**Sampling error**

Before taking on the notion of genetic drift in populations, let’s first take a look at sampling variation. Let’s consider the age-old coin tossing experiment.

Assume a fair coin with $p = \frac{1}{2}$.

If you sample many times the most likely single outcome = $\frac{1}{2}$ heads.

The overall most likely outcome $\neq \frac{1}{2}$ heads.

This is a binomial sampling problem.

\[
P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}
\]

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]

$n$ is the number of flips

$k$ is the number of successes

Let’s look at the probability of the following:

<table>
<thead>
<tr>
<th>$k$ heads from $n$ flips</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k = 5$ from $n = 10$</td>
<td>0.246</td>
</tr>
<tr>
<td>$k = 6$ from $n = 10$</td>
<td>0.205</td>
</tr>
</tbody>
</table>

So, the most likely single outcome is $\frac{1}{2}$ heads (with $p = 0.246$), the overall likelihood of observing something other than $\frac{1}{2}$ heads is higher ($p = 1 - 0.246 = 0.754$)

The good news is that as we increase the sample size the likelihood of observing something very close to the expected frequency, $E(p) = 0.5$, goes up. The probability of a given frequency of heads from $n$ flips of the coin is:

<table>
<thead>
<tr>
<th>N flips</th>
<th>p &lt;0.35</th>
<th>p = 0.35-0.45</th>
<th>p = 0.45-0.55</th>
<th>p = 0.55-0.65</th>
<th>p &lt;0.65</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.16</td>
<td>0.21</td>
<td>0.25</td>
<td>0.21</td>
<td>0.16</td>
<td>0.025</td>
</tr>
<tr>
<td>20</td>
<td>0.06</td>
<td>0.19</td>
<td>0.50</td>
<td>0.19</td>
<td>0.06</td>
<td>0.0125</td>
</tr>
<tr>
<td>50</td>
<td>0.002</td>
<td>0.16</td>
<td>0.68</td>
<td>0.16</td>
<td>0.002</td>
<td>0.005</td>
</tr>
</tbody>
</table>

It is clear that sample size $N$ is important. If we were to flip a coin 1000 times we would get very close to the expected frequency of 0.5 (but not exactly). Only in infinitely large samples is such error avoided.

Say we want to measure the frequency of some trait in a population, so we take a sample from that population. The frequencies in that sample will not be identical to the frequencies in the original population due to sampling error.
**Genetic Drift**

Now let's consider the problem in a diploid population. Remember that under HW conditions the frequencies do not change from one generation to the next. The gamete pool is in fact a random sample of the alleles in the population of reproducing adults. Under HW the problem of sampling error is avoided by assuming an idealized population where the sample size is infinite. If we specify that a population must have a finite size, the random union of gametes in each new generation will be subject to sampling errors. Allele frequencies will no longer remain the same among generations in the realistic case of finite population size.

The problem of sampling error of alleles in populations is a binomial problem. Suppose we have a natural population with the frequency of $A = 0.75$ and $a = 0.25$. Suppose only 200 individuals reproduce this year; the number of alleles will be 400. What is the probability that the frequency of the $A$ allele will be exactly 0.75 in the next generation? According to the binomial distribution the probability of actually getting exactly 0.75 in the next generation is only 0.05. Sampling errors from generation to generation are unavoidable.

**GENETIC DRIFT** is the accumulation of random sampling fluctuations in allele frequency over generations.

The magnitude of change in allele frequencies is inversely proportional to the sample size.

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

As we saw in the previous discussion of inbreeding the critical measure of population size is not the census size ($N$), but the effective population size ($N_e$). The magnitude of genetic drift in a population will thus be proportional to the inverse of the effective population size.

\[
\frac{1}{N_e}
\]

As $N_e$ increases, the expected magnitude of change decreases. Hence, the importance of genetic drift as a force for evolution changes with the effective population size. The figures below illustrate this relationship.
Note that the fluctuations in allele frequencies from generation to generation become less conspicuous as $N_e$ increases. In all the cases of drift, the ultimate fate of the system is the fixation of one or the other of the alleles. In the case of very small effective population sizes this occurs relatively quickly. The rate at which a population goes to fixation slows down with increasing population size. Although we do not see this outcome in the above populations with $N_e = 1000$, $10000$, and $50000$, had we run the simulation long enough, we would have observed them all go to fixation.

