1}{2} (\gamma_1 - \gamma_2) n(n - 1) + |s_1 - s_2 + \gamma_2 - \gamma_1| n \\
  \text{death rate:} & \quad \frac{1}{2} \gamma_2 n(n - 1).
  \end{align*}

- If $\gamma_2 \leq \gamma_1 < 2\gamma_2$, this process is certain to reach the UA in finite time. Otherwise, finite-time blow-up can occur.

- A graphical representation of the genealogy of the diploid model also exists under restrictive conditions.
Summary - Fecundity Variance Polymorphism

- Coalescent rates are background specific and therefore do not simply scale with the variance coefficient of the diffusion approximation.
Summary - Fecundity Variance Polymorphism

- Coalescent rates are background specific and therefore do not simply scale with the variance coefficient of the diffusion approximation.
- In diploid models, there are infinitely many combinations of fecundity distributions that have the same diffusion approximation but distinct genealogical processes.
Summary - Fecundity Variance Polymorphism

- Coalescent rates are background specific and therefore do not simply scale with the variance coefficient of the diffusion approximation.
- In diploid models, there are infinitely many combinations of fecundity distributions that have the same diffusion approximation but distinct genealogical processes.
- Fecundity variance polymorphism has a genome-wide impact on genealogies.
Summary - Fecundity Variance Polymorphism

- Coalescent rates are background specific and therefore do not simply scale with the variance coefficient of the diffusion approximation.
- In diploid models, there are infinitely many combinations of fecundity distributions that have the same diffusion approximation but distinct genealogical processes.
- Fecundity variance polymorphism has a genome-wide impact on genealogies.
- A graphical representation exists but can exhibit finite-time blow-up.
Summary - Fecundity Variance Polymorphism

- Coalescent rates are background specific and therefore do not simply scale with the variance coefficient of the diffusion approximation.

- In diploid models, there are infinitely many combinations of fecundity distributions that have the same diffusion approximation but distinct genealogical processes.

- Fecundity variance polymorphism has a genome-wide impact on genealogies.

- A graphical representation exists but can exhibit finite-time blow-up.

- Similar properties would be expected to hold in models that allow for variation in other traits that affect the coalescent rate:
  - sex ratio
  - fecundity and survival schedules
  - dispersal rates in subdivided populations
The Genealogical Consequences of Fluctuating Selection

Temporally-fluctuating selection has been studied using several approaches:

- Time series analysis of genotype or phenotype frequencies (Fisher & Ford 1947; Lynch 1987; O’Hara 2005).

- Dispersion indices of substitution processes (Gillespie 1979).

- Polymorphism in Poisson random field models (Huerta-Sanchez, Durrett & Bustamante 2008).

Figure 1. The scarlet tiger moth, Panaxia dominula. Top, the typical form (homozygote). Centre, t. medionigra (classical: CM) (heterozygote). Note the absence of the central yellow forewing spot in this specimen. Occasionally a small one is present. Bottom, t. limacuta (homozygote). Photograph: Phil Hurst (NHM). Used with permission of Academic Press and editor, The Linnean.
Suppose that fecundity depends on the environment in the following way,

\[
\mathbb{E}\left[ \eta_{ij}^{(\Delta)}(k; t) \right] = R \left( 1 + Y_i(t) + Y_j(t) \right)
\]

\[
\text{Var}\left[ \eta_{ij}^{(\Delta)}(k) \right] = \sigma^2 + O(N^{-1/2}),
\]

where \( \left( (Y_1^{(N)}(t), Y_2^{(N)}(t)) : t \geq 0 \right) \) is a sequence of IID random vectors satisfying

\[
\mathbb{E}\left[ Y_i^{(N)}(t) \right] = \frac{s_i}{2N}
\]

\[
\mathbb{E}\left[ Y_i^{(N)}(t) Y_j^{(N)}(t) \right] = \frac{\rho_{ij}}{2N}
\]

\[
\mathbb{E}\left[ \left( Y_i^{(N)}(t) \right)^{e_i} \left( Y_j^{(N)}(t) \right)^{e_j} \right] = O(N^{-2}) \quad \text{if } e_i + e_j > 2.
\]
The diffusion approximation has generator (Gillespie 1991):

\[
G_f(p) = \frac{1}{2} \left( v_1 + v_2 pq \right) pqf''(p) + \left( \mu_1 q - \mu_2 p + d_1 pq + d_2 pq(1 - 2p) \right) f'(p),
\]

where

- \( v_1 = \Psi + \sigma^2 / R^2 \) (demographic stochasticity)
- \( v_2 = \rho_{11} - 2\rho_{12} + \rho_{22} \) (environmental stochasticity)
- \( d_1 = s_1 - s_2 + \rho_{22} - \rho_{11} \) (directional selection)
- \( d_2 = v_2 \) (balancing selection)

