

# Diffusion Processes in Population Genetics

Jay Taylor

Hilary Term 2009

# Contents I

- 1 Introduction
- 2 Markov Chains - Review
- 3 Two Classic Models
  - Wright-Fisher Model
  - Moran Model
- 4 Diffusion Approximations and Markov Processes
  - Rationale
  - Generators
  - Diffusion Processes
  - Convergence of Stochastic Processes
- 5 Applications of Generators
  - Forward and Backward Equations
  - Stationary Distributions and Generators
  - Stationary Distributions of Diffusion Processes
- 6 Selection and Genetic Drift
  - Wright-Fisher Model with Selection
  - Hitting Probabilities
  - Fixation Probabilities of Selected Alleles
  - Substitution Rates and Molecular Evolution

# Contents II

## 7 Selection in Diploid Populations

- Diffusion Approximation
- Dominance and Fixation

## 8 Cannings Models

## 9 Effective Population Size

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient

(Derdeyn et al. 2004)

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

Variation differs between genes, populations and species, and is influenced by:

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

Variation differs between genes, populations and species, and is influenced by:

- **mutation**

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAACATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

Variation differs between genes, populations and species, and is influenced by:

- **mutation**
- **recombination**

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

Variation differs between genes, populations and species, and is influenced by:

- **mutation**
- **recombination**
- **demographic stochasticity**

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

Variation differs between genes, populations and species, and is influenced by:

- **mutation**
- **recombination**
- **demographic stochasticity**
- **natural selection**

The aim of **population genetics** is to understand the factors affecting genetic variation in populations.

**Example:** HIV-1 variation within a patient  
(Derdeyn et al. 2004)

vir1	AATAAAATTTGAA
vir2	AATAAAATTTGAA
vir3	AATA <b>GG</b> ATTTAAA
vir4	AATAAAATTTGAA
vir5	AATAAAATTT <b>AAA</b>
vir6	AATAAAATTTGAA
vir7	AATAAC <b>C</b> ATTTAAA
vir8	AATAAAATTT <b>AAA</b>
vir9	AATAAAATTT <b>AAA</b>
vir10	AATAAAATTT <b>AAG</b>

Variation differs between genes, populations and species, and is influenced by:

- **mutation**
- **recombination**
- **demographic stochasticity**
- **natural selection**
- **migration**

Population genetics provided the conceptual framework for the **Modern Synthesis** of Darwinian evolution and Mendelian genetics.

Population genetics provided the conceptual framework for the **Modern Synthesis** of Darwinian evolution and Mendelian genetics.

**Evolution by natural selection (Darwin, 1859):** Heritable traits that increase reproductive success will become more common in a population.

Population genetics provided the conceptual framework for the **Modern Synthesis** of Darwinian evolution and Mendelian genetics.

**Evolution by natural selection (Darwin, 1859):** Heritable traits that increase reproductive success will become more common in a population.

- **Variation within populations** - individuals have different traits (phenotypes).
- **Heritability** - offspring are similar to their parents.
- **Selection** - traits influence reproductive success and survivorship (**fitness**).

Population genetics provided the conceptual framework for the **Modern Synthesis** of Darwinian evolution and Mendelian genetics.

**Evolution by natural selection (Darwin, 1859):** Heritable traits that increase reproductive success will become more common in a population.

- **Variation within populations** - individuals have different traits (phenotypes).
- **Heritability** - offspring are similar to their parents.
- **Selection** - traits influence reproductive success and survivorship (**fitness**).

Examples of traits that can influence fitness:

- body size (thermoregulation)
- immunity/disease resistance
- ornaments (sexual selection)

To fully understand evolution, we need a mechanistic theory that explains how variation is created and inherited. Darwin struggled with these questions and proposed a model of **blending inheritance** - offspring traits are averages of the parental traits.

To fully understand evolution, we need a mechanistic theory that explains how variation is created and inherited. Darwin struggled with these questions and proposed a model of **blending inheritance** - offspring traits are averages of the parental traits.

**Mendelian genetics (Mendel, 1859)** provided a particulate theory of inheritance:

- Traits are determined by **genes**.
- There are finitely many kinds of each gene called **alleles**.
- Different alleles may produce different traits.
- Offspring are similar to their parents because they inherit their genes from their parents.

Mendel's particulate theory of inheritance was eventually explained by **molecular genetics**.

- Genetic information is stored by structures called chromosomes that are made up of **DNA**.
- DNA is a polymer - a molecule that is a sequence of nucleotides, TCAG.
- Genes are translated into **proteins** which in turn are responsible for the physical characteristics of organisms: size, color, behavior, immunity, etc.
- Genetic and phenotypic variation are generated by chemical changes to an individual's DNA called mutations.

Mendel's particulate theory of inheritance was eventually explained by **molecular genetics**.

- Genetic information is stored by structures called chromosomes that are made up of **DNA**.
- DNA is a polymer - a molecule that is a sequence of nucleotides, TCAG.
- Genes are translated into **proteins** which in turn are responsible for the physical characteristics of organisms: size, color, behavior, immunity, etc.
- Genetic and phenotypic variation are generated by chemical changes to an individual's DNA called mutations.

Most species are either:

- **haploid** - they have a single copy of each chromosome, or
- **diploid** - they have two copies of each chromosome, usually one inherited from each parent.

Rather than study changes in the frequencies of traits, population genetics focuses on the underlying genetic variation: How do allele frequencies change over time?

Rather than study changes in the frequencies of traits, population genetics focuses on the underlying genetic variation: How do allele frequencies change over time?



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

Rather than study changes in the frequencies of traits, population genetics focuses on the underlying genetic variation: How do allele frequencies change over time?

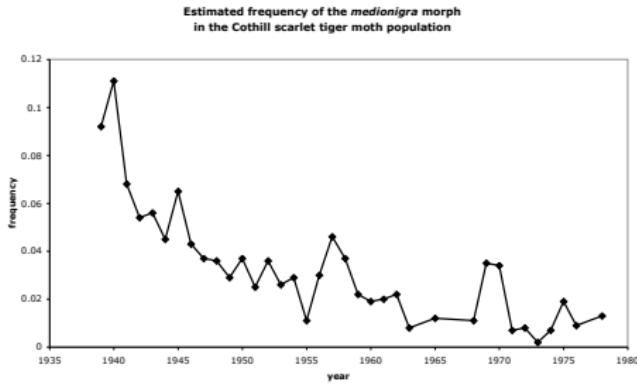


Figure 1. The scarlet tiger moth, *Panaxia dominula*. Top, the typical form (homozygote). Centre, f. *medionigra* (classical CM) (heterozygote). Note the absence of the central yellow forewing spot in this specimen. Occasionally a small one is present. Bottom, f. *bimacula* (homozygote). Photograph: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

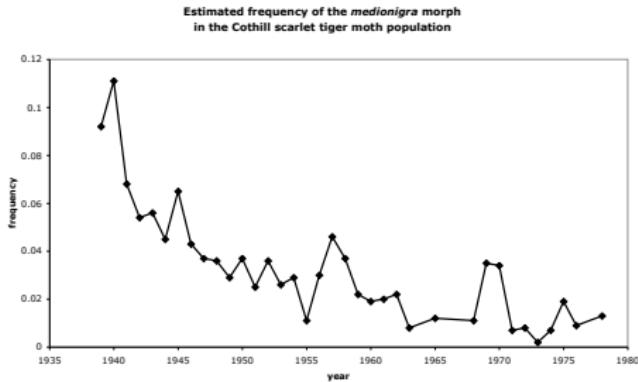
**Example:** Wing color variation in the scarlet tiger moth (*Callimorpha dominula*). Three different morphs occur and breeding experiments have shown that the wing color pattern is a simple Mendelian trait:

- *dominula* - AA
- *medionigra* - Aa
- *bimacula* - aa

The frequency of the *medionigra* morph in the Cothill Fen population in Oxfordshire was estimated from population samples collected annually between 1939 and 1979 (O'Hara, 2005).



The frequency of the *medionigra* morph in the Cothill Fen population in Oxfordshire was estimated from population samples collected annually between 1939 and 1979 (O'Hara, 2005).



Questions:

- Why is there a downward trend in the frequency?
- What accounts for the fluctuations around that trend?

## Markov Chains - A Brief Review

Let  $E = \{e_1, \dots, e_n\}$  be a finite set.

## Markov Chains - A Brief Review

Let  $E = \{e_1, \dots, e_n\}$  be a finite set.

### 1.) Discrete-time Markov chains:

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  $A \subset E$ :

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

In other words, if we know the present, then the past provides no additional information that can be used to predict the future.

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

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

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

$$\mathbb{P}\{X_{n+k} = e_j | X_n = e_i\} = (P^k)_{ij}.$$

## 2.) Continuous-time Markov chains:

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  $A \subset E$ :

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

## 2.) Continuous-time Markov chains:

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  $A \subset E$ :

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

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

$$\begin{aligned} q_{ij} &= \lim_{s \downarrow 0} \frac{1}{s} \mathbb{P}\{X_{t+s} = e_j | X_t = e_i\} \text{ if } j \neq i \\ q_{ii} &= - \sum_{j \neq i} q_{ij} \end{aligned}$$

## 2.) Continuous-time Markov chains:

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  $A \subset E$ :

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

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

$$\begin{aligned} q_{ij} &= \lim_{s \downarrow 0} \frac{1}{s} \mathbb{P}\{X_{t+s} = e_j | X_t = e_i\} \text{ if } j \neq i \\ q_{ii} &= - \sum_{j \neq i} q_{ij} \end{aligned}$$

The transition probabilities are given by the exponential of the rate matrix:

$$\mathbb{P}\{X_{t+s} = e_j | X_t = e_i\} = (P(s))_{ij},$$

where

$$P(s) = e^{Qs} \equiv \sum_{n \geq 0} \frac{1}{n!} Q^n s^n.$$

There are two equivalent ways to construct a continuous-time Markov chain.

(i) Given  $X_t = e_i$ , simulate a collection of independent, exponentially distributed random variables,  $\tau_j, j \neq i$ , where  $\tau_j$  has mean  $q_{ij}^{-1}$ . If

$$\tau \equiv \min\{\tau_j : j \neq i\} = \tau_k,$$

then

$$\begin{aligned} X_{t+s} &= e_i \text{ for } s \in [0, \tau) \\ X_{t+\tau} &= e_k. \end{aligned}$$

In other words,  $X_t$  jumps from state  $i$  to state  $j$  at (exponential) rate  $q_{ij}$ .

(ii) Alternatively, simulate one exponentially-distributed random variable,  $\tau$ , with mean  $(-q_{ii})^{-1}$ , and choose  $X_{t+\tau} \in E$  according to the distribution

$$p_{ij} \equiv \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} \text{ if } j \neq i$$

Notice that  $P = (p_{ij})$  is a stochastic matrix and so determines a discrete-time Markov chain. This chain is called the **embedded Markov chain** and is obtained by considering the continuous-time process at the jump times.

## Integrals and Expectations

If  $Z$  is a real-valued random variable with distribution  $P(dz)$ , then

$$\mathbb{P}\{Z \in A\} = \int_A P(dz),$$

for  $A \subset \mathbb{R}$ .

## Integrals and Expectations

If  $Z$  is a real-valued random variable with distribution  $P(dz)$ , then

$$\mathbb{P}\{Z \in A\} = \int_A P(dz),$$

for  $A \subset \mathbb{R}$ .

Likewise, by the **Law of the Unconscious Statistician**,

$$\begin{aligned}\mathbb{E}[f(Z)] &= \int_{\mathbb{R}} f(z)P(dz) \\ &= \int_{\mathbb{R}} f(z)p(z)dz \quad \text{if } Z \text{ has density } p(z) \\ &= \sum_{i=1}^N f(z_i)\mathbb{P}\{Z = z_i\} \quad \text{if } Z \text{ is discrete}\end{aligned}$$

## Two classic models in population genetics:

## Two classic models in population genetics:

### Wright-Fisher model

## Two classic models in population genetics:

### Wright-Fisher model

Assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.
- Generation  $t + 1$  is formed from generation  $t$  by choosing the parent of each individual uniformly at random and with replacement.
- Two alleles  $A_1$  and  $A_2$ .

## Two classic models in population genetics:

### Wright-Fisher model

Assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.
- Generation  $t + 1$  is formed from generation  $t$  by choosing the parent of each individual uniformly at random and with replacement.
- Two alleles  $A_1$  and  $A_2$ .

**Intuition:** Each parent gives birth to a large number of offspring, but only  $N$  offspring survive to form the next generation.

If  $X(t)$  denotes the number of individuals of type  $A_1$  in the  $t$ 'th generation, then  $(X(t), t \geq 0)$  is a discrete-time Markov chain. Furthermore, the distribution of  $X(t + 1)$  conditional on  $X(t) = k$  is  $\text{Binomial}(N, p)$ , where  $p = k/N$  is the frequency of  $A_1$ :

$$P\{X(t + 1) = m | X(t) = Np\} = \binom{N}{m} p^m (1 - p)^{N-m}.$$

If  $X(t)$  denotes the number of individuals of type  $A_1$  in the  $t$ 'th generation, then  $(X(t), t \geq 0)$  is a discrete-time Markov chain. Furthermore, the distribution of  $X(t+1)$  conditional on  $X(t) = k$  is  $\text{Binomial}(N, p)$ , where  $p = k/N$  is the frequency of  $A_1$ :

$$P\{X(t+1) = m | X(t) = Np\} = \binom{N}{m} p^m (1-p)^{N-m}.$$

However, usually we are only interested in the frequency of  $A_1$  and we define the Wright-Fisher process  $(p(t), t \geq 0)$  to be

$$p(t) \equiv \frac{X(t)}{N}.$$

Of course,  $(p(t) : t \geq 0)$  is also a discrete-time Markov chain.

If  $\delta = p(t+1) - p(t)$  is the change in the allele frequency over generation  $t$ , then

$$\mathbb{E}[\delta | p(t) = p] = 0$$

$$\mathbb{E}[\delta^2 | p(t) = p] = \frac{1}{N^2} Np(1-p) = \frac{p(1-p)}{N}.$$

If  $\delta = p(t+1) - p(t)$  is the change in the allele frequency over generation  $t$ , then

$$\mathbb{E}[\delta | p(t) = p] = 0$$

$$\mathbb{E}[\delta^2 | p(t) = p] = \frac{1}{N^2} Np(1-p) = \frac{p(1-p)}{N}.$$

Thus, two key properties of the **neutral** Wright-Fisher process are:

- The expected allele frequency is constant.
- The variance of the allele frequency fluctuations is inversely proportional to the population size.

One measure of genetic diversity at a **biallelic** locus is the quantity

$$H(p) = 2p(1 - p),$$

which is the probability that a sample of two individuals, chosen with replacement, contains both alleles.

One measure of genetic diversity at a **biallelic** locus is the quantity

$$H(p) = 2p(1 - p),$$

which is the probability that a sample of two individuals, chosen with replacement, contains both alleles.

To see how  $H(p)$  changes from generation to generation in the Wright-Fisher model, observe that

$$\begin{aligned}\mathbb{E}_p[H(p(1))] &= 2\mathbb{E}_p[p(1) - p(1)^2] \\ &= 2\mathbb{E}_p[p(1) + p^2 - 2pp(1) - (p(1) - p)^2] \\ &= (1 - 1/N)2p(1 - p) \\ &= (1 - 1/N)H(p(0)).\end{aligned}$$

By induction on  $t$ , it follows that

$$\mathbb{E}_p[H(p(t))] = (1 - 1/N)^t 2p(1 - p),$$

and so the expected diversity decreases geometrically at a rate that is inversely related to the  $N$ .

By induction on  $t$ , it follows that

$$\mathbb{E}_p[H(p(t))] = (1 - 1/N)^t 2p(1 - p),$$

and so the expected diversity decreases geometrically at a rate that is inversely related to the  $N$ .

The random process by which allele frequencies change is often called **genetic drift**. Genetic drift tends to remove genetic variation from populations. Furthermore, smaller populations loose variation more rapidly than larger populations.

Observe that  $p = 0$  and  $p = 1$  are **absorbing states** for the Wright-Fisher process. In the absence of mutation, an inevitable consequence of genetic drift is that one of the two alleles will be lost. When this happens, the surviving allele is said to be **fixed** in the population.

Observe that  $p = 0$  and  $p = 1$  are **absorbing states** for the Wright-Fisher process. In the absence of mutation, an inevitable consequence of genetic drift is that one of the two alleles will be lost. When this happens, the surviving allele is said to be **fixed** in the population.

**Theorem:** Let  $\tau = \inf\{t \geq 0 : p(t) \in \{0, 1\}\}$  be the first time that one of the two alleles is fixed in the population. Then  $\tau$  is almost surely finite, i.e.,  $\mathbb{P}\{\tau < \infty\} = 1$ .

Observe that  $p = 0$  and  $p = 1$  are **absorbing states** for the Wright-Fisher process. In the absence of mutation, an inevitable consequence of genetic drift is that one of the two alleles will be lost. When this happens, the surviving allele is said to be **fixed** in the population.

**Theorem:** Let  $\tau = \inf\{t \geq 0 : p(t) \in \{0, 1\}\}$  be the first time that one of the two alleles is fixed in the population. Then  $\tau$  is almost surely finite, i.e.,  $\mathbb{P}\{\tau < \infty\} = 1$ .

**Proof:** Let  $\kappa(p) = \mathbb{P}\{p(1) \in \{0, 1\} | p\}$  be the probability that  $A_1$  is either lost or fixed at time 1 given that its initial frequency is  $p$ . Then

$$\kappa(p) = p^N + (1 - p)^N > 2^{-N} \equiv \kappa > 0$$

for all  $p \in [0, 1]$ .

By the Markov property, it follows that

$$\mathbb{P}\{\tau > t\} < (1 - \kappa)^t,$$

and so

$$\mathbb{P}\{\tau = \infty\} = \lim_{t \rightarrow \infty} (1 - \kappa)^t = 0.$$

Thus, fixation occurs in finite time, as claimed.  $\square$

By the Markov property, it follows that

$$\mathbb{P}\{\tau > t\} < (1 - \kappa)^t,$$

and so

$$\mathbb{P}\{\tau = \infty\} = \lim_{t \rightarrow \infty} (1 - \kappa)^t = 0.$$

Thus, fixation occurs in finite time, as claimed.  $\square$

This proof also implies that the expected time to fixation,

$$\tau(p) \equiv \mathbb{E}_p[\tau],$$

is finite. Let us try to calculate  $\tau(p)$ .

Notice that by conditioning on the value of  $p(1)$  and using the Markov property, we obtain the identity

$$\tau(p) = 1 + \sum_q \mathbb{P}_p\{p(1) = q\}\tau(q).$$

Notice that by conditioning on the value of  $p(1)$  and using the Markov property, we obtain the identity

$$\tau(p) = 1 + \sum_q \mathbb{P}_p\{p(1) = q\}\tau(q).$$

This expresses  $\tau(p)$  as the solution to a certain linear equation. However, this equation can only be easily solved when  $N$  is small. On the other hand, because the variance of  $p(1)$  about  $p(0) = p$  is of order  $p(1 - p)/N$ , we can surmise that when  $N$  is large, the dominant terms in the sum are those in which  $q \approx p$ . This suggests the following approximation.

Suppose that we can expand  $\tau(q)$  in a Taylor series about  $p$ :

$$\begin{aligned}\tau(p) &= 1 + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \tau(p) + \tau'(p)\delta + \frac{1}{2}\tau''(p)\delta^2 \right) \\ &\quad + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \frac{1}{6}\tau'''(p)\delta^3 + O(\delta^4) \right)\end{aligned}$$

Suppose that we can expand  $\tau(q)$  in a Taylor series about  $p$ :

$$\begin{aligned}\tau(p) &= 1 + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \tau(p) + \tau'(p)\delta + \frac{1}{2}\tau''(p)\delta^2 \right) \\ &\quad + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \frac{1}{6}\tau'''(p)\delta^3 + O(\delta^4) \right) \\ &= 1 + \tau(p) + \mathbb{E}_p[\delta]\tau'(p) + \frac{1}{2}\mathbb{E}_p[\delta^2]\tau''(p) + O(N^{-2})\end{aligned}$$

Suppose that we can expand  $\tau(q)$  in a Taylor series about  $p$ :

$$\begin{aligned}
 \tau(p) &= 1 + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \tau(p) + \tau'(p)\delta + \frac{1}{2}\tau''(p)\delta^2 \right) \\
 &\quad + \sum_q \mathbb{P}_p\{p(1) = q\} \left( \frac{1}{6}\tau'''(p)\delta^3 + O(\delta^4) \right) \\
 &= 1 + \tau(p) + \mathbb{E}_p[\delta]\tau'(p) + \frac{1}{2}\mathbb{E}_p[\delta^2]\tau''(p) + O(N^{-2}) \\
 &= 1 + \tau(p) + \frac{p(1-p)}{2N}\tau''(p) + O(N^{-2}),
 \end{aligned}$$

where  $\delta = p(1) - p$ .

Consequently, for large  $N$ , we expect  $\tau(p)$  to approximately satisfy the differential equation,

$$\tau''(p) \approx \frac{-2N}{p(1-p)},$$

with boundary conditions  $\tau(0) = \tau(1) = 0$ .

Consequently, for large  $N$ , we expect  $\tau(p)$  to approximately satisfy the differential equation,

$$\tau''(p) \approx \frac{-2N}{p(1-p)},$$

with boundary conditions  $\tau(0) = \tau(1) = 0$ .

This can be solved explicitly and gives:

$$\tau(p) \approx -2N(p \log(p) + (1-p) \log(1-p)).$$

Thus, for the Wright-Fisher model, the expected time to fixation is of order  $O(2N)$  if neither allele is initially rare.

Suppose that  $f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  is smooth. For some applications, we would like to know how the expected value,

$$u(p, t) \equiv \mathbb{E}_p[f(p(t), t)],$$

changes through time. In principle, this quantity can be determined by raising the transition matrix of  $(p(t))$  to the appropriate power, but unless  $N$  is small, the required calculations are too cumbersome to be of use.

Suppose that  $f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$  is smooth. For some applications, we would like to know how the expected value,

$$u(p, t) \equiv \mathbb{E}_p[f(p(t), t)],$$

changes through time. In principle, this quantity can be determined by raising the transition matrix of  $(p(t))$  to the appropriate power, but unless  $N$  is small, the required calculations are too cumbersome to be of use.

As in the previous example, we can find an approximate solution when  $N$  is large. Let us introduce a new function,

$$U(p, t) \equiv u(p, Nt),$$

which is the expected value of  $f$  when time is measured in units of  $N$  generations.

Then, using the Markov property and supposing that  $U(p, t)$  can be expanded in a Taylor series in both variables,

$$U(p, t) = \sum_q \mathbb{P}_p\{p(1) = q\} U(q, t - 1/N)$$

Then, using the Markov property and supposing that  $U(p, t)$  can be expanded in a Taylor series in both variables,

$$\begin{aligned} U(p, t) &= \sum_q \mathbb{P}_p\{p(1) = q\} U(q, t - 1/N) \\ &= \sum_q \mathbb{P}_p\{p(1) = q\} \left( U(p, t) + (q - p)\partial_p U(p, t) \right. \\ &\quad \left. + \frac{1}{2}(q - p)^2 \partial_p^2 U(p, t) - \frac{1}{N} \partial_t U(p, t) \right) + O(N^{-2}) \end{aligned}$$

Then, using the Markov property and supposing that  $U(p, t)$  can be expanded in a Taylor series in both variables,

$$\begin{aligned}
 U(p, t) &= \sum_q \mathbb{P}_p\{p(1) = q\} U(q, t - 1/N) \\
 &= \sum_q \mathbb{P}_p\{p(1) = q\} \left( U(p, t) + (q - p)\partial_p U(p, t) \right. \\
 &\quad \left. + \frac{1}{2}(q - p)^2 \partial_p^2 U(p, t) - \frac{1}{N} \partial_t U(p, t) \right) + O(N^{-2}) \\
 &= U(p, t) + \frac{p(1-p)}{2N} \partial_p^2 U(p, t) - \frac{1}{N} \partial_t U(p, t) + O(N^{-2}).
 \end{aligned}$$

Thus, for large  $N$ , we expect  $U(p, t)$  to approximately satisfy the following partial differential equation:

$$\partial_t U(p, t) \approx \frac{1}{2}p(1-p)\partial_p^2 U(p, t), \quad (*)$$

with initial value  $U(0, p) = f(p)$ .  $\square$

Thus, for large  $N$ , we expect  $U(p, t)$  to approximately satisfy the following partial differential equation:

$$\partial_t U(p, t) \approx \frac{1}{2}p(1-p)\partial_p^2 U(p, t), \quad (*)$$

with initial value  $U(0, p) = f(p)$ .  $\square$

For example,  $U(p, t) = e^{-t}p(1-p)$  solves  $(*)$  when  $f(p) = p(1-p)$ , which can be compared with the exact solution that we found earlier:

$$\begin{aligned} u(p, Nt) &= \mathbb{E}_p[H(p(Nt))] = (1 - 1/N)^{Nt} p(1-p) \\ &\approx e^{-t}p(1-p). \end{aligned}$$

The second model that we will consider is the **Moran model**:

The second model that we will consider is the **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  $A_2$ .

