

**Diffusion Processes  
in Theoretical Population Genetics**

**Hilary Term 2008**

**Jay Taylor**

# Contents

1. Population Genetics: Background
2. The Wright-Fisher and Moran Models
3. Markov Processes and Diffusion Approximations
4. Applications of Generators: Stationary Distributions
5. Selection and Genetic Drift
6. Selection in Diploid Populations
7. Cannings Models and Effective Population Size

# Population Genetics - History and Aims

**Population genetics** is the study of genetic variation and the processes that generate and eliminate it. Two kinds of variation are of interest:

- **polymorphism** = variation within a population
- **divergence** = variation between populations (e.g., between different species)

Some of the processes that influence genetic variation are:

- **demographic stochasticity** - reproductive success and survival are in part random (independent of an individual's genes)
- **selection** - some genes do affect reproductive success and survival
- **mutation** - genes can spontaneously change
- **recombination** - sexual reproduction produces a mosaic of the two parental genomes

Population genetics emerged as a scientific discipline during the **Modern Synthesis** when Darwin's theory of evolution by natural selection was re-interpreted in terms of Mendelian genetics.

### **Evolution by natural selection (Darwin, 1859):**

Heritable traits that increase reproductive success will become more common in a population.

The key elements are:

- **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)
- beak/tooth morphology (foraging efficiency)
- 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.

**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.

Most species are either:

- **haploid** - they have a single copy of each chromosome (e.g., most bacteria), or
- **diploid** - they have two (homologous) copies of each chromosome, often one inherited from each parent.

Mendel's particulate theory of inheritance was further explained by **molecular genetics**, which emerged from Watson and Crick's description of the double helix structure of **DNA** in 1953.

DNA is a polymer - a molecule which consists of a sequence of nucleotides, A (adenine), T (thymine), C (cytosine), and G (guanine), that are joined together by sugar-phosphate bonds.

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:

---

|            |                      |
|------------|----------------------|
| parent:    | AATACGTGTAAC         |
| offspring: | AATACGTG <b>C</b> AC |

---

Here, an *A* in the parental gene has mutated to a *C* in the offspring.

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

**Simple Mendelian traits** are those which depend on a single gene. In contrast, many traits (body size, height) depend on many genes (many loci) and on the environment.

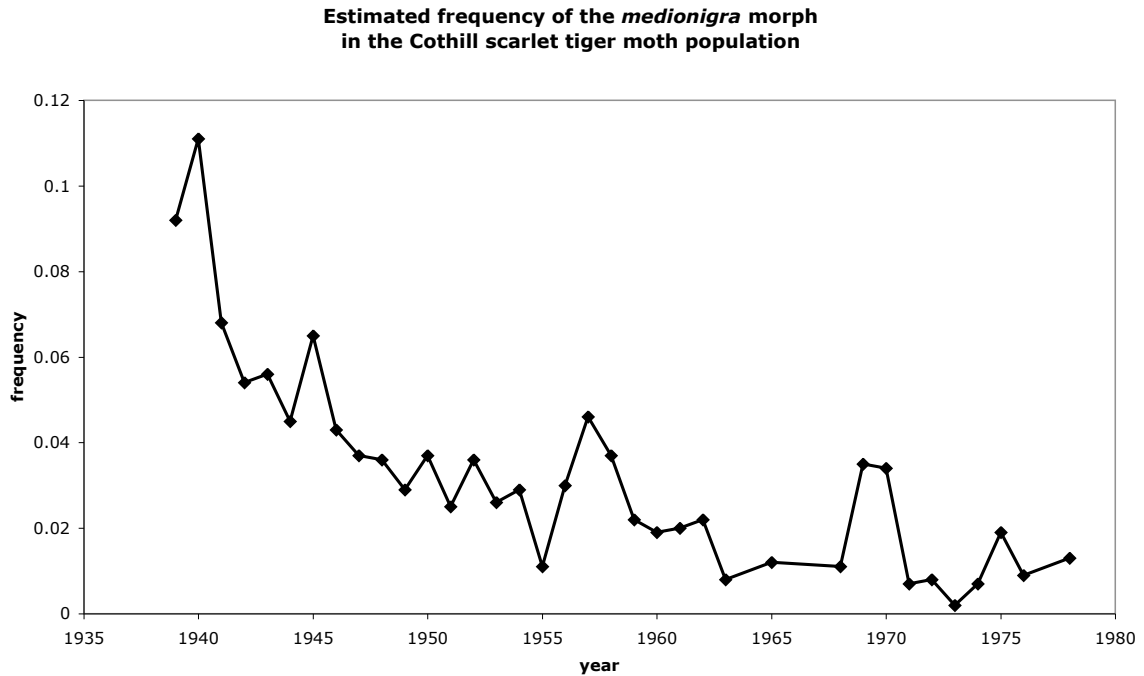
**Example:** Wing color variation in the scarlet tiger moth (*Panaxia dominula*). Three different morphs occur in the population at Cothill Fen (Oxfordshire) and breeding experiments have shown that the wing color pattern is a simple Mendelian trait:

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



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*.

The frequency of the *medionigra* morph in the Cothill population was estimated from population samples collected annually between 1939 and 1979.



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.

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  $A \subset E$ :

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

In other words, a Markov chain is dynamically sufficient - the present state contains all the information available up to then that can be used to predict the future.

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

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

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

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

## 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  $A \subset E$ :

$$\mathbf{P}[X_{t+s} \in A | (X_u : u \in [0, t])] = \mathbf{P}[X_{t+s} \in A | X_t].$$

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

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

To calculate the transition probabilities of a continuous-time chain, we need to exponentiate the rate matrix:

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

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

$$p_{ij} \equiv q_{ij}/(-q_{ii}) \text{ if } j \neq i$$

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 sampling the continuous-time process at the jump times.

## Two classic models in population genetics

Suppose that  $A_1$  and  $A_2$  are two alleles.

### *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.

If  $X(t)$  denotes the number of individuals of type  $A_1$  in the  $t$ 'th generation, then these assumptions define a discrete-time Markov chain  $(X(t), t \geq 0)$  with values in the set  $\{0, 1, \dots, N\}$ . The distribution of  $X(t + 1)$  conditional on  $X(t) = k$  is Binomial( $N, p$ ), where  $p = k/N$  is the frequency of  $A_1$ . Therefore the transition probabilities are:

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

## *Wright-Fisher model*

We are usually interested in the frequency of  $A_1$  rather than the number of copies, so we define the Wright-Fisher process  $(p(t), t \geq 0)$  to be

$$p(t) \equiv N^{-1}X(t).$$

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

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

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

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.

## *Wright-Fisher model*

One measure of genetic diversity at a **biallelic** locus is the quantity  $H(p) = p(1 - p)$ , which is the probability that a sample of two individuals, chosen with replacement, contains both alleles. In diploid populations, twice this quantity is often called the **heterozygosity**.

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

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

By induction on  $t$ , it follows that

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

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 lose variation more rapidly than larger populations.

## *Wright-Fisher model*

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.

*Thm.* 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.,  $P\{\tau < \infty\} = 1$ .

*Proof:* Let  $\kappa(p) = P\{p(1) \in \{0, 1\} | p\}$  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

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

and so

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

## *Wright-Fisher model*

This proof also implies that the expected time to fixation,

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

is finite (in fact, uniformly bounded) for all  $p \in [0, 1]$ . 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

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

This expresses  $\tau(p)$  as the solution to a certain linear equation. However, because the sum ranges over all values in the set  $\{0, 1/N, 2/N, \dots, 1\}$ , 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 only relevant terms in the sum are those in which  $q \approx p$ . This suggests the following approximation.

## *Wright-Fisher model*

Suppose that  $\tau(p)$  is ‘approximately’ smooth and expand  $\tau(q)$  in a Taylor series about  $p$ :

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

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  $\tau(0) = \tau(1) = 0$ . 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  $O(2N)$  when neither allele is rare.

## *Wright-Fisher model*

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

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

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 records the expected value of  $f$  when time is measured in units of  $N$  generations. Then we can again use the Markov property to show that  $U(p, t)$  satisfies the following equation:

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

## *Wright-Fisher model*

If we assume that  $U(p, t)$  can be approximated by a smooth function when  $N$  is large, then we can expand the right-hand side in a Taylor series in  $p$  and  $t$ :

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

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

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

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

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:

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

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\*).

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 (if  $X(t) = k$ ):

| parental genotype | deceased genotype | probability    | change in $X(t)$ |
|-------------------|-------------------|----------------|------------------|
| $A_1$             | $A_1$             | $(k/N)^2$      | 0                |
| $A_1$             | $A_2$             | $k(N - k)/N^2$ | +1               |
| $A_2$             | $A_1$             | $k(N - k)/N^2$ | -1               |
| $A_2$             | $A_2$             | $(1 - k/N)^2$  | 0                |

## *Moran model*

Notice that the transition probabilities only depend on  $X(t) = k$ . This implies that  $(X(t) : t \geq 0)$  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:

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

where  $p = k/N$  is the frequency of  $A_1$ .

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).

One of the virtues of the Moran model is that it is a **birth-death process**:  $p(t)$  can only change by increments of  $1/N$  (up or down).

## *Moran model*

To see how the birth-death property is useful, let us calculate the expected time to fixation for the Moran model as a function of the initial frequency:

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

where  $\tau = \inf\{t \geq 0 : p_t \in \{0, 1\}\}$  as before.

One way to solve this problem is to define a related discrete-time Markov chain,  $(\hat{p}(n); n \geq 0)$ , which records the frequency of  $A_1$  at every reproduction event. Notice that some reproductive events do not lead to a change in  $\hat{p}(n)$ , so  $\hat{p}(n)$  contains more data than the embedded Markov chain. The transition matrix,  $P$ , of  $\hat{p}(n)$  is defined by:

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

## *Moran model*

Our first task is to determine the expected number of steps,  $a_k$ , until  $\hat{p}(n)$  first absorbs at 0 or 1 when  $\hat{p}(0) = k/N$ . By conditioning on the possible values of  $\hat{p}(1)$  and using the Markov property, we obtain the identity:

$$a_k = 1 + \left(1 - \frac{2k(N-k)}{N^2}\right) a_k + \frac{k(N-k)}{N^2} (a_{k-1} + a_{k+1}),$$

which can be rearranged to the following recurrence equation (\*):

$$a_{k-1} - 2a_k + a_{k+1} = -\frac{N^2}{k(N-k)}.$$

Using the boundary conditions,

$$a_0 = a_N = 0,$$

(\*) can be solved explicitly to find

$$a_k = N \left( \sum_{i=1}^k \frac{N-k}{N-i} + \sum_{i=k+1}^{N-1} \frac{k}{i} \right).$$

## *Moran model*

The rationale for introducing the discrete-time process  $\hat{p}(n)$  is that it samples the continuous-time chain  $p(t)$  at times which are independent of  $\hat{p}(n)$ . Furthermore, the waiting times between samples are IID exponential random variables with mean  $N^{-1}$  (the expected time between successive reproduction events in the original Moran model). Consequently, if  $a$  is the absorption time for the chain  $\hat{p}(n)$ ,

$$a = \inf\{n \geq 0 : \hat{p}(n) \in \{0, 1\}\},$$

and if  $(S_n : n \geq 0)$  is an IID sequence of exponential random variables with mean  $N^{-1}$  which are independent of  $a$ , then

$$\tau \stackrel{D}{=} \sum_{n=1}^a S_n.$$

Therefore,

$$\begin{aligned} \tau(p) &= \mathbf{E}_p[\tau] = \mathbf{E}_p \left[ \sum_{n=1}^a S_n \right] = N^{-1} \mathbf{E}_p[a] = N^{-1} a_k \\ &= \sum_{i=1}^k \frac{N-k}{N-i} + \sum_{i=k+1}^{N-1} \frac{k}{i}. \end{aligned}$$

## *Moran model*

If  $N$  is large and we write  $k = Np$ , then  $\tau(p)$  has the following approximation:

$$\begin{aligned}\tau(p) &= N \left( (1-p) \sum_{i=1}^{Np} \frac{1}{N-i} + p \sum_{i=Np+1}^{N-1} \frac{1}{i} \right) \\ &\approx -N (p \log(p) + (1-p) \log(1-p)).\end{aligned}$$

Notice that:

- Unless  $p$  is either very close to 0 or 1,  $\tau(p)$  is of order  $N$ .
- $\tau(p)$  is maximized when  $p = 1/2$  and decreases as  $p$  approaches 0 or 1.
- The expected time to fixation in the Moran model is approximately half that of the Wright-Fisher model (when the population size is the same in both).
- The form of the large  $N$  approximation is the same in both models.

