Lecture Handout 15:

Expected loss of heterozygosity due to drift in a finite population:

$$H_{t+1} = \left(1 - \frac{1}{2N}\right) H_t$$

Number of generations until a proportion x of the original heterozygosity is left (x = $H_t/H_0$)

$$t = -2N \ln x$$

So half-life is 1.39 N, ie using x = 0.5

Once a locus becomes completely homozygous (fixed) for allele $A_1$ or allele $A_2$, it will continue to be homozygous for that allele unless the lost alleles are reintroduced. Thus where 2N = 4, 0 $A_2$ alleles or 4 $A_2$ alleles are called **absorbing states.**

TABLE 6.2 A probability transition matrix for a population of size two ($2N = 4$), where the values indicate the probability of $iA_2$ alleles in generation $t+1$, given $jA_2$ alleles in generation $t$.

| Generation $t+1$ | Generation $t$ | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| 0 | 1 | 0.3164 | 0.0625 | 0.0039 | 0 |
| 1 | 0 | 0.4219 | 0.25 | 0.0469 | 0 |
| 2 | 0 | 0.2109 | 0.375 | 0.2109 | 0 |
| 3 | 0 | 0.0469 | 0.25 | 0.4219 | 0 |
| 4 | 0 | 0.0039 | 0.0625 | 0.3164 | 1 |

**Allele frequency distribution** (or gene frequency distribution): distribution of allele frequencies over populations of the same size.

**TABLE 6.3** The distribution of allele frequencies and heterozygosity over generations for populations of size two ($2N = 4$) when $q_0 = 0.5$.

| Number of $A_2$ alleles | Generation | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | ... | $\infty$ |
| 0 | 0 | 0.0625 | 0.1660 | 0.2490 | 0.3117 | ... | 0.5 |
| 1 | 0 | 0.25 | 0.2109 | 0.1604 | 0.1205 | ... | 0.0 |
| 2 | 1 | 0.375 | 0.2461 | 0.1813 | 0.1356 | ... | 0.0 |
| 3 | 0 | 0.25 | 0.2109 | 0.1604 | 0.1205 | ... | 0.0 |
| 4 | 0 | 0.0625 | 0.1660 | 0.2490 | 0.3117 | ... | 0.5 |
| $q_t$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | ... | 0.5 |
| $H_t$ | 0.5 | 0.375 | 0.2812 | 0.2109 | 0.1582 | ... | 0.0 |

# Expected time of fixation for a rare neutral mutant is four times the population size.

$$T(q) = 4N$$



**Figure 6.6.** The smoothed distribution of populations becoming fixed for $A_2$ in each generation for three initial allele frequencies when $N = 20$.

