Chapter 2 – Genetic Drift & Mutation

The Wright-Fisher Model

Mutation

The Wright-Fisher model

\[ \text{2N haploid individuals} \xrightarrow{p_0} \text{\(\infty\) gametes} \xrightarrow{p_0} \text{2N individuals} \]

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

- **5 Match + Megaball:**
  - 1 in 175,711,536: $250,000**
  - 1 in 3,904,701: $10,000**
- **4 Match + 1 Megaball:**
  - 1 in 15,313: $150**
  - 1 in 689,065: $567
- **3 Match + 1 Megaball:**
  - 1 in 13,781: $150**
  - 1 in 306: $3
- **2 Match + 1 Megaball:**
  - 1 in 844: $4
  - 1 in 306: $3
- **1 Match + 1 Megaball:**
  - 1 in 141: $3
  - 1 in 75: $2
**Expectation**

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

<table>
<thead>
<tr>
<th></th>
<th>Net Ticket Sales*</th>
<th>Prizes*</th>
<th>Prizes as % of Ticket Sales</th>
<th>Admin Costs*</th>
<th>Admin as % of Ticket Sales</th>
<th>Net Proceeds to State*</th>
<th>Net Proceeds as % of Ticket Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Massachusetts</td>
<td>4,200</td>
<td>3,235</td>
<td>77.00%</td>
<td>83</td>
<td>2.00%</td>
<td>883</td>
<td>21.00%</td>
</tr>
</tbody>
</table>

* 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

\[
E[f_A(t+1)] = 2Nf_A(t)/2N = f_A(t)
\]
Genetic Drift

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

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 $f_A = f_a = 0.50$

![Histogram of allele distribution for $f_A = f_a = 0.50$.]

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

![Histogram of allele distribution for $f_A = 0.25, f_a = 0.75$.]
### The Moran model

- derived for a haploid population
  - but results for $2N$ 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

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

### The Wright-Fisher model

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

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

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

Matrix for $2N = 8$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.344</td>
<td>0.393</td>
<td>0.196</td>
<td>0.056</td>
<td>0.010</td>
<td>0.001</td>
<td>~0</td>
<td>~0</td>
<td>~0</td>
</tr>
<tr>
<td>2</td>
<td>0.100</td>
<td>0.267</td>
<td>0.311</td>
<td>0.208</td>
<td>0.087</td>
<td>0.023</td>
<td>0.004</td>
<td>~0</td>
<td>~0</td>
</tr>
<tr>
<td>3</td>
<td>0.023</td>
<td>0.112</td>
<td>0.235</td>
<td>0.282</td>
<td>0.211</td>
<td>0.101</td>
<td>0.030</td>
<td>0.005</td>
<td>~0</td>
</tr>
<tr>
<td>4</td>
<td>0.004</td>
<td>0.031</td>
<td>0.109</td>
<td>0.219</td>
<td>0.273</td>
<td>0.219</td>
<td>0.109</td>
<td>0.031</td>
<td>0.004</td>
</tr>
<tr>
<td>5</td>
<td>~0</td>
<td>0.005</td>
<td>0.030</td>
<td>0.101</td>
<td>0.211</td>
<td>0.282</td>
<td>0.235</td>
<td>0.112</td>
<td>0.023</td>
</tr>
<tr>
<td>6</td>
<td>~0</td>
<td>~0</td>
<td>0.004</td>
<td>0.023</td>
<td>0.087</td>
<td>0.208</td>
<td>0.311</td>
<td>0.267</td>
<td>0.100</td>
</tr>
<tr>
<td>7</td>
<td>~0</td>
<td>~0</td>
<td>~0</td>
<td>0.001</td>
<td>0.010</td>
<td>0.056</td>
<td>0.196</td>
<td>0.393</td>
<td>0.344</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

P = 100 populations

![Graph showing the distribution of A alleles across 100 populations.](image)
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

\[
P_{i \rightarrow j} = \frac{(2N)!}{j!(2N-j)!} f_A^j f_a^{2N-j}
\]

Matrix for 2N = 8

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.444</td>
<td>0.393</td>
<td>0.196</td>
<td>0.056</td>
<td>0.010</td>
<td>0.001</td>
<td>~0</td>
<td>~0</td>
<td>~0</td>
</tr>
<tr>
<td>2</td>
<td>0.100</td>
<td>0.267</td>
<td>0.311</td>
<td>0.208</td>
<td>0.087</td>
<td>0.023</td>
<td>0.004</td>
<td>~0</td>
<td>~0</td>
</tr>
<tr>
<td>3</td>
<td>0.023</td>
<td>0.112</td>
<td>0.235</td>
<td>0.282</td>
<td>0.211</td>
<td>0.101</td>
<td>0.030</td>
<td>0.005</td>
<td>~0</td>
</tr>
<tr>
<td>4</td>
<td>0.004</td>
<td>0.031</td>
<td>0.109</td>
<td>0.219</td>
<td>0.273</td>
<td>0.219</td>
<td>0.109</td>
<td>0.031</td>
<td>0.004</td>
</tr>
<tr>
<td>5</td>
<td>~0</td>
<td>0.005</td>
<td>0.030</td>
<td>0.101</td>
<td>0.211</td>
<td>0.282</td>
<td>0.235</td>
<td>0.112</td>
<td>0.023</td>
</tr>
<tr>
<td>6</td>
<td>~0</td>
<td>~0</td>
<td>0.004</td>
<td>0.023</td>
<td>0.087</td>
<td>0.208</td>
<td>0.311</td>
<td>0.267</td>
<td>0.100</td>
</tr>
<tr>
<td>7</td>
<td>~0</td>
<td>~0</td>
<td>~0</td>
<td>0.001</td>
<td>0.010</td>
<td>0.056</td>
<td>0.196</td>
<td>0.393</td>
<td>0.344</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Predicted distributions of allele frequencies in replicate populations of N = 16
The Diffusion Approximation

\[ \phi(p, x; t + \Delta t) - \phi(p, x; t) = \]

\[ -\left[ M(x)\phi(p, x; t) - M(x - \Delta x)\phi(p, x - \Delta x; t) \right] \]

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

"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 $N$. 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**
  \[ \bar{T}_{\text{fix}} = -4N \frac{(1 - p) \ln(1 - p)}{\ln(1 - p)} \]
  \[ \bar{T}_{\text{loss}} = -4N \frac{p \ln(p)}{1 - p} \]

- **persistence time (for polymorphism)**
  \[ \bar{T}_{\text{seg}} = -4N \left[ p \ln(p) + (1 - p) \ln(1 - p) \right] \]
Insights…

- Probability of fixation for a new, neutral mutation = $1/(2N)$
- Time to fixation = ~ $4N$
- Probability of loss = $1 - (1/(2N))$
- Time to loss = ~ $2 \ln(2N)$

* These are average times!