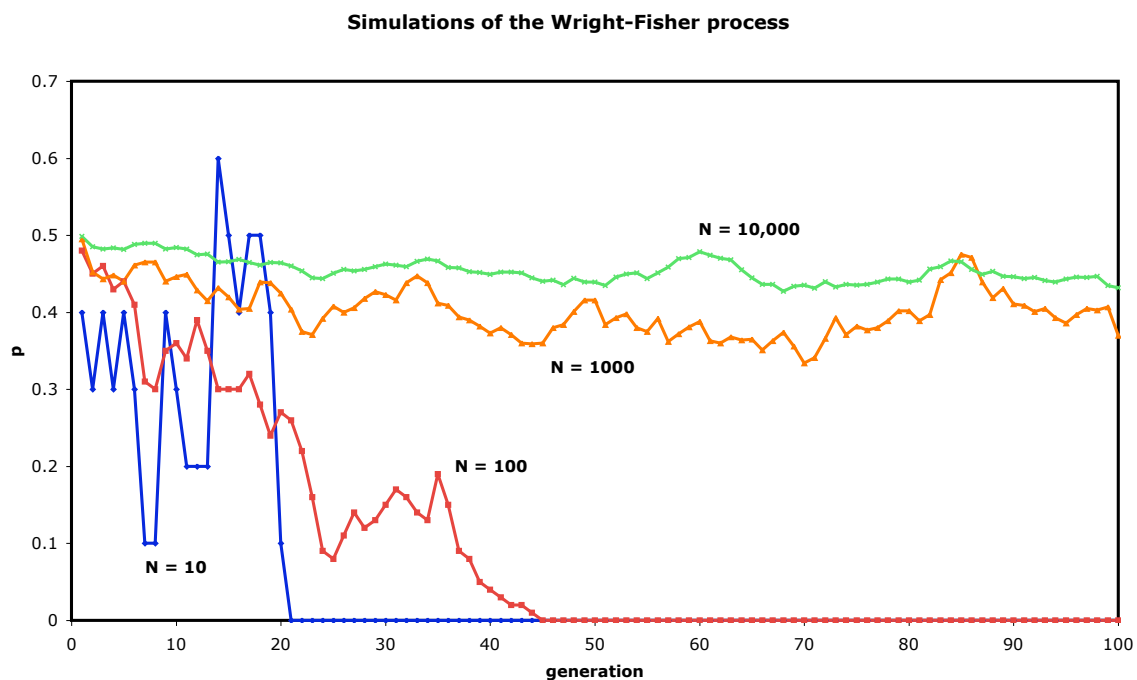
## Diffusion Approximations: Rationale

The preceding examples illustrate some of the difficulties associated with the Wright-Fisher and Moran models. Although it is easy to understand the mechanisms underlying these models, it is often difficult to use them to do explicit calculations. There are two issues:

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

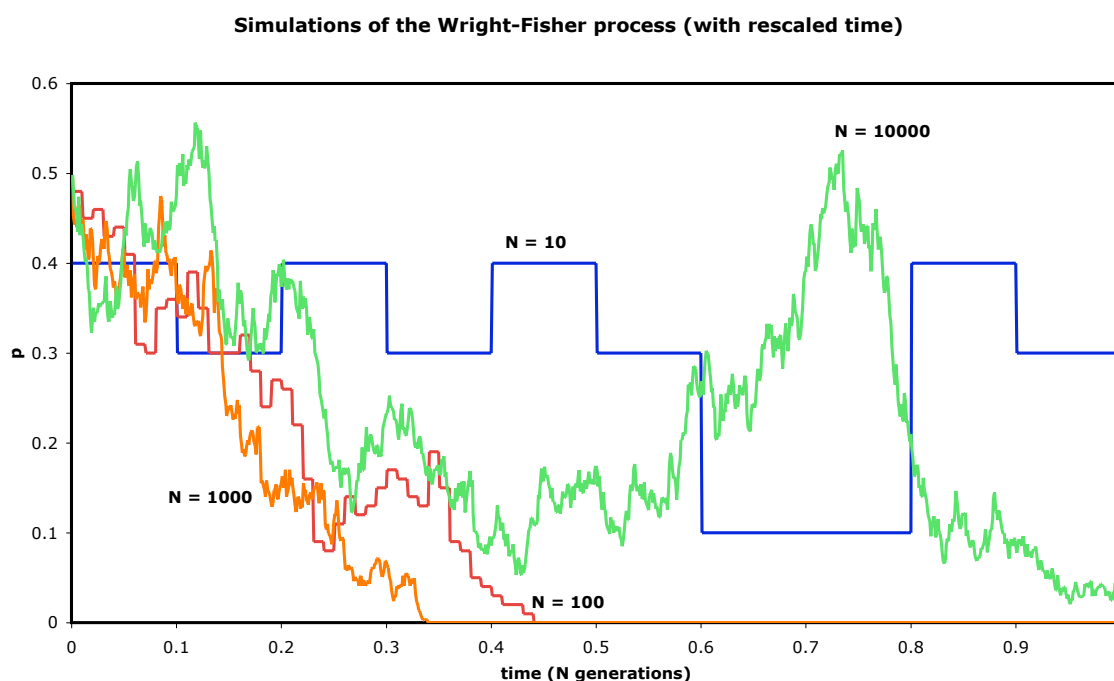
Nonetheless, in each of our examples, we were able to find relatively simple, approximate answers by assuming that the population size is very large and then neglecting terms of lower order. To understand why this approach works, we need to examine the asymptotic behavior of the Markov processes themselves as  $N$  tends to infinity.

The following figure shows a series of simulations of the Wright-Fisher model for 100 generations for  $N = 10, 100, 1000,$  and  $10,000$ .



Notice that the both the size of the fluctuations and the total change in  $p$  over 100 generations decrease as the population size is increased. This is consistent with our earlier observation that the variance in the change in  $p$  over one generation is  $p(1 - p)/N$ .

A different picture emerges if we plot each **sample path** against time measured in units of  $N$  generations. Here we have simulated Wright-Fisher processes with  $N = 10, 100, 1000$  or  $10,000$  for as many generations (so that the total rescaled time is 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. How can we make sense of this notion?

## Characterization of Markov Processes

Suppose that  $(X_t : t \geq 0)$  is an  $E$ -valued continuous-time Markov process. Here we will assume that  $E = (E, d)$  is a metric space and let  $\mathcal{B}(E)$  denote the Borel  $\sigma$ -algebra over  $E$ .

There are several ways to characterize the process  $X$ .

1.) *Transition function:*

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

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

where  $A \in \mathcal{B}(E)$ . Notice that for fixed  $t$  and  $x$ ,  $P(t, x, dy)$  is a probability distribution on  $E$ :

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

Suppose that  $E \subset \mathbb{R}^n$ . For some (but not all) Markov processes, the transition function has a density  $p(t, x, y)$ :

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

*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:

$$\begin{aligned} P(t + s, x, A) &= \mathbf{P}_x\{X_{t+s} \in A\} \\ &= \mathbf{E}_x[\mathbf{P}\{X_{t+s} \in A|X_t\}] \\ &= \mathbf{E}_x[\mathbf{P}_{X_t}\{X_s \in A\}] \\ &= \int_E \mathbf{P}_y\{X_s \in A\}\mathbf{P}_x\{X_t = dy\} \\ &= \int_E P(s, y, A)P(t, x, dy). \end{aligned}$$

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

## 2. Transition semigroup:

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

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

where  $x \in E$  and  $f \in B(E)$ .

The collection of operators  $(T_t : t \geq 0)$  is called the transition semigroup of  $X$  because the operators satisfy the *semigroup property*:

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

for every  $t, s \geq 0$ .

Notice that  $T_0 f(x) = \mathbf{E}_x[f(X_0)] = f(x)$ .

A Markov process is uniquely determined by its transition semigroup, but it too usually cannot be explicitly found.

### 3. Generator

The generator of a Markov process  $X$  is the operator  $G : B(E) \rightarrow B(E)$  defined by

$$Gf(x) = \frac{d}{dt} T_t f(x) |_{t=0} = \lim_{t \rightarrow 0} \frac{1}{t} (\mathbf{E}_x [f(X_t)] - f(x)),$$

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

It is often possible to write down the generator of a Markov process even when we cannot explicitly identify its transition function or semigroup.

*(Important caveat:* It is usually very difficult to identify  $\mathcal{D}(G)$  in its entirety, but theory allows us to work with sufficiently large subsets of the domain which are called cores.)

*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  $f : E \rightarrow \mathbb{R}$ ,

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

Thus, for a (finite state) Markov chain, the generator is both determined by and determines the rate matrix.

## 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 the simplest example of a **time change**.

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

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

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

# Generators and Convergence of Markov Processes

Suppose that for each  $N \geq 1$ ,  $X^N$  is an  $E$ -valued Markov process with generator  $G^N$ , and that  $X$  is an  $E$ -valued Markov process with generator  $G$ . We can often deduce that the processes  $X^N$  converge (in a sense that we will soon make more precise) to  $X$  by showing that the generators converge:

$$\lim_{N \rightarrow \infty} G^N f = Gf$$

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

Here, convergence of the functions  $G^N f$  to  $Gf$  should be uniform on  $E$ :

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

*Example:* Suppose that  $E$  is a finite set and that  $X^N$  and  $X$  are Markov chains with rate matrices  $Q^N$  and  $Q$ , respectively. If  $\lim_{N \rightarrow \infty} q_{ij}^N = q_{ij}$  for all  $i, j$ , then  $\lim_{N \rightarrow \infty} G^N f = Gf$  for every function  $f : E \rightarrow \mathbb{R}$ .

## Diffusion Approximation for the Moran Model

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

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

If the population size  $N$  is large, 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

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

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 the **Wright-Fisher diffusion**. It is just one representative of a much larger class of Markov processes.