q = 0.2
q = 0.5
q = 0.8

Proportion fixed for $A_2$

Generation
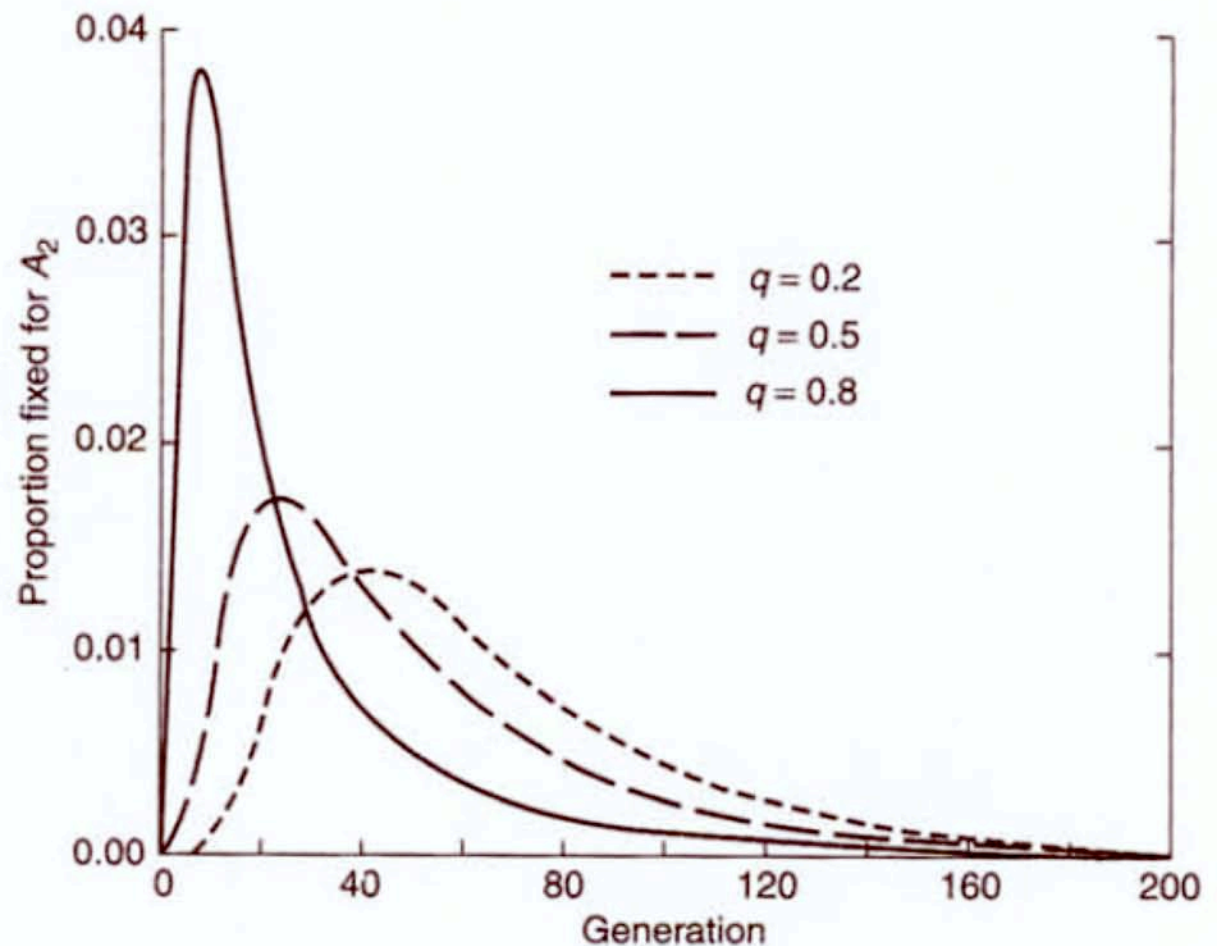
For calculating effective population size, three approaches have been used: inbreeding (loss of heterozygosity), variance and eigenvalue (= loss of allele frequency)

relating the increase in each of these factors to that of an ideal population
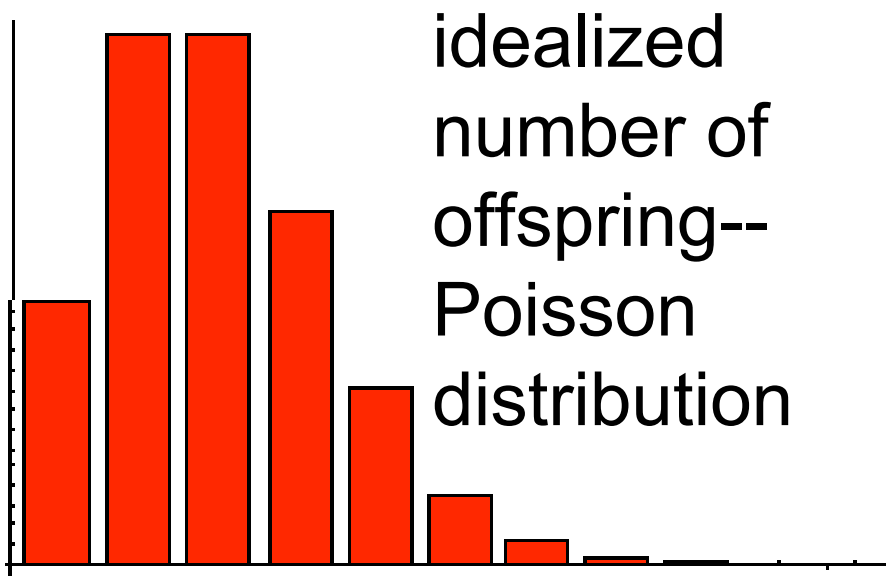
idealized number of offspring-- Poisson distribution