The second model that we will consider is the **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:

parental genotype	deceased genotype	probability	change in $X(t)$
$A_1$	$A_1$	$p^2$	0
$A_1$	$A_2$	$p(1 - p)$	+1
$A_2$	$A_1$	$p(1 - p)$	-1
$A_2$	$A_2$	$(1 - p)^2$	0

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

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:

parental genotype	deceased genotype	probability	change in $X(t)$
$A_1$	$A_1$	$p^2$	0
$A_1$	$A_2$	$p(1 - p)$	+1
$A_2$	$A_1$	$p(1 - p)$	-1
$A_2$	$A_2$	$(1 - p)^2$	0

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.

$(X(t) : t \geq 0)$  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:

$$\begin{aligned} Q_{k,k+1} &= Np(1-p) \\ Q_{k,k-1} &= Np(1-p) \\ Q_{k,k} &= -2Np(1-p) \\ Q_{k,j} &= 0 \text{ if } j \neq k-1, k, k+1 \end{aligned}$$

$(X(t) : t \geq 0)$  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:

$$\begin{aligned} Q_{k,k+1} &= Np(1-p) \\ Q_{k,k-1} &= Np(1-p) \\ Q_{k,k} &= -2Np(1-p) \\ Q_{k,j} &= 0 \text{ if } j \neq k-1, k, k+1 \end{aligned}$$

As before, we are mainly interested in the frequency process, so we define the Moran model  $(p(t) : t \geq 0)$  by setting  $p(t) = X(t)/N$ . This process has the same rate matrix as  $X(t)$  (only the set  $E$  changes).

For comparison with the Wright-Fisher model, let us calculate the expectation:

$$u(p, t) = \mathbb{E}_p[f(p(t), t)].$$

For comparison with the Wright-Fisher model, let us calculate the expectation:

$$u(p, t) = \mathbb{E}_p[f(p(t), t)].$$

First, using the Markov property and the fact that

$$\mathbb{P}_p \left\{ p(\delta t) = p \pm \frac{1}{N} \right\} = Np(1-p)\delta t + O((\delta t)^2),$$

For comparison with the Wright-Fisher model, let us calculate the expectation:

$$u(p, t) = \mathbb{E}_p[f(p(t), t)].$$

First, using the Markov property and the fact that

$$\mathbb{P}_p \left\{ p(\delta t) = p \pm \frac{1}{N} \right\} = Np(1-p)\delta t + O((\delta t)^2),$$

observe that:

$$u(p, t) = \mathbb{E}_p \left[ \mathbb{E} \left[ f(p(t), t) \middle| p(\delta t) \right] \right]$$

For comparison with the Wright-Fisher model, let us calculate the expectation:

$$u(p, t) = \mathbb{E}_p[f(p(t), t)].$$

First, using the Markov property and the fact that

$$\mathbb{P}_p \left\{ p(\delta t) = p \pm \frac{1}{N} \right\} = Np(1-p)\delta t + O((\delta t)^2),$$

observe that:

$$\begin{aligned} u(p, t) &= \mathbb{E}_p \left[ \mathbb{E} \left[ f(p(t), t) \middle| p(\delta t) \right] \right] \\ &= \mathbb{E}_p \left[ u(p(\delta t), t - \delta t) \right] \quad (\text{by the Markov property}) \end{aligned}$$

For comparison with the Wright-Fisher model, let us calculate the expectation:

$$u(p, t) = \mathbb{E}_p[f(p(t), t)].$$

First, using the Markov property and the fact that

$$\mathbb{P}_p \left\{ p(\delta t) = p \pm \frac{1}{N} \right\} = Np(1-p)\delta t + O((\delta t)^2),$$

observe that:

$$\begin{aligned} u(p, t) &= \mathbb{E}_p \left[ \mathbb{E} \left[ f(p(t), t) \middle| p(\delta t) \right] \right] \\ &= \mathbb{E}_p \left[ u(p(\delta t), t - \delta t) \right] \quad (\text{by the Markov property}) \\ &= Np(1-p)\delta t \cdot u(p + 1/N, t - \delta t) + \\ &\quad Np(1-p)\delta t \cdot u(p - 1/N, t - \delta t) + \\ &\quad \left( 1 - 2Np(1-p)\delta t \right) u(p, t - \delta t) + O((\delta t)^2) \end{aligned}$$

Next, if we expand the terms involving  $u$  in Taylor series, then we obtain:

$$u(p, t) =$$

$$Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) + \frac{1}{N} u_p(p, t - \delta t) + \frac{1}{2N^2} u_{pp}(p, t - \delta t) \right] +$$
$$Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) - \frac{1}{N} u_p(p, t - \delta t) + \frac{1}{2N^2} u_{pp}(p, t - \delta t) \right] +$$
$$\left( 1 - 2Np(1-p)\delta t \right) u(p, t - \delta t) + O(N^{-2}) + O((\delta t)^2)$$

Next, if we expand the terms involving  $u$  in Taylor series, then we obtain:

$$\begin{aligned}
 u(p, t) &= \\
 &Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) + \frac{1}{N}u_p(p, t - \delta t) + \frac{1}{2N^2}u_{pp}(p, t - \delta t) \right] + \\
 &Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) - \frac{1}{N}u_p(p, t - \delta t) + \frac{1}{2N^2}u_{pp}(p, t - \delta t) \right] + \\
 &\left( 1 - 2Np(1-p)\delta t \right)u(p, t - \delta t) + O(N^{-2}) + O((\delta t)^2) \\
 &= u(p, t - \delta t) + \delta t \cdot \frac{1}{N}p(1-p)u_{pp}(p, t - \delta t) + O(N^{-2}) + O((\delta t)^2).
 \end{aligned}$$

Next, if we expand the terms involving  $u$  in Taylor series, then we obtain:

$$\begin{aligned}
 u(p, t) &= \\
 &Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) + \frac{1}{N}u_p(p, t - \delta t) + \frac{1}{2N^2}u_{pp}(p, t - \delta t) \right] + \\
 &Np(1-p)\delta t \cdot \left[ u(p, t - \delta t) - \frac{1}{N}u_p(p, t - \delta t) + \frac{1}{2N^2}u_{pp}(p, t - \delta t) \right] + \\
 &\left( 1 - 2Np(1-p)\delta t \right)u(p, t - \delta t) + O(N^{-2}) + O((\delta t)^2) \\
 &= u(p, t - \delta t) + \delta t \cdot \frac{1}{N}p(1-p)u_{pp}(p, t - \delta t) + O(N^{-2}) + O((\delta t)^2).
 \end{aligned}$$

This can be rearranged to:

$$\frac{u(p, t) - u(p, t - \delta t)}{\delta t} = \frac{1}{N}p(1-p)u_{pp}(p, t - \delta t) + O(N^{-2}) + O(\delta t).$$

Then, letting  $\delta t \rightarrow 0$  gives:

$$\partial_t u(p, t) = \frac{1}{N} p(1-p) \partial_p^2 u(p, t) + O(N^{-2}).$$

Then, letting  $\delta t \rightarrow 0$  gives:

$$\partial_t u(p, t) = \frac{1}{N} p(1-p) \partial_p^2 u(p, t) + O(N^{-2}).$$

Furthermore, if we define  $U(p, t) = u(p, Nt)$  and let  $N \rightarrow \infty$ , then:

$$\partial_t U(p, t) = p(1-p) \partial_p^2 U(p, t).$$

Then, letting  $\delta t \rightarrow 0$  gives:

$$\partial_t u(p, t) = \frac{1}{N} p(1-p) \partial_p^2 u(p, t) + O(N^{-2}).$$

Furthermore, if we define  $U(p, t) = u(p, Nt)$  and let  $N \rightarrow \infty$ , then:

$$\partial_t U(p, t) = p(1-p) \partial_p^2 U(p, t).$$

Notice that:

- The rate of change of  $\mathbb{E}_p[f(p(t), t)]$  is of order  $N^{-1}$ .
- The quantity  $\mathbb{E}_p[f(p(t), t)]$  changes slowly when  $p$  is close to 0 or 1.
- The partial differential equation satisfied by  $U(p, t)$  under the Moran model is the same as that derived for the Wright-Fisher model apart from a factor of 1/2. (The Moran model evolves twice as fast as the Wright-Fisher model.)

## Diffusion Approximations: Rationale

**Problem:** The Wright-Fisher and Moran models are easy to understand, but explicit calculations are usually difficult:

- Exact calculations with the Wright-Fisher process often are impossible because the process can jump between any two frequencies.
- Explicit calculations can sometimes be carried out with the Moran model, but the answers are often difficult to interpret.

## Diffusion Approximations: Rationale

**Problem:** The Wright-Fisher and Moran models are easy to understand, but explicit calculations are usually difficult:

- Exact calculations with the Wright-Fisher process often are impossible because the process can jump between any two frequencies.
- Explicit calculations can sometimes be carried out with the Moran model, but the answers are often difficult to interpret.

**Remedy:** Accurate approximations can often be found by assuming that the population size is very large and then neglecting terms of higher order in  $N^{-1}$ .

## Diffusion Approximations: Rationale

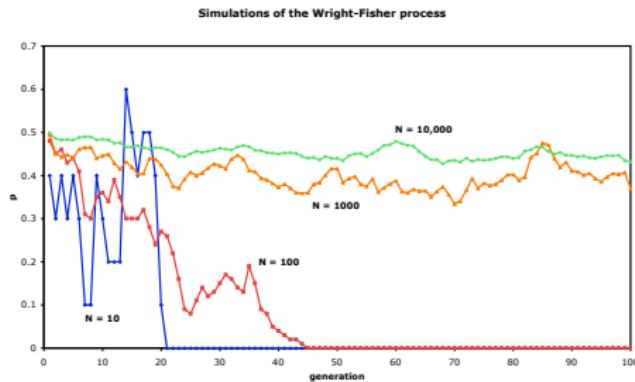
**Problem:** The Wright-Fisher and Moran models are easy to understand, but explicit calculations are usually difficult:

- Exact calculations with the Wright-Fisher process often are impossible because the process can jump between any two frequencies.
- Explicit calculations can sometimes be carried out with the Moran model, but the answers are often difficult to interpret.

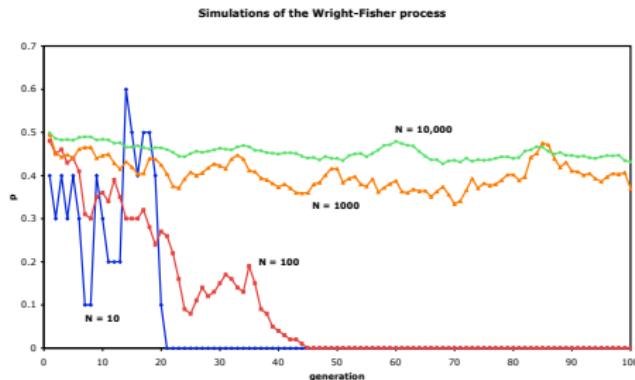
**Remedy:** Accurate approximations can often be found by assuming that the population size is very large and then neglecting terms of higher order in  $N^{-1}$ .

**Question:** Why does this approach work?

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).



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).

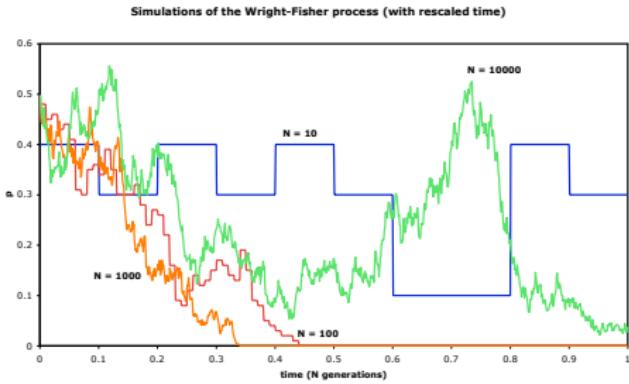


Notice that both the size of the fluctuations and the total change in  $p$  over 100 generations decrease as the population size is increased.

A different picture emerges if we plot each **sample path** against time measured in units of  $N$  generations. Here the total rescaled time is  $t = 1$ .



A different picture emerges if we plot each **sample path** against time measured in units of  $N$  generations. Here the total rescaled time is  $t = 1$ .



In this case, the typical jump sizes shrink as  $N$  increases, but the 'roughness' of the paths increases. This suggests that the following limit should exist:

$$\lim_{N \rightarrow \infty} (p^N(Nt) : t \geq 0) = (p(t) : t \geq 0),$$

where  $(p(t) : t \geq 0)$  is a stochastic process with continuous paths.

## Characterization of Markov Processes

Suppose that  $X = (X_t : t \geq 0)$  is a continuous-time Markov process in a metric space  $E = (E, d)$ .  $X$  can be characterized in several ways:

## Characterization of Markov Processes

Suppose that  $X = (X_t : t \geq 0)$  is a continuous-time Markov process in a metric space  $E = (E, d)$ .  $X$  can be characterized in several ways:

### 1.) Transition function:

The transition function  $P : [0, \infty) \times E \times \mathcal{B}(E) \rightarrow [0, 1]$  is defined by

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

where  $A \subset E$ . Strictly speaking, we require  $A \in \mathcal{B}(E)$ , the sigma algebra of Borel subsets of  $E$ .

## Characterization of Markov Processes

Suppose that  $X = (X_t : t \geq 0)$  is a continuous-time Markov process in a metric space  $E = (E, d)$ .  $X$  can be characterized in several ways:

### 1.) Transition function:

The transition function  $P : [0, \infty) \times E \times \mathcal{B}(E) \rightarrow [0, 1]$  is defined by

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

where  $A \subset E$ . Strictly speaking, we require  $A \in \mathcal{B}(E)$ , the sigma algebra of Borel subsets of  $E$ .

Notice that for fixed  $t$  and  $x$ ,  $P(t, x, dy)$  is a probability distribution on  $E$ :

$$P(t, x, A) = \int_A P(t, x, dy).$$

## Characterization of Markov Processes

Suppose that  $X = (X_t : t \geq 0)$  is a continuous-time Markov process in a metric space  $E = (E, d)$ .  $X$  can be characterized in several ways:

### 1.) Transition function:

The transition function  $P : [0, \infty) \times E \times \mathcal{B}(E) \rightarrow [0, 1]$  is defined by

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

where  $A \subset E$ . Strictly speaking, we require  $A \in \mathcal{B}(E)$ , the sigma algebra of Borel subsets of  $E$ .

Notice that for fixed  $t$  and  $x$ ,  $P(t, x, dy)$  is a probability distribution on  $E$ :

$$P(t, x, A) = \int_A P(t, x, dy).$$

**Remark:** A Markov process is uniquely determined by its transition function.

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Proof:** By the Markov property, we have:

$$P(t+s, x, A) = \mathbb{P}_x\{X_{t+s} \in A\}$$

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Proof:** By the Markov property, we have:

$$\begin{aligned} P(t+s, x, A) &= \mathbb{P}_x\{X_{t+s} \in A\} \\ &= \mathbb{E}_x[\mathbb{P}\{X_{t+s} \in A | X_t\}] \end{aligned}$$

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Proof:** By the Markov property, we have:

$$\begin{aligned} P(t+s, x, A) &= \mathbb{P}_x\{X_{t+s} \in A\} \\ &= \mathbb{E}_x[\mathbb{P}\{X_{t+s} \in A | X_t\}] \\ &= \int_E P(s, y, A)\mathbb{P}_x\{X_t = dy\} \quad (\text{using the Markov property}) \end{aligned}$$

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Proof:** By the Markov property, we have:

$$\begin{aligned} P(t+s, x, A) &= \mathbb{P}_x\{X_{t+s} \in A\} \\ &= \mathbb{E}_x[\mathbb{P}\{X_{t+s} \in A | X_t\}] \\ &= \int_E P(s, y, A)\mathbb{P}_x\{X_t = dy\} \quad (\text{using the Markov property}) \\ &= \int_E P(s, y, A)P(t, x, dy). \quad \square \end{aligned}$$

**Theorem:** The transition function  $P(t, x, dy)$  of a Markov process  $X$  satisfies the **Chapman-Kolmogorov equation**

$$P(t+s, x, A) = \int_E P(s, y, A)P(t, x, dy).$$

**Proof:** By the Markov property, we have:

$$\begin{aligned} P(t+s, x, A) &= \mathbb{P}_x\{X_{t+s} \in A\} \\ &= \mathbb{E}_x[\mathbb{P}\{X_{t+s} \in A | X_t\}] \\ &= \int_E P(s, y, A)\mathbb{P}_x\{X_t = dy\} \quad (\text{using the Markov property}) \\ &= \int_E P(s, y, A)P(t, x, dy). \quad \square \end{aligned}$$

In general, it is difficult to identify the transition function explicitly, although transition density functions sometimes have eigenfunction expansions.

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

### Remarks:

- $T_t$  is called an operator because it maps functions to functions:  $T_t f \in \bar{\mathcal{C}}(E)$ .

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

### Remarks:

- $T_t$  is called an operator because it maps functions to functions:  $T_t f \in \bar{\mathcal{C}}(E)$ .
- Notice that  $T_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)$

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

### Remarks:

- $T_t$  is called an operator because it maps functions to functions:  $T_t f \in \bar{\mathcal{C}}(E)$ .
- Notice that  $T_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)$ .
- The collection of operators  $(T_t : t \geq 0)$  satisfies the **semigroup property**:

$$T_{t+s} f(x) = T_t T_s f(x).$$

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

### Remarks:

- $T_t$  is called an operator because it maps functions to functions:  $T_t f \in \bar{\mathcal{C}}(E)$ .
- Notice that  $T_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)$ .
- The collection of operators  $(T_t : t \geq 0)$  satisfies the **semigroup property**:

$$T_{t+s} f(x) = T_t T_s f(x).$$

- A Markov process is uniquely determined by its transition semigroup.

## 2. Transition semigroup:

Let  $\bar{\mathcal{C}}(E)$  denote the space of bounded continuous functions  $f : E \rightarrow \mathbb{R}$ . For each  $t \geq 0$ , we can define an operator  $T_t : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  by setting

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

where  $x \in E$  and  $f \in \bar{\mathcal{C}}(E)$ .

### Remarks:

- $T_t$  is called an operator because it maps functions to functions:  $T_t f \in \bar{\mathcal{C}}(E)$ .
- Notice that  $T_0 f(x) = \mathbb{E}_x[f(X_0)] = f(x)$ .
- The collection of operators  $(T_t : t \geq 0)$  satisfies the **semigroup property**:

$$T_{t+s} f(x) = T_t T_s f(x).$$

- A Markov process is uniquely determined by its transition semigroup.
- Usually the transition semigroup is not known explicitly.

### 3. Infinitesimal Generator

The infinitesimal generator of a Markov process  $X$  is the operator  $G : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  defined by

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

provided the limit exists for all  $x \in E$  and convergence is uniform over  $E$ .

### 3. Infinitesimal Generator

The infinitesimal generator of a Markov process  $X$  is the operator  $G : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  defined by

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

provided the limit exists for all  $x \in E$  and convergence is uniform over  $E$ .

#### Remarks:

- The generator specifies the rate of change of the expected value of a function evaluated along sample paths of  $X$  started at  $x$ .

### 3. Infinitesimal Generator

The infinitesimal generator of a Markov process  $X$  is the operator  $G : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  defined by

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

provided the limit exists for all  $x \in E$  and convergence is uniform over  $E$ .

#### Remarks:

- The generator specifies the rate of change of the expected value of a function evaluated along sample paths of  $X$  started at  $x$ .
- The set of functions  $f$  for which  $Gf$  is defined is called the **domain** of  $G$  and is denoted  $\mathcal{D}(G)$ .

### 3. Infinitesimal Generator

The infinitesimal generator of a Markov process  $X$  is the operator  $G : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  defined by

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

provided the limit exists for all  $x \in E$  and convergence is uniform over  $E$ .

#### Remarks:

- The generator specifies the rate of change of the expected value of a function evaluated along sample paths of  $X$  started at  $x$ .
- The set of functions  $f$  for which  $Gf$  is defined is called the **domain** of  $G$  and is denoted  $\mathcal{D}(G)$ .
- A Markov process is uniquely determined by its generator.

### 3. Infinitesimal Generator

The infinitesimal generator of a Markov process  $X$  is the operator  $G : \bar{\mathcal{C}}(E) \rightarrow \bar{\mathcal{C}}(E)$  defined by

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

provided the limit exists for all  $x \in E$  and convergence is uniform over  $E$ .

#### Remarks:

- The generator specifies the rate of change of the expected value of a function evaluated along sample paths of  $X$  started at  $x$ .
- The set of functions  $f$  for which  $Gf$  is defined is called the **domain** of  $G$  and is denoted  $\mathcal{D}(G)$ .
- A Markov process is uniquely determined by its generator.
- The generator can often be found explicitly.

**Example:** Suppose that  $X$  is a continuous-time Markov chain on a finite set  $E = \{e_1, \dots, e_n\}$  with rate matrix  $Q$ . Then, for any function  $f : E \rightarrow \mathbb{R}$ ,

$$Gf(e_i) = \lim_{t \rightarrow 0} \frac{1}{t} (E_x[f(X_t)] - f(e_i))$$

**Example:** Suppose that  $X$  is a continuous-time Markov chain on a finite set  $E = \{e_1, \dots, e_n\}$  with rate matrix  $Q$ . Then, for any function  $f : E \rightarrow \mathbb{R}$ ,

$$\begin{aligned} Gf(e_i) &= \lim_{t \rightarrow 0} \frac{1}{t} (E_x[f(X_t)] - f(e_i)) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_j \mathbb{P}_{e_i}\{X_t = e_j\} (f(e_j) - f(e_i)) \end{aligned}$$

**Example:** Suppose that  $X$  is a continuous-time Markov chain on a finite set  $E = \{e_1, \dots, e_n\}$  with rate matrix  $Q$ . Then, for any function  $f : E \rightarrow \mathbb{R}$ ,

$$\begin{aligned} Gf(e_i) &= \lim_{t \rightarrow 0} \frac{1}{t} (E_x[f(X_t)] - f(e_i)) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_j \mathbb{P}_{e_i}\{X_t = e_j\} (f(e_j) - f(e_i)) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_{j \neq i} (q_{ij}t + o(t)) (f(e_j) - f(e_i)) \end{aligned}$$

**Example:** Suppose that  $X$  is a continuous-time Markov chain on a finite set  $E = \{e_1, \dots, e_n\}$  with rate matrix  $Q$ . Then, for any function  $f : E \rightarrow \mathbb{R}$ ,

$$\begin{aligned} Gf(e_i) &= \lim_{t \rightarrow 0} \frac{1}{t} (E_x[f(X_t)] - f(e_i)) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_j \mathbb{P}_{e_i}\{X_t = e_j\} (f(e_j) - f(e_i)) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_{j \neq i} (q_{ij}t + o(t)) (f(e_j) - f(e_i)) \\ &= \sum_{j \neq i} q_{ij} (f(e_j) - f(e_i)). \end{aligned}$$

Thus, for a continuous-time Markov chain, the generator is just a difference operator.

## Application: Generators and Time Changes

## 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**.

## 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:

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

## 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:

$$\begin{aligned}\hat{G}f(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]\end{aligned}$$

## 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:

$$\begin{aligned}\hat{G}f(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(X(\lambda t)) - f(x)] \\ &= \lambda \lim_{t \rightarrow 0} \frac{1}{\lambda t} \mathbb{E}_x[f(X(\lambda t)) - f(x)]\end{aligned}$$

## 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:

$$\begin{aligned}\hat{G}f(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(\hat{X}(t)) - f(x)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x[f(X(\lambda t)) - f(x)] \\ &= \lambda \lim_{t \rightarrow 0} \frac{1}{\lambda t} \mathbb{E}_x[f(X(\lambda t)) - f(x)] \\ &= \lambda Gf(x),\end{aligned}$$

i.e., rescaling a Markov process corresponds to multiplying its generator by a constant.

## Application: Generators and Convergence of Markov Processes

## Application: Generators and Convergence of Markov Processes

Suppose that

- $E$  is a metric space;
- 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  $G$ .

