Chapter 2 – Genetic Drift & Mutation

The Wright-Fisher Model

Mutation

The Wright-Fisher model

sample 2N gametes

2N haploid individuals ⇉ ∞ gametes → 2N individuals

p_0 p_0 p_1
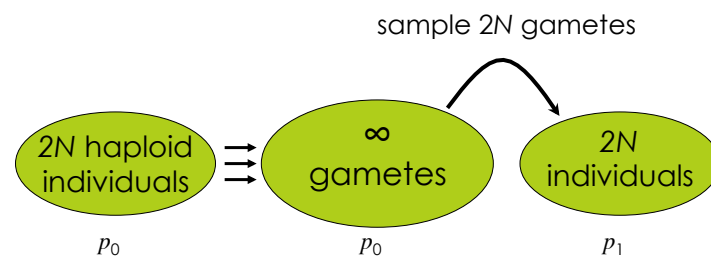
❖ effectively a haploid model, but with population size = 2N to mirror a diploid population

❖ random sampling of genotypes and random sampling of gametes essentially the same

http://psych.colorado.edu/~carey/hgss/hgssapplets/evolution/geneticdrift/GeneticDrift.html

http://darwin.eeb.uconn.edu/simulations/drift.html

Expectation

❖ What is the expected payoff from buying a $1 lottery ticket?

Expectation

❖ What is the expected payoff from buying a $1 lottery ticket?

Match	+	Megaball	Probability	Top Prize
5	+	1	1 in 175,711,536	JACKPOT
5			1 in 3,904,701	$250,000**
4	+	1	1 in 689,065	$10,000**
4			1 in 15,313	$150**
3	+	1	1 in 13,781	$150**
2	+	1	1 in 844	$10
3			1 in 306	$7
1	+	1	1 in 141	$3
0	+	1	1 in 75	$2

Expectation

❖ What is the expected payoff from buying a $1 lottery ticket?   ~$0.77

	Net Ticket Sales*	Prizes*	Prizes as % of Ticket Sales	Admin Costs*	Admin as % of Ticket Sales	Net Proceeds to State*	Proceeds as % of Ticket Sales
Massachusetts	4,200	3,235	77.00%	83	2.00%	883	21.00%

* values in millions

Expected Allele Frequency

❖ the **expected** allele frequency in the next generation is equal to the current allele frequency, but that's only what we expect on average

$$E\left[f_A(t+1)\right] = 2Nf_A(t)/2N = f_A(t)$$

Genetic Drift

❖ what happens when we relax the assumption of infinite population size?
  ◇ allele frequencies change due to random "genetic drift" (stochastic sampling)
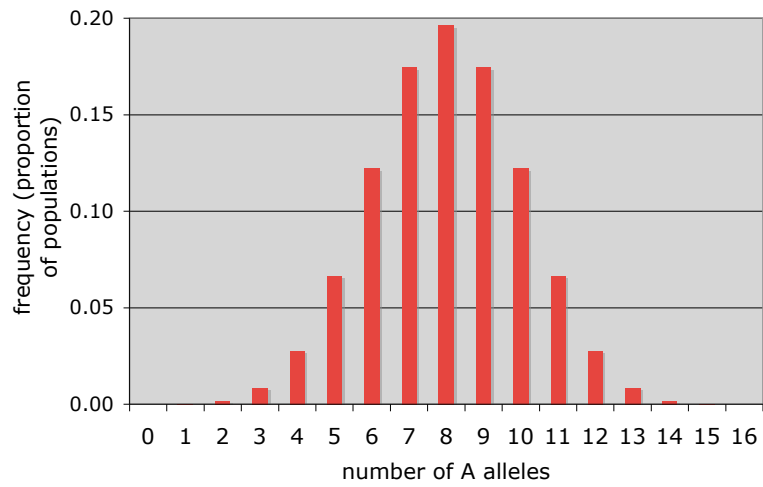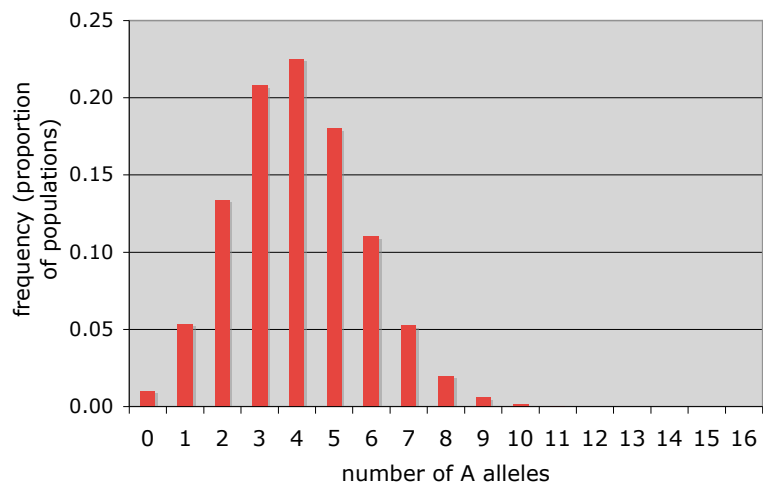
Genetic Drift

