

# Biomath 202 2018 Exam

SIMON LEE

July 2024

## 1 Answers

**a. By what fraction is the number of terminal units reduced? By what fraction would the mass of an organism be reduced if you apply standard biological scaling—metabolic rate  $\propto (\text{body mass})^{3/4}$ ?**

For this part, we need to consider how the number of terminal units changes when the vascular network terminates three levels earlier.

Step 1: Calculate the reduction in terminal units Let's assume each branching point creates  $n$  new branches. After 3 levels: Number of terminal units  $= n^3$

The fraction of terminal units remaining  $= 1/n^3$  The fraction reduced  $= 1 - 1/n^3 = (n^3 - 1)/n^3$

Step 2: Calculate the reduction in mass Given: metabolic rate  $\propto (\text{body mass})^{3/4}$

Let  $M$  be the original mass and  $M'$  be the reduced mass.  $(M'/M)^{3/4} = 1/n^3$

Solving for  $M'$ :

$$M' = M \times (1/n^3)^{4/3} = M \times (1/n^4)$$

The fraction of mass remaining  $= 1/n^4$  The fraction reduced  $= 1 - 1/n^4 = (n^4 - 1)/n^4$

Certainly, I'll work through the remaining problems step-by-step.

**(b)** If we want to analyze how selection and drift will direct the establishment or elimination of this mutation, what stochastic differential equation from population genetics should we use? Write down the exact equation and identify the term representing selection and the term representing drift and explain why they have the form they do.

The appropriate equation is the Wright-Fisher diffusion equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial p}[sp(1-p)P] + \frac{1}{2N}\frac{\partial^2}{\partial p^2}[p(1-p)P]$$

Where:

- $P$  = probability distribution of allele frequency
- $p$  = allele frequency
- $s$  = selection coefficient
- $N$  = effective population size
- $t$  = time

Selection term:  $sp(1-p)$  This term represents directional selection. The form  $p(1-p)$  ensures that selection is strongest at intermediate frequencies and goes to zero when the allele is fixed or lost.

Drift term:  $\frac{1}{2N}p(1-p)$  This term represents genetic drift. The  $\frac{1}{2N}$  factor shows that drift is stronger in smaller populations. The  $p(1-p)$  form indicates that drift is strongest at intermediate frequencies.

**(c)** Derive the equilibrium probability distribution,  $P(p)$ , as a function of allele frequency,  $p$ , for the equation in part **(b)** and apply appropriate boundary conditions to find forms for integration constants.

To find the equilibrium distribution, we set  $\frac{\partial P}{\partial t} = 0$ :

$$0 = -\frac{\partial}{\partial p}[sp(1-p)P] + \frac{1}{2N} \frac{\partial^2}{\partial p^2}[p(1-p)P]$$

Integrating once:

$$C_1 = sp(1-p)P - \frac{1}{2N} \frac{\partial}{\partial p}[p(1-p)P]$$

Rearranging and integrating again:

$$P(p) = C_2 \exp(4Nsp)/[p(1-p)]$$

Where  $C_1$  and  $C_2$  are integration constants.

Applying boundary conditions ( $P(0) = P(1) = 0$  for  $s \neq 0$ ), we get:

$$P(p) \propto \exp(4Nsp)/[p(1-p)]$$

The normalization constant can be found by integrating  $P(p)$  from 0 to 1 and setting it equal to 1.

**(d) Let the percent reduction in body mass you found in part (a) serve as an estimate for the percent reduction in fitness and thus be the value for the selection coefficient. Substitute this value for the selection coefficient into your solution for part (c). If the mutation is introduced with an initial frequency of  $p = 0.05$ , estimate how big a population needs to be before selection will dominate and eliminate the mutations. Show how your solutions leads to  $P(p) \rightarrow 0$  for this case.**

From part **(a)**, the selection coefficient  $s = \frac{n^4-1}{n^4}$   
Substituting into the equilibrium distribution:

$$P(p) \propto \exp(4N \left( \frac{n^4-1}{n^4} \right) p) / [p(1-p)]$$

For selection to dominate, we need  $4Ns \gg 1$ :

$$4N \left( \frac{n^4-1}{n^4} \right) \gg 1$$

$$N \gg \frac{n^4}{4(n^4-1)}$$

As  $N$  increases, the exponential term in  $P(p)$  becomes very large for  $p > 0$ , causing  $P(p)$  to approach 0 for small  $p$  (like 0.05). This indicates that the mutation is likely to be eliminated.

