Exercises for

"Introductory seminar on mathematical population genetics"

University of Vienna, WS 2020

Reinhard Bürger

1. The ABO blood group system can be described by three alleles, A, B, and O, which occur at a (single) gene locus. This gene (locus) encodes a *glycosyltransferase*, i.e., an enzyme that modifies the carbohydrate content of the red blood cell antigens. A and B are both dominant over O, and A and B are codominant, i.e., the phenotype (enzyme) of AB individuals differs from the A, B, and also O blood types.

   Denote the three allele frequencies by p_A, p_B, p_O, and compute the frequencies of the blood types A, B, AB, and O (assume Hardy-Weinberg proportions; explain why are there only these four blood types). In addition, assume that R_A, R_B, R_{AB}, and R_O are the frequencies of the blood types observed in a population. Calculate the allele frequencies. Is this possible without assuming Hardy-Weinberg proportions? How could one infer (in reality) if the population indeed is in Hardy-Weinberg proportions?

2. Prove the Hardy-Weinberg law for k alleles if random mating is assumed (and not just random union of gametes). In particular, compute P_{ij} for arbitrary i and j (see the instruction in the lecture notes).

3. Derive the Hardy-Weinberg law for a diallelic X-linked gene (see the instruction in the lecture notes).

4. Solve the linear fractional transformation

$$p' = \frac{\alpha + \beta p}{\gamma + \delta p}.$$

   Instruction:

   a) Start with the simple cases $\delta = 0$ and $\alpha\delta = \beta\gamma$.

b) Next, assume $\delta \neq 0$ and $\alpha\delta \neq \beta\gamma$ and show that two solutions of $p' = p$ are given by

$$p_\pm = (2\delta)^{-1}(\beta - \gamma \pm Q^{1/2}) \ ,$$

where $Q = (\beta - \gamma)^2 + 4\alpha\delta$. Consider only the case $Q > 0$, in which there are two distinct real solutions. Then $y = \frac{p-p_+}{p-p_-}$ satisfies the recursion $y' = \lambda y$ with

$$\lambda = \frac{\beta + \gamma - Q^{1/2}}{\beta + \gamma + Q^{1/2}}.$$

Now you can calculate $y(t)$ and then $p(t)$ as a functions of $y(0)$ and $p(0)$, respectively.

Show in addition:

c) If $\beta + \gamma > 0$ then $p(t) \to p_+$, $t \to \infty$.

d) If $\beta + \gamma < 0$ then $p(t) \to p_-$, $t \to \infty$.

e) If $\beta + \gamma = 0$ then $p(t)$ alternates between $p(0)$ and $p(1)$.

5. There are traits and genes with respect to which mating is generally not at random. If individuals with similar traits or the same genes mate with higher probability than at random, this is called assortative mating. Study the following instructive model.

   Assume that the genotypes $\mathcal{A}_1\mathcal{A}_1$ and $\mathcal{A}_1\mathcal{A}_2$ have the same phenotype, whereas $\mathcal{A}_2\mathcal{A}_2$ has a different one. Thus, $\mathcal{A}_1$ is (completely) dominant. Let ρ denote the fraction of individuals mating assortatively and assume that this fraction is the same for both phenotypes. Denote the (ordered) genotype frequencies by P_{ij} and the allele frequencies of $\mathcal{A}_1$ and $\mathcal{A}_2$ by p and $q = 1 - p$, respectively. Then the contribution of the random-mating group of individuals to the progeny of type $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$ will be $(1-\rho)p^2$, $(1-\rho)2pq$, and $(1-\rho)q^2$, respectively (explain this!).

   a) Compute the contribution of the group of assortatively mating individuals with genotype $\mathcal{A}_2\mathcal{A}_2$ to the progeny of type $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$.

   b) Compute the contribution of the group of assortatively mating individuals with the dominant phenotype to progeny of type $\mathcal{A}_1\mathcal{A}_1$, $\mathcal{A}_1\mathcal{A}_2$, and $\mathcal{A}_2\mathcal{A}_2$.

   c) Derive the recursion relations for P_{11}, P_{12}, and P_{22}. Conclude that the allele frequencies p and q remain constant.

d) We define the heterozygosity by $H = 2P_{12}$. Show (from c) that for complete assortative mating ($\rho = 1$), the heterozygosity $H = 2P_{12}$ evolves according to

$$H' = H\frac{p}{1 - P_{22}} = H\frac{p}{p + H/2} \ .$$

Thus, H approaches 0 as time increases, but very slowly. For instance, if $p(t = 0) = \frac{1}{2}$, then $H(t = 0) = \frac{1}{2}$, and the heterozygosity in generation n is $1/(n + 2)$.

e) If $\rho < 1$, the heterozygosity still decreases, but reaches a positive equilibrium value (you may use Exercise 4 to compute the limit $\lim_{t\to\infty} H(t)$).