**Remark:** This has the same form as the diffusion approximation for a diploid model of within-generation fecundity variance polymorphism.
The structured coalescent process has rates:

<table>
<thead>
<tr>
<th>Transition</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>two $A_1$ lineages coalesce</td>
<td>$\binom{n_1}{2} v_1/p$</td>
</tr>
<tr>
<td>two $A_2$ lineages coalesce</td>
<td>$\binom{n_2}{2} v_1/q$</td>
</tr>
<tr>
<td>a lineage mutates from $A_1$ to $A_2$</td>
<td>$n_1 \mu_1 (q/p)$</td>
</tr>
<tr>
<td>a lineage mutates from $A_2$ to $A_1$</td>
<td>$n_2 \mu_2 (p/q)$</td>
</tr>
<tr>
<td>a lineage recombines from $A_1$ to $A_2$</td>
<td>$n_1 rq$</td>
</tr>
<tr>
<td>a lineage recombines from $A_2$ to $A_1$</td>
<td>$n_2 rp$</td>
</tr>
</tbody>
</table>

- The within-background coalescent rates are not directly influenced by fluctuating selection.
- Environmental variation and fecundity variance polymorphism models that have the same diffusion have different coalescent processes.
Example: Symmetric Fluctuating Selection

Fluctuating selection reduces coalescent times and linked neutral variation even when it maintains excess selected variation:

- Heterozygosity at the selected site is increased (here, because heterozygotes have reduced fitness variance).
- Coalescence occurs more quickly in rapidly fluctuating backgrounds.
- Effectively, there are repeated soft sweeps in opposing directions.
Contrasting Effects of Balancing vs. Fluctuating Selection.

Balancing Selection:
- Coalescent times are greatly elevated at tightly linked sites.

Fluctuating Selection:
- Coalescent times are moderately reduced over an extensive range of linked sites.
Although fluctuating selection and fecundity variance polymorphism have the same diffusion approximation, they do so for different reasons:

**Within-generation fecundity variance:**
- Large differences in fecundity variance between backgrounds;
- Individual fecundity is uncorrelated within backgrounds.

**Fluctuating selection:**
- Small $O(N^{-1/2})$ differences in mean fecundity between backgrounds;
- Individual fecundity is correlated within genetic backgrounds.

**Question:** Are there ‘natural’ scenarios that interpolate between these two cases, i.e., by having larger differences in mean fecundity between backgrounds but weaker correlation within backgrounds?
Summary - Fluctuating Selection

The genealogical consequences of fluctuating selection are entirely mediated by the dynamics of the background frequencies.
Summary - Fluctuating Selection

- The genealogical consequences of fluctuating selection are entirely mediated by the dynamics of the background frequencies.
- Fluctuating selection can maintain selected variation at elevated levels but will also reduce the amount of neutral variation segregating at linked sites.

A possible signature for fluctuating selection would be a combination of excess amino acid variation against a background of reduced synonymous and silent variation, e.g., as in *Plasmodium falciparum*. Similar results hold in models with more general fitness schemes and with temporally-autocorrelated environments.
Summary - Fluctuating Selection

- The genealogical consequences of fluctuating selection are entirely mediated by the dynamics of the background frequencies.

- Fluctuating selection can maintain selected variation at elevated levels but will also reduce the amount of neutral variation segregating at linked sites.