Versus maximum variance in reproductive success

Time
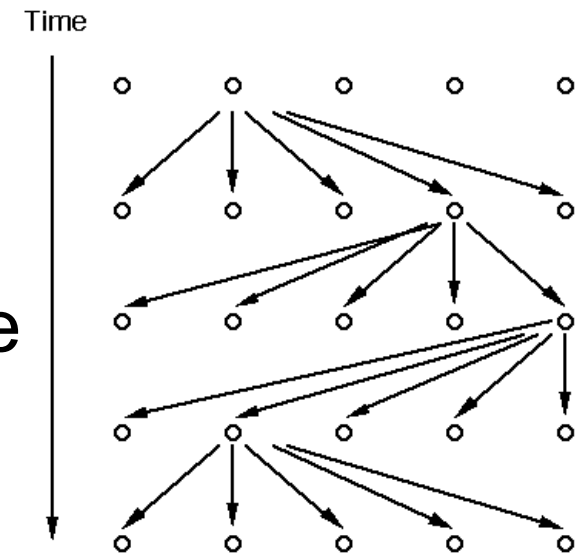
Demographic factors affect effective population size Ne, e.g., numbers of breeding males and females, variance in reproduction, and variance in numbers over generations.

**dioecy**: Having the male and female reproductive organs borne on separate individuals of the same species.

**monoecious** - Having the organs or flowers of both sexes borne on a single plant, as in corn and pines

N will equal Ne where the probability of an allele coming from a parent is 1/n for monoecious populations.

For dioecious organisms, half of gametes that give rise to progeny must come from each sex. So effective population size if numbers of males and females is the same is:

