

Estimation of Allele Frequency:

4 Approach to Equilibrium for an X-Linked Gene

4.1: Learning Objectives

By the end of this chapter, the reader should be able to set up and solve the recurrence relations describing X-linked allele frequency change across generations when male and female founding frequencies differ, and apply the results to a Nigerian example such as G6PD deficiency or red-green colour blindness.

4.2: Why X-Linked Genes Need Special Treatment

For an autosomal locus, both sexes carry two alleles, and — as shown in Chapter 3 — random mating brings the population to Hardy–Weinberg genotype proportions in a single generation, with allele frequencies remaining constant thereafter. For an **X-linked locus**, males carry only one X chromosome (hemizygous) while females carry two. If, for any reason (e.g., historical migration bringing together two populations with different allele frequencies, or differential selection in the two sexes), the allele frequency among males differs from that among females in a given generation, the two sexes do **not** reach a common equilibrium frequency instantly — it takes several generations, with the difference between the sexes shrinking geometrically.

This distinction matters clinically as well as theoretically. Because a male's single X-linked allele is always expressed (there is no second copy to mask a recessive allele), X-linked recessive

conditions such as G6PD deficiency, red-green colour blindness, and haemophilia are far more common, and more severe on average, in males than in females — females require two copies of the deficient allele (genotype-wise equivalent to the rare homozygote in Chapter 3) to be affected, whereas a male is affected with just one. This asymmetry is precisely why the two-sex recurrence developed in this chapter, rather than the single-sex Hardy–Weinberg model of Chapter 3, is required to correctly describe X-linked allele-frequency dynamics — and why pedigree analysis (Chapter 6) treats X-linked inheritance as visually and statistically distinct from autosomal inheritance.

4.3: Setting Up the Recurrence

Let $p_m(t)$ and $p_f(t)$ denote the frequency of a given X-linked allele (say, the deficient G6PD allele) among the X chromosomes carried by males and females, respectively, in generation t .

Sons receive their single X chromosome from their **mother only** (they receive a Y from their father). Hence the allele frequency among sons in generation $t+1$ equals the allele frequency among mothers' X chromosomes in generation t :

$$p_m(t + 1) = p_f(t)$$

Daughters receive one X from their mother and one X from their father. Assuming random mating with respect to this locus, the allele frequency among daughters' X chromosomes in generation $t+1$ is the average of the parental (mother and father) allele frequencies in generation t :

$$p_f(t + 1) = \frac{1}{2}p_f(t) + \frac{1}{2}p_m(t)$$

4.4: Solving the Recurrence: Approach to Equilibrium

Define the sex difference $d(t) = p_f(t) - p_m(t)$. Then:

$$\begin{aligned} d(t+1) &= p_f(t+1) - p_m(t+1) = \left[\frac{1}{2}p_f(t) + \frac{1}{2}p_m(t) \right] - p_f(t) \\ &= -\frac{1}{2}[p_f(t) - p_m(t)] = -\frac{1}{2}d(t) \end{aligned}$$

This is a simple geometric (first-order linear) recurrence, with solution:

$$d(t) = \left(-\frac{1}{2}\right)^t d(0)$$

So the difference between male and female allele frequencies **halves in magnitude and alternates in sign** each generation, decaying to zero as $t \rightarrow \infty$. The population therefore approaches a common equilibrium allele frequency, oscillating with diminishing amplitude — a classic **damped oscillation**.

The equilibrium frequency: Since males carry 1/3 of all X chromosomes in the population and females carry 2/3 (because for every one male X there are two female X's, on a per-chromosome count, in a population with an equal sex ratio: males contribute 1 X each, females 2 X each), the equilibrium (long-run, common) allele frequency is the weighted average:

$$p^\infty = (1/3)p_m(0) + (2/3)p_f(0)$$

This weighting (1/3 : 2/3) reflects the fact that two-thirds of all X chromosomes in the population reside in females and one-third in males.

Theorem 4.1: (Convergence of X-Linked Allele Frequencies)

Under the recurrence of Section 4.3, the sex-specific allele frequencies $p_m(t)$ and $p_f(t)$ both converge, as $t \rightarrow \infty$, to the common limit $p_\infty = (1/3)p_m(0) + (2/3)p_f(0)$, and the rate of convergence is geometric: $|d(t)| = |p_f(t) - p_m(t)| = (1/2)^t |d(0)|$.