## One-dimensional Diffusion Processes

Let  $E = [l, r] \subset \mathcal{R}$  and suppose that  $a : E \rightarrow [0, \infty)$  and  $b : E \rightarrow \mathcal{R}$  are bounded, continuous functions. Then there is a Markov process  $(X_t : t \geq 0)$  with continuous paths with values in  $E$  which corresponds to the generator

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

$X$  is called a one-dimensional diffusion process. The coefficients  $a(x)$  and  $b(x)$  are called the **infinitesimal variance** and **infinitesimal drift** of  $X$ . To see why, we need to consider the following calculations:

## *One-dimensional diffusions*

Notice that if  $f(x) = x$ , then

$$b(x) = Gf(x) = \lim_{t \downarrow 0} \frac{1}{t} \mathbf{E}_x [X_t - x],$$

i.e.,  $b(x)$  measures the tendency of  $X$  to drift upward or downward when it is at  $x$ .

Similarly, if  $f_1(x) = x^2$  and  $f_2(x) = x$ , then

$$\begin{aligned} a(x) &= Gf_1(x) - 2xGf_2(x) \\ &= \lim_{t \downarrow 0} \frac{1}{t} \mathbf{E}_x [(X_t^2 - x^2) - 2x(X_t - x)] \\ &= \lim_{t \downarrow 0} \frac{1}{t} \mathbf{E}_x [(X_t - x)^2], \end{aligned}$$

i.e.,  $a(x)$  measures the variance of the fluctuations of  $X_t$  about  $x$ .

*Example:* The infinitesimal drift and variance of the Wright-Fisher diffusion  $p(t)$  are  $b(p) \equiv 0$  and  $a(p) = p(1 - p)$ . Thus this process has no tendency to either increase or decrease, and the variance of the stochastic fluctuations is largest when  $p = 1/2$  and vanishes when either  $p = 0, 1$ .

# Multidimensional Diffusion Processes

There are also multivariate diffusion processes. Suppose that  $E \subset \mathbb{R}^n$  is a ‘nice’ subset (e.g., a ball) and let  $a_{ij} : E \rightarrow \mathbb{R}$  ( $i, j = 1, \dots, n$ ) and  $b_i : E \rightarrow \mathbb{R}$  ( $i = 1, \dots, n$ ) be bounded continuous functions such that the matrices

$$a(x) = (a_{ij}(x))$$

are non-negative definite for every  $x \in E$ . Then there is a diffusion process  $(X_t : t \geq 0)$  with values in  $E$  which corresponds to the generator

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

The matrix  $a(x)$  is called the infinitesimal variance-covariance matrix and measures the variances and covariances of the fluctuations of the coefficients of  $X_t$  about  $x$ . This is why we require  $a(x)$  to be non-negative definite.

## Diffusion Approximation for the Wright-Fisher Process

Unfortunately, we cannot apply the generator techniques directly to the Wright-Fisher process,  $(p^N(t) : t \geq 0)$ , since this is a discrete-time Markov chain. However, the figures we saw earlier suggest that this process also has a diffusion approximation. To make sense of this, consider the interpolated process

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

where  $\lfloor x \rfloor$  denotes the greatest integer less than  $x$ . In other words, to obtain  $\hat{p}^N(t)$ , we rescale time in the original Wright-Fisher process by a factor of  $N$  and we then introduce a piecewise linear interpolation.

The interpolated process  $\hat{p}^N$  is a continuous-time process, but it does not satisfy the Markov property because the intervals between jumps are not exponentially-distributed. For this reason,  $\hat{p}^N$  does not have a generator. Nonetheless, because the intervals between jumps become arbitrarily small as  $N$  tends to infinity, there is a sense in which  $\hat{p}^N$  ‘almost’ has a generator when  $N$  is large.

## *The Wright-Fisher Diffusion*

Morally, we could define:

$$\hat{G}^N f(p) = \left( \frac{1}{N-1} \right) \mathbf{E}_p [f(\hat{p}(1/N)) - f(p)],$$

where  $t = 1/N$  is the smallest time step over which the interpolated process changes. (If we take  $t$  any smaller, then the expression simply vanishes.)

As with the Moran model, we can find a large  $N$  approximation for this ‘generator’ by expanding the right-hand side in a Taylor series. If  $\delta = \hat{p}(1/N) - p$ , then

$$\begin{aligned} \hat{G}^N f(p) &= N \mathbf{E}_p [\delta f'(p) + \frac{1}{2} \delta^2 f''(p) + O(\delta^3)] \\ &= N \mathbf{E}_p[\delta] f'(p) + \frac{N}{2} \mathbf{E}_p[\delta^2] f''(p) + N \mathbf{E}_p[O(\delta^3)] \\ &= \frac{1}{2} p(1-p) f''(p) + O(N^{-1}), \end{aligned}$$

where we have used the fact that  $N\hat{p}(1/N)$  has a Binomial( $N, p$ ) distribution and

$$\begin{aligned} E[\delta] &= 0 \\ E[\delta^2] &= p(1-p)/N \\ E[\delta^n] &= O(N^{-2}) \text{ if } n \geq 3. \end{aligned}$$

## *The Wright-Fisher Diffusion*

It follows that:

$$\lim_{N \rightarrow \infty} \hat{G}^N f(p) = \frac{1}{2}p(1-p)f''(p),$$

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

Although  $\hat{G}^N$  formally is not a generator, it can be shown that this identity implies that the interpolated Wright-Fisher processes converge to a Wright-Fisher diffusion as  $N$  tends to infinity.

Recall that the diffusion approximation,  $(p_M(t))$ , for the Moran model had generator  $G_M f(p) = p(1-p)f''(p)$ , whereas the diffusion approximation,  $(p_{WF}(t))$ , for the Wright-Fisher model has generator  $G_{WF} f(p) = \frac{1}{2}p(1-p)f''(p)$ . Since  $G_{WF} f(p) = \frac{1}{2}G_M f(p)$ , we know that  $p_{WF}$  is a time change of  $p_M$ :

$$p_{WF}(t) = p_M(t/2). \tag{1}$$

Thus, apart from a rescaling of time, the Moran model and the Wright-Fisher model have very similar properties when  $N$  is large.

# Diffusion Approximations for Discrete-Time Markov Chains

Suppose that  $(X^N(n) : n \geq 0)$  is a discrete-time Markov chain with values in  $\mathbb{R}$  and let  $(\epsilon_N : N \geq 1)$  be a sequence of positive numbers which tends to 0. Let us define the interpolated process  $\hat{X}^N$  as before:

$$\hat{X}^N(t) = X^N(\lfloor \epsilon_N^{-1} t \rfloor).$$

If

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

then it can be shown that the interpolated processes  $\hat{X}^N$  converge to a diffusion process  $X$  with generator

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

$X$  is said to be a diffusion approximation for the Markov chain  $X^N$ .

## Convergence of Stochastic Processes

We still need to make more precise what it means for a sequence of Markov processes  $(X^N)$  to converge to a Markov process  $X$ . 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.

(However, it is possible to give a construction of the Moran model and the Wright-Fisher diffusion that does make this identity true. This is called the *lookdown process*.)

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

# Weak Convergence of Probability Distributions

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

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

Notice that

$$\lim_{N \rightarrow \infty} \mu_N(\{0\}) = 0 \neq 1 = \mu(\{0\}),$$

i.e., weak convergence does not imply the convergence of probabilities of individual sets.

## Convergence in Distribution

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:

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

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’.

*Example:* Let  $X_n, n \geq 1$  be a collection of IID random variables with mean 0 and variance 1, and let  $S_N = X_1 + \dots + X_N$ . Then the Central Limit Theorem asserts that the sequence of normalized partial sums,  $N^{-1/2}S_N$ , converges in distribution to a standard normal random variable.

## Convergence of Markov Processes

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** of  $X^N$  converge weakly to those of  $X$  if for every finite set  $0 \leq t_1 < t_2 < \cdots < t_n$ ,

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

as  $N$  tends to infinity. In particular, the convergence of the one-dimensional distributions,

$$X^N(t) \xrightarrow{d} X(t),$$

for every  $t \geq 0$  holds as a special case.

It is in this sense that Markov processes converge to a limit when their generators converge.

Convergence of Markov processes sometimes holds in an even stronger sense. We can think of a Markov process, not as a sequence of random variables, but as a single path-valued random variable. Usually these paths have some regularity properties, e.g., they are continuous, that allow us to impose a nice topology or metric on the space of all such paths. If so, we can sometimes show that these path-valued random variables converge in distribution.

## Applications of Generators

Having explained how we can use generators to characterize and approximate Markov processes, we will now put them to computational use as well.

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,

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

then

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

where we have used the semigroup property,  $T_{t+s} f = T_s (T_t f)$ , to pass from the first line to the second.

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 equation:

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

where  $G$  is understood to act on  $x$  on the right-hand side. The symbol  $\delta_x(dy)$  is a point mass at  $x$  which means that

$$\int_E \delta_x(dy)f(y) = f(x),$$

while 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

In the special case where  $X$  is a one-dimensional diffusion,  $p(t, x, y)$  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  $X$ .

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

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

## Brownian Motion

One-dimensional Brownian motion is the diffusion process  $B = (B_t : t \geq 0)$  with generator  $Gf(x) = \frac{1}{2}f''(x)$ . The Kolmogorov backward equation in this case is the heat equation,

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

which has solution

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(\frac{-(y-x)^2}{2t}\right).$$

Thus, Brownian motion has mean zero Gaussian increments:

$$B_t - B_s \stackrel{d}{=} N(0, t - s).$$

Similarly,  $d$ -dimensional Brownian motion is the diffusion process,  $B^{(d)} = (B_t^{(d)})$ , with generator:

$$Gf(x) = \frac{1}{2} \sum_{i=1}^d \partial_{x_i x_i} f(x).$$

$B^{(d)}$  has multivariate Gaussian increments with mean zero.

## Kolmogorov Forward Equation

The Kolmogorov backward equation is expressed in terms of derivatives with respect to the initial value of the process. There is a complementary equation which involves derivatives with respect to the final value.

Recall that the Chapman-Kolmogorov equation states that

$$p(t + s, x, y) = \int_E p(t, x, z)p(s, z, y)dz.$$

It follows that

$$\begin{aligned}\partial_t p(t + s, x, y) &= \partial_s p(t + s, x, y) \\ &= \int_E p(t, x, z) \partial_s p(s, z, y) dz \\ &= \int_E p(t, x, z) \left( \frac{1}{2} a(z) p''(s, z, y) + b(z) p'(s, z, y) \right) dz,\end{aligned}$$

where the  $'$  denotes a derivative with respect to  $z$ .

Suppose that  $E = (l, r)$ . Then the expression in the third line can be integrated-by-parts. The first integration gives

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

The second integration gives:

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

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  $p(t, x, y)$ :

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

In general, this will only have a solution if  $a(y)$  and  $b(y)$  are differentiable functions.

## Stationary Distributions

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 stationary, then

$$\begin{aligned}\mathbf{E}_\pi[f(X_0)] &= \int_E f(x)\pi(dx) = \\ \mathbf{E}_\pi[f(X_t)] &= \int_E \mathbf{E}_x[f(X_t)]\pi(dx)\end{aligned}$$

holds for all  $t \geq 0$ . We can rewrite this equation in terms of the semigroup

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

Thus, the right-hand side does not depend on  $t$ , and so

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

In particular, taking  $t = 0$  gives

$$\int_E G f(x)\pi(dx) = 0.$$

## Moments of Stationary Distributions

The identity,

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

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

*Moran model with mutation:* Consider a haploid population of size  $N$  in which reproduction is described by the Moran model. Suppose that there are two alleles in the population,  $A_1$  and  $A_2$ , and that each  $A_1$  allele mutates to  $A_2$  at rate  $\mu_2$  while each  $A_2$  allele mutates to  $A_1$  at rate  $\mu_1$ , and that 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

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

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

*Moran model:*

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  $A_1$ . As expected,  $\bar{p}$  increases with the mutation rate from  $A_2$  to  $A_1$  but decreases with the mutation rate from  $A_1$  to  $A_2$ .