## Application: Generators and Convergence of Markov Processes

Suppose that

- $E$  is a metric space;
- 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  $X^N$  converges to  $X$  by showing that the generators converge:

$$\lim_{N \rightarrow \infty} \sup_{x \in E} |G^N f(x) - Gf(x)| = 0.$$

for any function  $f \in \mathcal{D}(G)$ .

## Application: Generators and Convergence of Markov Processes

Suppose that

- $E$  is a metric space;
- 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  $X^N$  converges to  $X$  by showing that the generators converge:

$$\lim_{N \rightarrow \infty} \sup_{x \in E} |G^N f(x) - Gf(x)| = 0.$$

for any function  $f \in \mathcal{D}(G)$ .

**Remark:** In this case, the process  $X$  can be used as an approximation for  $X^N$  when  $N$  is large.

## Example: Diffusion Approximation for the Moran Model

## Example: Diffusion Approximation for the Moran Model

Now let's examine the generator of the Moran model  $p^N(t)$ :

$$\begin{aligned} G^N f(p) = & Np(1-p)(f(p+1/N) - f(p)) + \\ & Np(1-p)(f(p-1/N) - f(p)). \end{aligned}$$

## Example: Diffusion Approximation for the Moran Model

Now let's examine the generator of the Moran model  $p^N(t)$ :

$$\begin{aligned} G^N f(p) = & Np(1-p)(f(p+1/N) - f(p)) + \\ & Np(1-p)(f(p-1/N) - f(p)). \end{aligned}$$

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:

$$\begin{aligned} G^N f(p) = & Np(1-p) \left( \frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) + \\ & Np(1-p) \left( -\frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) \end{aligned}$$

## Example: Diffusion Approximation for the Moran Model

Now let's examine the generator of the Moran model  $p^N(t)$ :

$$\begin{aligned} G^N f(p) &= Np(1-p)(f(p+1/N) - f(p)) + \\ &\quad Np(1-p)(f(p-1/N) - f(p)). \end{aligned}$$

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:

$$\begin{aligned} G^N f(p) &= Np(1-p) \left( \frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) + \\ &\quad Np(1-p) \left( -\frac{1}{N} f'(p) + \frac{1}{2N^2} f''(p) + O(N^{-3}) \right) \\ &= \frac{1}{N} p(1-p) f''(p) + O(N^{-2}). \end{aligned}$$

This limit vanishes as  $N$  tends to infinity. However, if we consider the rescaled Moran model,  $\hat{p}^N(t) = p^N(Nt)$ , then

$$\begin{aligned}\lim_{N \rightarrow \infty} \hat{G}^N f(p) &= \lim_{N \rightarrow \infty} NG^N f(p) \\ &= p(1-p)f''(p),\end{aligned}$$

and the convergence is uniform in  $p \in [0, 1]$  if  $f$  is smooth.

This limit vanishes as  $N$  tends to infinity. However, if we consider the rescaled Moran model,  $\hat{p}^N(t) = p^N(Nt)$ , then

$$\begin{aligned}\lim_{N \rightarrow \infty} \hat{G}^N f(p) &= \lim_{N \rightarrow \infty} NG^N f(p) \\ &= p(1-p)f''(p),\end{aligned}$$

and the convergence is uniform in  $p \in [0, 1]$  if  $f$  is smooth.

These calculations suggest that the rescaled Moran models ( $p^N(N \cdot)$ ) converge to a Markov process ( $p(\cdot)$ ) with generator

$$Gf(p) = p(1-p)f''(p).$$

Such a process exists and is called a **Wright-Fisher diffusion**.

## One-dimensional Diffusion Processes

## 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:

## 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  $E$ ;

## 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 infinitesimal generator of  $X$  has the form

$$\begin{aligned} Gf(x) &= \lim_{t \rightarrow 0+} \frac{\mathbb{E}_x[f(X(t)] - f(x)}{t} \\ &= \frac{1}{2}a(x)f''(x) + b(x)f'(x) \end{aligned}$$

for every  $f \in \mathcal{C}_c^2(\mathbb{R})$ .

## 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 infinitesimal generator of  $X$  has the form

$$\begin{aligned} Gf(x) &= \lim_{t \rightarrow 0+} \frac{\mathbb{E}_x[f(X(t)] - f(x)}{t} \\ &= \frac{1}{2}a(x)f''(x) + b(x)f'(x) \end{aligned}$$

for every  $f \in \mathcal{C}_c^2(\mathbb{R})$ .

**Remark:** The infinitesimal variance  $a(x)$  must be non-negative.

**Interpretation of  $Gf(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x)$**

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

$$\mathbf{E}_x[X(t) - x] = b(x)t + o(t)$$

**Interpretation of  $Gf(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x)$**

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

$$\mathbf{E}_x[X(t) - x] = b(x)t + o(t)$$

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

$$\mathbf{E}_x[(X(t) - x)^2] = a(x)t + o(t)$$

**Interpretation of  $Gf(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x)$**

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

$$\mathbf{E}_x[X(t) - x] = b(x)t + o(t)$$

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

$$\mathbf{E}_x[(X(t) - x)^2] = a(x)t + o(t)$$

**Remark:** 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.

**Brownian motion** is the canonical example of a diffusion process.

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

- $B_{t+s} - B_t$  is independent of  $(B_u : u \in [0, t])$ .
- $B_{t+s} - B_t$  is normally distributed with mean 0 and variance  $s$ .
- $B$  has continuous sample paths.

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

- $B_{t+s} - B_t$  is independent of  $(B_u : u \in [0, t])$ .
- $B_{t+s} - B_t$  is normally distributed with mean 0 and variance  $s$ .
- $B$  has continuous sample paths.

Observe that:

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

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

- $B_{t+s} - B_t$  is independent of  $(B_u : u \in [0, t])$ .
- $B_{t+s} - B_t$  is normally distributed with mean 0 and variance  $s$ .
- $B$  has continuous sample paths.

Observe that:

$$\begin{aligned} Gf(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \left( \mathbb{E}_x[f(B_t)] - f(x) \right) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x \left[ f'(x)(B_t - x) + \frac{1}{2} f''(x)(B_t - x)^2 + \frac{1}{6} f'''(x)(B_t - x)^3 \right. \\ &\quad \left. + O((B_t - x))^4 \right] \end{aligned}$$

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

- $B_{t+s} - B_t$  is independent of  $(B_u : u \in [0, t])$ .
- $B_{t+s} - B_t$  is normally distributed with mean 0 and variance  $s$ .
- $B$  has continuous sample paths.

Observe that:

$$\begin{aligned} Gf(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \left( \mathbb{E}_x[f(B_t)] - f(x) \right) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x \left[ f'(x)(B_t - x) + \frac{1}{2} f''(x)(B_t - x)^2 + \frac{1}{6} f'''(x)(B_t - x)^3 \right. \\ &\quad \left. + O((B_t - x))^4 \right] \\ &= \frac{1}{2} f''(x). \end{aligned}$$

**Brownian motion** is the canonical example of a diffusion process.

Recall that a real-valued stochastic process  $B = (B_t : t \geq 0)$  is said to be a one-dimensional Brownian motion if:

- $B_{t+s} - B_t$  is independent of  $(B_u : u \in [0, t])$ .
- $B_{t+s} - B_t$  is normally distributed with mean 0 and variance  $s$ .
- $B$  has continuous sample paths.

Observe that:

$$\begin{aligned}
 Gf(x) &= \lim_{t \rightarrow 0} \frac{1}{t} \left( \mathbb{E}_x[f(B_t)] - f(x) \right) \\
 &= \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}_x \left[ f'(x)(B_t - x) + \frac{1}{2} f''(x)(B_t - x)^2 + \frac{1}{6} f'''(x)(B_t - x)^3 \right. \\
 &\quad \left. + O((B_t - x))^4 \right] \\
 &= \frac{1}{2} f''(x).
 \end{aligned}$$

Brownian motion has infinitesimal variance  $a(x) = 1$  and infinitesimal drift  $b(x) = 0$ .

## 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  $B_t$ .

## 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:

$$X_{t+\delta t} \approx X_t + b(X_t)\delta t + \sqrt{a(X_t)}(B_{t+\delta t} - B_t)$$

## Multidimensional Diffusion Processes

## 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:

## 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  $E$ ;

## 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 infinitesimal generator of  $X$  has the form

$$Gf(x) = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{ij} f(x) + \sum_{i=1}^n b_i(x) \partial_i f(x)$$

## 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 infinitesimal generator of  $X$  has the form

$$Gf(x) = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{ij} f(x) + \sum_{i=1}^n b_i(x) \partial_i f(x)$$

**Remark:** The matrix  $a(x) = (a_{ij}(x))$  is called the infinitesimal variance-covariance matrix of  $X$ . This must be non-negative definite for every  $x \in E$ .

## Diffusion Approximation for Discrete-time Markov Chains

## Diffusion Approximation for Discrete-time Markov Chains

The figures that we saw earlier suggest that the Wright-Fisher processes,  $p^N(n)$ , can also be approximated by a diffusion process when  $N$  is large. However, this cannot be justified using the same method that we applied to the Moran model.

## Diffusion Approximation for Discrete-time Markov Chains

The figures that we saw earlier suggest that the Wright-Fisher processes,  $p^N(n)$ , can also be approximated by a diffusion process when  $N$  is large. However, this cannot be justified using the same method that we applied to the Moran model.

- discrete-time processes do not have generators;
- discrete-time processes are only defined at integer times, whereas diffusion processes are defined at all times.

## Diffusion Approximation for Discrete-time Markov Chains

The figures that we saw earlier suggest that the Wright-Fisher processes,  $p^N(n)$ , can also be approximated by a diffusion process when  $N$  is large. However, this cannot be justified using the same method that we applied to the Moran model.

- discrete-time processes do not have generators;
- discrete-time processes are only defined at integer times, whereas diffusion processes are defined at all times.

We can avoid the second problem by considering the piecewise-constant processes:

$$\hat{p}^N(t) = p^N(\lfloor Nt \rfloor), t \geq 0$$

where  $\lfloor x \rfloor$  denotes the greatest integer less than  $x$ . However, these modified processes are not Markov processes.

**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:

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

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

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

**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:

$$\begin{aligned}\lim_{N \rightarrow \infty} \epsilon_N^{-1} \mathbb{E}_x[X^N(1) - x] &= b(x) \\ \lim_{N \rightarrow \infty} \epsilon_N^{-1} \mathbb{E}_x[(X^N(1) - x)^2] &= a(x) \\ \lim_{N \rightarrow \infty} \epsilon_N^{-1} \mathbb{E}_x[(X^N(1) - x)^n] &= 0 \text{ if } n \geq 3.\end{aligned}$$

Then, the interpolated processes  $(X^N(\lfloor \epsilon_N^{-1} t \rfloor) : t \geq 0)$  converge to the diffusion process  $X$  with generator

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

## Example: Diffusion approximation for the Wright-Fisher process

## Example: Diffusion approximation for the Wright-Fisher process

Let  $p^N = (p^N(n) : n \geq 0)$  be the Wright-Fisher process and set  $\epsilon_N = N^{-1}$ . Then,

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[p^N(1) - p] = 0$$

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

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

## Example: Diffusion approximation for the Wright-Fisher process

Let  $p^N = (p^N(n) : n \geq 0)$  be the Wright-Fisher process and set  $\epsilon_N = N^{-1}$ . Then,

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[p^N(1) - p] = 0$$

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

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

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

$$Gf(p) = \frac{1}{2}p(1 - p)f''(p).$$

## Comparison of the Wright-Fisher and Moran models:

## 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:  $G_Mf(p) = p(1-p)f''(p)$ .

## 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:  $G_Mf(p) = p(1-p)f''(p)$ .

Since  $G_Mf(p) = 2G_{WF}f(p)$ , 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:

$$p_M(t) = p_{WF}(2t).$$

## 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:  $G_Mf(p) = p(1-p)f''(p)$ .

Since  $G_Mf(p) = 2G_{WF}f(p)$ , 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.

## Convergence of Stochastic Processes

## Convergence of Stochastic Processes

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

## 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,

$$\lim_{N \rightarrow \infty} X^N(t) = X(t),$$

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

## 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,

$$\lim_{N \rightarrow \infty} X^N(t) = X(t),$$

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.

## Weak Convergence of Probability Distributions

## Weak Convergence of Probability Distributions

**Definition:** Let  $E = (E, d)$  be a metric space and let  $(\mu_N)$  be a sequence of probability distributions on  $E$ . We say that  $\mu_N$  converges **weakly** to a distribution  $\mu$  (also on  $E$ ) if for every bounded continuous function  $f : E \rightarrow \mathbb{R}$ ,

$$\lim_{N \rightarrow \infty} \int_E f(x) \mu_N(dx) = \int_E f(x) \mu(dx).$$

## Weak Convergence of Probability Distributions

**Definition:** Let  $E = (E, d)$  be a metric space and let  $(\mu_N)$  be a sequence of probability distributions on  $E$ . We say that  $\mu_N$  converges **weakly** to a distribution  $\mu$  (also on  $E$ ) if for every bounded continuous function  $f : E \rightarrow \mathbb{R}$ ,

$$\lim_{N \rightarrow \infty} \int_E f(x) \mu_N(dx) = \int_E f(x) \mu(dx).$$

**Example:** Let  $E = \mathbb{R}$ ,  $\mu_N = \delta_{1/N}$  and  $\mu = \delta_0$ . (Here  $\delta_x$  is the probability distribution that assigns all of its mass to the point  $x$ .) Then  $\mu_N$  converges weakly to  $\mu$ . Indeed, if  $f$  is any continuous function on  $\mathbb{R}$ , then

$$\begin{aligned} \lim_{N \rightarrow \infty} \int_E f(x) \mu_N(dx) &= \lim_{N \rightarrow \infty} f(1/N) = f(0) \\ &= \int_E f(x) \mu(dx). \end{aligned}$$

## Convergence in Distribution

## Convergence in Distribution

**Definition:** We say that a sequence of random variables,  $(X^N)$ , converges in distribution to a random variable  $X$  if the probability distributions  $\mu_N$  of the  $X^N$  converge weakly to the probability distribution  $\mu$  of  $X$ . This is equivalent to the condition:

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

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

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

as  $N$  tends to infinity, the  $d$  over the arrow standing for 'distribution'.

## Convergence of Markov Processes

## Convergence of Markov Processes

**Definition:** Suppose that  $(X^N)$  is a sequence of  $E$ -valued continuous-time Markov 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  $0 \leq t_1 < t_2 < \dots < t_n$ ,

$$(X^N(t_1), \dots, X^N(t_n)) \xrightarrow{d} (X(t_1), \dots, X(t_n)),$$

as  $N$  tends to infinity.

## Convergence of Markov Processes

**Definition:** Suppose that  $(X^N)$  is a sequence of  $E$ -valued continuous-time Markov 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  $0 \leq t_1 < t_2 < \dots < t_n$ ,

$$(X^N(t_1), \dots, X^N(t_n)) \xrightarrow{d} (X(t_1), \dots, X(t_n)),$$

as  $N$  tends to infinity.

### Remarks:

- Convergence of generators usually implies convergence of FDD's.
- Convergence to a diffusion approximation usually implies convergence of FDD's.
- There are also stronger forms of convergence involving entire sample paths.

## Applications of Generators

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

then

$$\partial_t u(t, x) = \lim_{s \rightarrow 0} \frac{1}{s} [T_{t+s} f(x) - T_t f(x)]$$

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

then

$$\begin{aligned}\partial_t u(t, x) &= \lim_{s \rightarrow 0} \frac{1}{s} [T_{t+s} f(x) - T_t f(x)] \\ &= \lim_{s \rightarrow 0} \frac{1}{s} [T_s (T_t f(x)) - T_t f(x)] \quad (\text{by the semigroup property})\end{aligned}$$

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

then

$$\begin{aligned}\partial_t u(t, x) &= \lim_{s \rightarrow 0} \frac{1}{s} [T_{t+s} f(x) - T_t f(x)] \\ &= \lim_{s \rightarrow 0} \frac{1}{s} [T_s (T_t f(x)) - T_t f(x)] \quad (\text{by the semigroup property}) \\ &= G(T_t f)(x)\end{aligned}$$

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

then

$$\begin{aligned}\partial_t u(t, x) &= \lim_{s \rightarrow 0} \frac{1}{s} [T_{t+s} f(x) - T_t f(x)] \\ &= \lim_{s \rightarrow 0} \frac{1}{s} [T_s (T_t f(x)) - T_t f(x)] \quad (\text{by the semigroup property}) \\ &= G(T_t f)(x) \\ &= Gu(t, x).\end{aligned}$$

## Applications of Generators

Let  $X = (X_t : t \geq 0)$  be a Markov process on  $E$  with semigroup  $(T_t : t \geq 0)$  and generator  $G$ , and suppose that  $f : E \rightarrow \mathbb{R}$  is continuous. If we define the function,

$$u(t, x) = T_t f(x) = \mathbb{E}_x[f(X_t)],$$

then

$$\begin{aligned}\partial_t u(t, x) &= \lim_{s \rightarrow 0} \frac{1}{s} [T_{t+s} f(x) - T_t f(x)] \\ &= \lim_{s \rightarrow 0} \frac{1}{s} [T_s (T_t f(x)) - T_t f(x)] \quad (\text{by the semigroup property}) \\ &= G(T_t f)(x) \\ &= Gu(t, x).\end{aligned}$$

**Key equation:**  $\partial_t T_t f(x) = G T_t f(x)$ .

Notice that if  $E \subset \mathbb{R}$  and if  $X$  has a transition density  $p(t, x, y)$ , then

$$u(t, x) = \int_E f(y)p(t, x, y)dy.$$

Notice that if  $E \subset \mathbb{R}$  and if  $X$  has a transition density  $p(t, x, y)$ , then

$$u(t, x) = \int_E f(y)p(t, x, y)dy.$$

Thus, provided we can interchange the integral with both  $\partial_t$  and  $G$ , then

$$\int_E f(y)\partial_t p(t, x, y)dy = \int_E f(y)Gp(t, x, y)dy.$$

Notice that if  $E \subset \mathbb{R}$  and if  $X$  has a transition density  $p(t, x, y)$ , then

$$u(t, x) = \int_E f(y)p(t, x, y)dy.$$

Thus, provided we can interchange the integral with both  $\partial_t$  and  $G$ , then

$$\int_E f(y)\partial_t p(t, x, y)dy = \int_E f(y)Gp(t, x, y)dy.$$

However, since this identity holds for all continuous functions  $f$ , it follows that the transition density satisfies the following initial value problem:

$$\begin{aligned}\partial_t p(t, x, y) &= Gp(t, x, y) \\ p(0, x, y) &= \delta_x(dy),\end{aligned}$$

where  $G$  is understood to act on  $x$  on the right-hand side and the identity  $p(0, x, y) = \delta_x(dy)$  means:

$$\lim_{t \rightarrow 0} \int_E f(y)p(t, x, y)dy = f(x).$$

## Kolmogorov Backward Equation

## Kolmogorov Backward Equation

In the special case where  $X$  is a one-dimensional diffusion, the transition density  $p(t, x, y)$  satisfies the **Kolmogorov backward equation**:

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2}a(x)\partial_{xx}p(t, x, y) + b(x)\partial_x p(t, x, y) \\ p(0, x, y) &= \delta_x(dy),\end{aligned}$$

where  $a(x)$  and  $b(x)$  are the infinitesimal variance and drift of  $X$ .

## Kolmogorov Backward Equation

In the special case where  $X$  is a one-dimensional diffusion, the transition density  $p(t, x, y)$  satisfies the **Kolmogorov backward equation**:

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2}a(x)\partial_{xx}p(t, x, y) + b(x)\partial_x p(t, x, y) \\ p(0, x, y) &= \delta_x(dy),\end{aligned}$$

where  $a(x)$  and  $b(x)$  are the infinitesimal variance and drift of  $X$ .

**Remark:** This is called the backward equation because the derivatives are taken with respect to the initial value of the process, i.e., looking backwards in time.

## Kolmogorov Forward Equation

## Kolmogorov Forward Equation

There is also an equation involving derivatives of the future value of the process. Suppose that  $X$  is a diffusion process on  $E = [l, r]$  with transition density  $p(t, x, y)$ .

## Kolmogorov Forward Equation

There is also an equation involving derivatives of the future value of the process. Suppose that  $X$  is a diffusion process on  $E = [l, r]$  with transition density  $p(t, x, y)$ .

Then, the Chapman-Kolmogorov equation implies that:

$$p(t+s, x, y) = \int_l^r p(t, x, z)p(s, z, y)dz.$$

## Kolmogorov Forward Equation

There is also an equation involving derivatives of the future value of the process. Suppose that  $X$  is a diffusion process on  $E = [l, r]$  with transition density  $p(t, x, y)$ .

Then, the Chapman-Kolmogorov equation implies that:

$$p(t+s, x, y) = \int_l^r p(t, x, z)p(s, z, y)dz.$$

It follows that

$$\partial_t p(t+s, x, y) = \partial_s p(t+s, x, y)$$

## Kolmogorov Forward Equation

There is also an equation involving derivatives of the future value of the process. Suppose that  $X$  is a diffusion process on  $E = [l, r]$  with transition density  $p(t, x, y)$ .

Then, the Chapman-Kolmogorov equation implies that:

$$p(t+s, x, y) = \int_l^r p(t, x, z)p(s, z, y)dz.$$

It follows that

$$\begin{aligned}\partial_t p(t+s, x, y) &= \partial_s p(t+s, x, y) \\ &= \int_l^r p(t, x, z)\partial_s p(s, z, y)dz\end{aligned}$$

## Kolmogorov Forward Equation

There is also an equation involving derivatives of the future value of the process. Suppose that  $X$  is a diffusion process on  $E = [l, r]$  with transition density  $p(t, x, y)$ .

Then, the Chapman-Kolmogorov equation implies that:

$$p(t+s, x, y) = \int_l^r p(t, x, z)p(s, z, y)dz.$$

It follows that

$$\begin{aligned}\partial_t p(t+s, x, y) &= \partial_s p(t+s, x, y) \\ &= \int_l^r p(t, x, z) \partial_s p(s, z, y) dz \\ &= \int_l^r p(t, x, z) \left( \frac{1}{2}a(z)p_{zz}(s, z, y) + b(z)p_z(s, z, y) \right) dz.\end{aligned}$$

Since our aim is to derive an equation involving derivatives with respect to the future value of the process, we use integration-by-parts. The first integration gives

Since our aim is to derive an equation involving derivatives with respect to the future value of the process, we use integration-by-parts. The first integration gives

$$\begin{aligned} & - \int_I^r \frac{1}{2} (a(z)p(t, x, z))' p'(s, z, y) dz \\ & - \int_I^r (b(z)p(t, x, z))' p(s, z, y) dz \\ & + \left( \frac{1}{2} a(z)p(t, x, z)p'(s, z, y) + b(z)p(t, x, z)p(s, z, y) \right) |_I^r. \end{aligned}$$

Since our aim is to derive an equation involving derivatives with respect to the future value of the process, we use integration-by-parts. The first integration gives

$$\begin{aligned}
 & - \int_I^r \frac{1}{2} (a(z)p(t, x, z))' p'(s, z, y) dz \\
 & - \int_I^r (b(z)p(t, x, z))' p(s, z, y) dz \\
 & + \left( \frac{1}{2} a(z)p(t, x, z)p'(s, z, y) + b(z)p(t, x, z)p(s, z, y) \right) |_I^r.
 \end{aligned}$$

The second integration gives:

$$\begin{aligned}
 & \int_I^r p(s, z, y) \left[ \frac{1}{2} (a(z)p(t, x, z))'' - (b(z)p(t, x, z))' \right] dz \\
 & + \left( \frac{1}{2} a(z)p(t, x, z)p'(s, z, y) + b(z)p(t, x, z)p(s, z, y) \right) |_I^r \\
 & - \frac{1}{2} (a(z)p(t, x, z))' p(s, z, y) |_I^r
 \end{aligned}$$

If we let  $s \rightarrow 0$ , then  $p(s, z, y) \rightarrow \delta_z(y)$  and so all three boundary terms vanish if  $z \neq l, r$ . This leads to the **Kolmogorov forward equation** for  $p(t, x, y)$ :

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2} \partial_{yy} (a(y)p(t, x, y)) - \partial_y (b(y)p(t, x, y)) \\ p(0, x, y) &= \delta_x(dy)\end{aligned}$$

If we let  $s \rightarrow 0$ , then  $p(s, z, y) \rightarrow \delta_z(y)$  and so all three boundary terms vanish if  $z \neq l, r$ . This leads to the **Kolmogorov forward equation** for  $p(t, x, y)$ :

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2} \partial_{yy} (a(y)p(t, x, y)) - \partial_y (b(y)p(t, x, y)) \\ p(0, x, y) &= \delta_x(dy)\end{aligned}$$

In general, the forward equation will only have a solution if the variance and drift coefficients,  $a(y)$  and  $b(y)$ , are smooth functions.