The probability that an allele will be fixed by drift is equal to its frequency in the population (i.e., $p$ or $q$). This has implication for the probability of fixation of new alleles. The frequency of a new allele in a population, and hence its probability of fixation, will be

$$\frac{1}{N_e}$$

This is as small as it can get. Thus, the fate of most new mutations in a population is loss due to drift. Note that even strong selection can’t overcome the effect of drift at this level.
Changes in allele frequency due to drift are unpredictable. Genetic drift is a random process. If we start at a frequency $p = q = 0.5$, the number of alleles fixed in either direction will be equal. Below is an example of 10 independent populations, with $N = 50$ and $p = q = 0.5$, that were run for 25 generations. It is clear that the frequencies in individual populations are erratic and unpredictable. In one case the specified allele was fixed and in another it was lost.

If you extend the above example over many more populations and many more generations, you would then be able to predict the outcome of genetic drift as an average over all the populations. If you took an infinitely large population, with $p = q = 0.5$, and split it up into a very large number of ideal populations with finite population sizes, after some number of generations there would be a distribution of allele frequencies. The panel of 10 populations above represents a sample from such a distribution.

As the number of generations increases, more and more populations will become fixed for one or the other allele.
The effects of drift are cumulative over time. Moreover, as an average over populations the effects of drift are predictable:

1. loss of variation within populations
2. gain of variation between populations

**Genetic Drift affects heterozygosity**

Let’s start with HW at generation $t$.

$$H_t = 2pq$$

The actual allele frequencies, $p$ and $q$, will change from generation to generation by some amount due to drift. Let’s call this amount $x$.

$$H_{t+1} = 2(p + x)(q - x)$$

$$H_{t+1} = 2pq + 2x(q - p) - 2x^2$$

What is $E(x)$? The average value of $x$ over time is expected to be 0, although the actual value from generation to generation will be either positive or negative. This means that $E(x^2)$ is always positive.

$$E(2pq + 2x(q - p) - 2x^2) = 2pq - 2x^2$$

Heterozygosity is expected reduced by genetic drift. Nice, eh?
**Bottlenecks and founder effects**

A single dramatic genetic drift event is often called a genetic bottleneck. A bottleneck occurs when a population undergoes an extraordinary reduction in size. Bottlenecks are common in natural populations and can be caused by external factors such as natural disasters, climatic changes, epidemics, etc. Such populations often recover from such events, but the random sampling of individuals that “pass through the bottleneck” changes the genetics of the population. Even if the population returns to the pre-bottleneck size, the genetics of the populations will be changed in two ways:

(i) the allele frequencies will be different from those in the pre-bottleneck population
(ii) the genetic diversity of the population will be reduced.

One of the most well-studied bottleneck events is one that occurred in the Northern Elephant seal (*Mirounga angustirostris*) at the end of the 19th century. Extreme hunting pressure motivated by a market for blubber oil reduced the effective population size of the species to about 20, pushing it to the very brink of extinction. However, the Northern Elephant seal is one of the great success stories of conservation. The species was protected under Mexican and US law in the early 20th century and since made a dramatic comeback. Recent estimates put the census size over 175,000 individuals.