Proof. The geometric decay of $d(t)$ was established directly above by solving the first-order linear recurrence $d(t+1) = -(1/2)d(t)$, giving $d(t) = (-1/2)^t d(0)$, from which $|d(t)| = (1/2)^t |d(0)| \rightarrow 0$ as $t \rightarrow \infty$ (since $1/2 < 1$). It remains to show both $p_m(t)$ and $p_f(t)$ converge to the stated weighted average. Define the population-wide (weighted) allele frequency $\bar{p}(t) = 1/3 p_m(t) + 2/3 p_f(t)$. Substituting the two recurrences of Section 4.3:

$$\begin{aligned}\bar{p}(t+1) &= 1/3 p_f(t) + 2/3 [1/2 p_f(t) + 1/2 p_m(t)] \\ &= 1/3 p_f(t) + 1/3 p_f(t) + 1/3 p_m(t) \\ &= 1/3 p_m(t) + 2/3 p_f(t) = \bar{p}(t),\end{aligned}$$

so $\bar{p}(t)$ is a **conserved quantity** (an invariant of the recurrence), constant at $\bar{p}(t) = \bar{p}(0) = p_\infty$ for all t . Since $p_m(t) = p_f(t) - d(t)$ and $d(t) \rightarrow 0$, while $\bar{p}(t) = 1/3 p_m(t) + 2/3 p_f(t)$ remains fixed at p_∞ , substituting $p_m(t) \rightarrow p_f(t)$ into the (constant) weighted average forces $p_f(t) \rightarrow p_\infty$ and hence $p_m(t) \rightarrow p_\infty$ as well.

Worked Example 4.2 (Convergence after a Northern-Southern Nigeria population mixing event)

Suppose historical migration brings together two Nigerian sub-populations with different long-standing G6PD allele frequencies: $p_m(0) = 0.30$ among males and $p_f(0) = 0.10$ among females (a large initial imbalance, perhaps reflecting differential historical malaria exposure or founder effects in the two source groups). By Theorem 4.1, the eventual common equilibrium frequency is $p_\infty = (1/3)(0.30) + (2/3)(0.10) = 0.10 + 0.0667 = 0.1667$, and the sex gap after t generations is $|d(t)| = (1/2)^t \times |0.10 - 0.30| = (1/2)^t \times 0.20$. After 5 generations, the residual gap is $(1/32)(0.20) \approx 0.006$ — practically negligible — illustrating how quickly (within a handful of generations) X-linked allele frequency differences between the sexes, and by extension between historically distinct but now interbreeding Nigerian sub-populations, are expected to equilibrate under simple random mating.

4.5: Worked Example 4.1 (G6PD deficiency, hypothetical founder scenario)

Suppose a historical scenario in which, at generation 0, the frequency of the G6PD-deficient allele is $p_m(0) = 0.30$ among males and $p_f(0) = 0.10$ among females (for instance, following the settlement of a population with a higher initial deficient-allele frequency among the founding men). Find the allele frequency among males and females at generations 1, 2, and 3, and the equilibrium frequency.

Solution.

Equilibrium: $p_\infty = (1/3)(0.30) + (2/3)(0.10) = 0.10 + 0.0667 = 0.1667$.

$d(0) = p_f(0) - p_m(0) = 0.10 - 0.30 = -0.20$.

Generation 1: $p_m(1) = p_f(0) = 0.10$. $p_f(1) = \frac{1}{2}(0.10) + \frac{1}{2}(0.30) = 0.20$. Check: $d(1) = 0.20 - 0.10 = 0.10 = -\frac{1}{2} \times (-0.20)$. ✓.

Generation 2: $p_m(2) = p_f(1) = 0.20$. $p_f(2) = \frac{1}{2}(0.20) + \frac{1}{2}(0.10) = 0.15$. $d(2) = 0.15 - 0.20 = -0.05 = -\frac{1}{2} \times (0.10)$. ✓

Generation 3: $p_m(3) = p_f(2) = 0.15$. $p_f(3) = \frac{1}{2}(0.15) + \frac{1}{2}(0.20) = 0.175$. $d(3) = 0.175 - 0.15 = 0.025$.

Observe the sequence of allele frequencies oscillating and converging toward 0.1667: males — 0.30, 0.10, 0.20, 0.15, 0.175, ...; females — 0.10, 0.20, 0.15, 0.175, ..., both closing in on $p_\infty = 0.1667$, with the male sequence lagging the female sequence by one generation, exactly as predicted by $p_m(t+1) = p_f(t)$.

4.6: Rate of Approach to Equilibrium

Because $|d(t)| = (1/2)^t |d(0)|$, the deviation from equilibrium is reduced by 50% every generation, and by more than 90% within about 4 generations ($(1/2)^4 = 1/16 \approx 0.0625$, i.e., 93.75% reduction). In practice, for humans (generation time roughly 25–30 years), the allele-frequency difference between the sexes at an X-linked locus becomes negligible within a century or so following any single perturbing event (migration, founder effect), which is one reason X-linked traits such as G6PD deficiency and red-green colour blindness are today observed at very similar underlying allele frequencies in both sexes within most stable, randomly mating Nigerian communities, even though the **phenotype** frequency of deficiency remains much higher in males ($\approx p$) than in females ($\approx p^2$) because of hemizyosity (Chapter 1).

