

SOLUTION TO
AN INTRODUCTION TO POPULATION GENETICS
THEORY

Author:

James F. Crow

Motoo Kimura

Solution by

Hanbin Lee

Contents

1	Models of Population Growth	2
2	Randomly Mating Populations	6

Chapter 1

Models of Population Growth

Problem 1.1. *In a population with discrete generations and with fitness w , how many generations are required to double the population number?*

Proof. By the definition of fitness in discrete generations, w is given by $w = \frac{N_{i+1}}{N_i}$. This gives $N_i = w^i N_0$. Since $N_i = 2N_0$ by the given condition, $i = \log 2 / \log w$. \square

Problem 1.2. *How long is required for the population to double with model 2?*

Proof. By the definition of fitness in continuous generations, w is given by $w = \frac{1}{N} \frac{dN}{dt}$. This gives $\frac{1}{N} dN = w dt$ followed by $\int_{N_0}^N \frac{1}{N} dN = \int_{t_0}^t w dt$. Therefore, $\log 2 = \log \frac{N}{N_0} = w(t - t_0)$. \square

Problem 1.3. *A population under model 3 has reached age stability. How long, in units of λ , will be required for the population to double? What is the effective generation length, defined as the unit that will give the same answer as problem 1?*

Proof. Assume that, like the textbook, every individual lives for 5 years.

Let $n_t = (n_{0t}, n_{1t}, \dots, n_{4t})^T$, (b_0, b_1, \dots, b_4) and (p_0, \dots, p_3) denote the number of individuals in each age, reproduction rates of each age and probability of survival of each age respectively. Then the following matrix equation holds:

$$n_t = \begin{pmatrix} n_{0,t} \\ n_{1,t} \\ n_{2,t} \\ n_{3,t} \\ n_{4,t} \end{pmatrix} = \begin{pmatrix} b_0 & b_1 & b_2 & b_3 & b_4 \\ p_0 & 0 & 0 & 0 & 0 \\ 0 & p_1 & 0 & 0 & 0 \\ 0 & 0 & p_2 & 0 & 0 \\ 0 & 0 & 0 & p_3 & 0 \end{pmatrix} \begin{pmatrix} n_{0,t-1} \\ n_{1,t-1} \\ n_{2,t-1} \\ n_{3,t-1} \\ n_{4,t-1} \end{pmatrix} = A \cdot n_{t-1}$$

Then, by induction, we obtain $n_t = A^t n_0$. To compute the power of the matrix A , we compute the characteristic polynomial and assume that it has at least one zero:

$$\det(A - \lambda I) = \lambda^5 - b_0 \lambda^4 - p_0 b_1 \lambda^3 - p_0 p_1 b_2 \lambda^2 - p_0 p_1 p_2 b_3 \lambda - p_0 p_1 p_2 p_3 b_4$$

Under suitable conditions, the polynomial above has a single positive largest (in terms of absolute value) zero λ^* and a corresponding eigenvector l^* .

For a large t , $(l^*)^t$ dominates all other power of eigenvalues, so $M^t n_0$, which is expressed as a linear combination of power of eigenvalues, can be approximated by

$$M^t n_0 \approx C M^t l = C(\lambda^*)^t l$$

where C is a constant determined by the initial condition.

N_t , the size of the population at time t , is given as the sum of entries of n_t . Therefore, $N_t = n_0 + n_1 + \dots + n_4 = C'(\lambda^*)^t$ by the above equation. Hence, using $N_t = 2N_0$, we get $t = \log 2 / \log \lambda$. \square

Problem 1.4. Suppose you know the age-specific death rates (the probability that an individual of age x will die during the next time unit). What is the life expectancy, that is, the mean length of life? What is the median length of life?

Proof. Let $\{p_i\}_{i \in \mathbb{Z}_{\geq 0}}$ be the probability that an individual will survive during age i . Then the probability of survival until age t is given as the product of p_i from 0 to $t - 1$. Thus, the expectation is given by

$$E(X) = \sum_{k=0}^{\infty} \left(k \cdot \prod_{i=0}^k p_i \right)$$

The median value can also be computed in a similar manner using formulas from basic probability theory. \square

Problem 1.5. Show that equation 1.6.8a is correct for any number of strains.