Obviously, Hardy-Weinberg proportions will neither be maintained (unless the population is monomorphic) nor approached.

6. Give a direct proof (not using the general diploid result) of the following special case of Fisher's Fundamental Theorem for the haploid selection equation (eq. (3.7) in the LN):

$$\Delta\overline{W} = \sigma_A^2/\overline{W} \ ,$$

where $\sigma_A^2 = \sum_i p_i(v_i - \bar{v})^2$ is the variance in fitness.

7. Determine the linear stability of all possible equilibria in the diploid two-allele selection model (see Sect. 3.3 in the LN). Assume $0 \le s < 1$ and $hs < 1$. It is recommended to consider the three cases of the lecture and the additional case $s = 0$.

8. Study the diploid 3-allele system with the following fitness matrix:

$$\mathsf{W} = \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} 1 & 3 & 2 \\ 3 & 1 & 3 \\ 2 & 3 & 1 \end{pmatrix} \end{array} \begin{array}{c} A_1 \\ A_2 \\ A_3 \end{array}$$

a) Determine all equilibria.

b) Determine the stability properties of all equilibria (at least linear or asymptotic stability).

c) Which equilibrium is globally asymptotically stable?

If you need background on dynamical system theory, then read Section A.1 in the Appendix of the LN, especially Theorem A.4. It describes the essence of a linear stability analysis.

9. Let W_{ij}, W_i, and $\overline{W}$ denote the fitness of genotype A_iA_j, the marginal fitness of allele A_i, and the mean fitness of the population, respectively. As in the lecture notes, define the additive, the dominance, and the (total) genetic variance by $\sigma_{\mathrm{A}}^2 = 2\sum_i p_i(W_i - \overline{W})^2$, $\sigma_{\mathrm{D}}^2 = \sum_{i,j} p_i p_j (W_{ij} - W_i - W_j + \overline{W})^2$, $\sigma_{\mathrm{G}}^2 = \sum_{i,j} p_i p_j (W_{ij} - \overline{W})^2$, respectively.

   a) Show that $\sigma_{\mathrm{G}}^2 = \sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{D}}^2$.

   b) Assume the diallelic selection model with fitnesses $W_{11} = 1$, $W_{12} = 1 - hs$, $W_{22} = 1 - s$ (Section 3.3 in the lecture notes). Derive explicit expressions for σ_{A}^2, σ_{D}^2, and σ_{G}^2.

   c) Investigate when $\Delta p = \sigma_{\mathrm{A}}^2/\overline{W}$ holds and compare the result with eqs. (3.12) in the lecture notes.

10. Study Remark 3.17 in the lecture notes about the generalized gradient system representation of the continuous-time selection equation. Derive equations (3.43), (3.44), (3.45), and (3.46).

11. Derive the differential equation for mutation and selection in a diploid population,

$$\dot{p}_i = p_i(m_i - \bar{m}) + \sum_j (p_j \tilde{\mu}_{ji} - p_i \tilde{\mu}_{ij}) \text{ for every } i$$

    (eq. 4.10 in the lecture notes) from the corresponding difference equation (eq. 4.9) by assuming that selection and mutation are weak. More precisely, assume $W_{ij} = 1 + \epsilon m_{ij}$ and $\mu_{ij} = \epsilon \tilde{\mu}_{ij}$, rescale time according to $\tau = \epsilon t$, let $\epsilon \to 0$ and proceed in a similar way as in the derivation of the continuous-time selection equation.

12. Consider the diallelic mutation-selection equation in discrete time with $\nu = 0$ (see eq. 4.14) and derive the equilibrium approximations (4.18) and (4.15b) in the lecture notes (the latter for $0 \le h \ll \sqrt{\mu/s}$).

13. Consider the diallelic mutation-selection equation in discrete time with $h = \frac{1}{2}$, $\mu > 0$ and $\nu > 0$.

    a) Prove that there is a uniquely determined equilibrium, and it is polymorphic.

    b) Prove that this equilibrium is globally asymptotically stable and show that convergence is monotone (hint: study $p' - p$).

14. Consider the mutation-selection model in continuous time with k alleles (eq. 4.10 in the lecture notes). Assume that the mutation rates satisfy the so-called

'house-of-cards' condition, i.e., there exist constants μ_j such that $\mu_{ij} = \mu_j$ for every pair $i \neq j$. In other words, mutation is parent independent. Prove that $V(p) = \frac{1}{2}\bar{m} + \sum_i \mu_i \log p_i$ is a Lyapunov function. (Hint: Write the differential equation in the form $\dot{p}_i = p_i(f_i - \bar{f})$ for suitable f_i.)