$$N_e = \frac{4 N_f N_m}{N_f + N_m}$$

For species where breeding males Nm < breeding females Nf, then Ne/N = $4(x_f - x_f^2)$ where $x_f = N_f/N$. Maximum Ne/N when $x_f = 0.5$
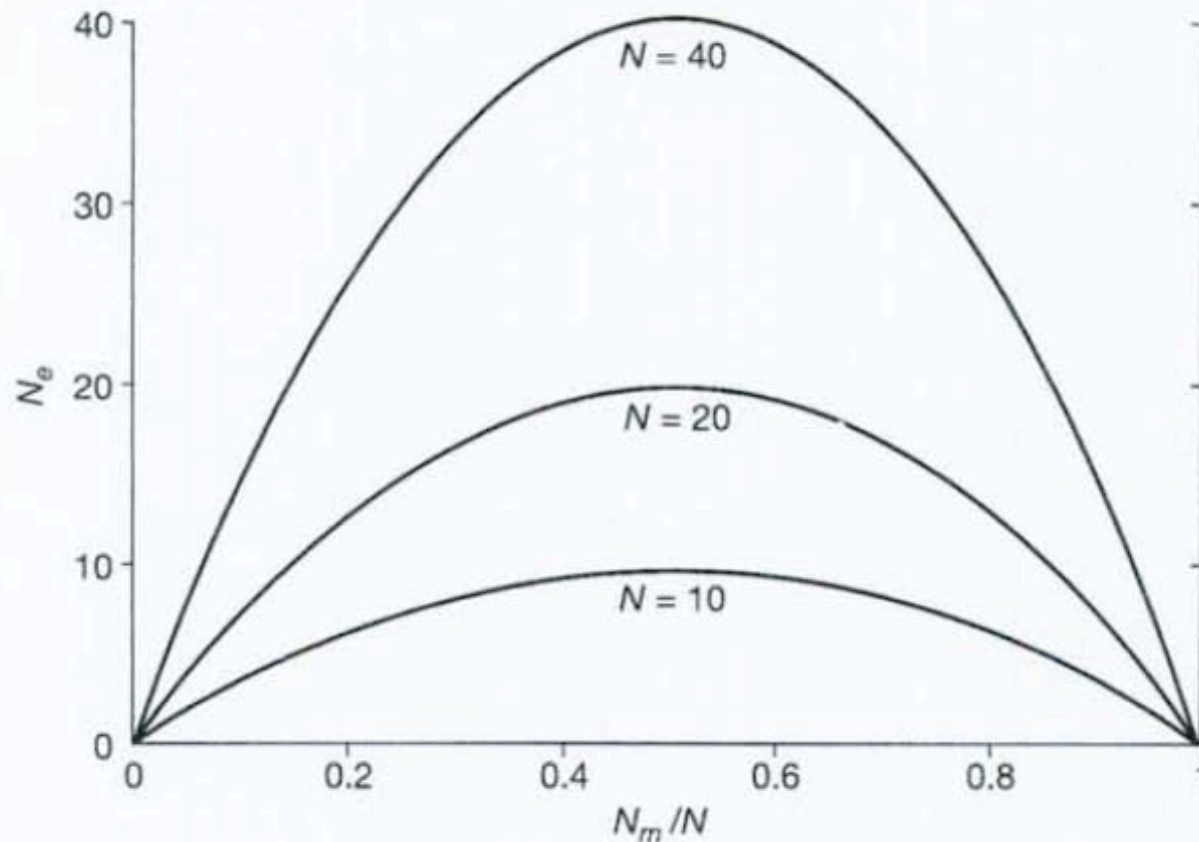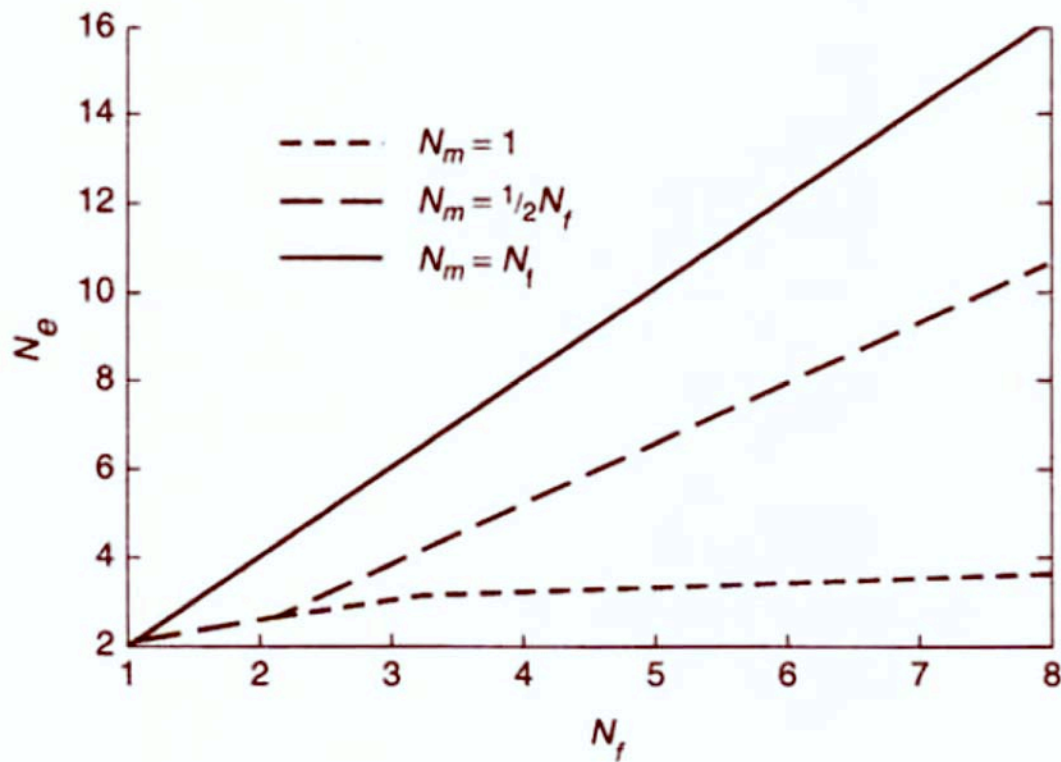


**Figure 6.8.** The effective population size as a function of the proportions of males, $N_m/N$, for three different total numbers of individuals.

# In species where males have harems,

$$N_e = \frac{4N_f}{N_f + 1}$$

## so Ne becomes 4

(1) For haplo-diploid loci, eg X chromosome, then Ne = N X 3/4 since males are haploid, & assuming males and females in equal numbers.

(2) Effective population sizes for mitochondrial DNA and for chloroplast DNA are: $N_e = N_{ef}/2$ or N/4 if males and females are in equal numbers. Since only a single "copy" is present and only females pass them on.

(3) Effective population sizes for the Y chromosome is $N_e = N_{em}/2$, or N/4 if males and females are in equal numbers. Since only a single copy is present and only males pass them on to other males.

This suggests that all other things being equal, loss of diversity and fixation will occur more quickly in X-chromosome loci than in autosomal, and even more quickly in mtDNA or (especially) Y chromosomes.

Variations in population size in time (across generations): Ne is estimated using the **harmonic mean (different from arithmetic mean)**. If Ne never gets too small, then:

$$N_e = \frac{t}{\sum \frac{1}{N_i}}$$

This gives heavy weighing to generations with lower numbers in the population.