- A possible signature for fluctuating selection would be a combination of excess amino acid variation against a background of reduced synonymous and silent variation, e.g., as in *Plasmodium falciparum*. 
Summary - Fluctuating Selection

- The genealogical consequences of fluctuating selection are entirely mediated by the dynamics of the background frequencies.
- Fluctuating selection can maintain selected variation at elevated levels but will also reduce the amount of neutral variation segregating at linked sites.
- A possible signature for fluctuating selection would be a combination of excess amino acid variation against a background of reduced synonymous and silent variation, e.g., as in *Plasmodium falciparum*.
- Similar results hold in models with more general fitness schemes and with temporally-autocorrelated environments.
Conservation of regulatory elements in primates and rodents (Keightley et al. 2005):

<table>
<thead>
<tr>
<th>Element</th>
<th>Primates</th>
<th>Rodents</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' intergenic</td>
<td>0.0016 (0.019)</td>
<td>0.17 (0.016)</td>
</tr>
<tr>
<td>3' intergenic</td>
<td>0.068 (0.018)</td>
<td>0.19 (0.018)</td>
</tr>
<tr>
<td>intron 1, 5' end</td>
<td>-0.0029 (0.019)</td>
<td>0.16 (0.018)</td>
</tr>
</tbody>
</table>

- **Selective constraint:**
  \[ C = 1 - \frac{\sum_i O_i}{\sum_i E_i}, \]
  where \( O_i \) and \( E_i \) are the observed and expected number of substitutions.

- **Conclusion:** Selection on regulatory regions is less effective in primates than in rodents.

- **Interpretation:** This is because primate effective population sizes are smaller than those of rodents, e.g., human \( N_e \approx 10^4 \), rat \( N_e \approx 10^5 \).
Combining Phylogenetics and Population Genetics

The Keightley et al. (2005) study combines phylogenetic and sequence data to explore population genetical processes.

- In this case, phylogenetics provides the tree: \{\{human, chimp\}, \{mouse, rat\}\}.
- Population genetics theory characterizes the interplay between genetic drift, mutation and selection, and is used to interpret the different degrees of divergence on different parts of the tree.
- Here these population genetical processes operate 'along' the branches of the tree, but not on the tree itself: the structure of the tree does not depend on drift/selection in any accessible way.

An important problem at the interface of phylogenetics and population genetics is:

- **How can we relate population genetical processes to divergence along independently-evolving lineages?**
Substitution Rates

The substitution rate is the rate at which new mutations are fixed in a population.

- Substitution rates depend on population size, mutation and selection.
- Divergence between populations or species occurs when different mutations are fixed (or nearly so) in these populations.
- Thus, substitution rates can sometimes be estimated from divergence.
- Even when selection cannot be observed directly, it can sometimes be inferred from the effect that it has on divergence.
A standard approximation for the substitution rate is (Kimura 1964):

$$\text{substitution rate} \approx N\mu \cdot u(1/N) \approx \mu \cdot u'(0)$$

- $N\mu$ is the expected number of new mutations entering the population per generation.
- $u(p)$ is the probability that an allele with initial frequency $p$ is fixed in the population.
- The initial frequency of a new mutant in a haploid population is $1/N$.
- This approximation is reasonable if each new mutation is likely to be lost or fixed before another mutation enters the population.
Absorption Probabilities: General Theory

Suppose that $X = (X(t) : t \geq 0)$ is a Markov process on $[0, 1]$ with generator $G$ that satisfies the following conditions:

- 0 and 1 are both absorbing states.
- The process is certain to absorb at one of these states in finite time.
Absorption Probabilities: General Theory

Suppose that \( X = (X(t) : t \geq 0) \) is a Markov process on \([0, 1]\) with generator \( G \) that satisfies the following conditions:

- 0 and 1 are both absorbing states.
- The process is certain to absorb at one of these states in finite time.

If we define the absorption probability

\[
  u(p) = \mathbb{P}_p \{ X \text{ eventually absorbs at } 1 \},
\]

then \( u(p) \) is the solution to the following boundary-value problem:

\[
  Gu(p) \equiv 0, \quad p \in [0, 1] \\
  u(0) = 0, \quad u(1) = 1.
\]
Absorption Probabilities: Diffusion Processes

For a diffusion process on \([0, 1]\), the absorption probability can be found by solving the following problem:

\[
\frac{1}{2} a(p) u''(p) + b(p) u'(p) \equiv 0, \quad p \in [0, 1]
\]

\[
u(0) = 0, \quad u(1) = 1.
\]
Absorption Probabilities: Diffusion Processes

For a diffusion process on \([0, 1]\), the absorption probability can be found by solving the following problem:

\[
\frac{1}{2} a(p) u''(p) + b(p) u'(p) \equiv 0, \quad p \in [0, 1]
\]

\[u(0) = 0, \quad u(1) = 1.\]