There also are multivariate versions of the forward and backward equations. If  $X$  is an  $n$ -dimensional diffusion with infinitesimal variance-covariance matrix  $a(x)$  and infinitesimal drift  $b(x)$ , then the transition density satisfies:

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{x_i x_j} p(t, x, y) + \sum_{i=1}^n b_i(x) \partial_{x_i} p(t, x, y) \\ p(0, x, y) &= \delta_x(dy) \quad \text{(backward equation)}\end{aligned}$$

There also are multivariate versions of the forward and backward equations. If  $X$  is an  $n$ -dimensional diffusion with infinitesimal variance-covariance matrix  $a(x)$  and infinitesimal drift  $b(x)$ , then the transition density satisfies:

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{x_i x_j} p(t, x, y) + \sum_{i=1}^n b_i(x) \partial_{x_i} p(t, x, y) \\ p(0, x, y) &= \delta_x(dy) \quad \text{(backward equation)}\end{aligned}$$

and

$$\begin{aligned}\partial_t p(t, x, y) &= \frac{1}{2} \sum_{i,j=1}^n \partial_{y_i y_j} \left( a_{ij}(y) p(t, x, y) \right) - \sum_{i=1}^n \partial_{y_i} \left( b_i(y) p(t, x, y) \right) \\ p(0, x, y) &= \delta_x(dy) \quad \text{(forward equation)}.\end{aligned}$$

## Stationary Distributions

## 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  $t \geq 0$ .

## 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

$$\begin{aligned}\mathbb{E}_\pi[f(X_0)] &= \mathbb{E}_\pi[f(X_t)], \text{ i.e.} \\ \int_E f(x)\pi(dx) &= \int_E \mathbb{E}_x[f(X_t)]\pi(dx)\end{aligned}$$

holds for all  $t \geq 0$ .

We can rewrite this equation in terms of the semigroup

$$\int_E f(x)\pi(dx) = \int_E T_t f(x)\pi(dx).$$

We can rewrite this equation in terms of the semigroup

$$\int_E f(x)\pi(dx) = \int_E T_t f(x)\pi(dx).$$

Thus, the right-hand side does not depend on  $t$ , and so if we can interchange differentiation and integration, then

$$0 = \partial_t \int_E T_t f(x)\pi(dx) = \int_E \partial_t T_t f(x)\pi(dx) = \int_E G T_t f(x)\pi(dx),$$

using the identity  $\partial_t T_t f = G T_t f$ .

In particular, taking  $t = 0$  gives the identity

$$\int_E Gf(x)\pi(dx) = 0,$$

for any function  $f \in \mathcal{D}(G)$ .

In particular, taking  $t = 0$  gives the identity

$$\int_E Gf(x)\pi(dx) = 0,$$

for any function  $f \in \mathcal{D}(G)$ .

This equation can sometimes be used to recursively calculate the moments of the stationary distribution  $\pi$ .

## Example: Moran model with mutation

## Example: Moran model with mutation

Assumptions:

- reproduction follows the Moran model, plus
- each  $A_1$  individual mutates to  $A_2$  at rate  $\mu_2$
- each  $A_2$  individual mutates to  $A_1$  at rate  $\mu_1$
- mutation occurs independently of reproduction

## Example: Moran model with mutation

Assumptions:

- reproduction follows the Moran model, plus
- each  $A_1$  individual mutates to  $A_2$  at rate  $\mu_2$
- each  $A_2$  individual mutates to  $A_1$  at rate  $\mu_1$
- mutation occurs independently of reproduction

If  $p(t)$  denotes the frequency of  $A_1$  at time  $t$ , then  $(p(t) : t > 0)$  is a continuous time Markov chain with values in  $K_N = \{0, 1/N, \dots, 1\}$  and generator

## Example: Moran model with mutation

Assumptions:

- reproduction follows the Moran model, plus
- each  $A_1$  individual mutates to  $A_2$  at rate  $\mu_2$
- each  $A_2$  individual mutates to  $A_1$  at rate  $\mu_1$
- mutation occurs independently of reproduction

If  $p(t)$  denotes the frequency of  $A_1$  at time  $t$ , then  $(p(t) : t > 0)$  is a continuous time Markov chain with values in  $K_N = \{0, 1/N, \dots, 1\}$  and generator

$$\begin{aligned} G\phi(p) &= N(p(1-p) + (1-p)\mu_1) [\phi(p+1/N) - \phi(p)] \\ &\quad + N(p(1-p) + p\mu_2) [\phi(p-1/N) - \phi(p)]. \end{aligned}$$

Although the stationary distribution  $\pi$  can be determined explicitly, it is easier to calculate the moments.

Although the stationary distribution  $\pi$  can be determined explicitly, it is easier to calculate the moments.

For example, to calculate the mean frequency of  $A_1$  in a stationary population, let  $\phi(p) = p$  and observe that

$$G\phi(p) = (1 - p)\mu_1 - p\mu_2 = \mu_1 - (\mu_1 + \mu_2)p.$$

Although the stationary distribution  $\pi$  can be determined explicitly, it is easier to calculate the moments.

For example, to calculate the mean frequency of  $A_1$  in a stationary population, let  $\phi(p) = p$  and observe that

$$G\phi(p) = (1 - p)\mu_1 - p\mu_2 = \mu_1 - (\mu_1 + \mu_2)p.$$

Then,

$$0 = \int_{K_N} G\phi(p)\pi(dp) = \mu_1 - (\mu_1 + \mu_2) \int_{K_N} p\pi(dp),$$

Although the stationary distribution  $\pi$  can be determined explicitly, it is easier to calculate the moments.

For example, to calculate the mean frequency of  $A_1$  in a stationary population, let  $\phi(p) = p$  and observe that

$$G\phi(p) = (1 - p)\mu_1 - p\mu_2 = \mu_1 - (\mu_1 + \mu_2)p.$$

Then,

$$0 = \int_{K_N} G\phi(p)\pi(dp) = \mu_1 - (\mu_1 + \mu_2) \int_{K_N} p\pi(dp),$$

and so

$$\bar{p} \equiv \int_{K_N} p\pi(dp) = \frac{\mu_1}{\mu_1 + \mu_2}$$

is the mean frequency of  $A_1$ .

We would also like to know how dispersed the stationary distribution is about its mean. To this end, let  $\phi(p) = p(1 - p)$  and calculate

$$\begin{aligned} G\phi(p) &= -2(\mu_1 + \mu_2 + 1/N)p(1 - p) + \mu_1 \\ &\quad + (\mu_2 - \mu_1)p - \frac{1}{N}(\mu_1 - (\mu_1 + \mu_2)p). \end{aligned}$$

We would also like to know how dispersed the stationary distribution is about its mean. To this end, let  $\phi(p) = p(1 - p)$  and calculate

$$\begin{aligned} G\phi(p) &= -2(\mu_1 + \mu_2 + 1/N)p(1 - p) + \mu_1 \\ &\quad + (\mu_2 - \mu_1)p - \frac{1}{N}(\mu_1 - (\mu_1 + \mu_2)p). \end{aligned}$$

Consequently,

$$0 = \int_{K_N} G\phi(p)\pi(dp)$$

We would also like to know how dispersed the stationary distribution is about its mean. To this end, let  $\phi(p) = p(1 - p)$  and calculate

$$\begin{aligned} G\phi(p) &= -2(\mu_1 + \mu_2 + 1/N)p(1 - p) + \mu_1 \\ &\quad + (\mu_2 - \mu_1)p - \frac{1}{N}(\mu_1 - (\mu_1 + \mu_2)p). \end{aligned}$$

Consequently,

$$\begin{aligned} 0 &= \int_{K_N} G\phi(p)\pi(dp) \\ &= -2(\mu_1 + \mu_2 + 1/N) \int_{K_N} p(1 - p)\pi(dp) + \int_{K_N} (\mu_1 + (\mu_2 - \mu_1)p)\pi(dp) \\ &\quad - \frac{1}{N} \int_{K_N} (\mu_1 - (\mu_1 + \mu_2)p)\pi(dp). \end{aligned}$$

Since  $\bar{p} = \mu_1/(\mu_1 + \mu_2)$ , we find:

$$\bar{H} \equiv \int_{K_N} 2p(1-p)\pi(dp)$$

Since  $\bar{p} = \mu_1/(\mu_1 + \mu_2)$ , we find:

$$\begin{aligned}\bar{H} &\equiv \int_{K_N} 2p(1-p)\pi(dp) \\ &= 2 \left( \frac{\mu_1 + (\mu_2 - \mu_1)\bar{p}}{2(\mu_1 + \mu_2 + 1/N)} \right)\end{aligned}$$

Since  $\bar{p} = \mu_1/(\mu_1 + \mu_2)$ , we find:

$$\begin{aligned}\bar{H} &\equiv \int_{K_N} 2p(1-p)\pi(dp) \\ &= 2 \left( \frac{\mu_1 + (\mu_2 - \mu_1)\bar{p}}{2(\mu_1 + \mu_2 + 1/N)} \right) \\ &= \frac{2\mu_1\mu_2}{(\mu_1 + \mu_2)(\mu_1 + \mu_2 + 1/N)}.\end{aligned}$$

Since  $\bar{p} = \mu_1/(\mu_1 + \mu_2)$ , we find:

$$\begin{aligned}
 \bar{H} &\equiv \int_{K_N} 2p(1-p)\pi(dp) \\
 &= 2 \left( \frac{\mu_1 + (\mu_2 - \mu_1)\bar{p}}{2(\mu_1 + \mu_2 + 1/N)} \right) \\
 &= \frac{2\mu_1\mu_2}{(\mu_1 + \mu_2)(\mu_1 + \mu_2 + 1/N)}.
 \end{aligned}$$

Notice that  $\bar{H} \approx 2\bar{p}(1 - \bar{p})$  if  $N\mu_1, N\mu_2 \gg 1$ . In this case,  $\pi(dp)$  is concentrated near  $\bar{p}$ , i.e., the process makes only small fluctuations around  $\bar{p}$ .

## Stationary Distributions of Diffusion Processes

## 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:

## 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  $x \in [l, r]$ ,

$$\lim_{t \rightarrow \infty} p(t, x, y) = \pi(y).$$

## 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  $x \in [l, r]$ ,

$$\lim_{t \rightarrow \infty} p(t, x, y) = \pi(y).$$

**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:

Under these conditions, we can use the forward equation to calculate:

$$0 = \partial_t \pi(y)$$

Under these conditions, we can use the forward equation to calculate:

$$\begin{aligned} 0 &= \partial_t \pi(y) \\ &= \partial_t \lim_{t \rightarrow \infty} p(t, x, y) \end{aligned}$$

Under these conditions, we can use the forward equation to calculate:

$$\begin{aligned} 0 &= \partial_t \pi(y) \\ &= \partial_t \lim_{t \rightarrow \infty} p(t, x, y) \\ &= \lim_{t \rightarrow \infty} \partial_t p(t, x, y) \quad (\text{if we can interchange the limit and differentiation}) \end{aligned}$$

Under these conditions, we can use the forward equation to calculate:

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

Under these conditions, we can use the forward equation to calculate:

$$\begin{aligned} 0 &= \partial_t \pi(y) \\ &= \partial_t \lim_{t \rightarrow \infty} p(t, x, y) \\ &= \lim_{t \rightarrow \infty} \partial_t p(t, x, y) \quad (\text{if we can interchange the limit and differentiation}) \\ &= \lim_{t \rightarrow \infty} \left( \frac{1}{2} \partial_{yy} (a(y)p(t, x, y)) - \partial_y (b(y)p(t, x, y)) \right) \\ &= \frac{1}{2} \partial_{yy} (a(y)\pi(y)) - \partial_y (b(y)\pi(y)) \\ &\quad (\text{again assuming that we can interchange the limit and differentiation}). \end{aligned}$$

Under these conditions, we can use the forward equation to calculate:

$$\begin{aligned}
 0 &= \partial_t \pi(y) \\
 &= \partial_t \lim_{t \rightarrow \infty} p(t, x, y) \\
 &= \lim_{t \rightarrow \infty} \partial_t p(t, x, y) \quad (\text{if we can interchange the limit and differentiation}) \\
 &= \lim_{t \rightarrow \infty} \left( \frac{1}{2} \partial_{yy} (a(y)p(t, x, y)) - \partial_y (b(y)p(t, x, y)) \right) \\
 &= \frac{1}{2} \partial_{yy} (a(y)\pi(y)) - \partial_y (b(y)\pi(y)) \\
 &\quad (\text{again assuming that we can interchange the limit and differentiation}).
 \end{aligned}$$

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).

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

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

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  $(l, r)$ .

Solving for  $\pi(x)$  gives:

$$\pi(x) = \frac{1}{Ca(x)} \exp \left( 2 \int_c^x \frac{b(y)}{a(y)} dy \right),$$

where the normalizing constant  $C < \infty$  must be chosen (if possible) so that

$$\int_l^r \pi(x) dx = 1.$$

## Example: Wright-Fisher Diffusion with Mutation

## Example: Wright-Fisher Diffusion with Mutation

Consider the Moran model with mutation that we examined earlier and suppose that the mutation rates are  $\mu_i = 2\theta_i/N$ . Then the generator of this process  $(p^N(t) : t \geq 0)$  is

$$\begin{aligned} G^N \phi(p) = & (Np(1-p) + 2\theta_1(1-p)) [\phi(p+1/N) - \phi(p)] \\ & + (Np(1-p) + 2\theta_2 p) [\phi(p-1/N) - \phi(p)]. \end{aligned}$$

## Example: Wright-Fisher Diffusion with Mutation

Consider the Moran model with mutation that we examined earlier and suppose that the mutation rates are  $\mu_i = 2\theta_i/N$ . Then the generator of this process  $(p^N(t) : t \geq 0)$  is

$$\begin{aligned} G^N \phi(p) &= (Np(1-p) + 2\theta_1(1-p)) [\phi(p+1/N) - \phi(p)] \\ &\quad + (Np(1-p) + 2\theta_2p) [\phi(p-1/N) - \phi(p)]. \end{aligned}$$

A simple Taylor series expansion shows that if  $\phi(p)$  is smooth, then

$$\begin{aligned} \lim_{N \rightarrow \infty} NG^N \phi(p) &= G\phi(p) \\ &= p(1-p)\phi''(p) + (2\theta_1(1-p) - 2\theta_2p)\phi'(p). \end{aligned}$$

Consequently, the rescaled processes converge to a diffusion process which is an example of a Wright-Fisher diffusion with mutation:

$$(p^N(Nt) : t \geq 0) \xrightarrow{d} (p(t) : t \geq 0).$$

Consequently, the rescaled processes converge to a diffusion process which is an example of a Wright-Fisher diffusion with mutation:

$$(p^N(Nt) : t \geq 0) \xrightarrow{d} (p(t) : t \geq 0).$$

Notice that the infinitesimal variance and drift of this diffusion are

$$\begin{aligned} a(p) &= 2p(1-p) \\ b(p) &= 2\theta_1(1-p) - 2\theta_2p, \end{aligned}$$

i.e., incorporating mutation changes the infinitesimal drift but not the infinitesimal variance.

The density of the stationary distribution of this process can be found by substituting  $a(p)$  and  $b(p)$  into the formula that we derived earlier:

The density of the stationary distribution of this process can be found by substituting  $a(p)$  and  $b(p)$  into the formula that we derived earlier:

$$\begin{aligned}\pi(p) &= \frac{1}{C} \frac{1}{2p(1-p)} \exp \left\{ 2 \int_c^p \frac{(2\theta_1(1-q) - 2\theta_2q)}{2q(1-q)} dq \right\} \\ &= \frac{1}{C} p^{2\theta_1-1} (1-p)^{2\theta_2-1}.\end{aligned}$$

In passing from the first to the second line, we have absorbed the 2 and the terms involving  $c$  into the normalizing constant.

The density of the stationary distribution of this process can be found by substituting  $a(p)$  and  $b(p)$  into the formula that we derived earlier:

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

In passing from the first to the second line, we have absorbed the 2 and the terms involving  $c$  into the normalizing constant.

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.

**Summary:** The stationary distribution of the Wright-Fisher diffusion with mutation is just the Beta distribution with parameters  $2\theta_1$  and  $2\theta_2$ , which has density

$$\begin{aligned}\pi(p) &= \frac{1}{B(2\theta_1, 2\theta_2)} p^{2\theta_1-1} (1-p)^{2\theta_2-1} \\ &= \frac{1}{B(2N\mu_1, 2N\mu_2)} p^{2N\mu_1-1} (1-p)^{2N\mu_2-1}.\end{aligned}$$

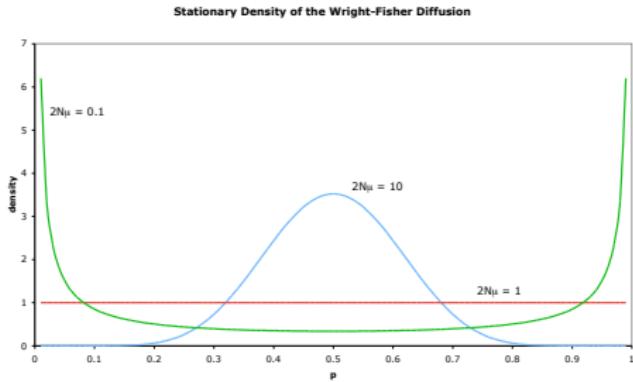
**Summary:** The stationary distribution of the Wright-Fisher diffusion with mutation is just the Beta distribution with parameters  $2\theta_1$  and  $2\theta_2$ , which has density

$$\begin{aligned}\pi(p) &= \frac{1}{B(2\theta_1, 2\theta_2)} p^{2\theta_1-1} (1-p)^{2\theta_2-1} \\ &= \frac{1}{B(2N\mu_1, 2N\mu_2)} p^{2N\mu_1-1} (1-p)^{2N\mu_2-1}.\end{aligned}$$

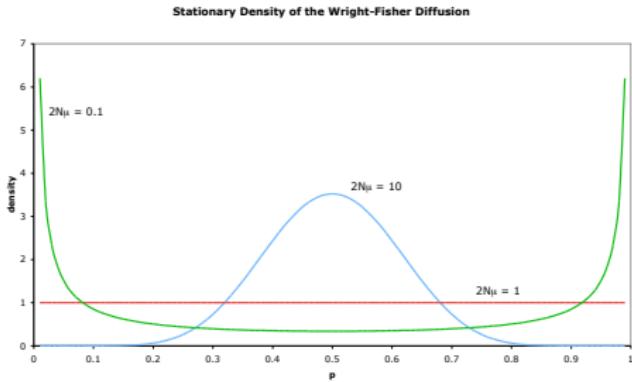
With this, we can directly calculate the mean  $\bar{p}$  and heterozygosity  $\bar{H}$ :

$$\begin{aligned}\bar{p} &= \int_0^1 p\pi(p)dp = \frac{\mu_1}{\mu_1 + \mu_2} \\ \bar{H} &= \int_0^1 2p(1-p)\pi(p)dp = \frac{2\mu_1\mu_2}{(\mu_1 + \mu_2)(\mu_1 + \mu_2 + 1/2N)}.\end{aligned}$$

The stationary distribution reflects the competing effects of genetic drift, which eliminates variation, and mutation, which generates variation.



The stationary distribution reflects the competing effects of genetic drift, which eliminates variation, and mutation, which generates variation.



- When  $N\mu_1, N\mu_2 > 1$ , mutation dominates drift and the stationary distribution is peaked about its mean (both alleles are common).
- When  $N\mu_1, N\mu_2 < 1$ , drift dominates mutation and the stationary distribution is bimodal, with peaks at the boundaries (one allele is common and one rare).

## Fixation and Stationary Distributions

## Fixation and Stationary Distributions

If there is no mutation ( $\theta_1 = \theta_2 = 0$ ), then  $b(p) = 0$  and

$$\pi(p) = \frac{1}{C} \frac{1}{2p(1-p)}.$$

## Fixation and Stationary Distributions

If there is no mutation ( $\theta_1 = \theta_2 = 0$ ), then  $b(p) = 0$  and

$$\pi(p) = \frac{1}{C} \frac{1}{2p(1-p)}.$$

Since

$$\int_0^1 \frac{1}{p(1-p)} dp = \infty,$$

it follows that  $\pi(p)$  is not the density of a stationary distribution.

## Fixation and Stationary Distributions

If there is no mutation ( $\theta_1 = \theta_2 = 0$ ), then  $b(p) = 0$  and

$$\pi(p) = \frac{1}{C} \frac{1}{2p(1-p)}.$$

Since

$$\int_0^1 \frac{1}{p(1-p)} dp = \infty,$$

it follows that  $\pi(p)$  is not the density of a stationary distribution.

Rather, this process is certain to absorb at one of the boundaries in finite time. In this case, the distributions,

$$\delta_0(dp) \text{ and } \delta_1(dp),$$

are both stationary, i.e., this process does not have a unique stationary distribution.

## Selection and Genetic Drift

## Selection and Genetic Drift

### Wright-Fisher model with selection:

Consider a population containing  $N$  haploid individuals and two alleles,  $A$  and  $a$ . Suppose that each adult in generation  $t + 1$  'chooses' its parent in generation  $t$  (independently and with replacement) with the following probabilities:

$$\begin{aligned} \text{A-type parent: } & \frac{p(1+s)}{p(1+s) + 1 - p} \\ \text{a-type parent: } & \frac{1-p}{p(1+s) + 1 - p}, \end{aligned}$$

where  $p$  is the frequency of  $A$ -type adults alive in generation  $t$ .

## Selection and Genetic Drift

### Wright-Fisher model with selection:

Consider a population containing  $N$  haploid individuals and two alleles,  $A$  and  $a$ . Suppose that each adult in generation  $t + 1$  'chooses' its parent in generation  $t$  (independently and with replacement) with the following probabilities:

$$\begin{aligned} \text{A-type parent: } & \frac{p(1+s)}{p(1+s) + 1 - p} \\ \text{a-type parent: } & \frac{1-p}{p(1+s) + 1 - p}, \end{aligned}$$

where  $p$  is the frequency of  $A$ -type adults alive in generation  $t$ .

$s$  is called the **selection coefficient** of  $A$  and quantifies the selective advantage ( $s > 0$ ) or disadvantage ( $s < 0$ ) of this allele relative to  $a$ . If  $s = 0$ , then the two alleles are said to be **neutral**. If  $s > 0$ , then  $A$  is said to be a **beneficial allele**, while if  $s < 0$ , then  $A$  is said to be **deleterious**.

To derive a diffusion approximation for this process, we must assume that the selection coefficient has the same order of magnitude as genetic drift, i.e., we set

$$s = \sigma/N.$$

To derive a diffusion approximation for this process, we must assume that the selection coefficient has the same order of magnitude as genetic drift, i.e., we set

$$s = \sigma/N.$$

In this case, using the moments of the binomial distribution, it is easy to show that:

$$\begin{aligned} N\mathbb{E}_p[p^N(1-p)] &= \sigma p(1-p) + O(N^{-1}) \\ N\mathbb{E}_p[(p^N(1-p))^2] &= p(1-p) + O(N^{-1}) \\ N\mathbb{E}_p[(p^N(1-p))^n] &= O(N^{-1}) \text{ if } n \geq 3. \end{aligned}$$

To derive a diffusion approximation for this process, we must assume that the selection coefficient has the same order of magnitude as genetic drift, i.e., we set

$$s = \sigma/N.$$

In this case, using the moments of the binomial distribution, it is easy to show that:

$$\begin{aligned} N\mathbb{E}_p[p^N(1-p)] &= \sigma p(1-p) + O(N^{-1}) \\ N\mathbb{E}_p[(p^N(1-p))^2] &= p(1-p) + O(N^{-1}) \\ N\mathbb{E}_p[(p^N(1-p))^n] &= O(N^{-1}) \text{ if } n \geq 3. \end{aligned}$$

These calculations show that the processes  $(p^N(\lfloor Nt \rfloor) : t \geq 0)$  have a diffusion approximation with generator

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) + \sigma p(1-p)f'(p).$$

This process is called a Wright-Fisher diffusion with selection.

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .
- Selection does not change the infinitesimal variance of the Wright-Fisher diffusion.

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .
- Selection does not change the infinitesimal variance of the Wright-Fisher diffusion.
- Selection does change the infinitesimal drift from 0 in the neutral diffusion to  $b(p) = \sigma p(1 - p)$ .

