Mathematical Modeling Lecture 7

Finite-population genetics: inheritance, drift, and fixation

Zan Ahmad

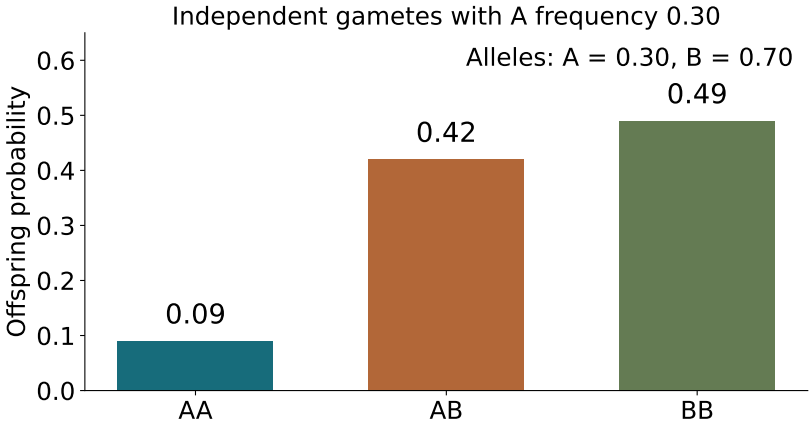
Fall 2024

Diploid accounting

$$n_{AA} + n_{AB} + n_{BB} = N, \qquad p = \frac{2n_{AA} + n_{AB}}{2N}.$$

Every heterozygote contributes one A copy and one B copy.

Independent gamete pairing



Independent gamete pairing

$$\Pr(AA) = p^2, \quad \Pr(AB) = 2p(1-p), \quad \Pr(BB) = (1-p)^2.$$

The heterozygote combines two ordered outcomes.

Probabilities and expected counts

At $p = 0.30$:

$$(p^2, 2p(1-p), (1-p)^2) = (0.09, 0.42, 0.49).$$

For 100 offspring, expected counts are $(9, 42, 49)$.

A finite random sample need not equal its expected counts.

Neutrality preserves the expected allele fraction

$$p^2 + \frac{1}{2}\,2p(1-p) = p.$$

No selection, mutation, or migration is included.

Constancy of expectation does not imply constancy of each realization.

A finite state variable

$$K_n \in \{0, 1, \ldots, 2N\}, \qquad p_n = K_n/(2N).$$

Offspring copies are sampled from the current parental pool.

The binomial transition law

$$K_{n+1} \mid p_n \sim \text{Binomial}(2N, p_n).$$

$$P_{ij} = \binom{2N}{j} (i/2N)^j (1 - i/2N)^{2N-j}.$$

One neutral generation

$$N = 25, \quad K_n = 15 \quad \Rightarrow \quad p_n = 0.30.$$

A possible next count is $K_{n+1} = 18$, giving $p_{n+1} = 0.36$.

The next generation samples with probability 0.36.

Conditional expectation and variance

$$\mathbb{E}[p_{n+1} \mid p_n] = p_n, \qquad \mathrm{Var}(p_{n+1} \mid p_n) = \frac{p_n(1-p_n)}{2N}.$$

Mean change is zero; random change need not be zero.

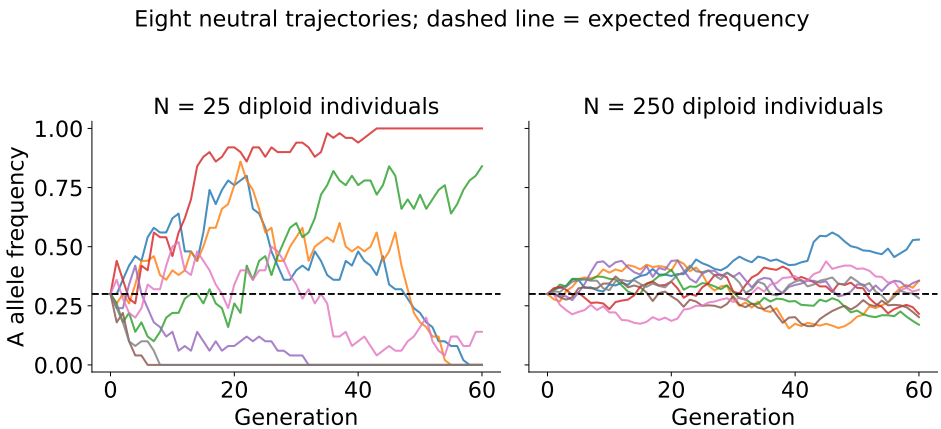
Population-size scaling of fluctuations

At $p = 0.30$:

$$\sigma_{25} = \sqrt{0.21/50} \simeq 0.06481, \qquad \sigma_{250} = \sqrt{0.21/500} \simeq 0.02049.$$

Tenfold larger N reduces standard deviation by $\sqrt{10}$.

A constant ensemble mean



A constant ensemble mean

$$\mathbb{E}[p_{n+1}] = \mathbb{E}[\mathbb{E}[p_{n+1} \mid p_n]] = \mathbb{E}[p_n] = p_0.$$

The plotted dashed line is p_0, not the mean of eight displayed runs.

Expected heterozygosity declines

$$\mathbb{E}[p_{n+1}(1-p_{n+1}) \mid p_n] = \left(1 - \frac{1}{2N}\right) p_n(1-p_n).$$

$$\mathbb{E}[H_n] = 2p_0(1-p_0)\left(1 - \frac{1}{2N}\right)^n.$$

Distributional broadening

$$\mathrm{Var}(p_n) = p_0(1-p_0)\left[1-\left(1-\frac{1}{2N}\right)^n\right].$$

The mean is fixed while probability moves toward the endpoints.

Absorbing endpoints

$$p_n = 0 \Rightarrow p_{n+1} = 0, \qquad p_n = 1 \Rightarrow p_{n+1} = 1.$$

Without mutation or migration, a missing allele cannot be resampled.

Neutral fixation probability

For a finite neutral chain, absorption occurs with probability one.

$$\Pr(p_\infty = 1) = \mathbb{E}[p_\infty] = p_0.$$

Fixation does not require a selective advantage.

Computational checks for a random model

Check integer counts, the $1/(2N)$ frequency grid, bounds, and absorbing states.

Compare repeated-draw moments with theoretical moments using sampling uncertainty.

A seed controls repeatability, not representativeness.

Mean dynamics and individual populations

$$p_{n+1} = p_n$$

reproduces the neutral expectation but removes all sampling variation.

Drift, loss, and fixation require the distributional model.