Two species that have suffered extreme bottlenecks due to commercial harvesting

<table>
<thead>
<tr>
<th>Northern elephant seal</th>
<th>Northern right whale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent population recovery</td>
<td>Poor population recovery</td>
</tr>
</tbody>
</table>
Numerous studies of genetic variation have revealed extremely little to absolutely no genetic variation. Given an effective population size of about 20 during the bottleneck, these results are not surprising. Comparisons with genetic variation in the Southern Elephant seal (M. leonina) reveals significantly higher genetic diversity. An expected consequence of such a reduction is inbreeding depression; yet the Northern Elephant seal seems to have recovered with very little difficulty. It is interesting to contrast this strong comeback with other cases, such as the northern right whale (Eubalaena glacialis) where there has been a similarly well documented bottleneck, but the species is struggling to recover.

What would be a possible explanation for differences between species that suffer similar bottleneck effects? Here is a hint: the Northern Elephant seal exhibits reproductive site fidelity, male-male competition for mates, and extreme polygamy).

Another type of extreme genetic drift is the founder effect. A founder effect occurs when a patch of newly available habitat, or an island, is colonized by a very small number of individuals from a species that was previously absent from that environment. Again, as all the genes in the new population came from the few individuals that founded the new population the drift effects will be large. The allele frequencies in the new population will be significantly different from those in the parental populations.

The population of Old Order Amish people of Lancaster county Pennsylvania number between 16 – 18,000 persons. This population descends from about 200 founders, who were members of an Anabaptist sect in Germany and immigrated to US in the early 1700’s. The strong effect of genetic drift associated with the founding of this population resulted in a much higher frequency of a genetic disease called Ellis-van Creveld syndrome. Because the people of this community rarely marry outside of it, inbreeding brings together recessive alleles of this disease more frequently than outside the community and the phenotypic affects of the disease such as polydactyly (extra fingers or toes) are observed at much higher frequencies among the Amish people.

Polydactyly caused by the homozygous recessive disease Ellis-van Creveld syndrome

Other symptoms of this disease include dwarfisms, abnormalities of the nails and teeth, and a hole between the two upper chambers of the heart.
The founder effect could have implications for the process of speciation. The Hawaiian Drosophila represents the most outstanding examples of a process of speciation closely associated with founder events. The estimated number of Drosophila species in the Hawaiian archipelago is over 1000 (only about 300 species in the rest of the world!). These species exhibit an exceptional amount of morphological diversity, yet genetic studies show them to be phylogenetically closely related. Most of the information about the process of speciation in Hawaiian Drosophila comes from studies of the picture wing group, which contains about 111 species; most of these are found on only 1 island!

The overall pattern of speciation is one of repeated colonization of younger islands by individuals from populations established on older islands. Most species occur on a single island and their most closely related species occurs on a neighbouring island. Shifts in mate recognition systems seem to be important to the process of speciation and appears linked to the pattern of colonization. Sister taxa on adjacent islands tend to occupy remarkably similar ecological niches, but differ dramatically in their reproductive behaviours.

There is controversy surrounding the relative importance to speciation of drift and inbreeding as compared with natural selection. However, dramatic drift affects, combined with inbreeding, will significantly change genotype frequencies; the combination of assortative mating and natural selection for certain phenotypes in such populations might be important. The drift effect might set off a cascade of changes, some driven by selection. Remember that these new populations are cut-off from exchange with the parental population for long periods of time. *Drift effects might have sped up the speciation process in the picture wings, but if so, it is likely to be generally very rare elsewhere.*
Keynotes:

- Genetic drift influences both allele frequency and genotype frequency.
- Drift decreases diversity within populations and increases diversity between populations.
- Under genetic drift, the rate to fixation is determined by \( N_e \) and the probability of fixation by \( p \).
- In specific cases the outcome of genetic drift is unpredictable.
- The effects of drift are predictable as an average over populations.
- Because drift reduces genetic variation in populations, a population’s ability to evolve in response to new selective pressures might be reduced (remember Trudy MacKay’s experiments). Alternatively, some believe that drift could actually increase the rate of speciation (e.g., Hawaiian Drosophila).
- Because the effect of drift is inversely proportional to the effective population size, its affects are particularly important in rare and endangered species.
- Founder effects may play an important role in some speciation events.