This can be done explicitly, giving

\[
u(p) = \frac{\int_0^p \exp \left( -2 \int_c^y \frac{b(z)}{a(z)} \, dz \right) \, dy}{\int_0^1 \exp \left( -2 \int_c^y \frac{b(z)}{a(z)} \, dz \right) \, dy}.
\]
Example: Fixation of Selected Alleles

Recall that the diffusion approximation for the haploid Wright-Fisher model with selection \((\sigma = Ns)\) has generator

\[
G_f(p) = \frac{1}{2} p(1 - p)f''(p) + \sigma p(1 - p)f'(p).
\]

Then, using the absorption probability formula for a diffusion process, the fixation probability of \(A_1\) is:

\[
u(p) = \begin{cases}
p & \text{if } s = 0 \text{ (neutral allele)} \\
\frac{(1 - e^{-2Ns})}{(1 - e^{-2Nsp})} & \text{if } s \neq 0 \text{ (selected allele)}
\end{cases}
\]

Exercise: Verify the preceding formula.
Usually we are interested in the fixation probability of a new mutation. In this case, the initial frequency of $A$ is $p = 1/N$ and we can calculate

$$u \left( \frac{1}{N} \right) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \quad (s \neq 0).$$

This function is plotted below as a function of the population size $N$. 

![Fixation Probabilities of New Mutants](image)
Approximate Substitution Rates:

Using the approximate formula for substitution rates, we obtain

\[ \rho = \begin{cases} 
\mu & \text{if } s = 0 \text{ (neutral allele)} \\
2Ns\mu / \left(1 - e^{-2Ns}\right) & \text{if } s \neq 0 \text{ (selected allele)} 
\end{cases} \]

- The neutral substitution rate equals the neutral mutation rate.
- The beneficial substitution rate is greater than the beneficial mutation rate and increases with population size.
- The deleterious substitution rate is less than the deleterious mutation rate and decreases with population size.
The Common Ancestor Process

The preceding results are based on an approximation that neglects overlapping mutations. However, we can also use coalescent theory to calculate substitution rates in settings where this assumptions fails.
The Common Ancestor Process

The preceding results are based on an approximation that neglects overlapping mutations. However, we can also use coalescent theory to calculate substitution rates in settings where this assumption fails.

Key ideas:

- At each time, the entire population is descended from a single individual, called the common ancestor, which was alive at some time in the past.
The Common Ancestor Process

The preceding results are based on an approximation that neglects overlapping mutations. However, we can also use coalescent theory to calculate substitution rates in settings where this assumption fails.

Key ideas:

- At each time, the entire population is descended from a single individual, called the common ancestor, which was alive at some time in the past.
- The sequence of common ancestors forms a unique lineage.
The Common Ancestor Process

The preceding results are based on an approximation that neglects overlapping mutations. However, we can also use coalescent theory to calculate substitution rates in settings where this assumption fails.

Key ideas:

- At each time, the entire population is descended from a single individual, called the common ancestor, which was alive at some time in the past.
- The sequence of common ancestors forms a unique lineage.
- Divergence between taxa that have been isolated for a very long time is dominated by mutations on the common ancestor lineages.
The Common Ancestor Process

The preceding results are based on an approximation that neglects overlapping mutations. However, we can also use coalescent theory to calculate substitution rates in settings where this assumptions fails.

Key ideas:

- At each time, the entire population is descended from a single individual, called the common ancestor, which was alive at some time in the past.
- The sequence of common ancestors forms a unique lineage.
- Divergence between taxa that have been isolated for a very long time is dominated by mutations on the common ancestor lineages.
- We will call the process governing the type of the common ancestor the substitution process to the common ancestor. This need not be a Markov process.
Coalescents and Common Ancestors

The structured coalescent and the common ancestor process are related in the following way.
Coalescents and Common Ancestors

The structured coalescent and the common ancestor process are related in the following way.

- If we follow any lineage sufficiently far into the past, it will eventually coalesce with the common ancestor lineage.
Coalescents and Common Ancestors

The structured coalescent and the common ancestor process are related in the following way.

