

# Homework 3 - P8149 Human Population Genetics

Yongyan Liu (yl6107)

Due: Monday, November 10, 2025

## Question 1 (3 marks)

Show that  $\bar{w} \leq 1$  (see Eq. 5.1 in your Lecture Notes). Assume the fitnesses are standardized relative to the maximum fitness.

### Solution

#### Proof:

Let  $W_{\max}$  be the maximum fitness among all genotypes. After standardization, we have:

$$w_i = \frac{W_i}{W_{\max}} \leq 1$$

The average fitness of the population is:

$$\bar{w} = \sum_i p_i w_i \leq \sum_i p_i \cdot 1 = \sum_i p_i = 1$$

Proved.

## Question 2 (1+2 = 3 marks)

The zygote-to-adult relative viabilities are 1, .5 and .25 for the  $AA$ ,  $Aa$ , and  $aa$  genotypes, respectively. Assume random mating.

- Starting with the zygotic frequency of  $A$  being .2, what are the genotype frequencies in the adult population?
- Suppose you do not know the fitnesses and you can only observe the adult population. A naive method to estimate viabilities is to use the ratios of observed genotypic frequencies to expected frequencies under Hardy-Weinberg equilibrium. Given that the observed genotype counts for  $AA$ ,  $Aa$ , and  $aa$  are 165, 562 and 339, respectively, what are the relative estimates of relative viabilities using the Hardy-Weinberg deviation technique?

## Solution

### Part (a)

Zygotic frequency of  $A$ :  $p = 0.2$ , so frequency of  $a$ :  $q = 1 - 0.2 = 0.8$

Assuming Hardy-Weinberg equilibrium:

$$P_{AA} = p^2 = 0.04, P_{Aa} = 2pq = 0.32, P_{aa} = q^2 = 0.64$$

According to the zygote-to-adult relative viabilities, let

$$T = P_{AA} * 1 + P'_{Aa} * 0.5 + P'_{aa} * 0.25 = 0.04 * 1 + 0.32 * 0.5 + 0.64 * 0.25 = 0.36$$

The adult genotype frequencies are

- $P'_{AA} = \frac{P_{AA} * 1}{T} = 0.1\bar{1}$
- $P'_{Aa} = \frac{P_{Aa} * 0.5}{T} = 0.4\bar{4}$
- $P'_{aa} = \frac{P_{aa} * 0.25}{T} = 0.4\bar{4}$ .

### Part (b)

Observed counts:  $n_{AA} = 165$ ,  $n_{Aa} = 562$ ,  $n_{aa} = 339$

Total:  $N = 165 + 562 + 339 = 1066$

So, the observed genotype frequencies in adults are

- $P'_{AA} = \frac{165}{1066} = 0.1548$
- $P'_{Aa} = \frac{562}{1066} = 0.5272$
- $P'_{aa} = \frac{339}{1066} = 0.3180$ .

And, the observed allele frequencies are  $p = P'_{AA} + \frac{P'_{Aa}}{2} = 0.4184$ ,  $q = P'_{aa} + \frac{P'_{Aa}}{2} = 0.5816$

Assuming Hardy-Weinberg equilibrium, the genotype frequencies in the zygotes are:

$$P_{AA} = p^2 = 0.1750, P_{Aa} = 2pq = 0.4867, P_{aa} = q^2 = 0.3383$$

So we can calculate the absolute viabilities:

- For AA:  $\frac{P'_{AA}}{P_{AA}} = 0.8842$
- For Aa:  $\frac{P'_{Aa}}{P_{Aa}} = 1.0833$
- For aa:  $\frac{P'_{aa}}{P_{aa}} = 0.9401$

Therefore, we can normalize them and get the relative viabilities by dividing the max absolute viabilities:

- For AA,  $\frac{0.8842}{1.0833} = 0.816$
- For Aa,  $\frac{1.0833}{1.0833} = 1$
- For aa,  $\frac{0.9401}{1.0833} = 0.868$ .