Theorem 4.2: (Monotonic Convergence of X-Linked Allele Frequency to Equilibrium)

Let $x_m(t)$ and $x_f(t)$ denote the male and female allele frequencies at an X-linked locus in generation t , related by the recurrence $x_m(t+1) = x_f(t)$ and $x_f(t+1) = [x_m(t) + x_f(t)]/2$ (Section 4.3). Then the difference $d(t) = x_m(t) - x_f(t)$ satisfies $d(t+1) = -d(t)/2$, and consequently $x_m(t)$ and $x_f(t)$ both converge monotonically in magnitude of oscillation to the common equilibrium value $x^* = [x_m(0) + 2x_f(0)]/3$.

Proof. From the recurrence, $d(t+1) = x_m(t+1) - x_f(t+1) = x_f(t) - [x_m(t) + x_f(t)]/2 = [x_f(t) - x_m(t)]/2 = -d(t)/2$. Hence $|d(t)| = |d(0)|/2^t \rightarrow 0$ as $t \rightarrow \infty$, so the male and female allele frequencies converge to a common limit. Because the weighted average $w(t) = [x_m(t) + 2x_f(t)]/3$ is conserved exactly across generations (a short calculation using the recurrence shows $w(t+1) = w(t)$), the common limit must equal the initial weighted average, $w(0) = [x_m(0) + 2x_f(0)]/3$. Since $|d(t)|$ shrinks by a constant factor of $1/2$ every generation without changing sign pattern indefinitely (it alternates sign but its magnitude is strictly decreasing), convergence to x^* is geometric, confirming and sharpening the rate-of-approach result of Section 4.6.

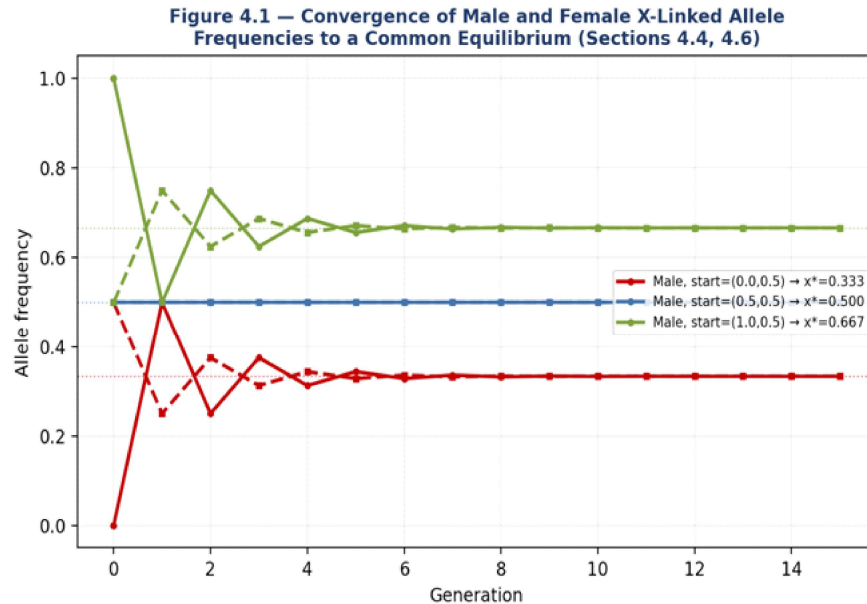


Figure 4.1: *Convergence of male and female X-linked allele frequencies to a common equilibrium (Sections 4.4, 4.6).*

4.7: Chapter Summary

- Sons inherit their single X allele from their mother; daughters inherit one X allele from each parent.
- This asymmetry produces the recurrence $p_m(t+1) = p_f(t)$, $p_f(t+1) = \frac{1}{2}p_f(t) + \frac{1}{2}p_m(t)$.
- The male–female allele–frequency difference decays geometrically as $(-1/2)^t$, oscillating in sign each generation.
- The equilibrium frequency is the $1/3 : 2/3$ weighted average of the male and female founding frequencies, reflecting the two-thirds share of X chromosomes carried by females.