❖ Binomial probability

$$\Pr\{j \text{ alleles of type } A\} =$$

$$\binom{2N}{j} f_A^j f_a^{2N-j} = \frac{(2N)!}{j!(2N-j)!} f_A^j f_a^{2N-j}$$

Random samples of 16 alleles for $f_A = f_a = 0.50$

Random samples of 16 for $f_A = 0.25, f_a = 0.75$

The Moran model

❖ derived for a haploid population
  ◇ but results for 2*N* haploid individuals are essentially the same as for *N* diploid individuals when assuming neutrality
❖ randomly select an adult to produce 1 offspring, then randomly select an adult to die
❖ results in allele frequency increments of ±1

The Wright-Fisher model

❖ transition probabilities obtained directly from binomial probability distribution with f_A and f_a adjusted each generation

$$T_{ij} = \begin{pmatrix} 2N \\ j \end{pmatrix} \left(\frac{i}{2N} \right)^j \left(\frac{2N-i}{2N} \right)^{2N-j} = \frac{(2N)!}{j!(2N-j)!} f_A^j f_a^{2N-j}$$

❖ variance in allele frequency change for a single generation $= f_A f_a / 2N$

 ◇ (= $2f_A f_a / 2N$ for 2*N* iterations of the Moran process)

Wright-Fisher transition probabilities

❖ binomial results in a matrix of transition
probabilities that can be multiplied to predict
the distribution of allele frequencies at any
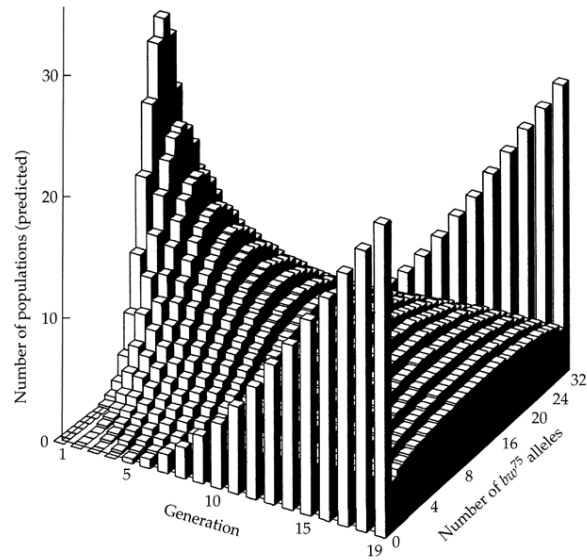number of generations in the future

to j

matrix for
$2N = 8$

from i

|	0	1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0	0	0
1	0.344	0.393	0.196	0.056	0.010	0.001	~0	~0	~0
2	0.100	0.267	0.311	0.208	0.087	0.023	0.004	~0	~0
3	0.023	0.112	0.235	0.282	0.211	0.101	0.030	0.005	~0
4	0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004
5	~0	0.005	0.030	0.101	0.211	0.282	0.235	0.112	0.023
6	~0	~0	0.004	0.023	0.087	0.208	0.311	0.267	0.100
7	~0	~0	~0	0.001	0.010	0.056	0.196	0.393	0.344
8	0	0	0	0	0	0	0	0	1