### Question 3 (2+1 = 3 marks)

The following fitnesses were estimated for humans living in a malarial environment in West Africa as a function of their genotype at the hemoglobin  $\beta$ -chain locus:

| genotype | <i>AA</i> | <i>AS</i> | <i>SS</i> | <i>AC</i> | <i>SC</i> | <i>CC</i> |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| fitness  | 0.9       | 1.0       | 0.2       | 0.9       | 0.71      | 1.31      |

- (a) Given these fitnesses and the assumption of random mating in an infinitely large population, calculate each population's average fitness and the average excess and rate of change of allele frequency for all three alleles for populations with the following gene pools:

|        | freq( <i>A</i> ) | freq( <i>S</i> ) | freq( <i>C</i> ) |
|--------|------------------|------------------|------------------|
| pop. 1 | .9               | .05              | .05              |
| pop. 2 | .85              | .05              | .10              |
| pop. 3 | .75              | .05              | .20              |

- (b) Do the initial allele frequencies affect the course of adaptive evolution in the above cases (i.e., are the signs in the changes of allele frequencies the same)?

### Solution

#### Part (a)

**Population 1** ( $p_A = 0.9$ ,  $p_S = 0.05$ ,  $p_C = 0.05$ ):

Assuming HWE:

- $P_{AA} = p_A^2 = 0.81$
- $P_{AS} = 2p_Ap_S = 0.09$
- $P_{AC} = 2p_Ap_C = 0.09$
- $P_{SS} = p_S^2 = 0.0025$
- $P_{SC} = 2p_Sp_C = 0.005$
- $P_{CC} = p_C^2 = 0.0025$

Average fitness:

$$\bar{w} = \sum_{X,Y \in \{A,S,C\}} P_{XY} w_{XY} = 0.81 \cdot 0.9 + 0.09 \cdot 1.0 + 0.09 \cdot 0.9 + 0.0025 \cdot 0.2 + 0.005 \cdot 0.71 + 0.0025 \cdot 1.31 = 0.907325$$

Average fitness of each allele:

- $\bar{w}_A = P_A w_{AA} + P_S w_{AS} + P_C w_{AC} = 0.9 \cdot 0.9 + 0.05 \cdot 1.0 + 0.05 \cdot 0.9 = 0.905$
- $\bar{w}_S = P_A w_{AS} + P_S w_{SS} + P_C w_{SC} = 0.9 \cdot 1.0 + 0.05 \cdot 0.2 + 0.05 \cdot 0.71 = 0.9455$
- $\bar{w}_C = P_A w_{AC} + P_S w_{SC} + P_C w_{CC} = 0.9 \cdot 0.9 + 0.05 \cdot 0.71 + 0.05 \cdot 1.31 = 0.911$

Average excess:

- $\alpha_A = \bar{w}_A - \bar{w} = 0.905 - 0.907325 = -0.002325$

- $\alpha_S = \bar{w}_S - \bar{w} = 0.9455 - 0.907325 = 0.038175$
- $\alpha_C = \bar{w}_C - \bar{w} = 0.911 - 0.907325 = 0.003675$

Rate of change:

- $\Delta p_A = \frac{p_A}{\bar{w}} \alpha_A = \frac{0.9}{0.907325} * -0.002325 = -0.00231$
- $\Delta p_S = \frac{p_S}{\bar{w}} \alpha_S = \frac{0.05}{0.907325} * 0.038175 = 0.00210$
- $\Delta p_C = \frac{p_C}{\bar{w}} \alpha_C = \frac{0.05}{0.907325} * 0.003675 = 0.00020$

**Population 2** ( $p_A = 0.85$ ,  $p_S = 0.05$ ,  $p_C = 0.10$ ):

Assuming HWE:

- $P_{AA} = p_A^2 = 0.7225$
- $P_{AS} = 2p_A p_S = 0.085$
- $P_{AC} = 2p_A p_C = 0.17$
- $P_{SS} = p_S^2 = 0.0025$
- $P_{SC} = 2p_S p_C = 0.01$
- $P_{CC} = p_C^2 = 0.01$

Average fitness:

$$\bar{w} = \sum_{X,Y \in \{A,S,C\}} P_{XY} * w_{XY} = 0.7225 * 0.9 + 0.085 * 1.0 + 0.17 * 0.9 + 0.0025 * 0.2 + 0.01 * 0.71 + 0.01 * 1.31 = 0.90895$$

Average fitness of each allele:

- $\bar{w}_A = P_A w_{AA} + P_S w_{AS} + P_C w_{AC} = 0.85 * 0.9 + 0.05 * 1.0 + 0.10 * 0.9 = 0.905$
- $\bar{w}_S = P_A w_{AS} + P_S w_{SS} + P_C w_{SC} = 0.85 * 1.0 + 0.05 * 0.2 + 0.10 * 0.71 = 0.931$
- $\bar{w}_C = P_A w_{AC} + P_S w_{SC} + P_C w_{CC} = 0.85 * 0.9 + 0.05 * 0.71 + 0.10 * 1.31 = 0.9315$

Average excess:

- $\alpha_A = \bar{w}_A - \bar{w} = 0.905 - 0.90895 = -0.00395$
- $\alpha_S = \bar{w}_S - \bar{w} = 0.931 - 0.90895 = 0.02205$
- $\alpha_C = \bar{w}_C - \bar{w} = 0.9315 - 0.90895 = 0.02255$

Rate of change:

- $\Delta p_A = \frac{p_A}{\bar{w}} \alpha_A = \frac{0.85}{0.90895} * -0.00395 = -0.00369$
- $\Delta p_S = \frac{p_S}{\bar{w}} \alpha_S = \frac{0.05}{0.90895} * 0.02205 = 0.00121$
- $\Delta p_C = \frac{p_C}{\bar{w}} \alpha_C = \frac{0.10}{0.90895} * 0.02255 = 0.00248$

**Population 3** ( $p_A = 0.75$ ,  $p_S = 0.05$ ,  $p_C = 0.20$ ):

Assuming HWE:

- $P_{AA} = p_A^2 = 0.5625$

- $P_{AS} = 2p_A p_S = 0.075$
- $P_{AC} = 2p_A p_C = 0.30$
- $P_{SS} = p_S^2 = 0.0025$
- $P_{SC} = 2p_S p_C = 0.02$
- $P_{CC} = p_C^2 = 0.04$

Average fitness:

$$\bar{w} = \sum_{X,Y \in \{A,S,C\}} P_{XY} w_{XY} = 0.5625 * 0.9 + 0.075 * 1.0 + 0.30 * 0.9 + 0.0025 * 0.2 + 0.02 * 0.71 + 0.04 * 1.31 = 0.91835$$

Average fitness of each allele:

- $\bar{w}_A = P_A w_{AA} + P_S w_{AS} + P_C w_{AC} = 0.75 * 0.9 + 0.05 * 1.0 + 0.20 * 0.9 = 0.905$
- $\bar{w}_S = P_A w_{AS} + P_S w_{SS} + P_C w_{SC} = 0.75 * 1.0 + 0.05 * 0.2 + 0.20 * 0.71 = 0.902$
- $\bar{w}_C = P_A w_{AC} + P_S w_{SC} + P_C w_{CC} = 0.75 * 0.9 + 0.05 * 0.71 + 0.20 * 1.31 = 0.9725$

Average excess:

- $\alpha_A = \bar{w}_A - \bar{w} = 0.905 - 0.91835 = -0.01335$
- $\alpha_S = \bar{w}_S - \bar{w} = 0.902 - 0.91835 = -0.01635$
- $\alpha_C = \bar{w}_C - \bar{w} = 0.9725 - 0.91835 = 0.05415$

Rate of change:

- $\Delta p_A = \frac{p_A}{\bar{w}} \alpha_A = \frac{0.75}{0.91835} * -0.01335 = -0.01091$
- $\Delta p_S = \frac{p_S}{\bar{w}} \alpha_S = \frac{0.05}{0.91835} * -0.01635 = -0.00089$
- $\Delta p_C = \frac{p_C}{\bar{w}} \alpha_C = \frac{0.20}{0.91835} * 0.05415 = 0.01180$

## Part (b)

Looking at the signs of  $\Delta p$  across all three populations:

- **Allele A:** negative  $\Delta p$  in all populations (consistently decreasing)
- **Allele S:** positive  $\Delta p$  in pop. 1 and pop. 2, but negative  $\Delta p$  in pop. 3
- **Allele C:** positive  $\Delta p$  in all populations (consistently increasing)

The initial allele frequencies affect the course of adaptive evolution. While alleles A and C show consistent directional changes (A decreasing, C increasing), allele S shows frequency-dependent behavior: it increases in populations 1 and 2 but decreases in population 3.