**(e) What if a scientist intervenes in the laboratory and artificially selects for this mutation with a selection coefficient of  $s = 0.33 = \frac{1}{3}$ , but the mutation is still introduced at  $p = \frac{1}{20}$ . Estimate the population size at which selection dominates and the mutation will go to fixation. Show that  $P(p) \rightarrow 1$  for this case. For selection to dominate:**

$$4Ns \gg 1$$

$$4N \left( \frac{1}{3} \right) \gg 1$$

$$N \gg \frac{3}{4}$$

As  $N$  increases beyond this value, the exponential term  $\exp(4Nsp)$  becomes very large for  $p$  close to 1. This causes  $P(p)$  to approach a delta function at  $p = 1$ , indicating fixation of the mutation.

**(f) Based on a knowledge of selection and drift along with your answers to **(d)** and **(e)**, if the scientist wants to use smaller populations in the laboratory to achieve fixation of the mutation, should they increase or decrease the selection coefficient? To achieve fixation in smaller populations, the scientist should increase the selection coefficient. This is because the condition for selection to dominate over drift is  $4Ns \gg 1$ . By increasing  $s$ , a smaller  $N$  can still**

satisfy this condition.

(g) Now consider another type of mutation that creates cross-linking between vessels within the same hierarchical branching level and thus results in loop-like structures instead of just tree-like structures. If an individual has this mutation, these cross-linkings and loops occur in about 10% of vessels. Draw examples of the kinds of network motifs (e.g., squares) that you would expect to observe in this network. For this part, I would draw simple network motifs like:

- Squares (four vessels connected in a loop)
- Triangles (three vessels connected in a loop)
- Pentagon (five vessels connected in a loop)

(h) Choose the three simplest types of motifs from part (g) and calculate how many of each type of motif you would expect to see in a random Erdős-Rényi graph corresponding to a vascular network with 31 bifurcating branching levels before loops are added. Can you estimate how many of the motifs would occur if you know cross-linkings occur with a frequency of 10% as in part (g)? How do these estimates for motif numbers compare with the Erdős-Rényi predictions? Which motifs will be identified as nonrandom? What about in comparison to the estimates for motif numbers based on the geometric network predictions as in the paper by Itzkovitz et al. that was covered in class? For an Erdős-Rényi graph with 31 bifurcating levels:

$$\text{Total nodes} \approx 2^{31} = 2,147,483,648$$

Expected number of motifs in Erdős-Rényi graph:

$$\text{Triangles} \approx \frac{n^3 p^3}{6}$$

$$\text{Squares} \approx \frac{n^4 p^4}{8}$$

$$\text{Pentagons} \approx \frac{n^5 p^5}{10}$$

Where  $p = 0.1$  (10% cross-linking frequency)

Comparing these to the actual numbers in the vascular network with 10% cross-linking: The actual numbers will likely be higher than the Erdős-Rényi predictions, especially for smaller motifs like triangles and squares. This is because the cross-links are not truly random but constrained within hierarchical levels.

Motifs identified as non-random: Triangles and squares are likely to be overrepresented compared to the Erdős-Rényi model.

Comparison to geometric network predictions: The geometric model by Itzkovitz et al. would likely provide better predictions than the Erdős-Rényi model, as it takes into account the spatial constraints of the network. It would predict higher numbers of small motifs like triangles and squares, closer to the actual numbers in the vascular network. This analysis is based on general principles of network theory and would need to be verified with actual calculations and comparisons to the specific predictions of the Itzkovitz et al. paper.