# Inbreeding reduces effective population size

$$N_e = \frac{N}{1+f}$$

In highly selfed plants, where f =1, inbreeding would make Ne = N/2

**TABLE 6.11** DNA sequence diversity for six loci in *L. crassa* populations with different levels of selfing (Charlesworth, 2003).

| Locus | Low | Intermediate | High |
|---|---|---|---|
| | | *Level of selfing* | |
| Adh1 | 0.036 | 0.000 | 0.000 |
| Adh2 | 0.008 | 0.006 | 0.014 |
| Adh3 | 0.017 | 0.007 | 0.000 |
| Gapc | 0.028 | 0.017 | 0.014 |
| Nir1 | 0.023 | 0.022 | 0.007 |
| PgiC | 0.000 | 0.013 | 0.011 |
| Mean | 0.019 | 0.013 | 0.008 |

Other factors can influence effective population size, Ne

- age structure (generation length)

- "neighborhood size" defined by gene flow

Estimates of effective population size can use temporal changes in allele frequencies (more variance due to drift if smaller Ne)

Both a founder effect and a bottleneck can lead to chance changes in genetic variation so that allele frequencies are different from those in the ancestral population, affecting Ne:

$$N_e = \frac{H_t}{2(H_t - H_{t+1})}$$

The greater the reduction in heterozygosity due to the founding event or the bottleneck, the lower the estimate of effective population size in the founder group or bottleneck generation

Both a founder effect and a bottleneck can quickly generate genetic distance between the ancestral and the new post-bottleneck or the newly founded population, since alleles are lost versus the ancestral population