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

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

*Moran model:*

Consequently,

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

and 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

It is sometimes possible to use the identity,

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

to identify the stationary distribution explicitly.

Suppose that  $X = (X_t : t \geq 0)$  is a diffusion process on  $(l, r)$  with infinitesimal variance  $a(x)$  and infinitesimal drift  $b(x)$ . Suppose that  $X$  has a stationary distribution  $\pi(x)$  on  $(l, r)$ . Then,

$$0 = \int_l^r \left( \frac{1}{2}a(x)f''(x) + b(x)f'(x) \right) \pi(x)dx,$$

is satisfied for every twice-differentiable function  $f$ . As in our derivation of the Kolmogorov forward equation, we can integrate-by-parts to obtain the identity

$$0 = \int_l^r f(x) \left( \frac{1}{2}(a(x)\pi(x))'' - (b(x)\pi(x))' \right) dx \\ + \text{boundary terms.}$$

## *Stationary Distributions*

Since this identity holds for every smooth  $f$ , it follows that  $\pi(x)$  must satisfy the differential 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. It turns out that, for  $\pi(x)$  to be integrable, i.e.,

$$\int_l^r \pi(x)dx < \infty,$$

we must choose  $C = 0$ . 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

$$\frac{\pi'(x)}{\pi(x)} = \frac{2b(x)}{a(x)} - \frac{a'(x)}{a(x)}.$$

## *Stationary Distributions*

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  $(l, r)$ . Therefore,

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

is the density of the stationary distribution of the diffusion  $X$  provided we can choose a normalizing constant  $C < \infty$  such that

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

Notice that this will be true if

$$C = \int_l^r \frac{1}{a(x)} \exp \left( 2 \int_c^x \frac{2b(y)}{a(y)} dy \right) dx < \infty,$$

and that different choices of  $c \in (l, r)$  simply change the value of the normalizing constant, but not the density  $\pi(x)$ .

## Ergodicity

It is often the case that if  $X$  is a Markov process with a unique stationary distribution  $\pi(dx)$ , then the one-dimensional distributions of  $X$  converge weakly to  $\pi(dx)$  as  $t$  tends to infinity, i.e.,

$$\lim_{t \rightarrow \infty} \mathbf{E}_x [f(X_t)] = \int_E f(x) \pi(dx),$$

for every initial value  $x \in E$  and every continuous, bounded function  $f : E \rightarrow \mathbb{R}$ . This is an example of an **ergodic** property.

Ergodicity is important because it implies that if a Markov process is run for a sufficiently long time, then its distribution will be close to the stationary distribution, irrespective of its initial condition. In this case, the stationary distribution can be used to make predictions about the probable state of a Markov process when we do not know its history.

## Stationary Solutions to the Forward Equation

If  $X$  is a one-dimensional diffusion with transition density  $p(t, x, y)$  and stationary distribution  $\pi(x)dx$ , then we can also use the ergodicity to deduce that  $\pi(x)$  is a stationary solution of the forward equation

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

Now, by the ergodic property, we expect that

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

for every  $x \in [l, r]$ . (This is actually a very strong form of ergodicity, because it asserts that not only do the distributions converge, but so do their densities.)

Thus, if we take the limit  $t \rightarrow \infty$  in every term in the forward equation (and we assume that the limit and the partial derivatives can be interchanged), then we obtain

$$0 = \partial_t \pi(y) = \frac{1}{2} \partial_{yy} (a(y)\pi(y)) - \partial_y (b(y)\pi(y)),$$

which is the same differential equation that we previously derived using the generator.

## 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

$$G^N \phi(p) = (Np(1-p) + 2\theta_1(1-p)) [\phi(p + 1/N) - \phi(p)] \\ + (Np(1-p) + 2\theta_2p) [\phi(p - 1/N) - \phi(p)].$$

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

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

Consequently, the rescaled process converges to a diffusion process with generator  $G$ , 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

$$a(p) = 2p(1-p) \\ b(p) = 2\theta_1(1-p) - 2\theta_2p.$$

## *Wright-Fisher diffusion*

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:

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

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

$\pi(p)$  is the density of the stationary distribution of the process only if it is integrable on  $[0, 1]$ . To see that this is true, observe that the normalizing constant,

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

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.

## *Wright-Fisher diffusion*

Thus the stationary distribution of the Wright-Fisher diffusion with generator

$$G\phi(p) = p(1-p)\phi''(p) + (2\theta_1(1-p) - 2\theta_2p)\phi'(p)$$

is just the Beta distribution with parameters  $2\theta_1$  and  $2\theta_2$ , which has density

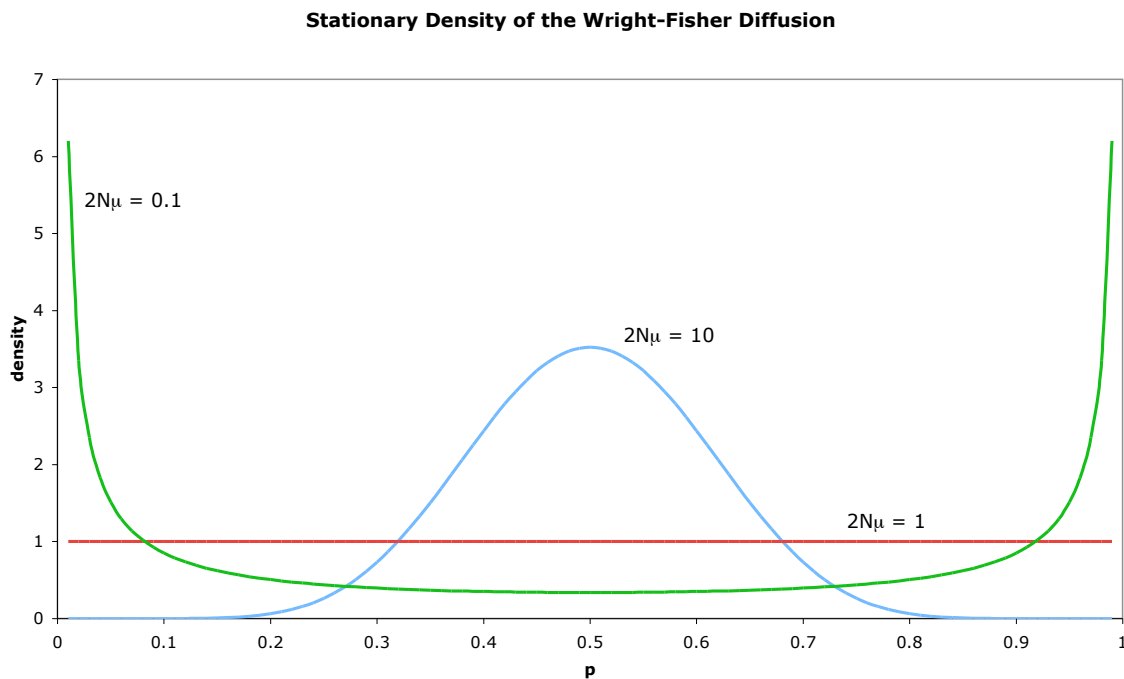
$$\begin{aligned}\pi(p) &= \frac{1}{B(2\theta_1, 2\theta_2)} p^{2\theta_1-1} (1-p)^{2\theta_2-1} \\ &= \frac{1}{B(2N\mu_1, 2N\mu_2)} p^{2N\mu_1-1} (1-p)^{2N\mu_2-1}.\end{aligned}$$

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

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

## *Wright-Fisher diffusion*

The stationary distribution reflects the competing effects of genetic drift, which eliminates variation, and mutation, which generates variation. When  $2N\mu_1$  and  $2N\mu_2$  are both greater than one, mutation dominates drift and  $\pi(p)dp$  is peaked about its mean. In contrast, when these quantities are less than one, then the stationary measure is bimodal, with peaks at the boundaries  $p = 0$  and  $p = 1$ .



## 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 for the Wright-Fisher diffusion with no mutation.

Rather, this process is certain to absorb at one of the boundaries in finite time, i.e., if there is no mutation, then one of the alleles is certain to be fixed in the population in finite time. Recall that this is also true of the Moran and Wright-Fisher models. In this case, the point masses,

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

are both stationary distributions of the Wright-Fisher diffusion, i.e., this process does not have a unique stationary distribution.

## Selection and Genetic Drift

So far we have considered only **neutral** population genetical models. In these models, survival and reproduction is independent of an individual's genotype. Our goal in this section is to relax this assumption by incorporating selective differences between individuals.

*Wright-Fisher model with selection:*

Suppose that a population containing  $N$  haploid individuals and two alleles,  $A$  and  $a$ , evolves according to the Wright-Fisher model with the following modification. Each adult in generation  $t + 1$  'chooses' its parent in generation  $t$  (independently and with replacement) with the following probabilities:

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

where  $k$  is the number of  $A$ -type adults alive in generation  $t$ .

## *Selection and Genetic Drift*

$s$  is called the **selection coefficient** of  $A$  and quantifies the selective advantage ( $s > 0$ ) or disadvantage ( $s < 0$ ) of this allele relative to  $a$ . This could be because:

- type  $A$  adults give birth on average to  $(1 + s)$ -times as many offspring as type  $a$  adults, or
- type  $A$  offspring are  $(1 + s)$ -times more likely to survive to adulthood than type  $a$  offspring.

It is customary to say that the  $A$  and  $a$  alleles have **relative fitnesses**  $1 + s : 1$ . This just means that the expected contribution of an  $A$ -type adult to the next generation is  $(1 + s)$ -times as great as that of an  $a$ -type adult.  $A$  is said to be a **beneficial allele** if  $s > 0$  and a **deleterious allele** if  $s < 0$ .

*Remark:* This is just the neutral Wright-Fisher model when  $s = 0$ .

## *Selection and Genetic Drift*

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, if the number of  $A$  alleles in generation  $t$  is  $Np$ , then the distribution of the number of  $A$  alleles in generation  $t + 1$  is Binomial with parameters  $N$  and  $p^*$ , where

$$p^* = \frac{p(1 + \sigma/N)}{p(1 + \sigma/N) + (1 - p)}.$$

Suppose that  $p^N(t)$  is the frequency of  $A$  in generation  $t$  and let  $\delta = p^N(1) - p^N(0)$ . Using the moments of the binomial distribution, it is easy to show that:

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

## *Selection and Genetic Drift*

These calculations show that when  $N$  is large, 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. Notice that:

- The infinitesimal variance is the same as in the neutral Wright-Fisher diffusion:  $a(p) = p(1-p)$ .

Thus, incorporating weak selection (i.e., of order  $1/N$ ) into the Wright-Fisher model does not change the dynamics of genetic drift in the diffusion approximation.

*Remark:* There are also diffusion approximations for models with strong selection (of order  $1/\sqrt{N}$ ) where the variance of the allele frequency fluctuations is influenced by selection.

## *Selection and Genetic Drift*

On the other hand, selection does change the infinitesimal drift from 0 in the neutral Wright-Fisher diffusion to:

- $b(p) = \sigma p(1 - p)$ .