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .
- Selection does not change the infinitesimal variance of the Wright-Fisher diffusion.
- Selection does change the infinitesimal drift from 0 in the neutral diffusion to  $b(p) = \sigma p(1 - p)$ .
- The frequency of a beneficial allele will tend to increase, while that of a deleterious allele will tend to decrease.

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .
- Selection does not change the infinitesimal variance of the Wright-Fisher diffusion.
- Selection does change the infinitesimal drift from 0 in the neutral diffusion to  $b(p) = \sigma p(1 - p)$ .
- The frequency of a beneficial allele will tend to increase, while that of a deleterious allele will tend to decrease.
- However, genetic drift can cause the frequency of a beneficial allele to decrease and the frequency of a deleterious allele to increase.

**Remarks:**

- The diffusion approximation is accurate when selection is weak:  $s \sim 1/N$ .
- Selection does not change the infinitesimal variance of the Wright-Fisher diffusion.
- Selection does change the infinitesimal drift from 0 in the neutral diffusion to  $b(p) = \sigma p(1 - p)$ .
- The frequency of a beneficial allele will tend to increase, while that of a deleterious allele will tend to decrease.
- However, genetic drift can cause the frequency of a beneficial allele to decrease and the frequency of a deleterious allele to increase.
- Selection has little effect on the population when  $A$  is either very common or very rare.

Recall that we say that an allele  $A$  is fixed in the population if all other alleles are lost. In a population initially containing two alleles, the time to fixation can be written as

$$\tau = \inf_{t \geq 0} \{p(t) = 0 \text{ or } 1\},$$

and it can be shown that

$$\mathbb{P}_p\{\tau < \infty\} = 1.$$

Recall that we say that an allele  $A$  is fixed in the population if all other alleles are lost. In a population initially containing two alleles, the time to fixation can be written as

$$\tau = \inf_{t \geq 0} \{p(t) = 0 \text{ or } 1\},$$

and it can be shown that

$$\mathbb{P}_p\{\tau < \infty\} = 1.$$

In other words, with probability one, one of the two alleles is certain to be fixed in the population at some finite (but random) time.

Recall that we say that an allele  $A$  is fixed in the population if all other alleles are lost. In a population initially containing two alleles, the time to fixation can be written as

$$\tau = \inf_{t \geq 0} \{p(t) = 0 \text{ or } 1\},$$

and it can be shown that

$$\mathbb{P}_p\{\tau < \infty\} = 1.$$

In other words, with probability one, one of the two alleles is certain to be fixed in the population at some finite (but random) time.

**Remark:**  $\tau$  is certain to be finite for a Wright-Fisher diffusion as long as there is no mutation (as is the case in our current model). If mutation is incorporated into the model, then fixation can only occur if the mutation rates are not too large.

Because  $\tau$  is almost surely finite, we can define the **fixation probability** of allele  $A$  to be

$$u(p) = \mathbb{P}_p \{p(\tau) = 1\}.$$

Because  $\tau$  is almost surely finite, we can define the **fixation probability** of allele  $A$  to be

$$u(p) = \mathbb{P}_p \{p(\tau) = 1\}.$$

Of course, the fixation probability of an allele depends on both its initial frequency and its selection coefficient. Intuitively, we would expect that:

- $u(p)$  is an increasing function of  $p$ : the more common an allele is, the more likely it is to be fixed in the population.
- $u(p)$  is an increasing function of  $\sigma$ : beneficial mutations are more likely to be fixed than deleterious mutations.

Because  $\tau$  is almost surely finite, we can define the **fixation probability** of allele  $A$  to be

$$u(p) = \mathbb{P}_p \{p(\tau) = 1\}.$$

Of course, the fixation probability of an allele depends on both its initial frequency and its selection coefficient. Intuitively, we would expect that:

- $u(p)$  is an increasing function of  $p$ : the more common an allele is, the more likely it is to be fixed in the population.
- $u(p)$  is an increasing function of  $\sigma$ : beneficial mutations are more likely to be fixed than deleterious mutations.

Our goal is to find an explicit formula for  $u(p)$  which illustrates the relationship between the selection coefficient, the initial frequency, and the probability that  $A$  is ultimately fixed in the population.

## Hitting Probabilities

## Hitting Probabilities

Suppose that  $X = (X(t) : t \geq 0)$  is a Markov process in  $[l, r] \subset \mathbb{R}$ , with generator  $G$ , and define

$$\tau = \inf_{t \geq 0} \{X(t) = l \text{ or } r\},$$

i.e.,  $\tau$  is the first time that the process hits  $l$  or  $r$ .

## Hitting Probabilities

Suppose that  $X = (X(t) : t \geq 0)$  is a Markov process in  $[l, r] \subset \mathbb{R}$ , with generator  $G$ , and define

$$\tau = \inf_{t \geq 0} \{X(t) = l \text{ or } r\},$$

i.e.,  $\tau$  is the first time that the process hits  $l$  or  $r$ .

We will make the following three assumptions about  $X$ :

- $\tau$  is almost surely finite:  $\mathbb{P}_x\{\tau < \infty\} = 1$ .
- At time  $\tau$ , either  $X(\tau) = l$  or  $X(\tau) = r$ .
- $l$  and  $r$  are both absorbing states for  $X$ .

## Hitting Probabilities

Suppose that  $X = (X(t) : t \geq 0)$  is a Markov process in  $[l, r] \subset \mathbb{R}$ , with generator  $G$ , and define

$$\tau = \inf_{t \geq 0} \{X(t) = l \text{ or } r\},$$

i.e.,  $\tau$  is the first time that the process hits  $l$  or  $r$ .

We will make the following three assumptions about  $X$ :

- $\tau$  is almost surely finite:  $\mathbb{P}_x\{\tau < \infty\} = 1$ .
- At time  $\tau$ , either  $X(\tau) = l$  or  $X(\tau) = r$ .
- $l$  and  $r$  are both absorbing states for  $X$ .

**Remark:** The second assumption holds whenever the sample paths of  $X$  are right-continuous.

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

Our aim is to derive an equation for  $u(x)$  involving the generator of  $X$ . We first observe that

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}$$

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

Our aim is to derive an equation for  $u(x)$  involving the generator of  $X$ . We first observe that

$$\begin{aligned} u(x) &= \mathbb{P}_x\{X(\tau) = r\} \\ &= \mathbb{P}_x\{X(\tau + t) = r\} \quad (r \text{ is an absorbing state}) \end{aligned}$$

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

Our aim is to derive an equation for  $u(x)$  involving the generator of  $X$ . We first observe that

$$\begin{aligned} u(x) &= \mathbb{P}_x\{X(\tau) = r\} \\ &= \mathbb{P}_x\{X(\tau + t) = r\} \quad (r \text{ is an absorbing state}) \\ &= \mathbb{E}_x[\mathbb{P}\{X(\tau + t) = r | X(t)\}] \end{aligned}$$

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

Our aim is to derive an equation for  $u(x)$  involving the generator of  $X$ . We first observe that

$$\begin{aligned} u(x) &= \mathbb{P}_x\{X(\tau) = r\} \\ &= \mathbb{P}_x\{X(\tau + t) = r\} \quad (r \text{ is an absorbing state}) \\ &= \mathbb{E}_x[\mathbb{P}\{X(\tau + t) = r | X(t)\}] \\ &= \int_E u(y) \mathbb{P}_x\{X(t) \in dy\} \quad (\text{by the Markov property}) \end{aligned}$$

When these assumptions are satisfied, we can ask how the probability that the process  $X$  hits  $r$  rather than  $l$  depends on its initial value:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\}.$$

Our aim is to derive an equation for  $u(x)$  involving the generator of  $X$ . We first observe that

$$\begin{aligned} u(x) &= \mathbb{P}_x\{X(\tau) = r\} \\ &= \mathbb{P}_x\{X(\tau + t) = r\} \quad (r \text{ is an absorbing state}) \\ &= \mathbb{E}_x[\mathbb{P}\{X(\tau + t) = r | X(t)\}] \\ &= \int_E u(y) \mathbb{P}_x\{X(t) \in dy\} \quad (\text{by the Markov property}) \\ &= \mathbb{E}_x[u(X(t))]. \end{aligned}$$

Next, we can use the definition of the generator to calculate

$$\begin{aligned} Gu(x) &= \lim_{t \downarrow 0} \frac{1}{t} (\mathbb{E}_x[u(X(t))] - u(x)) \\ &= 0, \end{aligned}$$

since  $u(x) = \mathbb{E}_x[u(X(t))]$  for every  $t \geq 0$ .

Next, we can use the definition of the generator to calculate

$$\begin{aligned} Gu(x) &= \lim_{t \downarrow 0} \frac{1}{t} (\mathbb{E}_x[u(X(t))] - u(x)) \\ &= 0, \end{aligned}$$

since  $u(x) = \mathbb{E}_x[u(X(t))]$  for every  $t \geq 0$ .

In addition, we know that

$$u(l) = 0 \text{ and } u(r) = 1.$$

This is true because if  $X$  starts at  $l$ , then it never hits  $r$ , while if  $X$  starts at  $r$ , then it certainly hits  $r$ .

Next, we can use the definition of the generator to calculate

$$\begin{aligned} Gu(x) &= \lim_{t \downarrow 0} \frac{1}{t} (\mathbb{E}_x[u(X(t))] - u(x)) \\ &= 0, \end{aligned}$$

since  $u(x) = \mathbb{E}_x[u(X(t))]$  for every  $t \geq 0$ .

In addition, we know that

$$u(l) = 0 \text{ and } u(r) = 1.$$

This is true because if  $X$  starts at  $l$ , then it never hits  $r$ , while if  $X$  starts at  $r$ , then it certainly hits  $r$ .

**Key Result:** The hitting probability  $u(x)$  of the process  $X$  can be found by solving the boundary value problem:

$$\begin{aligned} Gu(x) &\equiv 0, \quad x \in [l, r] \\ u(l) &= 0, \quad u(r) = 1. \end{aligned}$$

## Hitting Probabilities for Diffusions

## Hitting Probabilities for Diffusions

Suppose that  $X$  is a diffusion process on  $[l, r]$  with generator

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

Here we will stipulate that  $X$  absorbs at  $l$  or  $r$ , if necessary by stopping the sample paths of  $X$  whenever they hit  $l$  or  $r$ . (Such a process is called a **stopped process** and is still a Markov process.)

## Hitting Probabilities for Diffusions

Suppose that  $X$  is a diffusion process on  $[l, r]$  with generator

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

Here we will stipulate that  $X$  absorbs at  $l$  or  $r$ , if necessary by stopping the sample paths of  $X$  whenever they hit  $l$  or  $r$ . (Such a process is called a **stopped process** and is still a Markov process.)

We know that the hitting probability  $u(x)$  solves the equation

$$Gu(x) = \frac{1}{2}a(x)u''(x) + b(x)u'(x) = 0,$$

subject to the boundary conditions  $u(l) = 0$  and  $u(r) = 1$ .

This equation can be rearranged to

$$\frac{u''(x)}{u'(x)} = -2 \frac{b(x)}{a(x)},$$

This equation can be rearranged to

$$\frac{u''(x)}{u'(x)} = -2 \frac{b(x)}{a(x)},$$

which can be integrated to give

$$\ln(u'(x)) = -2 \int_c^x \frac{b(y)}{a(y)} dy + C_1,$$

where  $C_1$  is a constant of integration and  $c$  is an arbitrary point in  $(l, r)$ .

This equation can be rearranged to

$$\frac{u''(x)}{u'(x)} = -2 \frac{b(x)}{a(x)},$$

which can be integrated to give

$$\ln(u'(x)) = -2 \int_c^x \frac{b(y)}{a(y)} dy + C_1,$$

where  $C_1$  is a constant of integration and  $c$  is an arbitrary point in  $(l, r)$ .

Rearranging and integrating again gives

$$u(x) = C_2 + C_1 \int_c^x \exp \left( \int_c^y \frac{-2b(z)}{a(z)} dz \right) dy,$$

where  $C_2$  is a second constant of integration.

There are two boundary conditions to be satisfied, and these can be met by choosing suitable values for the constants  $C_1$  and  $C_2$ . This shows that the probability that the diffusion hits  $r$  before  $l$  is:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\} = \frac{\int_l^x \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}{\int_l^r \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}.$$

There are two boundary conditions to be satisfied, and these can be met by choosing suitable values for the constants  $C_1$  and  $C_2$ . This shows that the probability that the diffusion hits  $r$  before  $l$  is:

$$u(x) = \mathbb{P}_x\{X(\tau) = r\} = \frac{\int_l^x \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}{\int_l^r \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}.$$

Likewise, the probability that the diffusion hits  $l$  before  $r$  is simply  $1 - u(x)$  which is equal to:

$$1 - u(x) = \mathbb{P}_x\{X(\tau) = l\} = \frac{\int_x^r \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}{\int_l^r \exp\left(-2 \int_c^y \frac{b(z)}{a(z)} dz\right) dy}.$$

**Example:** Recall that the diffusion approximation for the Wright-Fisher model with selection has generator

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) + \sigma p(1-p)f'(p).$$

**Example:** Recall that the diffusion approximation for the Wright-Fisher model with selection has generator

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) + \sigma p(1-p)f'(p).$$

To apply the results on the preceding slide, note that  $a(p) = p(1-p)$  and  $b(p) = \sigma p(1-p)$ .

**Example:** Recall that the diffusion approximation for the Wright-Fisher model with selection has generator

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) + \sigma p(1-p)f'(p).$$

To apply the results on the preceding slide, note that  $a(p) = p(1-p)$  and  $b(p) = \sigma p(1-p)$ .

**Selected alleles:** If  $\sigma = Ns \neq 0$ , then the fixation probability of an allele  $A$  with initial frequency  $p$  is:

$$u(p) = \frac{1 - e^{-2\sigma p}}{1 - e^{-2\sigma}} = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}.$$

**Example:** Recall that the diffusion approximation for the Wright-Fisher model with selection has generator

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) + \sigma p(1-p)f'(p).$$

To apply the results on the preceding slide, note that  $a(p) = p(1-p)$  and  $b(p) = \sigma p(1-p)$ .

**Selected alleles:** If  $\sigma = Ns \neq 0$ , then the fixation probability of an allele  $A$  with initial frequency  $p$  is:

$$u(p) = \frac{1 - e^{-2\sigma p}}{1 - e^{-2\sigma}} = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}.$$

**Neutral alleles:** If  $A$  and  $a$  are neutral alleles ( $\sigma = 0$ ), then the fixation probability is equal to the initial frequency:

$$u(p) = p.$$

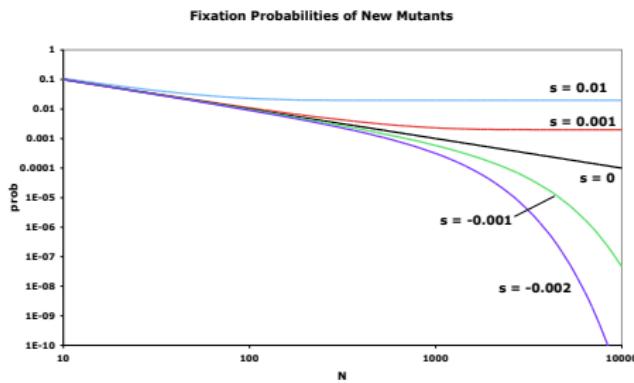
Usually we are interested in the fixation probability of a new mutation. In this case, the initial frequency of  $A$  is  $p = 1/N$  and we can calculate

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

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  $N$ .



Some specific cases.

Some specific cases.

**Deleterious alleles:**  $A$  is deleterious if  $s = -|s| < 0$ . Suppose that  $N|s| \gg 1$  and  $|s| \ll 1$ . Then,

$$u\left(\frac{1}{N}\right) = \frac{e^{2|s|} - 1}{e^{2N|s|} - 1} \approx 2|s|e^{-2N|s|},$$

and so the fixation probability of a deleterious allele is exponentially small and decreases as  $N|s|$  increases.

Some specific cases.

**Deleterious alleles:**  $A$  is deleterious if  $s = -|s| < 0$ . Suppose that  $N|s| \gg 1$  and  $|s| \ll 1$ . Then,

$$u\left(\frac{1}{N}\right) = \frac{e^{2|s|} - 1}{e^{2N|s|} - 1} \approx 2|s|e^{-2N|s|},$$

and so the fixation probability of a deleterious allele is exponentially small and decreases as  $N|s|$  increases.

### Intuition:

- Deleterious alleles can become fixed in finite populations through chance increases in their frequency caused by genetic drift.
- Many more such random events must occur for a deleterious allele to be fixed in a large population than in a small population.

**Beneficial alleles:**  $A$  is beneficial if  $s > 0$ . Suppose that  $Ns \gg 1$  and  $s \ll 1$ . Then,

$$u\left(\frac{1}{N}\right) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \approx 2s,$$

and thus the fixation probability of a beneficial allele is approximately equal to twice the selection coefficient. In particular, this probability is approximately independent of the population size as long as  $s \gg 1/N$ .

**Beneficial alleles:**  $A$  is beneficial if  $s > 0$ . Suppose that  $Ns \gg 1$  and  $s \ll 1$ . Then,

$$u\left(\frac{1}{N}\right) = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \approx 2s,$$

and thus the fixation probability of a beneficial allele is approximately equal to twice the selection coefficient. In particular, this probability is approximately independent of the population size as long as  $s \gg 1/N$ .

**Nearly neutral alleles:**  $A$  is nearly neutral if  $|Ns| \ll 1$ . In this case,

$$u\left(\frac{1}{N}\right) \approx \frac{1}{N} + s \approx \frac{1}{N},$$

and so the fixation probability of a nearly neutral allele is approximately equal to the reciprocal of the population size. (This relation is exact for neutral alleles:  $s = 0$ .)

We can draw the following main conclusions:

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.
- Mutations which have selection coefficients that are less than the reciprocal of the population size ( $|s| < N^{-1}$ ) behave like neutral mutations.

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.
- Mutations which have selection coefficients that are less than the reciprocal of the population size ( $|s| < N^{-1}$ ) behave like neutral mutations.
- Deleterious mutations are much more likely to be fixed in smaller populations.

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.
- Mutations which have selection coefficients that are less than the reciprocal of the population size ( $|s| < N^{-1}$ ) behave like neutral mutations.
- Deleterious mutations are much more likely to be fixed in smaller populations.
- Mean fitness can actually decrease in finite populations due to fixation of deleterious mutations.

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.
- Mutations which have selection coefficients that are less than the reciprocal of the population size ( $|s| < N^{-1}$ ) behave like neutral mutations.
- Deleterious mutations are much more likely to be fixed in smaller populations.
- Mean fitness can actually decrease in finite populations due to fixation of deleterious mutations.
- Selection is more effective in larger populations.

We can draw the following main conclusions:

- Most new mutations are lost from a population, even if they are beneficial.
- Mutations which have selection coefficients that are less than the reciprocal of the population size ( $|s| < N^{-1}$ ) behave like neutral mutations.
- Deleterious mutations are much more likely to be fixed in smaller populations.
- Mean fitness can actually decrease in finite populations due to fixation of deleterious mutations.
- Selection is more effective in larger populations.
- Fitness differences that are too small to measure in the field or the lab may still play an important role in evolution if  $|s| \geq 1/N$ .

## Substitution Rates

## Substitution Rates

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

## 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 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.

## 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 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.

Substitution rates are difficult to calculate exactly. However, we can find a good approximation if we assume that the mutation rate is low enough that each new mutation is likely to be lost or fixed before another mutation enters the population.

Under this assumption, we can approximate the substitution rate (per generation) by the expression

$$\rho \approx N\mu \cdot u \left( \frac{1}{N} \right),$$

where  $N\mu$  is the expected number of new mutations per generation, while  $u(1/N)$  is the probability that any one of these is fixed in the population. This is accurate when  $N\mu \ll 1$ .

Under this assumption, we can approximate the substitution rate (per generation) by the expression

$$\rho \approx N\mu \cdot u \left( \frac{1}{N} \right),$$

where  $N\mu$  is the expected number of new mutations per generation, while  $u(1/N)$  is the probability that any one of these is fixed in the population. This is accurate when  $N\mu \ll 1$ .

**Neutral substitution rate:** If all mutations are neutral, then

$$\rho = N\mu \cdot \frac{1}{N} = \mu,$$

and so the neutral substitution rate is simply equal to the mutation rate. In particular, the neutral substitution rate does not depend on the population size.

**Beneficial substitution rate:** If all mutations are beneficial, with selection coefficient  $s \gg 1/N$ , then

$$\rho \approx 2N\mu s,$$

and so the beneficial substitution rate is greater than the mutation rate and increases with population size.

**Beneficial substitution rate:** If all mutations are beneficial, with selection coefficient  $s \gg 1/N$ , then

$$\rho \approx 2N\mu s,$$

and so the beneficial substitution rate is greater than the mutation rate and increases with population size.

**Deleterious substitution rate:** If all mutations are deleterious, with selection coefficient  $s = -|s| \ll -1/N$ , then

$$\rho \approx 2N\mu|s|e^{-2N|s|},$$

and so the deleterious substitution rate is less than the mutation rate and decreases with population size.

**Beneficial substitution rate:** If all mutations are beneficial, with selection coefficient  $s \gg 1/N$ , then

$$\rho \approx 2N\mu s,$$

and so the beneficial substitution rate is greater than the mutation rate and increases with population size.

**Deleterious substitution rate:** If all mutations are deleterious, with selection coefficient  $s = -|s| \ll -1/N$ , then

$$\rho \approx 2N\mu|s|e^{-2N|s|},$$

and so the deleterious substitution rate is less than the mutation rate and decreases with population size.

**Important observation:** The substitution rate at a locus under selection is usually different from the mutation rate.

**Molecular Evolution:** Mutation, genetic drift and selection can all contribute to the genetic differences that are observed between species, and one of the central aims of population genetics is to assess the relative importance of these different processes.

**Molecular Evolution:** Mutation, genetic drift and selection can all contribute to the genetic differences that are observed between species, and one of the central aims of population genetics is to assess the relative importance of these different processes.

The answer to this question depends on the kind of substitution considered:

- (i) A **replacement** substitution is one that leads to a change in an amino acid in a protein.
- (ii) A **silent** substitution is one that changes only the DNA sequence.

**Molecular Evolution:** Mutation, genetic drift and selection can all contribute to the genetic differences that are observed between species, and one of the central aims of population genetics is to assess the relative importance of these different processes.

The answer to this question depends on the kind of substitution considered:

- (i) A **replacement** substitution is one that leads to a change in an amino acid in a protein.
- (ii) A **silent** substitution is one that changes only the DNA sequence.

Silent substitutions can occur in several kinds of sequence:

- intergenic regions
- pseudogenes (non-functional genes)
- introns (non-coding regions within eukaryotic genes)
- synonymous sites within genes (the genetic code is degenerate)

**Molecular Evolution:** Mutation, genetic drift and selection can all contribute to the genetic differences that are observed between species, and one of the central aims of population genetics is to assess the relative importance of these different processes.

The answer to this question depends on the kind of substitution considered:

- (i) A **replacement** substitution is one that leads to a change in an amino acid in a protein.
- (ii) A **silent** substitution is one that changes only the DNA sequence.

Silent substitutions can occur in several kinds of sequence:

- intergenic regions
- pseudogenes (non-functional genes)
- introns (non-coding regions within eukaryotic genes)
- synonymous sites within genes (the genetic code is degenerate)

About 98.5% of the human genome is non-coding.

Several observations suggest that most replacement mutations are deleterious, while most silent mutations are nearly neutral.

Several observations suggest that most replacement mutations are deleterious, while most silent mutations are nearly neutral.

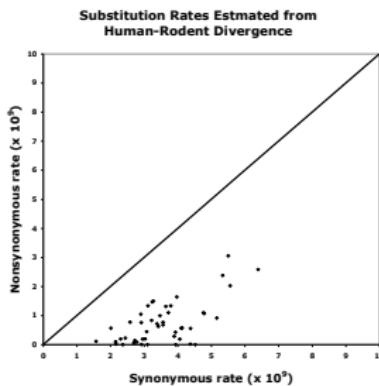
- The silent substitution rate is usually greater than the replacement substitution rate.
- Silent polymorphism within populations is usually greater than replacement polymorphism.
- The silent substitution rate per year is greater in species with shorter generation times, whereas the rate of replacement substitutions is only weakly correlated with generation length.

Several observations suggest that most replacement mutations are deleterious, while most silent mutations are nearly neutral.

- The silent substitution rate is usually greater than the replacement substitution rate.
- Silent polymorphism within populations is usually greater than replacement polymorphism.
- The silent substitution rate per year is greater in species with shorter generation times, whereas the rate of replacement substitutions is only weakly correlated with generation length.

**Important Caveat:** Some non-coding sequences are known to be functional and can be conserved between species. For example, gene expression is often regulated by sequences outside of genes.