P = 100 populations
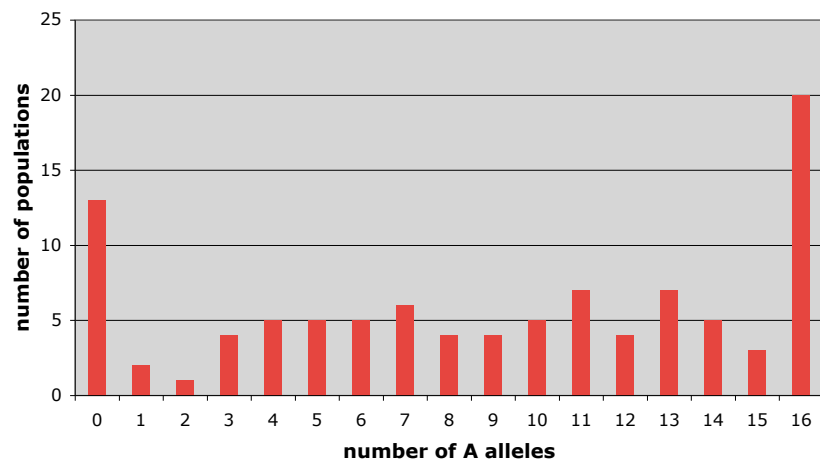
Predicted distributions of allele frequencies in replicate populations of *N* = 16

Random genetic drift in 107 *Drosophila* populations; 8 males, 8 females each generation

data from Buri (1956)
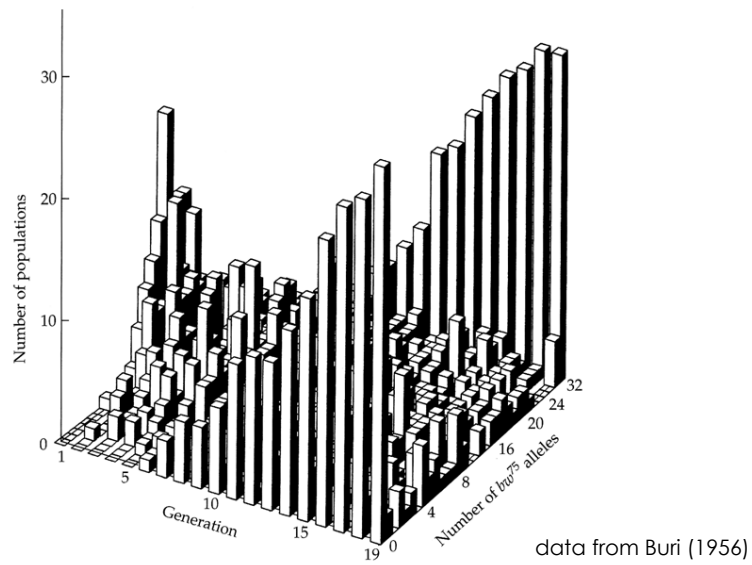
Questions

❖ How does time to fixation (or probability of fixation within a given number of generations) depend on **population size**?

❖ How does the probability of fixation for a given allele depend on its initial frequency?

❖ How likely is a new mutation to drift to fixation?