4.8: Computational Corner: Simulating Approach to X-Linked Equilibrium

R.

```
# Approach to equilibrium for an X-linked locus: iterate the
# recurrence
#  $q_{\text{female}}(t+1) = 0.5 * (q_{\text{female}}(t) + q_{\text{male}}(t))$ ;  $q_{\text{male}}(t+1) = q_{\text{female}}(t)$ 
iterate_xlinked <- function(q_f0, q_m0, generations = 15) {
  q_f <- numeric(generations + 1); q_m <- numeric(generations
+ 1)
  q_ff[1] <- q_f0; q_m[1] <- q_m0
  for (t in 1:generations) {
    q_ff[t + 1] <- 0.5 * (q_ff[t] + q_m[t])
    q_m[t + 1] <- q_ff[t]}
  data.frame(generation = 0:generations, q_female = q_f, q_male
= q_m)}

res <- iterate_xlinked(q_f0 = 0.05, q_m0 = 0.25) # e.g. G6PD
deficiency allele
print(res)
plot(res$generation, res$q_male, type = "b", col = "darkred",
xlab = "Generation", ylab = "Allele frequency (q)",
main = "Approach to X-linked equilibrium (G6PD example)")
lines(res$generation, res$q_female, type = "b", col
= "steelblue")
legend("topright", legend = c("Males", "Females"), col
= c("darkred", "steelblue"), lty = 1)
```

Python.

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
def iterate_xlinked(q_f0, q_m0, generations=15):
```

```
    q_f = [q_f0]; q_m = [q_m0]
```

```
    for _ in range(generations):
```

```
        q_f.append(0.5 * (q_f[-1] + q_m[-1]))
```

```
        q_m.append(q_f[-2])
```

```
    return np.arange(generations + 1), np.array(q_f),  
    np.array(q_m)
```

```
gens, q_f, q_m = iterate_xlinked(0.05, 0.25)
```

```
plt.plot(gens, q_m, "o-", label="Males", color="darkred")
```

```
plt.plot(gens, q_f, "o-", label="Females", color="steelblue")
```

```
plt.xlabel("Generation"); plt.ylabel("Allele frequency (q)")
```

```
plt.title("Approach to X-linked equilibrium (G6PD example)")
```

```
plt.legend()
```

Both scripts trace the damped oscillatory convergence of male and female allele frequencies to their common equilibrium value derived analytically in Section 4.4, converging geometrically at rate $(\frac{1}{2})^t$ exactly as the closed-form solution predicts — a useful visual check that the algebra of Section 4.4 is correct.

Mini-Project**Mini-Project 4: Projecting X-Linked Allele Frequency Trajectories**

Use the recurrence relation derived in Section 4.4 to project how quickly an X-linked allele frequency approaches equilibrium under different starting conditions.

Step 1: Implement the male/female allele-frequency recurrence from Section 4.3 as a function of generation number.

Step 2: Run the recurrence starting from three different initial male allele frequencies (e.g., 0.0, 0.5, and 1.0) with the same initial female frequency.

Step 3: Plot all three trajectories on the same axes and confirm they converge to the same equilibrium value.

Step 4: Compute the number of generations required for each trajectory to come within 1% of equilibrium, using the rate result from Section 4.6.

Step 5: Write a short paragraph explaining why the three trajectories converge to the same equilibrium despite different starting points.

Deliverable: *A trajectory plot (R or Python), a table of generations-to-convergence for each starting condition, and a short written explanation.*

4.9: Practice Questions

- Q1.** Suppose colour-blindness allele frequency at generation 0 is $p_m(0) = 0.08$ (males) and $p_f(0) = 0.08$ (females) in an Ibadan population. What happens to $p_m(t)$ and $p_f(t)$ for all future generations? Explain briefly.
- Q2.** In a founder population, $p_m(0) = 0.05$ and $p_f(0) = 0.25$. Compute $p_m(1)$, $p_f(1)$, $p_m(2)$, $p_f(2)$ and the equilibrium frequency p_* .
- Q3.** Show algebraically that $d(t) = p_f(t) - p_m(t)$ satisfies $d(t) = (1/2)^t d(0)$, starting from the two recurrence equations in Section 4.3.
- Q4.** After how many generations will $|d(t)|$ first fall below 1% of $|d(0)|$?
- Q5.** G6PD deficiency has an X-linked allele frequency of approximately 20% among males in some Nigerian populations, reflecting historical malaria selection pressure

(Chapter 3, Section 3.7). If a founding population began with $q_{\text{male}}(0) = 0.05$ and $q_{\text{female}}(0) = 0.05$ (equal in both sexes, e.g. after a migration event introducing the allele at low frequency into a previously naive population, followed by strong male-specific selective advantage that is not modelled by the simple neutral recurrence of Section 4.3), explain qualitatively why the *neutral* approach-to-equilibrium model of this chapter would be insufficient on its own to explain the allele reaching 20%, and what additional force would need to be incorporated.

- Q6.** Using the R or Python code of Section 4.6, plot the approach to equilibrium starting from $q_{\text{female}}(0) = 0.10$, $q_{\text{male}}(0) = 0.60$ (a large initial sex difference) and report after how many generations the two curves become visually indistinguishable on the plot.