Proof. Suppose that there are k strains and n_1, \dots, n_k individuals for each strain. The Malthusian parameters of each strain is given as r_1, \dots, r_k .

Now we use the same method in the textbook.

$$\begin{aligned} \frac{d \ln(p_1/(1-p_1))}{dt} &= \frac{d \ln(n_1/(N-n_1))}{dt} \\ &= \frac{dn_1}{dt} - \frac{d(N-n_1)}{dt} \\ &= \frac{d \ln n_1}{n_1 dt} - \frac{d \ln(N-n_1)}{(N-n_1) dt} \\ &= r_1 - \bar{r}_1 \end{aligned}$$

Here, \bar{r}_i denotes the mean parameter except for the i -th strain.

Notice also that

$$\begin{aligned}\frac{d \ln(p_1/(1-p_1))}{dt} &= \frac{\ln p_1}{dt} - \frac{\ln(1-p_1)}{dt} \\ &= \frac{dp_1}{p_1 dt} - \frac{d(1-p_1)}{(1-p_1)dt} \\ &= \frac{dp_1}{p_1(1-p_1)dt}\end{aligned}$$

Putting these two equations together gives

$$\begin{aligned}\frac{dp_1}{dt} &= (r_1 - \bar{r}_1)p_1(1-p_1) \\ &= ((N-n_1)r_1 - (n_2r_2 + \dots + n_kr_k)) \cdot p_1 \cdot \left(\frac{1}{N-n_1}\right) \cdot (1-p_1) \\ &= (Nr_1 - (n_1r_1 + \dots + n_kr_k)) \cdot p_1 \cdot \left(\frac{1}{N-n_1}\right) \cdot \left(\frac{N-n_1}{N}\right) \\ &= N(r_1 - \bar{r}) \cdot p_1 \cdot \left(\frac{1}{N-n_1}\right) \cdot \left(\frac{N-n_1}{N}\right) \\ &= p_1(r - \bar{r})\end{aligned}$$

□

Problem 1.6. *What are the median and mean length of life under model 2, expressed in terms of the death rate, d ?*

Proof. Suppose that N_0 individuals were born at a given time t_0 . Then the following equation holds:

$$\frac{dN_t}{dt} = -dN_t$$

The solution to this differential equation is

$$N = N_0 e^{-dt}$$

Thus, the number of dead individuals are

$$N_0 - N = N_0(1 - e^{-dt})$$

Therefore, the cumulative distribution function of survival time t is given as

$$F(t) = 1 - e^{-dt}$$

The resulting probability density function $f(t)$ is

$$f(t) = de^{-dt}$$

Hence, the expectation can be computed.

$$\int_0^\infty t \cdot de^{-dt} dt = \frac{1}{d}$$

□

Problem 1.7. Show that the time required to change the number from N_0 to N_t in a logistically growing population exceeds that in an unregulated population with the same intrinsic rate of increase by $\log[(K - N_0)/(K - N_t)]/r$.

Proof. By equation (1.6.3) given in the book,

$$t_l = \frac{1}{r} \log \frac{N_t(K - N_0)}{(K - N_t)N_0}$$

for logistically a growing population.

For a exponentially growing populations,

$$t_e = \frac{1}{r} \log \frac{N_t}{N_0}$$

Substracting t_e from t_l , we have

$$t_l - t_e = \frac{1}{r} \log \frac{K - N_0}{K - N_t}$$

□

Problem 1.8. Again considering a logistic population with carrying capacity K , what is the time required to go from a fraction x to a fraction y of this capacity?

Proof. Simply put $N_t = yK$ and $N_0 = xK$ to (1.6.3) of textbook.

$$\begin{aligned} t &= \frac{1}{r} \log \frac{yK(K - xK)}{(K - yK)xK} \\ &= \frac{1}{r} \log \frac{y(1 - x)}{(1 - y)x} \end{aligned}$$

□

Problem 1.9. One bacterium which reproduces by fission and follows a logistic growth pattern is introduced into each of several ponds. Show that the time required to fill a pond to half its capacity is proportional to the log of the carrying capacity.