Models of Genetic Drift

❖ Binomial Probability Distribution
❖ Markov Chains
❖ Diffusion Approximation
  ✧ based on continuous rather than discrete variables (i.e., time and allele frequencies)

Wright-Fisher transition probabilities

❖ binomial results in a matrix of transition probabilities that can be multiplied to predict the distribution of allele frequencies at any number of generations in the future

to j

matrix for
$2N = 8$

from i

	0	1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0	0	0
1	0.344	0.393	0.196	0.056	0.010	0.001	~0	~0	~0
2	0.100	0.267	0.311	0.208	0.087	0.023	0.004	~0	~0
3	0.023	0.112	0.235	0.282	0.211	0.101	0.030	0.005	~0
4	0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004
5	~0	0.005	0.030	0.101	0.211	0.282	0.235	0.112	0.023
6	~0	~0	0.004	0.023	0.087	0.208	0.311	0.267	0.100
7	~0	~0	~0	0.001	0.010	0.056	0.196	0.393	0.344
8	0	0	0	0	0	0	0	0	1

Predicted distributions of allele frequencies in replicate populations of $N = 16$

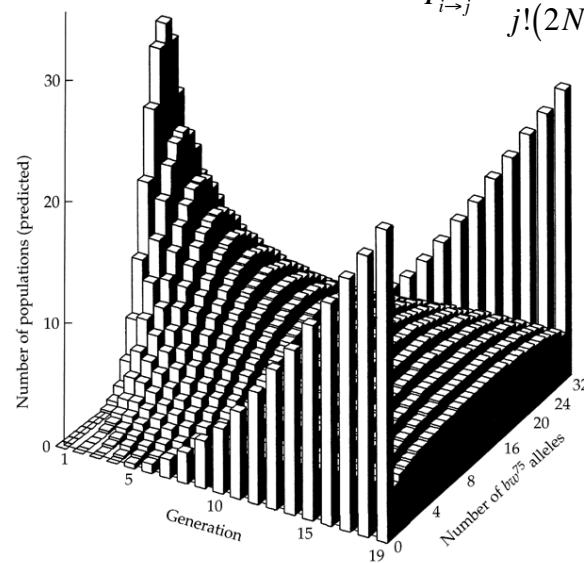$$P_{i \to j} = \frac{(2N)!}{j!(2N-j)!} f_A^j f_a^{2N-j}$$

Figure 3.10 Genetic drift modeled by a Markov chain. In this case, the sample size is two diploid genotypes ($2N = 2$) or four gametes per generation. Initial allele frequencies in all populations are $p = q = 0.5$. In one generation, sampling error shifts some proportion of the initial populations that contain two copies of each allele to states of zero, one, two, three, or four copies of one allele. Between generations one and two, sampling error again shifts some proportion of the initial populations to states of zero, one, two, three, or four copies of one allele. However, in generation one there are populations present with all allelic states. The arrows represent the possible allelic states produced by sampling error in the third generation for each of the states in the second generation. The bars in the histogram for the third generation are divided by horizontal lines to show the contributions of each second generation allelic state to the total frequency of populations with a given allelic state (some contributions are very small and are difficult to see). As the Markov process continues, the frequency distribution accumulates more and more of the populations at states of zero and four alleles, eventually reaching fixation or loss for all populations.
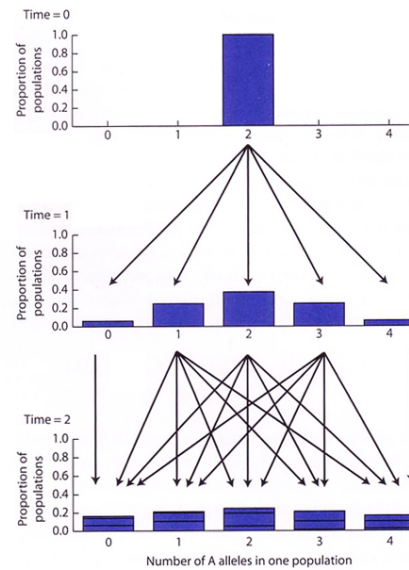
The Diffusion Approximation

$$\phi(p,x;t+\Delta t) - \phi(p,x;t) =$$
$$-\left[M(x)\phi(p,x;t) - M(x-\Delta x)\phi(p,x-\Delta x;t)\right]$$
$$+\frac{1}{2}\left\{\left[V(x+\Delta x)\phi(p,x+\Delta x;t) - V(x)\phi(p,x;t)\right]\right.$$
$$\left. -\left[V(x)\phi(p,x;t) - V(x-\Delta x)\phi(p,x-\Delta x;t)\right]\right\}$$