- If we follow any lineage sufficiently far into the past, it will eventually coalesce with the common ancestor lineage.
- Thus, the stationary distribution of the common ancestor lineage is equal to the stationary distribution of the structured coalescent restricted to a single lineage.
Coalescents and Common Ancestors

The structured coalescent and the common ancestor process are related in the following way.

- If we follow any lineage sufficiently far into the past, it will eventually coalesce with the common ancestor lineage.
- Thus, the stationary distribution of the common ancestor lineage is equal to the stationary distribution of the structured coalescent restricted to a single lineage.
- The substitution process to the common ancestor lineage can be identified by time-reversing the single-lineage structured coalescent with respect to this stationary distribution.
The Type Process of a Single Ancestral Lineage

When \( n = 1 \), the variable \((n_1, n_2, p)\) can be replaced by \((z, p)\), where \( z \in \{1, 2\} \) denotes the type of the single ancestral lineage being tracked into the past. The generator of this process can be written as:

\[
\begin{align*}
\Gamma f(1, p) &= \mu_1 \left( \frac{1 - p}{p} \right) [f(2, p) - f(1, p)] + Gf(1, p) \\
\Gamma f(2, p) &= \mu_2 \left( \frac{p}{1 - p} \right) [f(1, p) - f(2, p)] + Gf(2, p),
\end{align*}
\]

where \( G \) is the generator of the diffusion approximation:

\[
Gf(p) = \frac{1}{2} pqf''(p) + \left( \mu_1 q - \mu_2 p + \sigma(p)pq \right) f'(p).
\]
**Task 1:** Our first objective is to identify the stationary distribution, \( \pi(z, p)dp \) of the ancestral process \((z_t, p_t)\). This is characterized by the equation:

\[
\int_0^1 \Gamma f(1, p) \pi(1, p)dp + \int_0^1 \Gamma f(2, p) \pi(2, p)dp = 0 \quad (*)
\]

To solve this, write

\[
\pi(1, p) = h(p)\pi(p)
\]
\[
\pi(2, p) = (1 - h(p))\pi(p),
\]

where

- \( h(p) \) is the conditional probability that the common ancestor is \( A_1 \) given that the frequency of that allele is \( p \).
- \( \pi(p) \) is the density of the stationary distribution of the frequency of \( A_1 \).
Provided that both mutation rates $\mu_1$ and $\mu_2$ are positive, the Wright-Fisher diffusion with mutation and selection has a unique stationary distribution with density

$$\pi(p) = C p^{2\mu_1-1} (1 - p)^{2\mu_2-1} \exp \left( 2 \int_0^p \sigma(q) dq \right),$$

where $C$ is a normalizing constant.

Then, using equation (*) as well as the properties of $\pi$, it can be shown that $h(p)$ is the unique solution to the following boundary value problem:

$$Gh(p) - \left( \mu_2 \left( \frac{p}{1 - p} \right) + \mu_1 \left( \frac{1 - p}{p} \right) \right) h(p) = -\mu_2 \left( \frac{p}{1 - p} \right),$$

$$h(0) = 0, \ h(1) = 1.$$
**Task 2:** We can identify the common ancestor process by time reversal of the process $(z_t, p_t)$ with respect to the stationary measure $\pi(z, p)dp$.

Recalling the work that we did on time reversal of Markov processes, we know that the generator $\tilde{\Gamma}$ of the time-reversed process can be characterized by the following condition:

$$
\sum_{z=1,2} \int_0^1 \Gamma \phi(z, p) \psi(z, p) \pi(z, p) dp = \sum_{z=1,2} \int_0^1 \phi(z, p) \tilde{\Gamma} \psi(z, p) \pi(z, p) dp,
$$

where $\phi, \psi : \{1, 2\} \times [0, 1] \rightarrow \mathbb{R}$ are twice differentiable as functions of $p$. 