The following plot shows synonymous and nonsynonymous substitution rates estimated from comparisons of human and rodent genes (Li (1997), pp. 180-181). In every case the nonsynonymous substitution rate is less than the synonymous substitution rate.



The average nonsynonymous and synonymous substitution rates in these genes are:

Nonsyn: 0.74 (0.67) vs. Syn: 3.51 (1.01).

## Some applications of substitution rates:

## Some applications of substitution rates:

### Estimating mutation rates:

- The fact that the neutral substitution rate equals the mutation rate can be used to estimate the mutation rate from sequence data.
- If two species last shared a common ancestor  $T$  generations ago, then the number of neutral substitutions per site that will have occurred since they split will be Poisson distributed with mean  $2\mu T$ .
- $T$  can be estimated from fossil or biogeographical data.

## Some applications of substitution rates:

### Estimating mutation rates:

- The fact that the neutral substitution rate equals the mutation rate can be used to estimate the mutation rate from sequence data.
- If two species last shared a common ancestor  $T$  generations ago, then the number of neutral substitutions per site that will have occurred since they split will be Poisson distributed with mean  $2\mu T$ .
- $T$  can be estimated from fossil or biogeographical data.

### Dating species divergence:

- If the mutation rate is known, then genetic data can be used to estimate how long ago two species diverged.
- We must assume that substitutions are neutral.
- We must also assume that the mutation rate is constant (**molecular clock**).

## Selection in Diploid Populations

## Selection in Diploid Populations

Recall that a diploid species contains two copies of each chromosome. Our goal in this section is to study a model of selection and drift in a diploid population.

## Selection in Diploid Populations

Recall that a diploid species contains two copies of each chromosome. Our goal in this section is to study a model of selection and drift in a diploid population.

To model selection in such a population, we need to assign a fitness to each diploid genotype. Here we will adopt the convention that the relative fitness of the  $A_2A_2$  homozygote is 1:

genotype	relative fitness
$A_1A_1$	$1 + s$
$A_1A_2$	$1 + hs$
$A_2A_2$	1

## Selection in Diploid Populations

Recall that a diploid species contains two copies of each chromosome. Our goal in this section is to study a model of selection and drift in a diploid population.

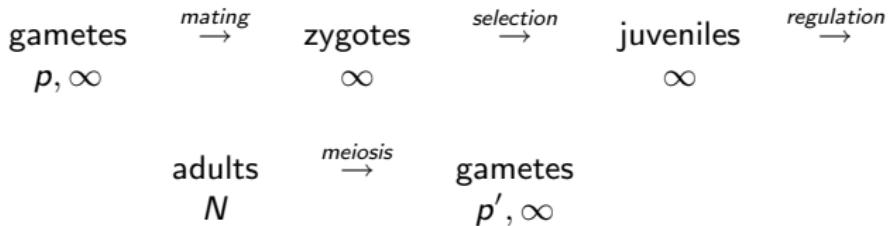
To model selection in such a population, we need to assign a fitness to each diploid genotype. Here we will adopt the convention that the relative fitness of the  $A_2A_2$  homozygote is 1:

genotype	relative fitness
$A_1A_1$	$1 + s$
$A_1A_2$	$1 + hs$
$A_2A_2$	1

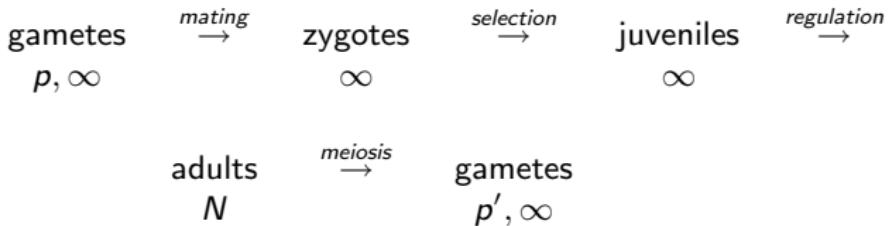
In this scheme,  $s$  is called the selection coefficient of the  $A_1A_1$  homozygote, and  $h$  is called the **degree of dominance** of  $A_1$ .

In principle, we need to know the frequencies of each of the three diploid genotypes. However, if the population is random mating, then we can formulate a model which tracks only the changes in the allele frequencies. We will assume that the population has a life history summarized by the following diagram:

In principle, we need to know the frequencies of each of the three diploid genotypes. However, if the population is random mating, then we can formulate a model which tracks only the changes in the allele frequencies. We will assume that the population has a life history summarized by the following diagram:



In principle, we need to know the frequencies of each of the three diploid genotypes. However, if the population is random mating, then we can formulate a model which tracks only the changes in the allele frequencies. We will assume that the population has a life history summarized by the following diagram:



In other words, the diploid adults produce an effectively infinite number of haploid gametes, which combine at random to form diploid zygotes. These zygotes then undergo selection while developing into juveniles. Finally, population regulation (e.g., due to competition for territories) allows only  $N$  juveniles to survive to adulthood.

We can study this model in terms of the changes in the **gametic** frequencies of  $A_1$  from generation to generation. However, to determine the transition probabilities for  $p \rightarrow p'$ , we will need to examine how mating, selection and population regulation alter the genotypic frequencies at intermediate stages.

We can study this model in terms of the changes in the **gametic** frequencies of  $A_1$  from generation to generation. However, to determine the transition probabilities for  $p \rightarrow p'$ , we will need to examine how mating, selection and population regulation alter the genotypic frequencies at intermediate stages.

Suppose that the gametic frequency of  $A_1$  in generation  $t$  is  $p^N(t) = p$ .

We can study this model in terms of the changes in the **gametic** frequencies of  $A_1$  from generation to generation. However, to determine the transition probabilities for  $p \rightarrow p'$ , we will need to examine how mating, selection and population regulation alter the genotypic frequencies at intermediate stages.

Suppose that the gametic frequency of  $A_1$  in generation  $t$  is  $p^N(t) = p$ .

**Random mating:** Because mating is random and the number of gametes is assumed to be infinite, the frequencies of the diploid genotypes immediately following mating are in **Hardy-Weinberg** equilibrium:

genotype	frequency
$A_1A_1$	$p_{11} = p^2$
$A_1A_2$	$p_{12} = 2p(1 - p)$
$A_2A_2$	$p_{22} = (1 - p)^2$

**Selection:** Selection causes the frequency of each genotype to change in proportion to its relative fitness. That is, if  $p_{ij}$  is the frequency of  $A_iA_j$  before selection, then the frequency  $p_{ij}^*$  of this genotype after selection is

$$p_{ij}^* = p_{ij} \left( \frac{w_{ij}}{\bar{w}} \right),$$

where  $w_{ij}$  is the relative fitness of  $A_iA_j$  and  $\bar{w}$  is the mean fitness of the population:

$$\begin{aligned}\bar{w} &= p^2(1+s) + 2p(1-p)(1+hs) + (1-p)^2 \\ &= 1 + p^2s + 2p(1-p)hs.\end{aligned}$$

**Selection:** Selection causes the frequency of each genotype to change in proportion to its relative fitness. That is, if  $p_{ij}$  is the frequency of  $A_iA_j$  before selection, then the frequency  $p_{ij}^*$  of this genotype after selection is

$$p_{ij}^* = p_{ij} \left( \frac{w_{ij}}{\bar{w}} \right),$$

where  $w_{ij}$  is the relative fitness of  $A_iA_j$  and  $\bar{w}$  is the mean fitness of the population:

$$\begin{aligned}\bar{w} &= p^2(1+s) + 2p(1-p)(1+hs) + (1-p)^2 \\ &= 1 + p^2s + 2p(1-p)hs.\end{aligned}$$

Consulting the table of relative fitnesses on a previous slide, we find:

genotype	frequency after selection
$A_1A_1$	$p_{11}^* = p_{11}(1+s)/\bar{w}$
$A_1A_2$	$p_{12}^* = p_{12}(1+hs)/\bar{w}$
$A_2A_2$	$p_{22}^* = p_{22}/\bar{w}$

**Population regulation:** We will assume that population regulation acts in a manner similar to the Wright-Fisher scheme: the  $N$  adults are sampled uniformly at random (independently and with replacement) from the juvenile cohort. However, because the species is diploid, we are actually sampling two genes per individual.

**Population regulation:** We will assume that population regulation acts in a manner similar to the Wright-Fisher scheme: the  $N$  adults are sampled uniformly at random (independently and with replacement) from the juvenile cohort. However, because the species is diploid, we are actually sampling two genes per individual.

Suppose that  $p'_{ij}$  denotes the frequency of  $A_i A_j$  genotypes following population regulation. Then, the numbers of adults of each of the three genotypes has a Multinomial distribution:

$$N(p'_{11}, p'_{12}, p'_{22}) \sim \text{Multinomial}(N, p^*_{11}, p^*_{12}, p^*_{22})$$

**Population regulation:** We will assume that population regulation acts in a manner similar to the Wright-Fisher scheme: the  $N$  adults are sampled uniformly at random (independently and with replacement) from the juvenile cohort. However, because the species is diploid, we are actually sampling two genes per individual.

Suppose that  $p'_{ij}$  denotes the frequency of  $A_i A_j$  genotypes following population regulation. Then, the numbers of adults of each of the three genotypes has a Multinomial distribution:

$$N(p'_{11}, p'_{12}, p'_{22}) \sim \text{Multinomial}(N, p^*_{11}, p^*_{12}, p^*_{22})$$

**Meiosis:** The final stage is meiosis, during which each adult produces an effectively infinite number of haploid gametes. Whereas  $A_1 A_1$  adults produce only  $A_1$  gametes and  $A_2 A_2$  adults produce only  $A_2$  gametes,  $A_1 A_2$  adults produce an equal mixture of  $A_1$  and  $A_2$  gametes. It follows that the gametic frequency of  $A_1$  in generation  $t + 1$  is equal to:

$$p^N(t+1) = p' = p'_{11} + \frac{1}{2}p'_{12}.$$

To derive a diffusion approximation for the model, we must assume that the strength of selection is of order  $O(1/N)$ :  $s = \sigma/N$ . Then, if  $\delta = p' - p$  is the change in the gametic frequency of  $A_1$  over one generation, we have:

To derive a diffusion approximation for the model, we must assume that the strength of selection is of order  $O(1/N)$ :  $s = \sigma/N$ . Then, if  $\delta = p' - p$  is the change in the gametic frequency of  $A_1$  over one generation, we have:

$$2N\mathbb{E}_p[\delta] = 2\sigma(h + (1 - 2h)p)p(1 - p) + O(N^{-1})$$

$$2N\mathbb{E}_p[\delta^2] = p(1 - p) + O(N^{-1})$$

$$2N\mathbb{E}_p[\delta^n] = O(N^{-1}) \text{ if } n \geq 3.$$

To derive a diffusion approximation for the model, we must assume that the strength of selection is of order  $O(1/N)$ :  $s = \sigma/N$ . Then, if  $\delta = p' - p$  is the change in the gametic frequency of  $A_1$  over one generation, we have:

$$\begin{aligned} 2N\mathbb{E}_p[\delta] &= 2\sigma(h + (1 - 2h)p)p(1 - p) + O(N^{-1}) \\ 2N\mathbb{E}_p[\delta^2] &= p(1 - p) + O(N^{-1}) \\ 2N\mathbb{E}_p[\delta^n] &= O(N^{-1}) \text{ if } n \geq 3. \end{aligned}$$

It follows that the processes  $(p^N(\lfloor 2Nt \rfloor) : t \geq 0)$  converge to a Wright-Fisher diffusion with generator

$$Gf(p) = \frac{1}{2}p(1 - p)f''(p) + 2\sigma(h + (1 - 2h)p)p(1 - p)f'(p).$$

**Remarks:**

**Remarks:**

- We have rescaled time by a factor of  $2N$  rather than  $N$  because there are  $2N$  genes in a diploid population with  $N$  individuals.
- The infinitesimal variance of the diffusion approximation is the same as that for a haploid Wright-Fisher model with  $N$  individuals:

$$a(p) = p(1 - p)$$

**Remarks:**

- We have rescaled time by a factor of  $2N$  rather than  $N$  because there are  $2N$  genes in a diploid population with  $N$  individuals.
- The infinitesimal variance of the diffusion approximation is the same as that for a haploid Wright-Fisher model with  $N$  individuals:

$$a(p) = p(1 - p)$$

This can be explained by noting that because selection is weak (of order  $O(1/N)$ ) and mating is random, the two alleles carried by each individual are nearly independent of one another. Thus, sampling  $N$  individuals at random in this model is essentially equivalent to sampling  $2N$  individuals at random.

- The infinitesimal drift of the diffusion approximation is:

$$\begin{aligned}b(p) &= 2\sigma(h + (1 - 2h)p)p(1 - p) \\&= \sigma(p)p(1 - p).\end{aligned}$$

- $\sigma(p) \equiv 2\sigma(h + (1 - 2h)p)$  is the selection coefficient of allele  $A_1$  relative to  $A_2$ .
- Selection in diploid populations is usually **frequency-dependent**.

- The infinitesimal drift of the diffusion approximation is:

$$\begin{aligned} b(p) &= 2\sigma(h + (1 - 2h)p)p(1 - p) \\ &= \sigma(p)p(1 - p). \end{aligned}$$

- $\sigma(p) \equiv 2\sigma(h + (1 - 2h)p)$  is the selection coefficient of allele  $A_1$  relative to  $A_2$ .
- Selection in diploid populations is usually **frequency-dependent**.

The **marginal fitness** of an allele is equal to the average of the fitnesses of the genotypes containing that allele, weighted by the frequencies of those genotypes:

$$\begin{aligned} w_{A_1} &= pw_{11} + (1 - p)w_{12} = 1 + ps + (1 - p)hs \\ w_{A_2} &= (1 - p)w_{22} + pw_{12} = 1 + phs \\ w_{A_1} - w_{A_2} &= (h + p(1 - 2h))s \end{aligned}$$

Let's consider some specific cases.

Let's consider some specific cases.

**Genic selection:** If  $h = 1/2$ , then the selection coefficient

$$\sigma(p) = \sigma$$

does not depend on the allele frequency, and the diffusion is equivalent to that derived for a haploid population in which the relative fitnesses of the alleles are  $1 + \sigma/N : 1$ .

Let's consider some specific cases.

**Genic selection:** If  $h = 1/2$ , then the selection coefficient

$$\sigma(p) = \sigma$$

does not depend on the allele frequency, and the diffusion is equivalent to that derived for a haploid population in which the relative fitnesses of the alleles are  $1 + \sigma/N : 1$ .

In this case, we say the fitness of a diploid genotype is an additive function of the number of copies of  $A_1$  that it contains:

$A_1A_1$	$A_1A_2$	$A_2A_2$
$1 + s$	$1 + s/2$	1

In other words, each copy of  $A_1$  adds  $s/2$  to the relative fitness of the genotype.

**$A_1$  is dominant:** If  $h \in (1/2, 1]$ , then  $A_1$  is said to be dominant to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_1A_1$  homozygote than to the fitness of the  $A_2A_2$  homozygote:

**$A_1$  is dominant:** If  $h \in (1/2, 1]$ , then  $A_1$  is said to be dominant to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_1A_1$  homozygote than to the fitness of the  $A_2A_2$  homozygote:

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ p^2 & 2p(1-p) & (1-p)^2 \\ \hline 1+s & 1+hs & 1 \end{array}$$

$A_1$  is dominant: If  $h \in (1/2, 1]$ , then  $A_1$  is said to be dominant to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_1A_1$  homozygote than to the fitness of the  $A_2A_2$  homozygote:

$$\begin{array}{ccc}
 A_1A_1 & A_1A_2 & A_2A_2 \\
 p^2 & 2p(1-p) & (1-p)^2 \\
 1+s & 1+hs & 1
 \end{array}$$

In this case, the selection coefficient is a decreasing function of  $p$ :

$$\sigma(p) = 2\sigma(h + (1 - 2h)p),$$

i.e., the selective advantage of  $A_1$  is greater when  $A_1$  is rare than when  $A_1$  is common.

**$A_1$  is dominant:** If  $h \in (1/2, 1]$ , then  $A_1$  is said to be dominant to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_1A_1$  homozygote than to the fitness of the  $A_2A_2$  homozygote:

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ p^2 & 2p(1-p) & (1-p)^2 \\ 1+s & 1+hs & 1 \end{array}$$

In this case, the selection coefficient is a decreasing function of  $p$ :

$$\sigma(p) = 2\sigma(h + (1-2h)p),$$

i.e., the selective advantage of  $A_1$  is greater when  $A_1$  is rare than when  $A_1$  is common.

If  $h = 1$ , then  $A_1$  is said to be **completely dominant** to  $A_2$ .

With  $h = 1$ , the marginal fitnesses are:

$$\begin{aligned}w_{A_1} &= p(1+s) + (1-p)(1+s) = (1+s) \\w_{A_2} &= p(1+s) + (1-p) = 1+ps.\end{aligned}$$

With  $h = 1$ , the marginal fitnesses are:

$$\begin{aligned}w_{A_1} &= p(1+s) + (1-p)(1+s) = (1+s) \\w_{A_2} &= p(1+s) + (1-p) = 1+ps.\end{aligned}$$

Thus, when  $A_1$  is rare,  $w_{A_1} = 1+s$  while  $w_{A_2} \approx 1$ , and so the selection coefficient is approximately  $s$ .

With  $h = 1$ , the marginal fitnesses are:

$$\begin{aligned}w_{A_1} &= p(1+s) + (1-p)(1+s) = (1+s) \\w_{A_2} &= p(1+s) + (1-p) = 1+ps.\end{aligned}$$

Thus, when  $A_1$  is rare,  $w_{A_1} = 1+s$  while  $w_{A_2} \approx 1$ , and so the selection coefficient is approximately  $s$ .

In contrast, when  $A_1$  is common,  $w_{A_2} \approx w_{A_1} = 1+s$  and so the selection coefficient is approximately 0.

With  $h = 1$ , the marginal fitnesses are:

$$\begin{aligned}w_{A_1} &= p(1+s) + (1-p)(1+s) = (1+s) \\w_{A_2} &= p(1+s) + (1-p) = 1+ps.\end{aligned}$$

Thus, when  $A_1$  is rare,  $w_{A_1} = 1+s$  while  $w_{A_2} \approx 1$ , and so the selection coefficient is approximately  $s$ .

In contrast, when  $A_1$  is common,  $w_{A_2} \approx w_{A_1} = 1+s$  and so the selection coefficient is approximately 0.

In other words, under complete dominance, the marginal fitness of the dominant allele is independent of its frequency. On the other hand, the marginal fitness of  $A_2$  does depend on  $p$  because the fitness of any particular copy of  $A_2$  depends on whether it occurs within a homozygote or heterozygote.

$A_1$  is recessive: If  $h \in [0, 1/2)$ , then  $A_1$  is said to be recessive to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_2A_2$  homozygote than to the fitness of the  $A_1A_1$  homozygote. In this case, the selection coefficient  $\sigma(p)$  is an increasing function of  $p$ .

$A_1$  is recessive: If  $h \in [0, 1/2)$ , then  $A_1$  is said to be recessive to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_2A_2$  homozygote than to the fitness of the  $A_1A_1$  homozygote. In this case, the selection coefficient  $\sigma(p)$  is an increasing function of  $p$ .

If  $h = 0$ , then  $A_1$  is completely recessive to  $A_2$  and the marginal fitnesses are

$$\begin{aligned} w_{A_1} &= p(1 + s) + (1 - p) = 1 + ps \\ w_{A_2} &= p + (1 - p) = 1. \end{aligned}$$

$A_1$  is recessive: If  $h \in [0, 1/2)$ , then  $A_1$  is said to be recessive to  $A_2$  because the fitness of the heterozygote is closer to that of the  $A_2A_2$  homozygote than to the fitness of the  $A_1A_1$  homozygote. In this case, the selection coefficient  $\sigma(p)$  is an increasing function of  $p$ .

If  $h = 0$ , then  $A_1$  is completely recessive to  $A_2$  and the marginal fitnesses are

$$\begin{aligned} w_{A_1} &= p(1 + s) + (1 - p) = 1 + ps \\ w_{A_2} &= p + (1 - p) = 1. \end{aligned}$$

Thus, the marginal fitness of  $A_1$  increases with frequency because more copies of  $A_1$  are incorporated into homozygotes, while the marginal fitness of  $A_2$  does not depend on the frequency.

**Overdominance:** If  $\sigma > 0$  and  $h > 1$ , then the fitness of the heterozygote is greater than the fitness of either homozygote and the two alleles are said to be overdominant.

**Overdominance:** If  $\sigma > 0$  and  $h > 1$ , then the fitness of the heterozygote is greater than the fitness of either homozygote and the two alleles are said to be overdominant.

In this case, there is an intermediate frequency,

$$p^* = \frac{h}{2h-1} \in (0, 1),$$

such that

- $\sigma(p^*) = 2\sigma(h + (1-2h)p^*) = 0$  (both alleles are equally fit)
- $\sigma(p) > 0$  if  $p < p^*$  ( $A_1$  is more fit)
- $\sigma(p) < 0$  if  $p > p^*$  ( $A_2$  is more fit)

**Overdominance:** If  $\sigma > 0$  and  $h > 1$ , then the fitness of the heterozygote is greater than the fitness of either homozygote and the two alleles are said to be overdominant.

In this case, there is an intermediate frequency,

$$p^* = \frac{h}{2h-1} \in (0, 1),$$

such that

- $\sigma(p^*) = 2\sigma(h + (1-2h)p^*) = 0$  (both alleles are equally fit)
- $\sigma(p) > 0$  if  $p < p^*$  ( $A_1$  is more fit)
- $\sigma(p) < 0$  if  $p > p^*$  ( $A_2$  is more fit)

Thus,  $A_1$  tends to rise in frequency when rare and tends to decrease in frequency when common. This kind of selection is called **balancing selection** and maintains genetic variation in the population.

**Example:** The classic example of overdominance is the **sickle cell** mutation that is prevalent in some human populations with a high incidence of malaria infections. This is an amino-acid changing mutation which causes hemoglobin molecules to clump together and deform red blood cells.

**Example:** The classic example of overdominance is the **sickle cell** mutation that is prevalent in some human populations with a high incidence of malaria infections. This is an amino-acid changing mutation which causes hemoglobin molecules to clump together and deform red blood cells.

There are two alleles -  $A$  which is the non-sickle-cell ('wild type') allele and  $S$  which causes sickling of red blood cells. The diploid genotypes and their phenotypes are:

**Example:** The classic example of overdominance is the **sickle cell** mutation that is prevalent in some human populations with a high incidence of malaria infections. This is an amino-acid changing mutation which causes hemoglobin molecules to clump together and deform red blood cells.

There are two alleles -  $A$  which is the non-sickle-cell ('wild type') allele and  $S$  which causes sickling of red blood cells. The diploid genotypes and their phenotypes are:

- $AA$ : These individuals have normal hemoglobin, but are susceptible to malaria infections (which can be fatal in children and pregnant women).
- $AS$ : These individuals have a mild form of anemia but are very resistant to malaria infection.
- $SS$ : These individuals have a very severe anemia.

In regions with a high incidence of malaria, the benefits of the resistance to malaria conferred by the *AS* genotype outweigh the costs of the mild anemia, and *AS* heterozygotes have higher fitness than either homozygote.

In regions with a high incidence of malaria, the benefits of the resistance to malaria conferred by the *AS* genotype outweigh the costs of the mild anemia, and *AS* heterozygotes have higher fitness than either homozygote.

The viabilities of the three genotypes in malarial regions have been estimated to be (see Cavalli-Sforza and Bodmer, 1971):

<i>SS</i>	<i>AS</i>	<i>AA</i>
0.2	1.1	1

In regions with a high incidence of malaria, the benefits of the resistance to malaria conferred by the *AS* genotype outweigh the costs of the mild anemia, and *AS* heterozygotes have higher fitness than either homozygote.

The viabilities of the three genotypes in malarial regions have been estimated to be (see Cavalli-Sforza and Bodmer, 1971):

<i>SS</i>	<i>AS</i>	<i>AA</i>
0.2	1.1	1

Thus, in the notation of our model,  $\sigma \approx -0.8$  and  $h \approx -0.125$ . This predicts an equilibrium frequency for *S* of  $p^* = 0.1$ , whereas the observed frequency is about 0.09 averaged across West Africa.

In regions with a high incidence of malaria, the benefits of the resistance to malaria conferred by the *AS* genotype outweigh the costs of the mild anemia, and *AS* heterozygotes have higher fitness than either homozygote.