❖ "the discussion may seem quite mysterious, but please do not be discouraged because a detailed understanding is not necessary to understand the rest of the chapter or anything later in the book"

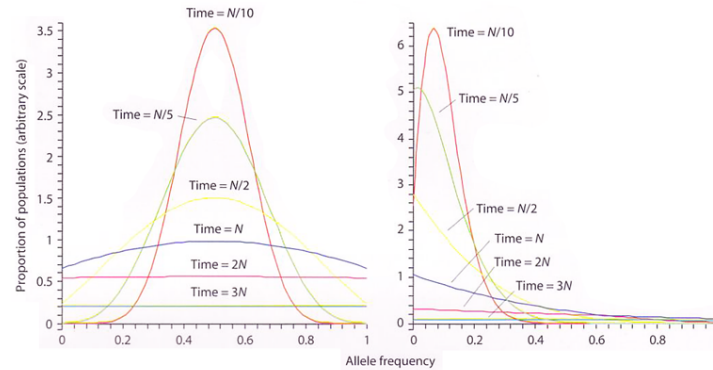## Expectations based on the diffusion approximation

Plate 3.13 Probability densities of allele frequency for many replicate populations predicted using the diffusion equation. The initial allele frequency is 0.5 on the left and 0.1 on the right. Each curve represents the probability that a single population would have a given allele frequency after some interval of time has passed. The area under each curve is the proportion of alleles that are not fixed. Time is scaled in multiples of the effective population size, N. Both small and large populations have identically shaped distributions, although small populations reach fixation and loss in less time than large populations. The populations that have reached fixation or loss are not shown for each curve.

Insights…

❖ time to fixation and time to loss

$$\bar{T}_{fix} = -4N\frac{(1-p)\ln(1-p)}{p}$$

$$\bar{T}_{loss} = -4N\frac{p\ln(p)}{1-p}$$

❖ persistence time (for polymorphism)

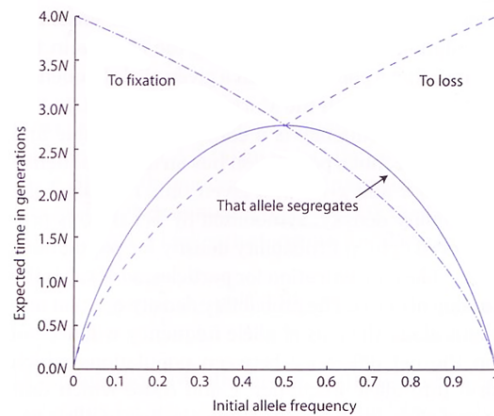$$\bar{T}_{segregate} = -4N\left[p\ln(p)+(1-p)\ln(1-p)\right]$$

12

Figure 3.14   Average time that an allele segregates, takes to reach fixation, or takes to reach loss depending on its initial frequency when under the influence of genetic drift alone. Alleles remain segregating (persist) for an average of 2.8N generations when their initial frequency is $^{1}/_{2}$. Fixation or loss takes up to an average of 4N generations when alleles are initially very rare or nearly fixed, respectively. Since these are average times, alleles in individual populations experience longer and shorter fixation, loss, and segregation times. Time is scaled in multiples of the population size.

Insights…

❖ probability of fixation for a new, neutral mutation = $1/(2N)$

  ✧ time to fixation = $\sim 4N$

❖ probability of loss = $1 - (1/(2N))$

  ✧ time to loss = $\sim 2 \ln(2N)$

* these are *average* times!