Common Ancestor Process: Generator

A bit of algebra shows that this condition will be satisfied if

\[
\tilde{\Gamma}_\psi(1, p) = G\psi(1, p) + p(1 - p) \left( \frac{h'(p)}{h(p)} \right) \psi'(1, p) + \mu_2 \left( \frac{p(1 - h(p))}{(1 - p)h(p)} \right) (\psi(2, p) - \psi(1, p))
\]

\[
\tilde{\Gamma}_\psi(2, p) = G\psi(2, p) - p(1 - p) \left( \frac{h'(p)}{1 - h(p)} \right) \psi'(2, p) + \mu_1 \left( \frac{(1 - p)h(p)}{p(1 - h(p))} \right) (\psi(1, p) - \psi(2, p)).
\]

Remark: Once we know \( h(p) \), the substitution process is also fully determined.
Common Ancestor Process: Substitution Rates

Conditional on $p$, the substitution rates to the common ancestor are given by:

\[
\begin{align*}
\mu_{1}^{CA}(p) &= \mu_1 \left( \frac{h(p)}{p} \frac{1 - p}{1 - h(p)} \right) \\
\mu_{2}^{CA}(p) &= \mu_2 \left( \frac{p}{h(p)} \frac{1 - h(p)}{1 - p} \right).
\end{align*}
\]

- If $\sigma(p) \equiv 0$, then $h(p) \equiv p$ and the substitution rates are just the mutation rates.
- If $\sigma(p) > 0$ for all $p \in [0, 1]$, then $h(p) > p$ for all $p \in [0, 1]$.
- It follows that the beneficial substitution rate is greater than the neutral mutation rate, while the deleterious substitution rate is smaller.
Suppose that the selection coefficient $\sigma(p) = \sigma$ does not depend on frequency. Then the generator of the diffusion approximation is

$$G\phi(p) = \frac{1}{2} p(1 - p)\phi''(p) + (\mu_1(1 - p) - \mu_2 p + \sigma p(1 - p))\phi'(p),$$

and to find $h(p)$ we need to solve the boundary value problem

$$Gh(p) - \left(\mu_2 \left(\frac{p}{1 - p}\right) + \mu_1 \left(\frac{1 - p}{p}\right)\right) h(p) = -\mu_2 \left(\frac{p}{1 - p}\right)$$

$$h(0) = 0, \ h(1) = 1.$$
Solution by perturbation expansion

If \( h(p) \) is expanded in a power series in \( \sigma \),

\[
h(p) = p + \sum_{n \geq 1} h_n(p) \sigma^n,
\]

then \( h_n(p) \) satisfies the following recursion,

\[
\frac{1}{2} p(1 - p) h_n''(p) + (\mu_1 (1 - p) - \mu_2 p) h_n'(p) - \left( \mu_2 \left( \frac{p}{1 - p} \right) + \mu_1 \left( \frac{1 - p}{p} \right) \right) h_n(p) = -p(1 - p) h_{n-1}'(p),
\]

subject to the conditions \( h_n(0) = h_n(1) = 0 \).
Writing $\beta(p)$ for the incomplete Beta function with parameters $2\mu_1$ and $2\mu_2$,

$$\beta(p) = \int_0^p q^{2\mu_1-1}(1 - q)^{2\mu_2-1} dq,$$

the solution to each equation in the recursion is:

$$h_n(p) = \frac{1}{\beta'(p)} \left[ \frac{\beta(p)}{\beta(1)} \int_0^1 \beta'(q) h_{n-1}(q) dq - \int_0^p \beta'(q) h_{n-1}(q) dq \right].$$

Furthermore, we can use this formula to show that the function $H(p) = \beta(p) h(p)$ satisfies a first-order differential equation:

$$\frac{d}{dp} H(p) = 1 + \sigma \frac{\beta'(p)}{\beta(1)} \int_0^1 \beta'(q) h_{n-1}(q) dq - \sigma H(p) H(0) = 0.$$
Writing $\beta(p)$ for the incomplete Beta function with parameters $2\mu_1$ and $2\mu_2$,

$$\beta(p) = \int_0^p q^{2\mu_1-1}(1-q)^{2\mu_2-1}dq,$$

the solution to each equation in the recursion is:

$$h_n(p) = \frac{1}{\beta'(p)} \left[ \frac{\beta(p)}{\beta(1)} \int_0^1 \beta'(q)h_{n-1}(q)dq - \int_0^p \beta'(q)h_{n-1}(q)dq \right].$$