Proof. Simply put $N_t = (1/2)K$ to (1.6.3) of textbook.

$$\begin{aligned} t &= \frac{1}{r} \log \frac{(1/2) \cdot K(K - 1)}{(K - (1/2)K) \cdot 1} \\ &\approx \frac{1}{r} \log K \end{aligned}$$

□

Chapter 2

Randomly Mating Populations

In all problems, unless the contrary is stated, assume random mating.

Problem 2.1. *In a population there are 8 times as many heterozygotes as homozygous recessives. What is the frequency of the recessive gene?*

Proof. Let p and q be the frequency of each allele. Then, $2pq = 8 \cdot q^2 \Rightarrow p = 4q$ so $q = \frac{1}{5}$. \square

Problem 2.2. *Show that, for a very rare recessive gene, the proportion of heterozygous carriers is approximately twice the frequency of the recessive gene.*

Proof. The proportion of heterozygotes is $2pq$. The frequency of recessive gene is q . Then the ratio of these two values is $\frac{2pq}{q} = 2p \approx 2$. \square

Problem 2.3. *If 16% of the population are Rh- (dd), what fraction of the Rh+ population (DD and Dd) are homozygous?*

Proof. We obtain $q = \sqrt{16/100} = 0.04$. The value that we are computing is given by

$$\frac{p^2}{p^2 + 2pq} = \frac{p}{p + 2q} = \frac{0.96}{0.96 + 2 \cdot 0.04} \approx 0.92$$

\square

Problem 2.4. *From the data in problem 3, what fraction of the children from a large group of families where both parents were Rh+ would be expected to be Rh+?*

Proof. The frequency of gamete with allele d is $\frac{2pq}{2pq + 2pq + 2p^2} = \frac{pq}{2pq + p^2} \approx 0.038$. Then the frequency of Rh+ progeny is $1 - 0.038^2 = 0.996$. \square

Problem 2.5. *Show that if the A and B antigens of the ABO blood group system were caused by two dominant genes, independently inherited, the product of the frequency of A and B should equal the product of O and AB.*

Proof. Let a_1, a_2 be the frequency of allele A, a and b_1, b_2 be the frequency of allele B, b . Then the frequencies of A and B are $(a_1^2 + 2a_1a_2) \cdot b_2^2$ and $(b_1^2 + 2b_1b_2) \cdot a_2^2$. The frequencies of O and AB are $a_2^2 \cdot b_2^2$ and $(a_1^2 + 2a_1a_2) \cdot (b_1^2 + 2b_1b_2)$. Therefore, the products of these values are equal. \square

Problem 2.6. *What is the maximum proportion of heterozygotes with two alleles? With three alleles? With n alleles?*

Proof. **i) Biallelic** This is a simple example of the inequality of arithmetic and geometric means. Under $p + q = 1$, the maximum value of $2pq$ is $1/4$ since $2\sqrt{pq} \leq p + q$.

ii) Triallelic We need to compute the maximum value of $1 - (p^2 + q^2 + r^2)$ under the constraint $p + q + r = 1$. Let $f(p, q, r) = 1 - (p^2 + q^2 + r^2)$ and $g(p, q, r) = p + q + r$. Then we have to attain the maximum of $f(p, q, r)$ under $g(p, q, r) = 1$. We proceed with the Lagrange multiplier criterion.

$$\nabla f = \lambda \nabla g \Leftrightarrow (-2) \cdot (p, q, r) = \lambda \cdot (1, 1, 1)$$

Now we have $(p, q, r) = -\frac{\lambda}{2}(1, 1, 1)$. Substituting these values to $p + q + r = 1$ yields $-\frac{3\lambda}{2} = 1$ resulting $\lambda = -\frac{2}{3}$ so $p = q = r = \frac{1}{3}$. Thus, we have $\frac{2}{3}$ as the maximum value.

iii) n -allelic The proof is identical to that of case ii) so we omit the details. \square