The viabilities of the three genotypes in malarial regions have been estimated to be (see Cavalli-Sforza and Bodmer, 1971):

<i>SS</i>	<i>AS</i>	<i>AA</i>
0.2	1.1	1

Thus, in the notation of our model,  $\sigma \approx -0.8$  and  $h \approx -0.125$ . This predicts an equilibrium frequency for *S* of  $p^* = 0.1$ , whereas the observed frequency is about 0.09 averaged across West Africa.

In contrast, in regions with little or no malaria, the sickle cell mutation is deleterious and is usually very rare.

**Underdominance:** If  $\sigma < 0$  and  $h > 1$ , then the fitness of the heterozygote is less than that of either homozygote and the two alleles are said to be underdominant.

**Underdominance:** If  $\sigma < 0$  and  $h > 1$ , then the fitness of the heterozygote is less than that of either homozygote and the two alleles are said to be underdominant.

In this case, there is still an intermediate frequency,

$$p^* = \frac{h}{2h-1} \in (0, 1),$$

such that  $\sigma(p^*) = 0$ , but now

- $\sigma(p) < 0$  if  $p < p^*$  ( $A_2$  is more fit)
- $\sigma(p) > 0$  if  $p > p^*$  ( $A_1$  is more fit)

**Underdominance:** If  $\sigma < 0$  and  $h > 1$ , then the fitness of the heterozygote is less than that of either homozygote and the two alleles are said to be underdominant.

In this case, there is still an intermediate frequency,

$$p^* = \frac{h}{2h-1} \in (0, 1),$$

such that  $\sigma(p^*) = 0$ , but now

- $\sigma(p) < 0$  if  $p < p^*$  ( $A_2$  is more fit)
- $\sigma(p) > 0$  if  $p > p^*$  ( $A_1$  is more fit)

Thus,  $A_1$  tends to decrease in frequency when rare, but increases when common. This kind of selection favors common alleles and removes polymorphism from the population.

**Underdominance:** If  $\sigma < 0$  and  $h > 1$ , then the fitness of the heterozygote is less than that of either homozygote and the two alleles are said to be underdominant.

In this case, there is still an intermediate frequency,

$$p^* = \frac{h}{2h-1} \in (0, 1),$$

such that  $\sigma(p^*) = 0$ , but now

- $\sigma(p) < 0$  if  $p < p^*$  ( $A_2$  is more fit)
- $\sigma(p) > 0$  if  $p > p^*$  ( $A_1$  is more fit)

Thus,  $A_1$  tends to decrease in frequency when rare, but increases when common. This kind of selection favors common alleles and removes polymorphism from the population.

**Example:** Chromosomal rearrangements are often underdominant due to problems that occur during meiosis. These may sometimes play a role in speciation.

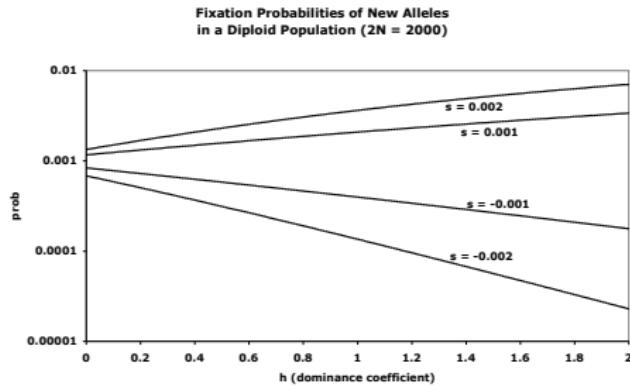
We can use our analysis of the hitting probabilities of a diffusion process to calculate the fixation probabilities of alleles in a diploid population:

$$u(p) \equiv \mathbb{P}_p\{A_1 \text{ is fixed}\} = \frac{\int_0^p e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}{\int_0^1 e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}.$$

We can use our analysis of the hitting probabilities of a diffusion process to calculate the fixation probabilities of alleles in a diploid population:

$$u(p) \equiv \mathbb{P}_p\{A_1 \text{ is fixed}\} = \frac{\int_0^p e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}{\int_0^1 e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}.$$

The fixation probabilities of a new allele ( $p = 1/2N$ ) are shown as functions of the dominance coefficient  $h$  in the figure below.



This figure makes several important points:

This figure makes several important points:

- Dominant beneficial mutations are more likely to be fixed than recessive beneficial mutations.

This figure makes several important points:

- Dominant beneficial mutations are more likely to be fixed than recessive beneficial mutations.
- Recessive deleterious mutations are more likely to be fixed than dominant deleterious mutations.

This figure makes several important points:

- Dominant beneficial mutations are more likely to be fixed than recessive beneficial mutations.
- Recessive deleterious mutations are more likely to be fixed than dominant deleterious mutations.
- Overdominance increases the fixation probabilities of rare alleles.

This figure makes several important points:

- Dominant beneficial mutations are more likely to be fixed than recessive beneficial mutations.
- Recessive deleterious mutations are more likely to be fixed than dominant deleterious mutations.
- Overdominance increases the fixation probabilities of rare alleles.
- Underdominance decreases the fixation probabilities of rare alleles.

This figure makes several important points:

- Dominant beneficial mutations are more likely to be fixed than recessive beneficial mutations.
- Recessive deleterious mutations are more likely to be fixed than dominant deleterious mutations.
- Overdominance increases the fixation probabilities of rare alleles.
- Underdominance decreases the fixation probabilities of rare alleles.

It has been observed that deleterious mutations are more likely to be recessive than dominant. This may be because many deleterious mutations are **loss-of-function mutations**: the mutation prevents the gene from being expressed or inactivates the protein. In this case, the wild type allele in a heterozygote may produce sufficient protein to complement the inactive allele. Such alleles will be more likely to be fixed in a population than they would if they were dominant.

# Cannings Models and Diffusion Approximations

## Cannings Models and Diffusion Approximations

Earlier, we showed that the diffusion approximations for the Wright-Fisher model and the Moran model are equivalent, apart from a change in time scale:

$$\begin{aligned} Gf(p) &= \frac{1}{2}p(1-p)f''(p) \quad (\text{Wright-Fisher model}) \\ Gf(p) &= p(1-p)f''(p) \quad (\text{Moran model}). \end{aligned}$$

## Cannings Models and Diffusion Approximations

Earlier, we showed that the diffusion approximations for the Wright-Fisher model and the Moran model are equivalent, apart from a change in time scale:

$$\begin{aligned} Gf(p) &= \frac{1}{2}p(1-p)f''(p) \quad (\text{Wright-Fisher model}) \\ Gf(p) &= p(1-p)f''(p) \quad (\text{Moran model}). \end{aligned}$$

Our goal in this section is to show that the Wright-Fisher diffusion is also the diffusion approximation for a much wider class of neutral population genetics models, called **Cannings models**. These will have the property that the numbers of surviving offspring of the individuals alive in the previous generation are exchangeable:

$$(\zeta_1, \dots, \zeta_N) \stackrel{d}{=} (\zeta_{\sigma(1)}, \dots, \zeta_{\sigma(N)}),$$

where  $\sigma$  is any permutation of  $\{1, \dots, N\}$ .

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

- Constant population size:  $N$  haploid adults.

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.
- Each adult gives birth to a random number of offspring, and the numbers of offspring born to the  $N$  adults are IID random variables, with the same distribution as some random variable  $\eta$ .

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.
- Each adult gives birth to a random number of offspring, and the numbers of offspring born to the  $N$  adults are IID random variables, with the same distribution as some random variable  $\eta$ .
- $N$  of the offspring are sampled without replacement to develop into the adults of the next generation; all other individuals (adult and offspring) die.

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure. To make these processes more explicit, we will consider models that satisfy the following assumptions:

- Constant population size:  $N$  haploid adults.
- Non-overlapping generations.
- Each adult gives birth to a random number of offspring, and the numbers of offspring born to the  $N$  adults are IID random variables, with the same distribution as some random variable  $\eta$ .
- $N$  of the offspring are sampled without replacement to develop into the adults of the next generation; all other individuals (adult and offspring) die.
- Notice that the numbers of surviving offspring are exchangeable but not independent.

This model explicitly prescribes the **fecundity** of each adult, and delays population regulation until after birth. For example, this could apply to a haploid organism that sheds spores which only survive if they land on an unoccupied territory.

This model explicitly prescribes the **fecundity** of each adult, and delays population regulation until after birth. For example, this could apply to a haploid organism that sheds spores which only survive if they land on an unoccupied territory.

We will make several assumptions about the offspring distribution  $\eta$ :

- $\mathbb{P}\{\eta = 0\} = 0$ , i.e., each adult gives birth to at least one offspring;
- $\mathbb{P}\{\eta \leq M\} = 1$ , i.e.,  $\eta$  is bounded. This implies:
- $m = \mathbb{E}[\eta] < \infty$ ;
- $\sigma^2 = \mathbb{E}[(\eta - m)^2] < \infty$ .

This model explicitly prescribes the **fecundity** of each adult, and delays population regulation until after birth. For example, this could apply to a haploid organism that sheds spores which only survive if they land on an unoccupied territory.

We will make several assumptions about the offspring distribution  $\eta$ :

- $\mathbb{P}\{\eta = 0\} = 0$ , i.e., each adult gives birth to at least one offspring;
- $\mathbb{P}\{\eta \leq M\} = 1$ , i.e.,  $\eta$  is bounded. This implies:
- $m = \mathbb{E}[\eta] < \infty$ ;
- $\sigma^2 = \mathbb{E}[(\eta - m)^2] < \infty$ .

**Note:** The first assumption can be relaxed provided we assume that  $m > 1$ . However, then we must modify the model to deal with times when fewer than  $N$  offspring are born. This happens exceptionally rarely when  $N$  is large and  $m > 1$ , and so it is still possible to derive a diffusion approximation.

Suppose that there are two alleles,  $A$  and  $a$ , present in the population, and let  $p$  denote the frequency of  $A$ . Our first problem is to determine the distribution of the frequency of  $A$  in the next generation. Let us call this  $p'$ .

Suppose that there are two alleles,  $A$  and  $a$ , present in the population, and let  $p$  denote the frequency of  $A$ . Our first problem is to determine the distribution of the frequency of  $A$  in the next generation. Let us call this  $p'$ .

Let  $\eta_1, \dots, \eta_N$  be the offspring numbers of these  $N$  adults, and suppose that we arbitrarily assign the labels  $1, \dots, Np$  to the type  $A$  adults, and  $Np + 1, \dots, N$  to the type  $a$  adults.

Suppose that there are two alleles,  $A$  and  $a$ , present in the population, and let  $p$  denote the frequency of  $A$ . Our first problem is to determine the distribution of the frequency of  $A$  in the next generation. Let us call this  $p'$ .

Let  $\eta_1, \dots, \eta_N$  be the offspring numbers of these  $N$  adults, and suppose that we arbitrarily assign the labels  $1, \dots, Np$  to the type  $A$  adults, and  $Np + 1, \dots, N$  to the type  $a$  adults.

The total number of type  $A$  offspring is

$$Y = \sum_{i=1}^{Np} \eta_i,$$

while the total number of offspring of either type is

$$Z = \sum_{i=1}^N \eta_i.$$

Since the adults of the next generation are obtained by sampling  $N$  individuals without replacement from these  $Z$  offspring, it follows that the conditional distribution of the number of  $A$  offspring surviving to adulthood given  $Y$  and  $Z$  is hypergeometric,

$$X \sim H(Z, Y, N),$$

and then  $p' = X/N$ .

Since the adults of the next generation are obtained by sampling  $N$  individuals without replacement from these  $Z$  offspring, it follows that the conditional distribution of the number of  $A$  offspring surviving to adulthood given  $Y$  and  $Z$  is hypergeometric,

$$X \sim H(Z, Y, N),$$

and then  $p' = X/N$ .

We will need the following facts about the moments of the hypergeometric distribution. If a random variable  $S$  has hypergeometric distribution  $H(N, m, n)$ , then

$$\begin{aligned}\mathbb{E}[S] &= n \left( \frac{m}{N} \right) \\ \mathbb{E}[(S - \mathbb{E}[S])^2] &= n \left( \frac{N-n}{N-1} \right) \left( \frac{m}{N} \right) \left( 1 - \frac{m}{N} \right) \\ \mathbb{E}[(S - \mathbb{E}[S])^e] &= O(n^{e-2}) \text{ if } e \geq 3.\end{aligned}$$

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbb{E}_p[(p' - p)^n]$  as  $N$  tends to infinity.

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbb{E}_p[(p' - p)^n]$  as  $N$  tends to infinity.

First consider the case  $n = 1$ . By conditioning on  $Y$  and  $Z$ , we have:

$$\mathbb{E}_p[p'] = \mathbb{E}_p\left[\frac{1}{N}X\right] = \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}[X|Y, Z]\right]$$

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbb{E}_p[(p' - p)^n]$  as  $N$  tends to infinity.

First consider the case  $n = 1$ . By conditioning on  $Y$  and  $Z$ , we have:

$$\begin{aligned}\mathbb{E}_p[p'] &= \mathbb{E}_p\left[\frac{1}{N}X\right] = \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}[X|Y, Z]\right] \\ &= \mathbb{E}_p\left[\frac{1}{N}\left(N\frac{Y}{Z}\right)\right] = \mathbb{E}_p\left[\frac{Y}{Z}\right].\end{aligned}$$

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbb{E}_p[(p' - p)^n]$  as  $N$  tends to infinity.

First consider the case  $n = 1$ . By conditioning on  $Y$  and  $Z$ , we have:

$$\begin{aligned}\mathbb{E}_p[p'] &= \mathbb{E}_p\left[\frac{1}{N}X\right] = \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}[X|Y, Z]\right] \\ &= \mathbb{E}_p\left[\frac{1}{N}\left(N\frac{Y}{Z}\right)\right] = \mathbb{E}_p\left[\frac{Y}{Z}\right].\end{aligned}$$

Using the fact that

$$\mathbb{E}\left[\frac{\eta_i}{Z}\right] = \mathbb{E}\left[\frac{\eta_1}{Z}\right] \quad (\text{by exchangeability of the } \eta_i)$$

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbb{E}_p[(p' - p)^n]$  as  $N$  tends to infinity.

First consider the case  $n = 1$ . By conditioning on  $Y$  and  $Z$ , we have:

$$\begin{aligned}\mathbb{E}_p[p'] &= \mathbb{E}_p\left[\frac{1}{N}X\right] = \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}[X|Y, Z]\right] \\ &= \mathbb{E}_p\left[\frac{1}{N}\left(N\frac{Y}{Z}\right)\right] = \mathbb{E}_p\left[\frac{Y}{Z}\right].\end{aligned}$$

Using the fact that

$$\mathbb{E}\left[\frac{\eta_i}{Z}\right] = \mathbb{E}\left[\frac{\eta_1}{Z}\right] \quad (\text{by exchangeability of the } \eta_i)$$

this last term can be rewritten as:

$$\mathbb{E}_p\left[\frac{Y}{Z}\right] = \mathbb{E}_p\left[\frac{1}{Z}\sum_{i=1}^{Np}\eta_i\right] = \sum_{i=1}^{Np}\mathbb{E}\left[\frac{\eta_i}{Z}\right] = Np\mathbb{E}\left[\frac{\eta_1}{Z}\right].$$

Similarly,

$$1 = \mathbb{E}_p \left[ \frac{Z}{\bar{Z}} \right] = \mathbb{E}_p \left[ \frac{1}{\bar{Z}} \sum_{i=1}^N \eta_i \right] = \sum_{i=1}^N \mathbb{E} \left[ \frac{\eta_i}{\bar{Z}} \right] = N \mathbb{E} \left[ \frac{\eta_1}{\bar{Z}} \right],$$

Similarly,

$$1 = \mathbb{E}_p \left[ \frac{Z}{Z} \right] = \mathbb{E}_p \left[ \frac{1}{Z} \sum_{i=1}^N \eta_i \right] = \sum_{i=1}^N \mathbb{E} \left[ \frac{\eta_i}{Z} \right] = N \mathbb{E} \left[ \frac{\eta_1}{Z} \right],$$

which implies that

$$\mathbb{E} \left[ \frac{\eta_1}{Z} \right] = \frac{1}{N}.$$

Similarly,

$$1 = \mathbb{E}_p \left[ \frac{Z}{Z} \right] = \mathbb{E}_p \left[ \frac{1}{Z} \sum_{i=1}^N \eta_i \right] = \sum_{i=1}^N \mathbb{E} \left[ \frac{\eta_i}{Z} \right] = N \mathbb{E} \left[ \frac{\eta_1}{Z} \right],$$

which implies that

$$\mathbb{E} \left[ \frac{\eta_1}{Z} \right] = \frac{1}{N}.$$

Substituting this result into the identities obtained on the previous page shows that:

$$\mathbb{E}_p [p'] = Np \mathbb{E} \left[ \frac{\eta_1}{Z} \right] = Np \cdot \frac{1}{N} = p,$$

Similarly,

$$1 = \mathbb{E}_p \left[ \frac{Z}{Z} \right] = \mathbb{E}_p \left[ \frac{1}{Z} \sum_{i=1}^N \eta_i \right] = \sum_{i=1}^N \mathbb{E} \left[ \frac{\eta_i}{Z} \right] = N \mathbb{E} \left[ \frac{\eta_1}{Z} \right],$$

which implies that

$$\mathbb{E} \left[ \frac{\eta_1}{Z} \right] = \frac{1}{N}.$$

Substituting this result into the identities obtained on the previous page shows that:

$$\mathbb{E}_p [p'] = Np \mathbb{E} \left[ \frac{\eta_1}{Z} \right] = Np \cdot \frac{1}{N} = p,$$

and so

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p [p' - p] = N(p - p) = 0.$$

Our next task is to find the limit of  $N\mathbb{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

Our next task is to find the limit of  $N\mathbb{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

$$N\mathbb{E}_p[(p' - p)^2] = N\mathbb{E}_p \left[ \frac{1}{N^2} (X - Np)^2 \right]$$

Our next task is to find the limit of  $N\mathbb{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

$$\begin{aligned} N\mathbb{E}_p[(p' - p)^2] &= N\mathbb{E}_p \left[ \frac{1}{N^2} (X - Np)^2 \right] \\ &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E}[(X - Np)^2 | Y, Z] \right] \end{aligned}$$

Our next task is to find the limit of  $N\mathbb{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

$$\begin{aligned} N\mathbb{E}_p[(p' - p)^2] &= N\mathbb{E}_p \left[ \frac{1}{N^2} (X - Np)^2 \right] \\ &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E}[(X - Np)^2 | Y, Z] \right] \\ &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E} \left[ \left( X - \mathbb{E}[X|Y, Z] + \mathbb{E}[X|Y, Z] - Np \right)^2 | Y, Z \right] \right] \end{aligned}$$

Our next task is to find the limit of  $N\mathbb{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

$$\begin{aligned}
 N\mathbb{E}_p[(p' - p)^2] &= N\mathbb{E}_p\left[\frac{1}{N^2}(X - Np)^2\right] \\
 &= \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}[(X - Np)^2|Y, Z]\right] \\
 &= \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}\left[\left(X - \mathbb{E}[X|Y, Z] + \mathbb{E}[X|Y, Z] - Np\right)^2|Y, Z\right]\right] \\
 &= \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}\left[\left(X - \mathbb{E}[X|Y, Z]\right)^2|Y, Z\right]\right] \\
 &\quad + \mathbb{E}_p\left[\frac{2}{N}\mathbb{E}\left[\left(X - \mathbb{E}[X|Y, Z]\right)\left(\mathbb{E}[X|Y, Z] - Np\right)|Y, Z\right]\right] \\
 &\quad + \mathbb{E}_p\left[\frac{1}{N}\mathbb{E}\left[\left(\mathbb{E}[X|Y, Z] - Np\right)^2|Y, Z\right]\right] \\
 &\equiv K_1 + K_2 + K_3.
 \end{aligned}$$

Beginning with  $K_1$ , observe that the expression inside the  $\mathbb{E}_p[\cdot]$  in this term is just  $1/N$  times the variance of this hypergeometric distribution,

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2|Y, Z] = N \left( \frac{Z - N}{Z - 1} \right) \left( \frac{Y}{Z} \right) \left( 1 - \frac{Y}{Z} \right).$$

Beginning with  $K_1$ , observe that the expression inside the  $\mathbb{E}_p[\cdot]$  in this term is just  $1/N$  times the variance of this hypergeometric distribution,

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2 | Y, Z] = N \left( \frac{Z - N}{Z - 1} \right) \left( \frac{Y}{Z} \right) \left( 1 - \frac{Y}{Z} \right).$$

Consequently,

$$K_1 = \mathbb{E}_p \left[ \left( \frac{Z - N}{Z - 1} \right) \left( \frac{Y}{Z} \right) \left( 1 - \frac{Y}{Z} \right) \right]$$

Beginning with  $K_1$ , observe that the expression inside the  $\mathbb{E}_p[\cdot]$  in this term is just  $1/N$  times the variance of this hypergeometric distribution,

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2|Y, Z] = N \left( \frac{Z-N}{Z-1} \right) \left( \frac{Y}{Z} \right) \left( 1 - \frac{Y}{Z} \right).$$

Consequently,

$$\begin{aligned} K_1 &= \mathbb{E}_p \left[ \left( \frac{Z-N}{Z-1} \right) \left( \frac{Y}{Z} \right) \left( 1 - \frac{Y}{Z} \right) \right] \\ &= \mathbb{E}_p \left[ \left( \frac{\hat{Z}-1}{\hat{Z}-1/N} \right) \left( \frac{\hat{Y}}{\hat{Z}} \right) \left( 1 - \frac{\hat{Y}}{\hat{Z}} \right) \right], \end{aligned}$$

where we have introduced the random variables  $\hat{Y} = Y/N$  and  $\hat{Z} = Z/N$ .

Now, because we have assumed that  $m = \mathbb{E}[\eta] < \infty$ , the strong law of large numbers tells us that

$$\hat{Y} \rightarrow mp \text{ a.s.} \quad \text{and} \quad \hat{Z} \rightarrow m \text{ a.s.}$$

as  $N$  tends to infinity.

Now, because we have assumed that  $m = \mathbb{E}[\eta] < \infty$ , the strong law of large numbers tells us that

$$\hat{Y} \rightarrow mp \text{ a.s.} \quad \text{and} \quad \hat{Z} \rightarrow m \text{ a.s.}$$

as  $N$  tends to infinity.

Furthermore, because we have also assumed that  $\eta \geq 1$ , we know that  $\hat{Z} \geq 1$  which implies that the expression within the expectation is bounded above by 1. Thus, we can bring the limit inside the expectation:

$$\begin{aligned}\lim_{N \rightarrow \infty} K_1 &= \lim_{N \rightarrow \infty} \mathbb{E}_p \left[ \left( \frac{\hat{Z} - 1}{\hat{Z} - 1/N} \right) \left( \frac{\hat{Y}}{\hat{Z}} \right) \left( 1 - \frac{\hat{Y}}{\hat{Z}} \right) \right] \\ &= \left( \frac{m - 1}{m} \right) \left( \frac{mp}{m} \right) \left( 1 - \frac{mp}{m} \right) \\ &= (1 - 1/m)p(1 - p).\end{aligned}$$

Next consider  $K_2$ . Because the expression

$$\mathbb{E}[X|Y, Z] - Np = N\left(\frac{Y}{Z} - p\right)$$

is a deterministic function of  $Y$  and  $Z$  (i.e., if we know  $Y$  and  $Z$ , then we know this expression exactly), it can be pulled outside of the conditional expectation in  $K_2$ .

Next consider  $K_2$ . Because the expression

$$\mathbb{E}[X|Y, Z] - Np = N\left(\frac{Y}{Z} - p\right)$$

is a deterministic function of  $Y$  and  $Z$  (i.e., if we know  $Y$  and  $Z$ , then we know this expression exactly), it can be pulled outside of the conditional expectation in  $K_2$ .

This gives

$$K_2 = \mathbb{E}_p \left[ \frac{2}{N} \mathbb{E}[(X - \mathbb{E}[X|Y, Z])(\mathbb{E}[X|Y, Z] - Np)|Y, Z] \right]$$

Next consider  $K_2$ . Because the expression

$$\mathbb{E}[X|Y, Z] - Np = N\left(\frac{Y}{Z} - p\right)$$

is a deterministic function of  $Y$  and  $Z$  (i.e., if we know  $Y$  and  $Z$ , then we know this expression exactly), it can be pulled outside of the conditional expectation in  $K_2$ .