Recall that the drift measures the tendency of the diffusion to either increase or decrease on average:

$$b(p) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbf{E}_p [p(t) - p].$$

Thus, the frequency of a beneficial allele will tend to increase, while that of a deleterious allele will tend to decrease. Furthermore, the average rate of increase or decrease is proportional to the strength of selection,  $\sigma$ , and the amount of genetic variation,  $p(1 - p)$ , in the population. Because of this, selection will have little effect on the population when  $A$  is either very common or very rare.

## *Selection and Genetic Drift*

One of the most important questions that we can ask about selection in finite populations concerns the fixation probabilities of selected alleles. Recall that we say that the population becomes fixed for the allele  $A$  if all other alleles are lost from the population. Thus, 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

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

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.

## *Selection and Genetic Drift*

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

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

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

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:  $\mathbf{P}_x\{\tau < \infty\} = 1$ .
- At time  $\tau$ , either  $X(\tau) = l$  or  $X(\tau) = r$ .
- $l$  and  $r$  are both absorbing states for the Markov process.

*Remark:* The second assumption means that we will only consider processes with sample paths that don't jump around too wildly (e.g., diffusions, finite Markov chains).

## *Hitting Probabilities*

When these assumptions are satisfied, we can define the function

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

which is just the probability that the process  $X$  hits  $r$  rather than  $l$ .

Our aim is to derive an equation for  $u(x)$  involving the generator of the Markov process  $X$ . We first observe that because  $l$  and  $r$  are absorbing states, it follows that

$$X(\tau + t) = X(\tau)$$

for all  $t \geq 0$ .

Consequently,

$$\begin{aligned} u(x) &= \mathbf{P}_x\{X(\tau) = r\} \\ &= \mathbf{P}_x\{X(\tau + t) = r\} \\ &= \mathbf{E}_x[\mathbf{P}\{X(\tau + t) = r | X(t)\}] \\ &= \int_E u(y) \mathbf{P}_x\{X(t) \in dy\} \\ &= \mathbf{E}_x[u(X(t))], \end{aligned}$$

where we have used the Markov property to pass from the third to the fourth line.

## *Hitting Probabilities*

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

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

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

In addition, we know that

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

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

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 assume that  $X$  absorbs at  $l$  or  $r$ , but even if this is not the case, we can construct a modified version of  $X$  which does have this property. (The modified process is called a **stopped process**.)

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

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

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

This equation can be rearranged and 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  $(l, r)$ .

## *Hitting probabilities for diffusions*

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 absorbs at  $r$  is:

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

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

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

## Selection and Fixation

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

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

i.e.,  $a(p) = p(1-p)$  and  $b(p) = \sigma p(1-p)$ .

We can use our results on the hitting probabilities of one-dimensional diffusions to derive one of the most important formulas of theoretical population genetics.

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

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

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

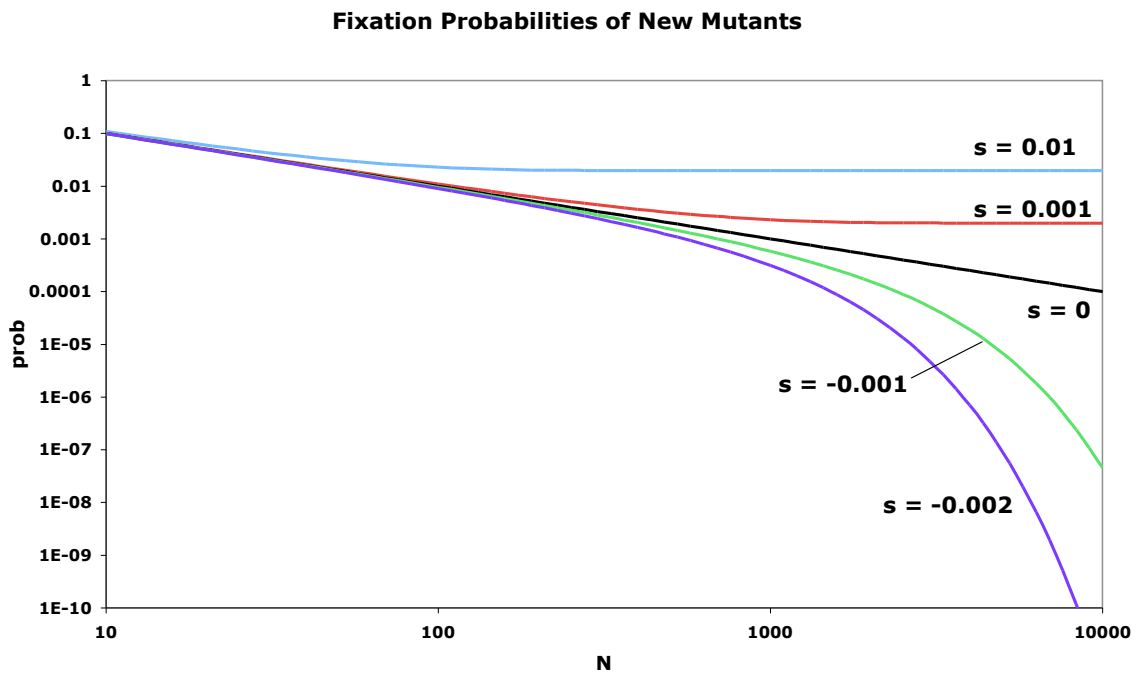
$$u(p) = p.$$

## *Selection and fixation*

Usually we are interested in the fixation probability of a new mutant. 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  $N$ .



## *Selection and fixation*

Let's consider 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 either the selective disadvantage  $|s|$  increases or the population size  $N$  increases.

*Intuition:* Deleterious alleles can become fixed in finite populations through chance increases in their frequency caused by genetic drift. However, many more such random events are required for fixation in large populations than in small populations, and so the probability of fixation rapidly decreases as  $N$  increases.

## *Selection and fixation*

**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 (in a haploid population) 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:  $s = 0$ .)

## *Selection and fixation*

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  $|s|$  that are less than the reciprocal of the population size behave almost 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

Even when selection cannot be directly observed, it can sometimes be inferred from DNA sequence data. This is because selection influences the rate at which different populations or species diverge from one another.

To understand the relationship between divergence, mutation, genetic drift and selection, we need to consider **substitution rates**. Recall that most new mutations are lost. However, some rise to fixation, and whenever this happens in one population but not in another it contributes to the genetic divergence between the two populations. Such events are called substitutions, and the rate at which they occur is called the substitution rate.

Substitution rates are difficult to calculate exactly. However, we can find a good approximation if we assume that recurrent mutations at a site are independent. This is a reasonable assumption whenever the mutation rate is low enough that each new mutation is likely to be lost or fixed before another mutation enters the population.

## *Substitution rates*

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 expected to be a good approximation if  $N\mu \ll 1$ .

**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 is independent of the population size.

## *Substitution rates*

**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.

## *Substitution rates*

These results have several applications.

**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 seems to depend on the kind of genetic variation considered. A **replacement** substitution is one that leads to a change in an amino acid in a protein, while 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

About 98.5% of the human genome is non-coding.

## *Substitution rates*

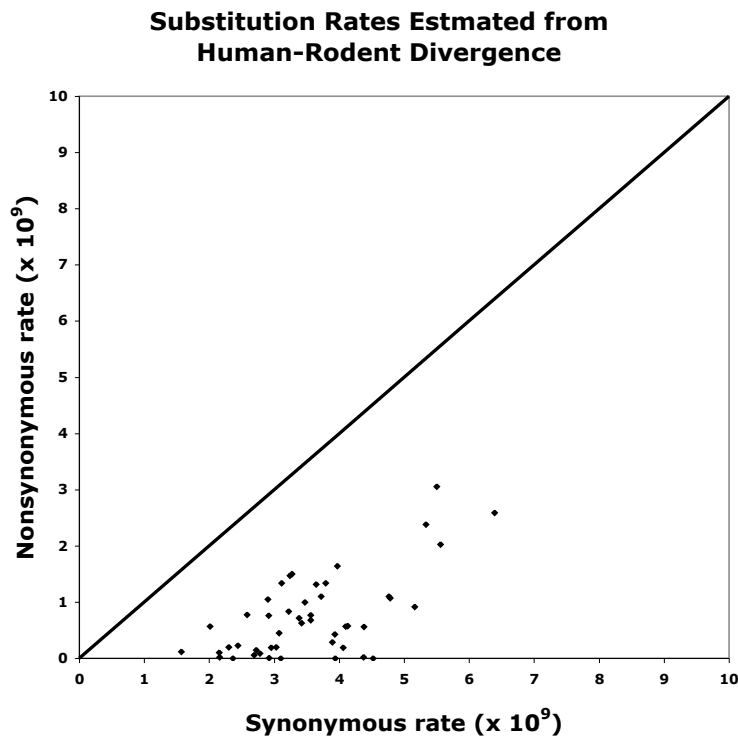
Several observations suggest that most replacement mutations are deleterious, while most silent mutations are nearly neutral.

- 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.

The disparity between silent and replacement substitution rates has been found in comparisons between pseudogenes and functional genes, between introns and exons within the same gene, and between synonymous and nonsynonymous sites within the same gene. (This is also true of polymorphism.)

## *Substitution rates*

The following plot shows synonymous and nonsynonymous substitution rates estimated from comparisons of human and rodent genes (based on Li (1997), pp. 180-181). Although there is considerable rate variation between genes, 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).

## *Substitution rates*

The third observation concerns the **generation time effect**. There is some evidence that the mutation rate per generation is relative conserved between taxa. Because species with shorter generation times have more generations per year, both the mutation rate and the neutral substitution rate (when measured per year) are expected to be higher in species with shorter generations. This is roughly what we see when we consider silent substitutions.

The weaker relationship between generation time and the amino acid substitution rate can be explained if we postulate that:

- Generation time is negatively correlated with population size.
- Most amino acid-changing mutations are weakly deleterious.

In this case, species with shorter generation times and larger population sizes have higher mutation rates, but lower fixation probabilities for replacement mutations. These competing processes will weaken the generation time effect.

## *Substitution rates*

The hypothesis that most silent substitutions are effectively neutral while most replacement substitutions are moderately deleterious was proposed by T. Ohta and is sometimes called the **nearly neutral theory**. It is consistent with the belief that:

- Most silent mutations are neutral because they have little or no effect on biological traits.
- Most replacement mutations are somewhat deleterious because they affect protein function.

However, there are some important caveats. One is that non-coding DNA may have functions of which we are unaware. It is clear that some non-coding DNA plays an important role in regulation of gene expression, and there is evidence that some of the most important changes in evolution have come about through mutations to regulatory DNA rather than coding DNA.

There is also evidence that adaptive evolution of proteins, perhaps in response to environmental variation, may occur more frequently than the nearly neutral theory allows.

## *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. For example, if we know that two species last shared a common ancestor  $T$  generations ago, then the expected number of neutral substitutions per site that will have occurred since they split will be  $2\mu T$ . In some cases,  $T$  can be inferred from fossil or biogeographical information, and then  $\mu$  can be estimated from observed levels of sequence divergence in sequences thought to be evolving neutrally.

**Dating species divergence:** Conversely, if the mutation rate is known (and constant), then genetic data can be used to estimate how long ago two species diverged if we assume that all substitutions are neutral. In this case, we say that substitutions obey a **molecular clock**.

## Selection in Diploid Populations

In this section, we will examine a model of selection and drift in a diploid population. Recall that a species is diploid if individuals have two copies of each chromosome. Thus, if there are two alleles present in a diploid population, say  $A_1$  and  $A_2$ , then there will be three possible genotypes:  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$ .