Furthermore, we can use this formula to show that the function $H(p) = \beta(p)h(p)$ satisfies a first-order differential equation:

$$\partial_p H(p) = 1 + \frac{\sigma \beta'(p)}{\beta(1)} \int_0^1 H(q)dq - \sigma H(p)$$

$$H(0) = 0.$$
Solving for $H(p)$, we obtain the following expression for $h(p)$:

$$h(p) = p + 2\sigma \int_0^p (\tilde{p} - q) e^{2\sigma(q-p)} \left(\frac{q}{p}\right)^{2\mu_1} \left(\frac{1-q}{1-p}\right)^{2\mu_2} dq,$$

where

$$\tilde{p} = \frac{\int_0^1 e^{2\sigma q} q^{2\mu_1+1} (1-q)^{2\mu_2} dq}{\int_0^1 e^{2\sigma q} q^{2\mu_1} (1-q)^{2\mu_2} dq}$$

is the conditional probability that a sample of three individuals contains only $A_1$ alleles given that it contains at least two $A_1$ alleles.

**Question:** In effect, the perturbation approach works because it finds an integrating factor for the original second-order differential equation. Can this approach be extended to the general case with frequency-dependent selection?
Notice that the relative substitution rate from $A_1$ to $A_2$:

- decreases with $\sigma$ and $p$;
- increases with $\mu_1 = \mu_2 = \mu$. 
Mean Substitution Rates under Genic Selection

Mean substitution rates can be defined as:

\[
\mu_{CA1} = \frac{\int_0^1 \mu_{CA1}(p) h(p) \pi(p) dp}{\int_0^1 h(p) \pi(p) dp}
\]

\[
\mu_{CA2} = \frac{\int_0^1 \mu_{CA2}(p)(1 - h(p)) \pi(p) dp}{\int_0^1 (1 - h(p)) \pi(p) dp}
\]

In general, the usual approximation underestimates both substitution rates.

The underestimate is greater for deleterious substitution.

The error increases both with \( \mu \) and \( \sigma \).
Comparison with the ASG

The common ancestor process for the Wright-Fisher distribution with genic selection has also been characterized with the ancestral selection graph.

Fearnhead (2002) shows that, at stationarity, the probability that the common ancestor is of type $A_1$ and that there are $n$ virtual lineages is

$$
\pi(1; n) \equiv \left( \prod_{i=1}^{n} \lambda_i \right) \int_{0}^{1} p(1-p)^n \pi(p) dp,
$$

where the constants $\lambda_n$ satisfy the recursion:

$$
\lambda_{n-1} = \frac{2s}{n + 2(\mu_2 + \mu_1) + 2s - (n + \mu_2)\lambda_n}
$$

$$
\lim_{n \to \infty} \lambda_n = 0.
$$
A direct calculation shows that

\[ h(p) = p + \sum_{n \geq 1} \left( \prod_{i=1}^{n} \lambda_i \right) p(1 - p)^n. \]

This identity has several useful implications.

\[
\pi_1 = \int_{0}^{1} h(p) \pi(p) dp = \sum_{n \geq 0} \int_{0}^{1} \left( \prod_{i=1}^{n} \lambda_i \right) p(1 - p)^n \pi(p) dp \\
\quad = \sum_{n \geq 0} \pi(1, n).
\]

In other words, the marginal distribution of the common ancestor is the same under both representations.

**Problem:** Can one prove correspondence for the substitution process as well?
We can also give an explicit formula for the $\lambda_n$:

$$
\lambda_n = -\left( \frac{V^{(n)}(1)}{V^{(n-1)}(1)} \right),
$$

where $V(p) = (h(p) - p)/p$ and $V^{(n)}(p) = \frac{d^n}{dp^n} V(p)$.

In particular, this allows us to calculate the first constant:

$$
\lambda_1 = \frac{s}{1 + 2\mu_2}(1 - \bar{p}).
$$

Remark: This result can be used to recursively calculate the $\{\lambda_n\}$ starting at $n = 1$, without having to make the approximation $\lambda_n \approx 0$ for some large $n$. 
Common Ancestor Process in a Diploid Population

Assumptions:

- Diploid population containing \( N \) individuals.
- Random mating and Wright-Fisher resampling.
- The relative fitnesses of the genotypes \( A_1A_1, A_1A_2, \) and \( A_2A_2 \) are:

\[
1 + 2s : 1 + 2ds : 1.
\]

- Here \( d \) is the dominance coefficient.

