Biomath 202 2016 Exam

SIMON LEE

1 Answers

a. Let n_A be the number of individuals in population A and n_B be the number in population B after G generations. The relative abundance is then:

$$\frac{n_B}{n_A} = \frac{n_B}{n_A}$$

b. The power law for relative abundance simplifies to:

$$\frac{n_B}{n_A} = \left(\frac{\sigma}{r}\right)^G$$

where σ is the survivorship of type B ($1/2$ in this case) and r is the reproductive output ratio of B to A (2 in this case).

c. After 1 generation with $\sigma = 4$ and $r = 2$:

$$\frac{n_B}{n_A} = \left(\frac{4}{2}\right)^1$$
$$= 2^1 = 2$$

After 10 generations with $\sigma = 16$ and $r = 10$:

$$\frac{n_B}{n_A} = \left(\frac{16}{10}\right)^{10}$$
$$= 1.6^{10} \approx 104.9$$

For the relative mean fitness, type B starts with an advantage but loses it over time to type A. The exact values depend on the initial population sizes which are not given.

d. The key is whether $\frac{\sigma}{r} < 1$ or > 1. If $\frac{\sigma}{r} < 1$, then $(\frac{\sigma}{r})^G \to 0$ as G increases, so type A will dominate. If $\frac{\sigma}{r} > 1$, then $(\frac{\sigma}{r})^G \to \infty$ as G increases, so type B will dominate. Fixation occurs for the dominant type at large timescales. The result only holds if the infection persists indefinitely. If it only lasts a few generations, the less fit type could still persist in the population.

e. Initially:

$$n_A = 100$$
$$n_B = 50$$
$$\sigma = 1.2$$
$$r = 1.2$$

Other effects like mutations, genetic drift, population size changes, etc. could influence the evolutionary dynamics beyond just relative fitness. These factors could be equally important for both A and B, but it depends on the specifics (no definitive determination can be made without more info). If $r \neq 100$, then the initial reproductive advantage of B is not actually 2 as stated earlier.

f. The most useful model would capture the key dynamics in a simple form. One option is:

$$\frac{dn_B}{dn_A} = \left(\frac{\sigma}{r}\right)^G \left(1 + \sum_i c_i E_i\right)$$

where E_i are additional effects and c_i are coefficients for their relative importance. The sum represents the combination of other evolutionary factors.

(g) To solve part g, we start from the equation given:

$$\frac{dN(t)}{dt} = \tilde{r}(t)N(t)$$

where $\tilde{r}(t) = g(r, \sigma, t)$ and the intrinsic rate of population growth is time dependent.
Solving this equation:

$$\frac{dN(t)}{dt} = \tilde{r}(t)N(t)$$

$$\frac{dN(t)}{N(t)} = \tilde{r}(t)dt$$

$$\int \frac{dN(t)}{N(t)} = \int \tilde{r}(t)dt$$

$$\ln(N(t)) = \int \tilde{r}(t)dt + C$$

$$N(t) = e^{\int \tilde{r}(t)dt + C}$$

$$N(t) = e^{\int \tilde{r}(t)dt} e^C$$

$$N(t) = N(0)e^{\int \tilde{r}(t)dt}$$

In the last step, we used the initial condition $N(0)$ to determine the constant of integration $e^C = N(0)$. Therefore, the exact expression for $\tilde{r}(t)$ is:

$$\tilde{r}(t) = g(r, \sigma, t)$$

This matches the answer given in part g.
Now, let's solve the equation for constant r:

$$\frac{dN(t)}{dt} = rN(t)$$

$$\frac{dN(t)}{N(t)} = rdt$$

$$\int \frac{dN(t)}{N(t)} = \int rdt$$

$$\ln(N(t)) = rt + C$$

$$N(t) = e^{rt + C}$$

$$N(t) = e^{rt} e^C$$

$$N(t) = N(0)e^{rt}$$

This is the standard exponential growth equation for constant r.
Comparing the two results: - For time-dependent $\tilde{r}(t)$, we have $N(t) = N(0)e^{\int \tilde{r}(t)dt}$ - For constant r, we have $N(t) = N(0)e^{rt}$
The time-dependent case generalizes the constant case by replacing rt with $\int \tilde{r}(t)dt$. This allows for more complex growth dynamics where the growth rate can vary over time.

Both expressions have the same basic exponential form, but the time-dependent case has a more general exponent that can capture a wider range of growth behaviors.

h. To convert the equation into discrete form, we evaluate the derivative at two points: $t = 0$ and $t = G$. Let $\tilde{N}(t) = N(0)e^{\tilde{r}(r,\sigma,t)t}$.

At $t = 0$:

$$\frac{1}{N(t)}\frac{dN(t)}{dt}\bigg|_{t=0} = \tilde{r}(r,\sigma,t)|_{t=0}$$

$$\frac{1}{N(0)}\frac{dN(0)}{dt} = \tilde{r}(r,\sigma,0)$$

At $t = G$:

$$\frac{1}{N(t)}\frac{dN(t)}{dt}\bigg|_{t=G} = \tilde{r}(r,\sigma,t)|_{t=G}$$

$$\frac{1}{\tilde{N}(G)}\frac{d\tilde{N}(G)}{dt} = \tilde{r}(r,\sigma,G)$$

Expanding the exponential in $\tilde{N}(t) = N(0)e^{\tilde{r}(r,\sigma,t)t}$:

$$\tilde{N}(t) = N(0)\left(1 + \tilde{r}(r,\sigma,t)t + \frac{(\tilde{r}(r,\sigma,t)t)^2}{2!} + \cdots\right)$$

$$\approx N(0)(1 + \tilde{r}(r,\sigma,t)t)$$

This is a first-order asymptotic approximation for small $\tilde{r}(r,\sigma,t)t$.
Now solving for $\tilde{r}(t)$:

$$\frac{dN(t)}{dt} = \tilde{r}_{asymp}(t)N(t)$$

$$\frac{dN(0)}{dt} = \tilde{r}_{asymp}(0)N(0)$$

$$\tilde{r}(r,\sigma,t) = \tilde{r}_{asymp}(r,\sigma,t)$$

This matches the result from part g. If it did not match, it could be because the expansion of $\tilde{r}_{asymp}(t)$ was not taken to high enough order to capture the full time dependence.