This gives

$$\begin{aligned} K_2 &= \mathbb{E}_p \left[ \frac{2}{N} \mathbb{E}[(X - \mathbb{E}[X|Y, Z])(\mathbb{E}[X|Y, Z] - Np)|Y, Z] \right] \\ &= \mathbb{E}_p \left[ 2 \left( \frac{Y}{Z} - p \right) \mathbb{E}[X - \mathbb{E}[X|Y, Z]|Y, Z] \right] \end{aligned}$$

Next consider  $K_2$ . Because the expression

$$\mathbb{E}[X|Y, Z] - Np = N\left(\frac{Y}{Z} - p\right)$$

is a deterministic function of  $Y$  and  $Z$  (i.e., if we know  $Y$  and  $Z$ , then we know this expression exactly), it can be pulled outside of the conditional expectation in  $K_2$ .

This gives

$$\begin{aligned} K_2 &= \mathbb{E}_p \left[ \frac{2}{N} \mathbb{E}[(X - \mathbb{E}[X|Y, Z])(\mathbb{E}[X|Y, Z] - Np)|Y, Z] \right] \\ &= \mathbb{E}_p \left[ 2 \left( \frac{Y}{Z} - p \right) \mathbb{E}[X - \mathbb{E}[X|Y, Z]|Y, Z] \right] \\ &= 0, \end{aligned}$$

which vanishes because  $\mathbb{E}[X - \mathbb{E}[X|Y, Z]|Y, Z] = 0$ .

Last, observe that because  $\mathbb{E}[X|Y, Z] = NY/Z$  and  $\mathbb{E}_p[Y/Z] = p$ , we can write

Last, observe that because  $\mathbb{E}[X|Y, Z] = NY/Z$  and  $\mathbb{E}_p[Y/Z] = p$ , we can write

$$K_3 = \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E} \left[ (\mathbb{E}[X|Y, Z] - Np)^2 | Y, Z \right] \right]$$

Last, observe that because  $\mathbb{E}[X|Y, Z] = NY/Z$  and  $\mathbb{E}_p[Y/Z] = p$ , we can write

$$\begin{aligned} K_3 &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E} \left[ (\mathbb{E}[X|Y, Z] - Np)^2 | Y, Z \right] \right] \\ &= \mathbb{E}_p \left[ N \left( \frac{Y}{Z} - p \right)^2 \right] \end{aligned}$$

Last, observe that because  $\mathbb{E}[X|Y, Z] = NY/Z$  and  $\mathbb{E}_p[Y/Z] = p$ , we can write

$$\begin{aligned} K_3 &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E} \left[ (\mathbb{E}[X|Y, Z] - Np)^2 | Y, Z \right] \right] \\ &= \mathbb{E}_p \left[ N \left( \frac{Y}{Z} - p \right)^2 \right] \\ &= N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right). \end{aligned}$$

Last, observe that because  $\mathbb{E}[X|Y, Z] = NY/Z$  and  $\mathbb{E}_p[Y/Z] = p$ , we can write

$$\begin{aligned} K_3 &= \mathbb{E}_p \left[ \frac{1}{N} \mathbb{E} \left[ (\mathbb{E}[X|Y, Z] - Np)^2 | Y, Z \right] \right] \\ &= \mathbb{E}_p \left[ N \left( \frac{Y}{Z} - p \right)^2 \right] \\ &= N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right). \end{aligned}$$

To calculate the first term in parentheses, we will again exploit the exchangeability of  $(\eta_1, \dots, \eta_N)$ . Specifically, observe that

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{Z^2} \eta_i^2 \right] &= \mathbb{E} \left[ \frac{1}{Z^2} \eta_1^2 \right] \quad i = 1, \dots, N \\ \mathbb{E} \left[ \frac{1}{Z^2} \eta_i \eta_j \right] &= \mathbb{E} \left[ \frac{1}{Z^2} \eta_1 \eta_2 \right] \quad i \neq j = 1, \dots, N. \end{aligned}$$

It follows that

$$\mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right]$$

It follows that

$$\begin{aligned}\mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\ &= \mathbb{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right]\end{aligned}$$

It follows that

$$\begin{aligned}
 \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\
 &= \mathbb{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right] \\
 &= Np \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + Np(Np-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right].
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\
 &= \mathbb{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right] \\
 &= Np \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + Np(Np-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right].
 \end{aligned}$$

Exchangeability can be used to further simplify this expression. Observe that:

$$1 = \mathbb{E} \left[ \left( \frac{Z}{Z} \right)^2 \right] = \mathbb{E} \left[ \frac{1}{Z^2} \left( \sum_{i=1}^N \eta_i \right)^2 \right]$$

It follows that

$$\begin{aligned}
 \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\
 &= \mathbb{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right] \\
 &= Np \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + Np(Np-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right].
 \end{aligned}$$

Exchangeability can be used to further simplify this expression. Observe that:

$$\begin{aligned}
 1 &= \mathbb{E} \left[ \left( \frac{Z}{Z} \right)^2 \right] = \mathbb{E} \left[ \frac{1}{Z^2} \left( \sum_{i=1}^N \eta_i \right)^2 \right] \\
 &= N \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + N(N-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right]
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbb{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\
 &= \mathbb{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right] \\
 &= Np \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + Np(Np-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right].
 \end{aligned}$$

Exchangeability can be used to further simplify this expression. Observe that:

$$\begin{aligned}
 1 &= \mathbb{E} \left[ \left( \frac{Z}{Z} \right)^2 \right] = \mathbb{E} \left[ \frac{1}{Z^2} \left( \sum_{i=1}^N \eta_i \right)^2 \right] \\
 &= N \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] + N(N-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] \\
 &= N \left( \frac{v_N}{N^2} \right) + N(N-1) \mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right],
 \end{aligned}$$

where  $v_N \equiv N^2 \mathbb{E}[\eta_1^2/Z^2]$ .

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$N \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$N \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

whence

$$K_3 = N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right)$$

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$N \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

whence

$$\begin{aligned} K_3 &= N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right) \\ &= v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} - p \right) \end{aligned}$$

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$N \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

whence

$$\begin{aligned} K_3 &= N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right) \\ &= v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} - p \right) \\ &= v_N p(1-p) - p(1-p) + O(N^{-1}) \end{aligned}$$

This implies that

$$\mathbb{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$N \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

whence

$$\begin{aligned} K_3 &= N \left( \mathbb{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right) \\ &= v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} - p \right) \\ &= v_N p(1-p) - p(1-p) + O(N^{-1}) \\ &= (v_N - 1)p(1-p) + O(N^{-1}). \end{aligned}$$

Finally, observe that

$$v_N \equiv N^2 \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] = \mathbb{E} \left[ \frac{\eta_1^2}{\hat{Z}^2} \right],$$

where  $\hat{Z} = Z/N$ .

Finally, observe that

$$v_N \equiv N^2 \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] = \mathbb{E} \left[ \frac{\eta_1^2}{\hat{Z}^2} \right],$$

where  $\hat{Z} = Z/N$ . To evaluate the limit as  $N \rightarrow \infty$ , recall that  $\hat{Z} \rightarrow m$  (a.s.), and note that

$$\frac{\eta_1^2}{\hat{Z}^2} \leq M^2 < \infty \quad (\text{since } \hat{Z} > 1 \text{ and } \eta_1 \leq M)$$

Finally, observe that

$$v_N \equiv N^2 \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] = \mathbb{E} \left[ \frac{\eta_1^2}{\hat{Z}^2} \right],$$

where  $\hat{Z} = Z/N$ . To evaluate the limit as  $N \rightarrow \infty$ , recall that  $\hat{Z} \rightarrow m$  (a.s.), and note that

$$\frac{\eta_1^2}{\hat{Z}^2} \leq M^2 < \infty \quad (\text{since } \hat{Z} > 1 \text{ and } \eta_1 \leq M)$$

Thus, as before, we can bring the limit inside the expectation, obtaining

$$\lim_{N \rightarrow \infty} v_N = \frac{1}{m^2} \mathbb{E}[\eta_1^2] \equiv \frac{v}{m^2},$$

Finally, observe that

$$v_N \equiv N^2 \mathbb{E} \left[ \frac{\eta_1^2}{Z^2} \right] = \mathbb{E} \left[ \frac{\eta_1^2}{\hat{Z}^2} \right],$$

where  $\hat{Z} = Z/N$ . To evaluate the limit as  $N \rightarrow \infty$ , recall that  $\hat{Z} \rightarrow m$  (a.s.), and note that

$$\frac{\eta_1^2}{\hat{Z}^2} \leq M^2 < \infty \quad (\text{since } \hat{Z} > 1 \text{ and } \eta_1 \leq M)$$

Thus, as before, we can bring the limit inside the expectation, obtaining

$$\lim_{N \rightarrow \infty} v_N = \frac{1}{m^2} \mathbb{E}[\eta_1^2] \equiv \frac{v}{m^2},$$

and so

$$\lim_{N \rightarrow \infty} K_3 = \left( \frac{v}{m^2} - 1 \right) p(1-p) = \frac{\sigma^2}{m^2},$$

where  $\sigma^2 = v - m^2$  is the variance of  $\eta$ .

Having evaluated each of the terms  $K_1$ ,  $K_2$ , and  $K_3$ , we can calculate

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p [\delta^2] = K_1 + K_2 + K_3$$

Having evaluated each of the terms  $K_1$ ,  $K_2$ , and  $K_3$ , we can calculate

$$\begin{aligned}\lim_{N \rightarrow \infty} N \mathbb{E}_p [\delta^2] &= K_1 + K_2 + K_3 \\ &= (1 - 1/m)p(1 - p) + 0 + \frac{\sigma^2}{m^2}p(1 - p)\end{aligned}$$

Having evaluated each of the terms  $K_1$ ,  $K_2$ , and  $K_3$ , we can calculate

$$\begin{aligned}\lim_{N \rightarrow \infty} N \mathbb{E}_p [\delta^2] &= K_1 + K_2 + K_3 \\ &= (1 - 1/m)p(1 - p) + 0 + \frac{\sigma^2}{m^2}p(1 - p) \\ &= \left( \frac{\sigma^2}{m^2} + 1 - \frac{1}{m} \right) p(1 - p).\end{aligned}$$

To show that third and higher order moments of  $\delta$  vanish in the limit, we calculate

$$N\mathbb{E}_p[\delta^n] = N\mathbb{E}_p\left[\frac{1}{N^n}\mathbb{E}[\delta^n|Y, Z]\right]$$

To show that third and higher order moments of  $\delta$  vanish in the limit, we calculate

$$\begin{aligned} N\mathbb{E}_p[\delta^n] &= N\mathbb{E}_p\left[\frac{1}{N^n}\mathbb{E}[\delta^n|Y, Z]\right] \\ &= \mathbb{E}_p\left[\frac{1}{N^{n-1}}O(N^{n-2})\right] \end{aligned}$$

To show that third and higher order moments of  $\delta$  vanish in the limit, we calculate

$$\begin{aligned} N\mathbb{E}_p[\delta^n] &= N\mathbb{E}_p\left[\frac{1}{N^n}\mathbb{E}[\delta^n|Y, Z]\right] \\ &= \mathbb{E}_p\left[\frac{1}{N^{n-1}}O(N^{n-2})\right] \\ &= N^{-1} \quad \text{when } n \geq 3, \end{aligned}$$

which vanishes as  $N$  tends to infinity.

To show that third and higher order moments of  $\delta$  vanish in the limit, we calculate

$$\begin{aligned}
 N\mathbb{E}_p[\delta^n] &= N\mathbb{E}_p\left[\frac{1}{N^n}\mathbb{E}[\delta^n|Y, Z]\right] \\
 &= \mathbb{E}_p\left[\frac{1}{N^{n-1}}O(N^{n-2})\right] \\
 &= N^{-1} \quad \text{when } n \geq 3,
 \end{aligned}$$

which vanishes as  $N$  tends to infinity.

**Remark:** We are implicitly using the fact that the bound in  $O(N^{n-2})$  is uniform in  $Y$  and  $Z$ .

To summarize, we have shown that:

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta] = 0$$

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^2] = \left(1 - \frac{1}{m} + \frac{\sigma^2}{m^2}\right) p(1-p)$$

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^n] = 0 \text{ if } n \geq 3.$$

To summarize, we have shown that:

$$\begin{aligned}\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta] &= 0 \\ \lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^2] &= \left(1 - \frac{1}{m} + \frac{\sigma^2}{m^2}\right) p(1-p) \\ \lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^n] &= 0 \text{ if } n \geq 3.\end{aligned}$$

Consequently, we know that when  $N$  is large, the process  $(p^N(\lfloor Nt \rfloor) : t \geq 0)$  can be approximated by the diffusion with generator

$$Gf(p) = \frac{1}{2} \left(1 - \frac{1}{m} + \frac{\sigma^2}{m^2}\right) p(1-p) f''(p).$$

To summarize, we have shown that:

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta] = 0$$

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^2] = \left(1 - \frac{1}{m} + \frac{\sigma^2}{m^2}\right) p(1-p)$$

$$\lim_{N \rightarrow \infty} N \mathbb{E}_p[\delta^n] = 0 \text{ if } n \geq 3.$$

Consequently, we know that when  $N$  is large, the process  $(p^N(\lfloor Nt \rfloor) : t \geq 0)$  can be approximated by the diffusion with generator

$$Gf(p) = \frac{1}{2} \left(1 - \frac{1}{m} + \frac{\sigma^2}{m^2}\right) p(1-p)f''(p).$$

This is just a neutral Wright-Fisher diffusion run at speed  $1 - \frac{1}{m} + \frac{\sigma^2}{m^2}$ .

**Remarks:**

- Cannings models can be approximated by the Wright-Fisher diffusion when  $N$  is large.
- In fact, this is true for a much larger class of models - the key conditions are exchangeability and finite variance of the offspring numbers.
- The diffusion approximation is said to be **robust** - the fine details of the model only influence the limit through a scalar time change.
- We can think of the diffusion approximation as a functional version of the Central Limit Theorem.

# Effective Population Size

## Effective Population Size

Suppose that

- $(p_{WF}^N(t) : t \geq 0)$  is the Wright-Fisher model for a population of size  $N$
- $(p_C^N(t) : t \geq 0)$  is a Cannings model for a population of size  $N$

## Effective Population Size

Suppose that

- $(p_{WF}^N(t) : t \geq 0)$  is the Wright-Fisher model for a population of size  $N$
- $(p_C^N(t) : t \geq 0)$  is a Cannings model for a population of size  $N$

If  $N$  is large, then the process  $(p_{WF}^N(\lfloor t \rfloor) : t \geq 0)$  can be approximated by the Wright-Fisher diffusion with generator

$$Gf(p) = \frac{1}{2N}p(1-p)f''(p).$$

## Effective Population Size

Suppose that

- $(p_{WF}^N(t) : t \geq 0)$  is the Wright-Fisher model for a population of size  $N$
- $(p_C^N(t) : t \geq 0)$  is a Cannings model for a population of size  $N$

If  $N$  is large, then the process  $(p_{WF}^N(\lfloor t \rfloor) : t \geq 0)$  can be approximated by the Wright-Fisher diffusion with generator

$$Gf(p) = \frac{1}{2N} p(1-p)f''(p).$$

Likewise, we know that the Cannings model  $(p_C^N(\lfloor t \rfloor) : t \geq 0)$  can be approximated by the Wright-Fisher diffusion with generator

$$Gf(p) = \frac{1}{2N} \left( 1 - \frac{1}{m} + \frac{\sigma^2}{m^2} \right) p(1-p)f''(p).$$

Notice that the Wright-Fisher model ( $p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0$ ) is approximated by the same diffusion process as the Cannings model if

$$N_e = \frac{N}{1 - \frac{1}{m} + \frac{\sigma^2}{m^2}}.$$

Notice that the Wright-Fisher model  $(p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0)$  is approximated by the same diffusion process as the Cannings model if

$$N_e = \frac{N}{1 - \frac{1}{m} + \frac{\sigma^2}{m^2}}.$$

This implies that the two models have similar behavior when  $N$  is large:

$$(p_C^N(\lfloor t \rfloor) : t \geq 0) \approx (p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0).$$

Notice that the Wright-Fisher model  $(p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0)$  is approximated by the same diffusion process as the Cannings model if

$$N_e = \frac{N}{1 - \frac{1}{m} + \frac{\sigma^2}{m^2}}.$$

This implies that the two models have similar behavior when  $N$  is large:

$$(p_C^N(\lfloor t \rfloor) : t \geq 0) \approx (p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0).$$

The quantity  $N_e$  is said to be the **effective population size** of the population with census population size  $N$  described by the Cannings model.

The most important aspect of the effective population size is that it determines the rate at which genetic drift will remove variation from the population. In general, the smaller the effective population size is, the more rapidly genetic variation will be lost.

The most important aspect of the effective population size is that it determines the rate at which genetic drift will remove variation from the population. In general, the smaller the effective population size is, the more rapidly genetic variation will be lost.

For the Cannings model, notice that

- $N_e = N$  when  $\sigma^2 = m$  (as for the Poisson distribution)
- $N_e \approx N$  when  $m^2 \gg \sigma^2$  and  $m \gg 1$
- $N_e \approx \frac{N}{\sigma^2}$  if  $\sigma^2 \gg m^2$

The most important aspect of the effective population size is that it determines the rate at which genetic drift will remove variation from the population. In general, the smaller the effective population size is, the more rapidly genetic variation will be lost.

For the Cannings model, notice that

- $N_e = N$  when  $\sigma^2 = m$  (as for the Poisson distribution)
- $N_e \approx N$  when  $m^2 \gg \sigma^2$  and  $m \gg 1$
- $N_e \approx \frac{N}{\sigma^2}$  if  $\sigma^2 \gg m^2$

In particular, the effective population size may be much smaller than the census population size when there is large variance in reproductive success between adults.

Recall (from the first problem set) that the generator of the diffusion approximation for a Wright-Fisher model with stochastically varying population size is:

$$G\phi(p) = \frac{1}{2N_e} p(1-p)\phi''(p),$$

where

$$\begin{aligned} N_e &= \frac{1}{\mathbb{E}\left[\frac{1}{N}\right]} = \frac{1}{\sum_i \frac{q_i}{N_i}} \\ q_i &= \mathbb{P}\{N = N_i\} \quad (\text{assumed IID across generations}) \end{aligned}$$

Recall (from the first problem set) that the generator of the diffusion approximation for a Wright-Fisher model with stochastically varying population size is:

$$G\phi(p) = \frac{1}{2N_e} p(1-p)\phi''(p),$$

where

$$\begin{aligned} N_e &= \frac{1}{\mathbb{E}\left[\frac{1}{N}\right]} = \frac{1}{\sum_i \frac{q_i}{N_i}} \\ q_i &= \mathbb{P}\{N = N_i\} \quad (\text{assumed IID across generations}) \end{aligned}$$

This shows that:

- The effective population size of a population with fluctuating population sizes is the harmonic mean of the census population size.
- Since the harmonic mean is dominated by small values,  $N_e$  is very sensitive to periods when the population size is small. Such episodes are called **bottlenecks**.

**Estimation:** The effective population size can be estimated from genetic data if we have an independent estimate of the mutation rate per generation (e.g., from divergence between species).

**Estimation:** The effective population size can be estimated from genetic data if we have an independent estimate of the mutation rate per generation (e.g., from divergence between species).

For example, suppose that the population contains two alleles,  $A$  and  $a$ , and that the frequency of  $A$  can be modeled by a Wright-Fisher diffusion with generator

$$G\phi(p) = \frac{1}{2N_e} p(1-p)\phi''(p) + \mu(1-2p)\phi'(p).$$

**Estimation:** The effective population size can be estimated from genetic data if we have an independent estimate of the mutation rate per generation (e.g., from divergence between species).

For example, suppose that the population contains two alleles,  $A$  and  $a$ , and that the frequency of  $A$  can be modeled by a Wright-Fisher diffusion with generator

$$G\phi(p) = \frac{1}{2N_e} p(1-p)\phi''(p) + \mu(1-2p)\phi'(p).$$

Earlier, we showed that the stationary distribution for this process is just the Beta distribution with parameters  $2\theta$  and  $2\theta$ , where  $\theta = N_e\mu$  in a haploid population or  $\theta = 2N_e\mu$  in a diploid population. Consequently, the expected heterozygosity is

$$\bar{H} = 2 \int_0^1 p(1-p)\pi(p)dp = \frac{2\theta}{4\theta + 1}.$$

It follows that a moment estimator for  $\theta$  is

$$\hat{\theta} = \frac{1}{2} \left( \frac{H}{1 - 2H} \right).$$

It follows that a moment estimator for  $\theta$  is

$$\hat{\theta} = \frac{1}{2} \left( \frac{H}{1 - 2H} \right).$$

Thus, if we know  $\mu$ , then we can use this result to estimate  $N_e$ :

$$\hat{N}_e = \frac{1}{2\mu} \left( \frac{H}{1 - 2H} \right) \quad (\text{haploid population})$$

$$\hat{N}_e = \frac{1}{4\mu} \left( \frac{H}{1 - 2H} \right) \quad (\text{diploid population})$$

It follows that a moment estimator for  $\theta$  is

$$\hat{\theta} = \frac{1}{2} \left( \frac{H}{1 - 2H} \right).$$

Thus, if we know  $\mu$ , then we can use this result to estimate  $N_e$ :

$$\hat{N}_e = \frac{1}{2\mu} \left( \frac{H}{1 - 2H} \right) \quad (\text{haploid population})$$

$$\hat{N}_e = \frac{1}{4\mu} \left( \frac{H}{1 - 2H} \right) \quad (\text{diploid population})$$

$H$  can be estimated from sequence data by calculating the sample probability that two individuals have different nucleotides at the same site in a gene.

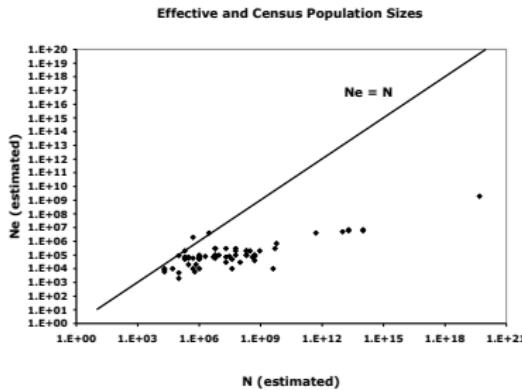
**Remark:** In practice, more sophisticated methods can be used to obtain better estimates of the effective population. Coalescent theory and computationally-intensive methods play an important role in this area.

**Remark:** In practice, more sophisticated methods can be used to obtain better estimates of the effective population. Coalescent theory and computationally-intensive methods play an important role in this area.

As a general rule, the effective population size is usually smaller, sometimes by several orders of magnitude, than the census population size. Some examples include:

Organism	$N_e$
<i>E. coli</i> (bacterium)	$10^8 - 10^9$
<i>D. melanogaster</i> (fruit fly)	$10^6 - 10^7$
house mouse	$10^5 - 10^6$
humans (global)	$10^4$
HIV (within host)	$10^3$

The reduction of the effective population size relative to the census population size can also be seen in the following figure.



These data were compiled by Nei and Graur (1984), using protein diversity ( $H$ ) averaged over 20 or more proteins to estimate  $N_e$  for 77 different species. These estimates assume neutrality, which would be violated if most amino acid variation is weakly deleterious.

**Caveat:** These estimates of  $N_e$  assume that the Wright-Fisher diffusion and Kingman's coalescent) are suitable models of genetic drift. If not, then the concept of the effective population size is not well-defined.

**Caveat:** These estimates of  $N_e$  assume that the Wright-Fisher diffusion and Kingman's coalescent) are suitable models of genetic drift. If not, then the concept of the effective population size is not well-defined.

Some complications that we have ignored in this course are:

- Non-equilibrium population dynamics (e.g., bottlenecks).
- Geographical structure of populations.
- Lineage-specific mutation rates.
- Strong selection and environmental variation.
- Hitchhiking of neutral variation with selected alleles at linked loci.

Different models are needed to address these effects.

# References I



R. Durrett.

*Probability Models for DNA Sequence Evolution.*

Springer-Verlag, 2002.



S. N. Ethier and T. G. Kurtz.

*Markov Processes: Characterization and Convergence.*

John Wiley & Sons, New York, N.Y., 1986.



W.J. Ewens.

*Mathematical Population Genetics. Biomathematics, Vol 9.*

Springer, New York, 2'nd edition, 2004.



R. A. Fisher.

*The Genetical Theory of Natural Selection. A Complete Variorum Edition.*

Oxford University Press, 1999.

# References II

-  J. H. Gillespie.  
*The Causes of Molecular Evolution.*  
Oxford University Press, Oxford, 1991.
-  J. H. Gillespie.  
*Population Genetics. A Concise Guide.*  
Johns Hopkins University Press, 2'nd edition, 2004.
-  D. L. Hartl and A. G. Clark.  
*Principles of Population Genetics.*  
Sinauer, 4'th edition, 2007.
-  W.-H. Li.  
*Molecular Evolution.*  
Sinauer, 1997.