## Question 4 (2 marks)

Assume that a typical gene in the human genome has a mutation rate of  $2 \times 10^{-6}$  per gene per cell division. If one generation involves an average of 36 cell divisions, what is the mutation rate per gene per generation?

## Solution

For small mutation rates, the mutation rate per generation is approximately:

$$\mu_{\text{gen}} \approx n \cdot \mu_{\text{cell}} = 36 \times 2 \times 10^{-6} = 72 \times 10^{-6} = 7.2 \times 10^{-5}$$

## Question 5 (2 marks)

For the two-way neutral mutation model, show that  $\hat{q} \approx 0.91$  for any values of  $u$  and  $v$  as long as  $u = 10v$ . What is the equilibrium value if  $u = 100v$ ?

## Solution

In the two-way neutral mutation model, at equilibrium, the frequency of allele  $a$  is:  $\hat{q} = \frac{u}{u+v}$

**When  $u = 10v$ :**

$$\hat{q} = \frac{10v}{10v+v} = \frac{10v}{11v} = \frac{10}{11} \approx 0.91$$

This result is independent of the actual values of  $u$  and  $v$ ; it only depends on their ratio.

**When  $u = 100v$ :**

$$\hat{q} = \frac{100v}{100v+v} = \frac{100v}{101v} = \frac{100}{101} \approx 0.99$$

## Question 6 (1+2 = 3 marks)

A population of 20 individuals were genotyped at a particular locus. Calculate the observed and expected heterozygosities in each of the following situations:

- (a) The locus had only two alleles,  $A_1$  and  $A_2$ ; 10 individuals were  $A_1A_1$  and 10 were  $A_1A_2$ .
- (b) All 20 individuals were heterozygous at the locus and each allele in the heterozygotes was present only once.

## Solution

**Part (a)**

**Observed heterozygosity:**

$$H_{\text{obs}} = \frac{\text{number of heterozygotes}}{\text{total individuals}} = \frac{10}{20} = 0.5$$

**Expected heterozygosity:**

$$\text{Allele frequencies: } p_1 = \frac{2(10) + 10}{40} = \frac{30}{40} = 0.75, p_2 = \frac{10 + 0}{40} = \frac{10}{40} = 0.25$$

$$H_{\text{exp}} = 2p_1p_2 = 2(0.75)(0.25) = 0.375$$

**Part (b)**

**Observed heterozygosity:**

$$H_{\text{obs}} = \frac{20}{20} = 1.0$$

**Expected heterozygosity:**

Since each allele appears exactly once, there are  $k = 40$  unique alleles, each with frequency  $p_i = \frac{1}{40}$ .

$$H_{\text{exp}} = 1 - \sum_{i=1}^{40} p_i^2 = 1 - 40 \times \left(\frac{1}{40}\right)^2 = 1 - \frac{40}{1600} = 1 - 0.025 = 0.975$$

From this example, we can see that the observed heterozygosity and the expected heterozygosity can be very different.

**Question 7 (1+1+2 = 4 marks)**

The following 5 DNA sequences were sequenced at 15 nucleotide sites.

```
A A T C G A G A C T T T A G C
A T T C C A G A T T T A A G C
A T T C C A G A T T T A A G C
A A T C G A G A C T T T A G C
T A T C G A G A C T A T C C C
```

- (a) What are the number of segregating sites in the sample?
- (b) Estimate the nucleotide polymorphism in the sample.
- (c) Estimate the nucleotide diversity in the sample.

**Solution**

**Part (a): Number of segregating sites**

Sites with variation (from left to right): No. 1, 2, 5, 9, 11, 12, 13, 14

**Answer:**  $S = 8$  segregating sites.

**Part (b): Nucleotide polymorphism**

Nucleotide polymorphism:  $S_n^* = \frac{S_n}{B} = \frac{8}{15} = 0.53\bar{3}$

Watterson's estimator:

$$\theta = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}} = \frac{8}{\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}} = 3.84$$

**Part (c): Nucleotide diversity**

Number of differences for each nucleotide sides from left to right are 4 6 0 0 6 0 0 0 6 0 4 6 4 4 0, and the total is 40.

Number of pairwise comparisons:  $\binom{5}{2} = 10$

$$\hat{\theta}_{\pi} = \frac{40}{10} = 4.0$$