$$D_t = -\tfrac{1}{2} \ln \left( \frac{1 - H_0}{1 - H_t} \right)$$

$$D_t = -\tfrac{1}{2} \ln \left[ \frac{1 - H_0}{1 - H_0(1 - \frac{1}{2N_e})^t} \right]$$
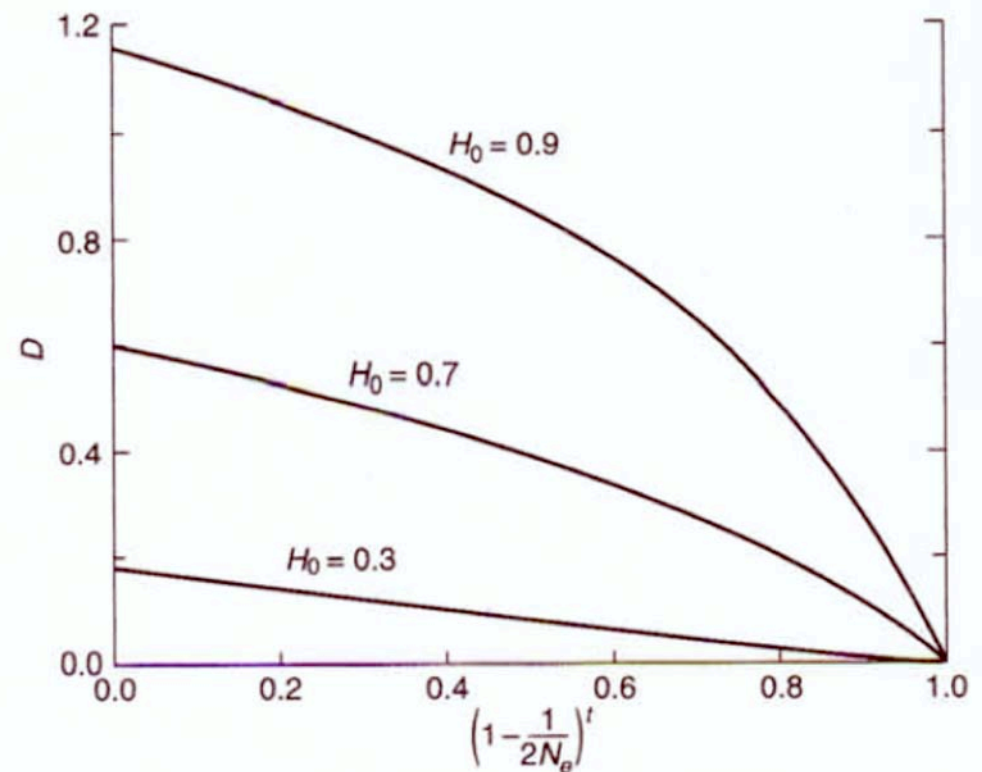
**Figure 6.12.** The probability of polymorphism, given different numbers of founders in a sample for three different allele frequencies.
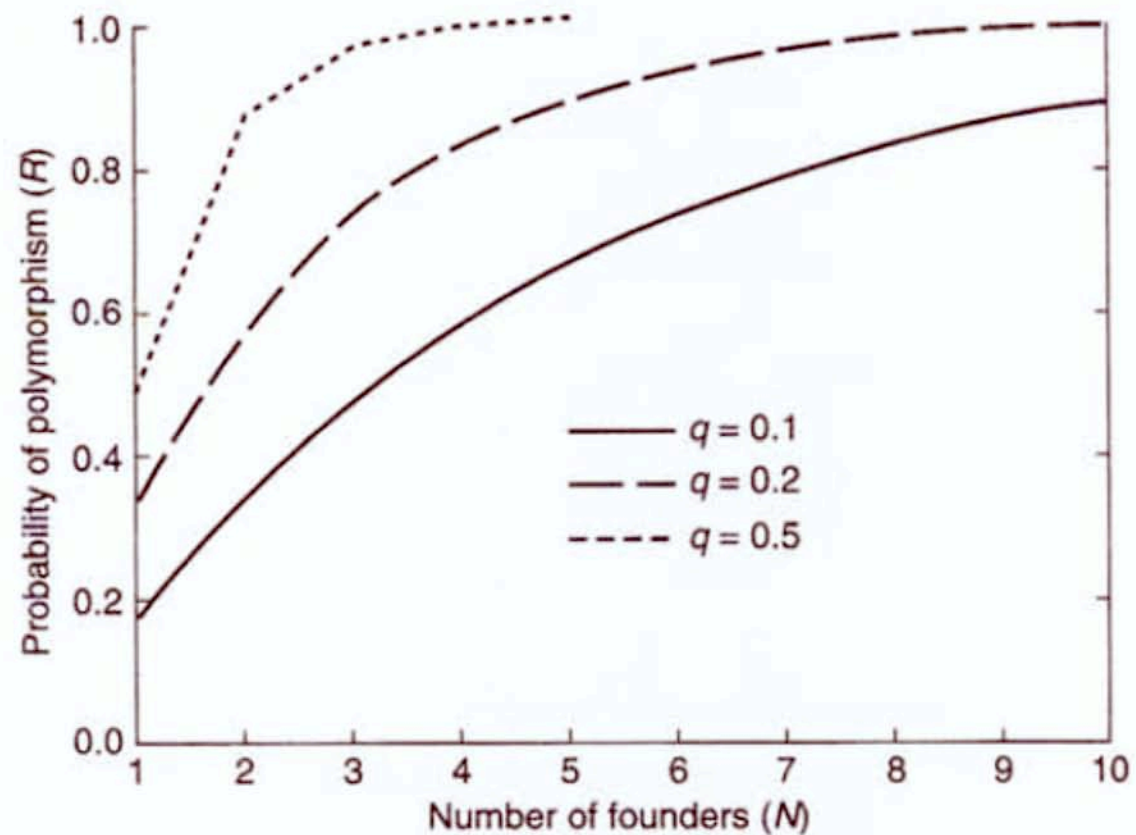
Selection in finite populations.

In an infinite population (no genetic drift), a **favorable** allele always increases in frequency asymptotically to approach **fixation**

In a finite population, the probability of fixation depends on initial frequency of allele, the amount of selection, and the finite population size. The effects of genetic drift may preclude even a favorable allele from becoming fixed.

Selection in finite populations.

In an infinite population (no genetic drift), a **detrimental** allele always decreases in frequency asymptotically to approach **loss**

In a finite population, the effects of genetic drift may by chance lead even a detrimental allele to become fixed, especially when the detrimental effects are small. (Led to Ohta's "nearly neutral" model of molecular evolution).

# Simulation: fixation of a slightly detrimental allele in a small population.



**Figure 6.16.** The smoothed distribution of allele frequency for a population of size 20 and an initial allele frequency of 0.5 with selection such that the fitnesses of $A_1A_1$, $A_1A_2$, and $A_2A_2$ are 1.1, 1.05, and 1.0, respectively, after 1, 5, and 20 generations.