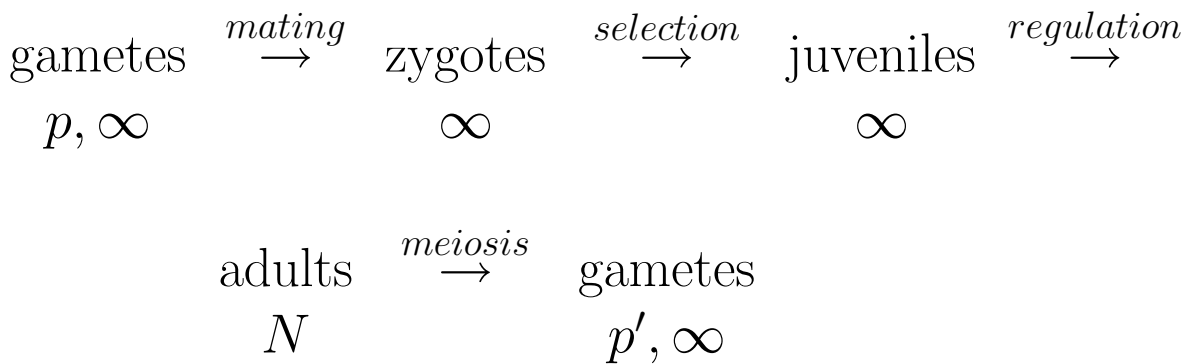
To model selection in such a population, we need to assign a fitness to each diploid genotype. Here we will adopt the convention that the relative fitness of the  $A_2A_2$  homozygote is 1:

| 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  $A_1$  or the **heterozygous effect**.

## *Selection in Diploid Populations*

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.

## *Selection in Diploid Populations*

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 in Diploid Populations*

*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<br>(after selection)      |
|----------|-------------------------------------|
| $A_1A_1$ | $p_{11}^* = p_{11}(1 + s)/\bar{w}$  |
| $A_1A_2$ | $p_{12}^* = p_{12}(1 + hs)/\bar{w}$ |
| $A_2A_2$ | $p_{22}^* = p_{22}/\bar{w}$         |

## *Selection in Diploid Populations*

*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. What is new is that both the juveniles and the adults are diploid, so each time we sample an individual, we are sampling two genes.

Suppose that  $p'_{ij}$  denotes the frequency of  $A_iA_j$  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  $A_1A_1$  adults produce only  $A_1$  gametes and  $A_2A_2$  adults produce only  $A_2$  gametes,  $A_1A_2$  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:

$$p^N(t + 1) = p' = p'_{11} + \frac{1}{2}p'_{12}.$$

## *Selection in Diploid Populations*

As usual, to derive a diffusion approximation for the model, we will need to 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 can use the calculations on the preceding slides and the properties of the Multinomial distribution to show that:

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

It follows that the processes  $(p^N(\lfloor 2Nt \rfloor) : t \geq 0)$  converge to a Wright-Fisher diffusion with generator

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

## *Selection in Diploid Populations*

*Remark:* Notice that we have rescaled time by a factor of  $2N$  rather than  $N$ . This is because there are  $2N$  genes in a diploid population with  $N$  individuals. What is surprising (perhaps) is that the diffusion approximation for this process has the same infinitesimal variance as the diffusion approximation for a haploid Wright-Fisher model with  $N$  individuals in which we rescale time by a factor of  $N$ :

- $a(p) = p(1-p)$ .

This is the case despite the fact that we only sample  $N$  individuals.

This result can be explained by noting that because selection is weak (of order  $O(1/N)$ ), the genotypic frequencies in the juveniles just prior to population regulation deviate from Hardy-Weinberg equilibrium only by very small amounts. Consequently, the two alleles carried by each individual are nearly independent of one another, so that sampling  $N$  individuals at random is nearly equivalent to sampling  $2N$  individuals at random.

## *Selection in Diploid Populations*

The infinitesimal drift of the diffusion approximation is:

- $b(p) = 2\sigma(h + (1 - 2h)p)p(1 - p)$ .

Notice that this can be written as  $b(p) = \sigma(p)p(1 - p)$ , where we define

$$\sigma(p) \equiv 2\sigma(h + (1 - 2h)p)$$

to be the frequency-dependent selection coefficient of allele  $A_1$  relative to  $A_2$ .

What this shows is that selection in (random mating) diploid populations is **frequency-dependent**:

- The **marginal** fitness of an allele depends on its frequency.

This is because the marginal fitness of an allele is equal to the genotypic fitnesses weighted by the frequency with which the allele occurs as part of each diploid genotype. However, under random mating (and any realistic mating scheme), the genotype frequencies are determined by the allele frequencies.

## *Selection in Diploid Populations*

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 that selection is genic because the fitness of a diploid genotype is an additive function of the number of copies of  $A_1$  that it contains:

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ 1 + s & 1 + s/2 & 1 \end{array}$$

In other words, each copy of  $A_1$  adds  $s/2$  to the relative fitness of the genotype.

## *Selection in Diploid Populations*

**$A_1$  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. When  $h = 1$ ,  $A_1$  is said to be completely dominant and the selection coefficient,

$$\sigma(p) = 2\sigma(1 - p),$$

is a decreasing function of  $p$ .

This can be explained by examining the relationship between  $p$  and the marginal fitness of each allele. Recall that the genotypic frequencies (before selection) and the relative fitnesses are:

|          |             |             |
|----------|-------------|-------------|
| $A_1A_1$ | $A_1A_2$    | $A_2A_2$    |
| $p^2$    | $2p(1 - p)$ | $(1 - p)^2$ |
| $1 + s$  | $1 + s$     | $1$         |

## *Selection in Diploid Populations*

Consequently, the marginal fitnesses are:

$$w_{A_1} = p(1 + s) + (1 - p)(1 + s) = (1 + s)$$

$$w_{A_2} = p(1 + s) + (1 - p) = 1 + ps.$$

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.

## *Selection in Diploid Populations*

**$A_1$  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. When  $h = 0$ ,  $A_1$  is said to be completely recessive and the selection coefficient,

$$\sigma(p) = 2\sigma p,$$

is an increasing function of  $p$ .

This too can be explained by the relationship between the marginal fitnesses of the two alleles and the frequency of  $A_1$ :

$$\begin{aligned}w_{A_1} &= p(1 + s) + (1 - p) = 1 + ps \\w_{A_2} &= p + (1 - p) = 1.\end{aligned}$$

Thus, when  $A_1$  is rare, its marginal fitness is approximately equal to that of  $A_2$ , while when  $A_1$  is common, its marginal fitness is approximately  $1 + s$  compared with 1 for  $A_2$ .

## *Selection in Diploid Populations*

**Overdominance:** If  $\sigma > 0$  and  $h > 1$  (or  $\sigma < 0$  and  $h < 0$ ), 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,  $\sigma(p^*) = 0$  when

$$p^* = \frac{h}{2h - 1} \in (0, 1),$$

and

- $\sigma(p) > 0$  if  $p < p^*$
- $\sigma(p) < 0$  if  $p > p^*$ .

Thus,  $A_1$  tends to rise in frequency when rare, but fall in frequency when common. This kind of selection is called **balancing selection** and will increase polymorphism.

## *Selection in Diploid Populations*

*Example:* The classical 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, usually fatal anemia.

## *Selection in Diploid Populations*

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.

## *Selection in Diploid Populations*

**Underdominance:** If  $\sigma < 0$  and  $h > 1$  (or  $\sigma > 0$  and  $h < 0$ ), 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, it is still true that  $\sigma(p^*) = 0$  when

$$p^* = \frac{h}{2h - 1} \in (0, 1);$$

however,

- $\sigma(p) < 0$  if  $p < p^*$
- $\sigma(p) > 0$  if  $p > p^*$ .

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.

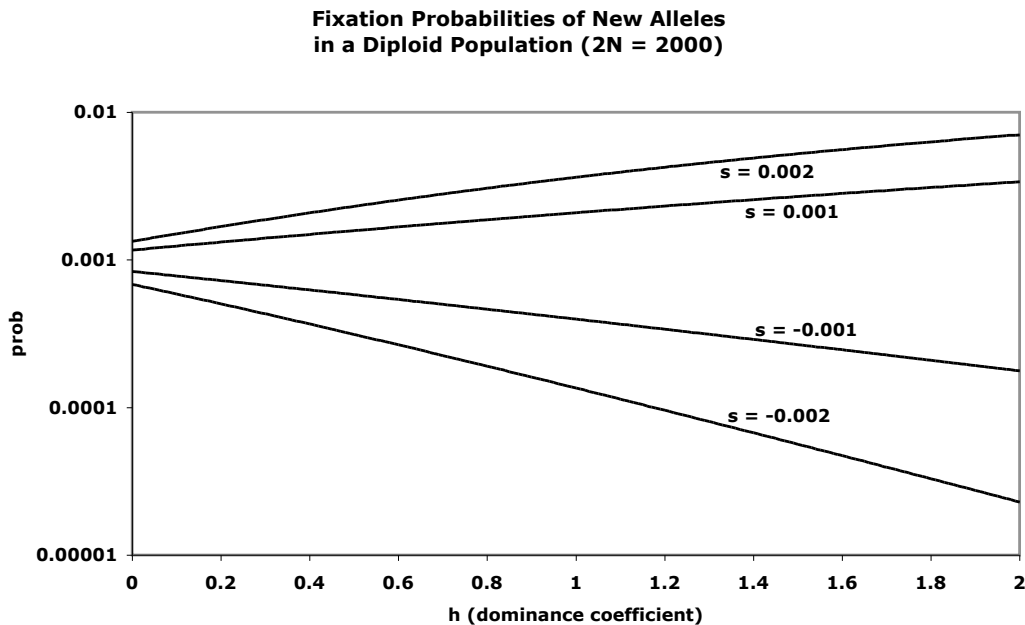
*Example:* Chromosomal rearrangements are sometimes underdominant due to defects in meiosis that occur when chromosomes with different structures attempt to pair up.

## *Selection in Diploid Populations*

We can use our analysis of the hitting probabilities of a diffusion process to calculate the fixation probabilities of alleles in a diploid population:

$$u(p) \equiv \mathbf{P}_p\{A_1 \text{ is fixed}\} = \frac{\int_0^p e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}{\int_0^1 e^{-4\sigma hq - 2\sigma(1-2h)q^2} dq}.$$

The fixation probabilities of a new allele ( $p = 1/2N$ ) are shown as functions of the dominance coefficient  $h$  in the figure below.



## *Selection in Diploid Populations*

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

Earlier, we showed that the diffusion approximations for the Wright-Fisher process and the Moran model are equivalent, apart from a change in time scale:

$$Gf(p) = \frac{1}{2}p(1-p)f''(p) \quad (\text{Wright-Fisher model})$$

$$Gf(p) = p(1-p)f''(p) \quad (\text{Moran model}).$$

Our goal in this section is to show that this is true of a wider class of population genetics models. The models that we will consider are called **Cannings models** after the person who introduced them. Cannings models have the following property. Suppose that there are  $N$  adults alive at any time, which we label  $1, \dots, N$ , and let  $\zeta_i$  denote the number of adults descended from individual  $i$  following a reproduction event. Then, in a Cannings model, the distribution of the vector  $(\zeta_1, \dots, \zeta_N)$  is exchangeable:

$$(\zeta_1, \dots, \zeta_N) \stackrel{d}{=} (\zeta_{\sigma(1)}, \dots, \zeta_{\sigma(N)}),$$

where  $\sigma$  is any permutation of  $\{1, \dots, N\}$ .