If time is measured in units of \( 2N \) generations and \( s = \sigma/2N \), then the diffusion approximation has generator

\[
Gf(p) = \frac{1}{2}pqf''(p) + \left( \mu_1 q - \mu_2 p + 2\sigma(d - (2d - 1)p)pq \right)f'(p).
\]
Notice that selection can be frequency-dependent in this model.

- If $d = 1/2$, then the selection coefficient is simply $\sigma$ (no frequency dependence).
- If $d > 1$ and $s > 0$, then rare alleles are favored (balancing selection).
- If $d < 0$ and $s > 0$, then common alleles are favored (disruptive selection).

Unfortunately, there are fewer analytical results in this case. However, the differential equation specifying $h(p)$ can be solved numerically using the shooting method.

**Caveat:** Because the equation is singular, some care is required when solving it numerically.
In this case the relative substitution rate from $A_1$ to $A_2$:

- increases with $d$ when $A_1$ is common;
- decreases with $d$ when $A_2$ is common.
Mean Substitution Rates in a Diploid Population

- The mean deleterious substitution rate increases with $d$.
- In contrast, the beneficial substitution rate is either decreasing or unimodal as a function of $d$.
- The deleterious substitution rates are again underestimated by the usual approximation, but the relationship to the beneficial substitution rates is more complex.
Acknowledgements:

- Alison Etheridge, Simon Myers (Oxford)
- Nick Barton (IST Austria)
- Ellen Baake (University of Bielefeld)
Convergence of Stochastic Processes

**Question:** What does it mean for a sequence of Markov processes \((X^N)\) to converge to a Markov process \(X\)?

**A negative answer:** In general, this does not mean that the actual values of these processes converge, i.e., the identity,

\[
\lim_{N \to \infty} X^N(t) = X(t),
\]

usually is not even well-defined. This is because the processes \(X^N\) and \(X\) are typically constructed on separate probability spaces.

Rather, when we discuss convergence of Markov processes, we usually have convergence of certain probability distributions in mind.
Convergence in Distribution

**Definition:** We say that a sequence of random variables, $(X^N)$, converges in distribution to a random variable $X$ if the following condition,

$$\lim_{N \to \infty} \mathbb{E}[f(X^N)] = \mathbb{E}[f(X)],$$

is satisfied for every bounded, continuous function $f : E \to \mathbb{R}$. In this case, we write that

$$X^N \xrightarrow{d} X$$

as $N$ tends to infinity, the $d$ over the arrow standing for ‘distribution’.
Example: Suppose that $X$ and $\{X^{(N)} : N \geq 1\}$ are random variables with the following (degenerate) distributions:

$$P\{X = 0\} = P\{X^{(N)} = 1/N\} = 1.$$  

Then $X^{(N)}$ converges in distribution to $X$ as $N \to \infty$. Indeed, if $f : \mathbb{R} \to \mathbb{R}$ is any continuous function, then

$$\lim_{N \to \infty} E[f(X^N)] = \lim_{N \to \infty} f(1/N) = f(0) = E[f(X)].$$

Caveat: Weak convergence does not necessarily imply convergence of probabilities. For example, if $A = \{0\}$, then

$$P\{X \in A\} = 1 \neq 0 = \lim_{N \to \infty} P\{X^{(N)} \in A\}.$$
Convergence of Markov Processes

**Definition:** Suppose that \((X^N)\) is a sequence of \(E\)-valued continuous-time stochastic processes and let \(X\) be another \(E\)-valued Markov process. We say that the finite-dimensional distributions (FDD's) of \(X^N\) converge weakly to those of \(X\) if for every positive integer \(N\) and every finite set \(0 \leq t_1 < t_2 < \cdots < t_n\),

\[
(X^N(t_1), \ldots, X^N(t_n)) \xrightarrow{d} (X(t_1), \ldots, X(t_n)),
\]

as \(N\) tends to infinity.

**Remarks:**
- Convergence to a diffusion approximation usually implies convergence of FDD’s.
- The processes \(X^N\) are not required to be Markov.
- There are also stronger forms of convergence in which we think of the process as a path-valued random variable.