i. Converting to discrete form with endpoints at $t = G$ and $t = G + 1$:
At $t = G$:

$$\frac{1}{N(t)}\frac{dN(t)}{dt}\bigg|_{t=G} = \tilde{r}(t)|_{t=G}$$

$$\frac{1}{N(G)}\frac{dN(G)}{dt} = \tilde{r}(G)$$

At $t = G + 1$:

$$\frac{1}{N(t)}\frac{dN(t)}{dt}\bigg|_{t=G+1} = \tilde{r}(t)|_{t=G+1}$$

$$\frac{1}{N(G+1)}\frac{dN(G+1)}{dt} = \tilde{r}(G+1)$$

Solving for $\frac{dN(t)}{dt} = \tilde{r}(t)N(t)$:

$$\frac{dN(t)}{dt} = \tilde{r}(G)N(G) \quad \text{at } t = G$$

$$\frac{dN(t)}{dt} = \tilde{r}(G+1)N(G+1) \quad \text{at } t = G+1$$

This matches the result from h. in the asymptotic limits.

Comparing all three derivations for constant r:

- Part g. gives the exact expression for $\tilde{r}(t)$

- Part h. is an asymptotic approximation valid for small $\tilde{r}(r, \sigma, t)t$

- Part i. is a discrete version that matches h. in the asymptotic limit

The discrete form in i. is most meaningful for comparison to the original calculation, as it captures the generation-to-generation change without approximation. The asymptotic expressions provide useful simplified models when applicable.

(j) To answer part j, we need to understand what the function from the paper by Itzkovitz and Alon represents and how it relates to clustering.

The function they discuss is the likelihood that two nodes would connect, which was a decreasing or decaying function of the distance R between the nodes:

$$\text{likelihood of connection} \propto \text{decreasing function of } R$$

In other words, nodes that are closer together (smaller R) are more likely to be connected than nodes that are far apart (larger R).

This function is a measure of clustering because clustering refers to the tendency for nodes to form tightly connected groups. If nodes that are close together are more likely to connect, this will naturally lead to clustering in the network.

The clustering coefficient C is defined as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of wedges}}$$

where a triangle is a set of three nodes that are all connected to each other, and a wedge is a set of three nodes where two pairs are connected.

If the likelihood of connection decreases with distance, then triangles (which require all three nodes to be close together) will be more common relative to wedges. This means the numerator in the clustering coefficient will be larger relative to the denominator, resulting in a higher clustering coefficient.

As the network grows in size (total number of nodes P increases), the average distance between nodes will generally increase. If the likelihood of connection continues to decrease with distance, then the relative proportion of triangles to wedges will decrease, causing the clustering coefficient to decrease.

Therefore, the decreasing function of distance leads to a clustering coefficient that decreases with network size:

$$C \propto \text{decreasing function of } P$$

The exact form of this decrease depends on the specific decreasing function of R and how R scales with P, but the general trend is that clustering decreases as the network grows, due to the decreased likelihood of long-range connections.

In terms of comparing to geometric networks like the Erdős-Rényi random network, the key difference is that in the geometric networks, the likelihood of connection explicitly depends on node distance, while in the Erdős-Rényi network, all connections are equally likely regardless of distance. This leads to the geometric networks having higher clustering that decreases with size, while the Erdős-Rényi network has low clustering that is independent of size.

(k) To answer part k, we need to consider how the number of wedge motifs, triangle motifs, and clustering coefficient vary with the total number of nodes P in the network.

Let w be the number of wedge motifs, t be the number of triangle motifs, and s be the number of "susceptible motifs" (motifs with at least one susceptible individual as a node).

The clustering coefficient is defined as:

$$C = \frac{3t}{w}$$

If we assume s is directly proportional to P, then:

$$s = aP$$

where a is some constant of proportionality.

Now, if w, t, and s all vary with P, we can write:

$$w = f(P)$$
$$t = g(P)$$
$$s = aP$$

where $f(P)$ and $g(P)$ are some functions of P.

The variation of the clustering coefficient with network size is then:

$$C(P) = \frac{3g(P)}{f(P)}$$

The exact dependence depends on the specific forms of $f(P)$ and $g(P)$.

If the number of susceptible individuals is independent of network size, then s would be constant:

$$s = b$$

where b is some constant value.

In this case, the clustering coefficient would still vary as:

$$C(P) = \frac{3g(P)}{f(P)}$$

but the number of "susceptible motifs" would not scale with P.

Without knowing more about the specific forms of $f(P)$ and $g(P)$, it's difficult to make definitive statements about how C, w, t, and s depend on P. The key point is that the clustering coefficient depends on the relative scaling of wedge and triangle motifs with network size, while the number of "susceptible motifs" depends on how the number of susceptible individuals varies with network size.