## *Cannings models*

Another way to state this property is that the numbers of surviving offspring are independent of the labels of the parents. Notice that both the Wright-Fisher model and the Moran model have this property.

One criticism of the Wright-Fisher model is that the biological processes underlying birth and population regulation are obscure (but see question 2 on the first problem set). 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.

## *Cannings models*

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  $\eta$ :

- $\mathbf{P}\{\eta = 0\} = 0$ , i.e., each adult gives birth to at least one offspring;
- $\mathbf{P}\{\eta \leq M\} = 1$ , i.e.,  $\eta$  is bounded. This implies:
- $m = \mathbf{E}[\eta] < \infty$ ;
- $\sigma^2 = \mathbf{E}[(\eta - m)^2] < \infty$ .

*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.

## *Cannings models*

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. Then 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

$$Z = \sum_{i=1}^N \eta_i.$$

## *Cannings models*

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

$$\begin{aligned}\mathbf{E}[S] &= n \binom{m}{N} \\ \mathbf{E}[(S - \mathbf{E}[S])^2] &= n \binom{N-n}{N-1} \binom{m}{N} \left(1 - \frac{m}{N}\right) \\ \mathbf{E}[(S - \mathbf{E}[S])^e] &= O(n^{e-2}) \text{ if } e \geq 3.\end{aligned}$$

## *Cannings models*

To derive a diffusion approximation for this model, we need to calculate the limits of the expectations  $N\mathbf{E}_p[\delta^n]$  as  $N$  tends to infinity, where  $\delta = p' - p$ .

First consider the case  $n = 1$ . By conditioning on  $Y$  and  $Z$ , we have:

$$\begin{aligned}\mathbf{E}_p[p'] &= \mathbf{E}_p \left[ \frac{1}{N} X \right] \\ &= \mathbf{E}_p \left[ \frac{1}{N} \mathbf{E}[X|Y, Z] \right] \\ &= \mathbf{E}_p \left[ \frac{1}{N} \left( N \frac{Y}{Z} \right) \right] = \mathbf{E}_p \left[ \frac{Y}{Z} \right].\end{aligned}$$

Using the fact that

$$\mathbf{E} \left[ \frac{\eta_i}{Z} \right] = \mathbf{E} \left[ \frac{\eta_1}{Z} \right]$$

for all  $i = 1, \dots, N$ , this last term can be rewritten as:

$$\mathbf{E}_p \left[ \frac{Y}{Z} \right] = \mathbf{E}_p \left[ \frac{1}{Z} \sum_{i=1}^{Np} \eta_i \right] = \sum_{i=1}^{Np} \mathbf{E} \left[ \frac{\eta_i}{Z} \right] = Np \mathbf{E} \left[ \frac{\eta_1}{Z} \right].$$

## *Cannings models*

Similarly,

$$1 = \mathbf{E}_p \left[ \frac{Z}{Z} \right] = \mathbf{E}_p \left[ \frac{1}{Z} \sum_{i=1}^N \eta_i \right] = \sum_{i=1}^N \mathbf{E} \left[ \frac{\eta_i}{Z} \right] = N \mathbf{E} \left[ \frac{\eta_1}{Z} \right],$$

which implies that

$$\mathbf{E} \left[ \frac{\eta_1}{Z} \right] = \frac{1}{N}.$$

Substituting this result into the identities obtained on the previous page shows that:

$$\mathbf{E}_p [p'] = Np \mathbf{E} \left[ \frac{\eta_1}{Z} \right] = Np \cdot \frac{1}{N} = p,$$

and so

$$\lim_{N \rightarrow \infty} N \mathbf{E}_p [\delta] = N(p - p) = 0.$$

*Remark:* The crucial identity,

$$\mathbf{E} \left[ \frac{\eta_1}{Z} \right] = \frac{1}{N} = \frac{\mathbf{E}[\eta_1]}{\mathbf{E}[Z]},$$

is a consequence of the exchangeability of the random variables  $\eta_i$  and holds even if the  $\eta_i$  are not independent.

## *Cannings models*

Our next task is to find the limit of  $N\mathbf{E}_p[\delta^2]$  as  $N$  tends to infinity. As above, we can simplify the calculation by conditioning on  $Y$  and  $Z$ :

$$\begin{aligned} N\mathbf{E}_p[\delta^2] &= N\mathbf{E}_p[(p' - p)^2] = N\mathbf{E}_p\left[\frac{1}{N^2}(X - Np)^2\right] \\ &= \mathbf{E}_p\left[\frac{1}{N}\mathbf{E}[(X - Np)^2|Y, Z]\right] \\ &= \mathbf{E}_p\left[\frac{1}{N}\mathbf{E}[(X - \mathbf{E}[X|Y, Z] + \mathbf{E}[X|Y, Z] - Np)^2|Y, Z]\right] \\ &= \mathbf{E}_p\left[\frac{1}{N}\mathbf{E}[(X - \mathbf{E}[X|Y, Z])^2|Y, Z]\right] \\ &\quad + \mathbf{E}_p\left[\frac{2}{N}\mathbf{E}[(X - \mathbf{E}[X|Y, Z])(\mathbf{E}[X|Y, Z] - Np)|Y, Z]\right] \\ &\quad + \mathbf{E}_p\left[\frac{1}{N}\mathbf{E}[(\mathbf{E}[X|Y, Z] - Np)^2|Y, Z]\right] \\ &\equiv K_1 + K_2 + K_3. \end{aligned}$$

This sum can be evaluated term-by-term using the fact that the conditional distribution of  $X$  given  $Y$  and  $Z$  is hypergeometric  $H(Z, Y, N)$ .

## *Cannings models*

Beginning with  $K_1$ , observe that the expression inside the  $\mathbf{E}_p[\cdot]$  in this term is just  $1/N$  times the variance of this hypergeometric distribution,

$$\mathbf{E}[(X - \mathbf{E}[X|Y, Z])^2|Y, Z] = N \binom{Z-N}{Z-1} \binom{Y}{Z} \left(1 - \frac{Y}{Z}\right).$$

Consequently,

$$\begin{aligned} K_1 &= \mathbf{E}_p \left[ \binom{Z-N}{Z-1} \binom{Y}{Z} \left(1 - \frac{Y}{Z}\right) \right] \\ &= \mathbf{E}_p \left[ \binom{\hat{Z}-1}{\hat{Z}-1/N} \binom{\hat{Y}}{\hat{Z}} \left(1 - \frac{\hat{Y}}{\hat{Z}}\right) \right], \end{aligned}$$

where we have introduced the random variables  $\hat{Y} = Y/N$  and  $\hat{Z} = Z/N$ .

Now, because we have assumed that  $m = \mathbf{E}[\eta] < \infty$ , 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.

## *Cannings models*

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:

$$\begin{aligned}\lim_{N \rightarrow \infty} K_1 &= \left( \frac{m-1}{m} \right) \binom{mp}{m} \left( 1 - \frac{mp}{m} \right) \\ &= (1 - 1/m)p(1 - p).\end{aligned}$$

We next consider  $K_2$ . Because the expression

$$\mathbf{E}[X|Y, Z] - Np = N \left( \frac{Y}{Z} - p \right)$$

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

$$\begin{aligned}K_2 &= \mathbf{E}_p \left[ 2 \left( \frac{Y}{Z} - p \right) \mathbf{E}[X - \mathbf{E}[X|Y, Z]|Y, Z] \right] \\ &= 0,\end{aligned}$$

which vanishes because  $\mathbf{E}[X - \mathbf{E}[X|Y, Z]|Y, Z] = 0$ .

## *Cannings models*

Lastly, observe that because  $\mathbf{E}[X|Y, Z] = NY/Z$  and  $\mathbf{E}_p[Y/Z] = p$  (as shown earlier), we can write

$$\begin{aligned} K_3 &= \mathbf{E}_p \left[ \frac{1}{N} \mathbf{E} [ (\mathbf{E}[X|Y, Z] - Np)^2 | Y, Z ] \right] \\ &= \mathbf{E}_p \left[ N \left( \frac{Y}{Z} - p \right)^2 \right] \\ &= N \left( \mathbf{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right). \end{aligned}$$

To calculate the first term in parentheses, we will again exploit the exchangeability of  $(\eta_1, \dots, \eta_N)$ . Specifically, observe that

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{Z^2} \eta_i^2 \right] &= \mathbf{E} \left[ \frac{1}{Z^2} \eta_1^2 \right] \quad i = 1, \dots, N \\ \mathbf{E} \left[ \frac{1}{Z^2} \eta_i \eta_j \right] &= \mathbf{E} \left[ \frac{1}{Z^2} \eta_1 \eta_2 \right] \quad i \neq j = 1, \dots, N. \end{aligned}$$

## *Cannings models*

It follows that

$$\begin{aligned}
 \mathbf{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] &= \mathbf{E}_p \left[ \frac{1}{Z^2} \left( \sum_{i=1}^{Np} \eta_i \right)^2 \right] \\
 &= \mathbf{E}_p \left[ \frac{1}{Z^2} \sum_{i=1}^{Np} \eta_i^2 + \frac{1}{Z^2} \sum_{i \neq j}^{Np} \eta_i \eta_j \right] \\
 &= Np \mathbf{E} \left[ \frac{\eta_1^2}{Z^2} \right] + Np(Np - 1) \mathbf{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right].
 \end{aligned}$$

Exchangeability can be used to further simplify this expression. Observe that:

$$\begin{aligned}
 1 &= \mathbf{E} \left[ \left( \frac{Z}{Z} \right)^2 \right] = \mathbf{E} \left[ \frac{1}{Z^2} \left( \sum_{i=1}^N \eta_i \right)^2 \right] \\
 &= N \mathbf{E} \left[ \frac{\eta_1^2}{Z^2} \right] + N(N - 1) \mathbf{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] \\
 &= N \left( \frac{v_N}{N^2} \right) + N(N - 1) \mathbf{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right],
 \end{aligned}$$

where we have defined  $v_N = N^2 \mathbf{E}[\eta_1^2/Z^2]$ .

## *Cannings models*

This implies that

$$\mathbf{E} \left[ \frac{\eta_1 \eta_2}{Z^2} \right] = \frac{1}{N(N-1)} \left( 1 - \frac{v_N}{N} \right).$$

Substituting this expression into the first set of identities on the preceding page shows that

$$\mathbf{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = Np \left( \frac{v_N}{N^2} \right) + \left( \frac{Np(Np-1)}{N(N-1)} \right) \left( 1 - \frac{v_N}{N} \right),$$

whence

$$N\mathbf{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] = v_N p \left( 1 - \frac{k-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} \right),$$

and

$$\begin{aligned} K_3 &= N \left( \mathbf{E}_p \left[ \left( \frac{Y}{Z} \right)^2 \right] - p^2 \right) \\ &= v_N p \left( 1 - \frac{Np-1}{N-1} \right) + Np \left( \frac{Np-1}{N-1} - p \right) \\ &= v_N p(1-p) - p(1-p) + O(N^{-1}) \\ &= (v_N - 1)p(1-p) + O(N^{-1}). \end{aligned}$$

## *Cannings models*

Finally, observe that

$$v_N \equiv N^2 \mathbf{E} \left[ \frac{\eta_1^2}{Z^2} \right] = \mathbf{E} \left[ \frac{\eta_1^2}{\hat{Z}^2} \right],$$

where we have defined  $\hat{Z} = Z/N$  as before. To evaluate the limit, recall that

$$\hat{Z} \rightarrow m \quad \text{a.s.},$$

by the strong law of large numbers, while the fact that  $\hat{Z} \geq 1$  and  $\eta_1 \leq M$  shows that

$$\frac{\eta_1^2}{\hat{Z}^2} \leq M^2 < \infty.$$

As above, we can pull the limit inside the expectation, obtaining

$$\lim_{N \rightarrow \infty} v_N = \frac{1}{m^2} \mathbf{E}[\eta_1^2] \equiv \frac{v}{m^2},$$

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  $\eta$ .

## *Cannings models*

Having evaluated each of the terms  $K_1$ ,  $K_2$ , and  $K_3$ , we can calculate

$$\begin{aligned}\lim_{N \rightarrow \infty} N \mathbf{E}_p[\delta^2] &= K_1 + K_2 + K_3 \\ &= (1 - 1/m)p(1 - p) + 0 + \frac{\sigma^2}{m^2}p(1 - p) \\ &= \left( \frac{\sigma^2}{m^2} + 1 - \frac{1}{m} \right) p(1 - p).\end{aligned}$$

To show that third and higher order moments of  $\delta$  vanish in the limit, we calculate

$$\begin{aligned}N \mathbf{E}_p[\delta^n] &= N \mathbf{E}_p \left[ \frac{1}{N^n} \mathbf{E}[\delta^n | Y, Z] \right] \\ &= \mathbf{E}_p \left[ \frac{1}{N^{n-1}} O(N^{n-2}) \right] \\ &= N^{-1},\end{aligned}$$

which vanishes as  $N$  tends to infinity. (Here we are implicitly using the fact that the bound in  $O(N^{n-2})$  is uniform in  $Y$  and  $Z$ .)

## *Cannings models*

To summarize, we have shown that:

$$\begin{aligned}\lim_{N \rightarrow \infty} N \mathbf{E}_p[\delta] &= 0 \\ \lim_{N \rightarrow \infty} N \mathbf{E}_p[\delta^2] &= \left( \frac{\sigma^2}{m^2} + 1 - \frac{1}{m} \right) p(1-p) \\ \lim_{N \rightarrow \infty} N \mathbf{E}_p[\delta^n] &= 0 \text{ if } n \geq 3.\end{aligned}$$

Consequently, we know that when  $N$  is large, the process  $(p^N(\lfloor Nt \rfloor) : t \geq 0)$  can be approximated by a diffusion with generator

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

This is just a neutral Wright-Fisher diffusion run at speed  $\frac{\sigma^2}{m^2} + 1 - \frac{1}{m}$ .

## *Cannings models*

The significance of this result is that it shows that we can approximate a large class of population genetic models by the Wright-Fisher diffusion when the population size is large. In fact, this is true for an even larger class of models than we have considered here - the key requirements are that the numbers of descendants of each adult are exchangeable and have finite variance. This suggests that the diffusion approximation is **robust**:

- The fine details of the model (in this case, the offspring distribution  $\eta$ ) only influence the limit through a scalar time change.

Thus, even though we can't formulate an exact model of a real population, we can at least have some confidence that the Wright-Fisher diffusion is a reasonable description of genetic drift.

## Effective Population Size

Suppose that  $p_C^N = (p_C^N(t) : t \geq 0)$  is the Cannings model studied in the preceding slides and let  $p_{WF}^N = (p_{WF}^N(t) : t \geq 0)$  be the classical Wright-Fisher model for a haploid population of size  $N$ .

If  $N$  is large, then the (unscaled) 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 a Wright-Fisher diffusion with generator

$$Gf(p) = \frac{1}{2N} \left( \frac{\sigma^2}{m^2} + 1 - \frac{1}{m} \right) p(1-p)f''(p).$$

*Remark:* In both cases we are running the discrete-time process at its original rate and so we must rescale the rate of the diffusion approximation by  $1/N$ .

## *Effective Population Size*

Notice that the Cannings model  $(p_C^N(\lfloor t \rfloor) : t \geq 0)$  and the Wright-Fisher model  $(p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0)$  will have the same diffusion approximation if we define

$$N_e = \frac{N}{\frac{\sigma^2}{m^2} + 1 - \frac{1}{m}}.$$

However, if both processes can be approximated by the same diffusion process, it follows that they are in fact similar to each other when  $N$  is sufficiently large:

$$(p_C^N(\lfloor t \rfloor) : t \geq 0) \approx (p_{WF}^{N_e}(\lfloor t \rfloor) : t \geq 0).$$

In other words, any population of size  $N$  which is described by a Cannings model is approximately equivalent to an ideal Wright-Fisher population with a possibly different population size,  $N_e$ . The quantity  $N_e$  that makes this correspondence true is called the **effective population size** of the original population.

## *Effective Population Size*

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

- $N_e \approx N$  when  $m^2 \gg \sigma^2$  and  $m \gg 1$ ,

i.e., if every adult gives birth to a large, comparable number of offspring, then the Cannings model is already ‘close’ to the Wright-Fisher model of the same size.

Also,

- $N_e \approx \frac{N}{\sigma^2}$  if  $\sigma^2 \gg m^2$ .

Thus, the greater the variance in reproductive success between adults, the smaller the effective population size will be. In particular, the effective population size may be much smaller than the census population size.

## *Effective Population Size*

In question 3 of the first problem set, you considered a generalization of the Wright-Fisher model which allowed the (actual) population size to fluctuate independently from generation to generation with distribution:

$$\mathbf{P}\{N = N_i\} = q_i.$$

One consequence of the diffusion approximation derived in that problem is that the effective population size of a population with fluctuating population sizes is the harmonic mean population size:

$$N_e = \frac{1}{\mathbf{E}\left[\frac{1}{N}\right]} = \frac{1}{\sum_i \frac{q_i}{N_i}}.$$

In particular, this shows that the effective population size is very sensitive to periods when the population size is small, even if these are rare. Such events are called **bottlenecks**.

## *Effective Population Size*

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

$$G\phi(p) = \frac{1}{2}p(1-p)\phi''(p) + \theta(1-2p)\phi'(p).$$

Here,  $\theta = N_e\mu$  in a haploid population or  $\theta = 2N_e\mu$  in a diploid population, and  $\mu$  is the mutation rate between  $A$  and  $a$ , which we assume to be symmetric.

Earlier, we showed that the stationary distribution for this process is just the Beta distribution with parameters  $2\theta$  and  $2\theta$ . Consequently, the expected level of diversity in a stationary population is

$$\bar{H} = 2 \int_0^1 p(1-p)\pi(p)dp = \frac{2\theta}{4\theta + 1}.$$

## *Effective Population Size*

It follows that a moment estimator for  $\theta$  is

$$\hat{\theta} = \frac{1}{2} \left( \frac{H}{1 - 2H} \right).$$

Furthermore, if we know  $\mu$ , then we can use this result to estimate  $N_e$ . For a haploid population,  $\theta = N_e\mu$  and so

$$\hat{N}_e = \frac{1}{2\mu} \left( \frac{H}{1 - 2H} \right).$$

Likewise, for a diploid population,  $\theta = 2N_e\mu$  and so

$$\hat{N}_e = \frac{1}{4\mu} \left( \frac{H}{1 - 2H} \right).$$

$H$  can be estimated from sequence data by calculating the sample probability that two individuals have different nucleotides at the same site in a gene. If this can be done for a large number of sites and if different sites evolve independently of one another (as is usually assumed), then this procedure will give an accurate estimate of  $H$  even for samples of modest size (e.g., 10 individuals).

## *Effective Population Size*

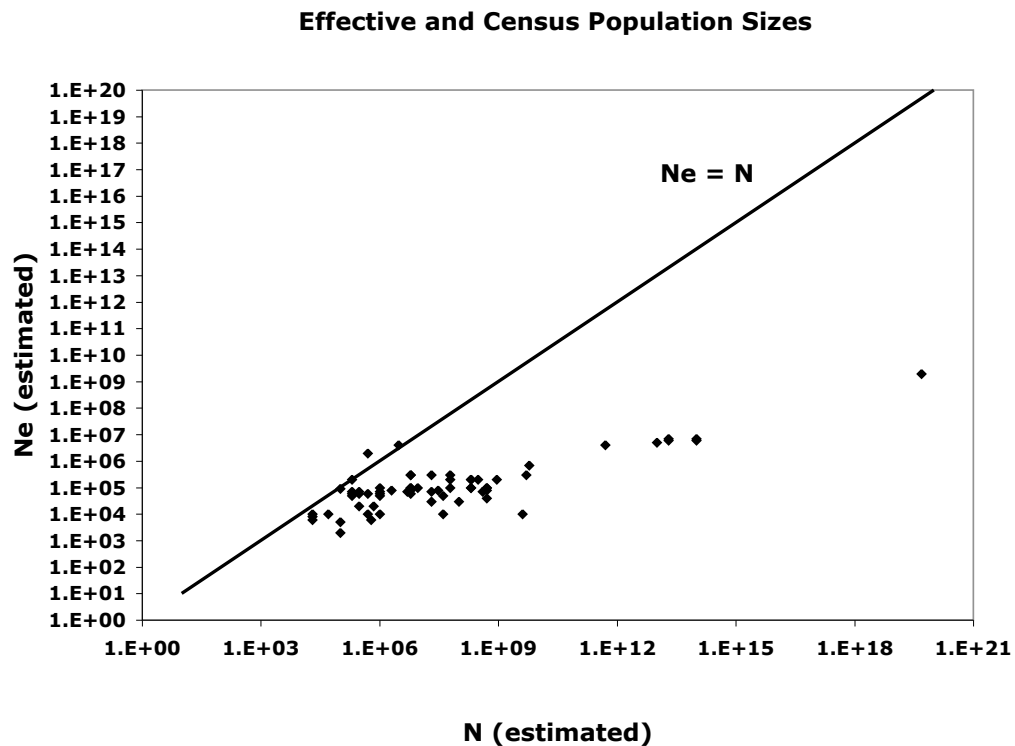
*Remark:* In practice, more sophisticated methods can be used to obtain better estimates of the effective population. Coalescent theory and computationally-intensive methods play an important role in this area.

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)                  | $10^3$        |

## *Effective Population Size*

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.

## *Effective Population Size*

*Caveat:* These estimates of  $N_e$  beg the bigger question, which is whether the Wright-Fisher diffusion (and Kingman's coalescent) are suitable models of genetic drift. If not, then the meaning of  $N_e$  becomes unclear.

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

- [1] R. Durrett. *Probability Models for DNA Sequence Evolution*. Springer-Verlag, 2002.
- [2] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, New York, N.Y., 1986.
- [3] W.J. Ewens. *Mathematical Population Genetics. Biomathematics, Vol 9*. Springer, New York, 2'nd edition, 2004.
- [4] R. A. Fisher. *The Genetical Theory of Natural Selection. A Complete Variorum Edition*. Oxford University Press, 1999.
- [5] J. H. Gillespie. *The Causes of Molecular Evolution*. Oxford University Press, Oxford, 1991.
- [6] J. H. Gillespie. *Population Genetics. A Concise Guide*. Johns Hopkins University Press, 2'nd edition, 2004.
- [7] D. L. Hartl and A. G. Clark. *Principles of Population Genetics*. Sinauer, 4'th edition, 2007.
- [8] W.-H. Li. *Molecular Evolution*. Sinauer, 1997.