

Comprehensive Exam 2016

Biomath 202: Structure, Function, and Evolution of Biological Systems

Consider a population exposed to a new infection. Within this population, two types of mutations arise, leading to two types of individuals. The first type (A) is immune to the infection and has a reproductive output (i.e., number of offspring per individual per generation or the fitness) of r individuals. The second type (B) is susceptible to the infection and has a reproductive output of r before any infections have occurred in that population. Infected individuals of the second type (B) have a $1/\sigma$ chance of survivorship, but if they survive, their reproductive output is doubled (i.e., it would increase from r to $2r$ for a surviving individual in the first generation). This doubling could be due to changes in behavior or changes in physiology. Moreover, at birth each individual inherits the realized reproductive output of their parent. That is, if their parent was infected but survived and then had 4 offspring, each child would then have 4 offspring by inheritance if it avoided infection and 8 ($=2*4$) offspring if it was infected and survived. Finally, the infection is rampant and essentially always occurs in early development so that all individuals of type B are infected before they reproduce.

- Derive expressions for the relative abundance (number of individuals) and the relative mean fitness of population B relative to population A after G generations. You can assume all individuals in the same generation reproduce at the same time (i.e., discrete generations) and die after they reproduce. In principle your expression could depend on r , σ , G , and 2, but in practice you may find that some of these are absent from your final expression. Why?
- When the initial population size of A and B individuals are equal and the mutations arise at the same time, your expression for relative abundance from a. should simplify and be expressible as a pure power law. What is the power for relative abundance? Is the power a constant or a variable?
- Using your expression from part b., when $\sigma=4$ and $r=2$, what is the relative abundance of population B to population A after 1 generation and after 10 generations? When $\sigma=16$ and $r=10$, what is the relative abundance of population B to population A after 1 generation and after 10 generations? What about the relative mean fitness of type B versus A?
- Does relative abundance or relative mean fitness change in time such that type A gains or loses an advantage to type B in either case in c., and if so, after how many generations? Which population is likely to go to fixation at large time scales if the infection persists? What does your result imply if the infection only lasts one or a few generations?
- When the number of individuals in each of populations A and B is initially 100 and $\sigma=50$ and $r=1.2$, what other effects (i.e., ones not captured by relative fitness) might matter for the evolution of the two populations and which population goes to fixation? Are these potential other effects equally important for both populations A and B? What if $r=100$?
- To combine relative fitness with additional effects considered in part e., write down the equation from class that would be most useful for modeling this?

- g. Recall from class that for populations like type A with a constant reproductive output, r , it follows that the abundance (number of individuals), $N(t)$, at time t , can be related to the simplest model of exponential population growth as

$$\frac{dN(t)}{dt} = \tilde{r}N(t) \quad \text{or} \quad N(t) = N(0)e^{\tilde{r}t}$$

such that the intrinsic rate of population growth is $\tilde{r} = \ln(r)$ in terms of reproductive output, r , from above. We will now derive analogous expressions for type B populations in three different ways. First, from part a. you should be able to write down an expression for $N(t) = f(r, \sigma, t)N(0)$, where time t can be thought of as the number of generations, and you should be able to rewrite this equation in the form $N(t) = N(0)e^{g(r, \sigma, t)t}$. From this it follows that $\tilde{r}(t) = g(r, \sigma, t)$ where the intrinsic rate of population growth is now time dependent. Write down the exact expression for $\tilde{r}(t)$.

- h. A second way to derive an expression for $\tilde{r}(t)$ is by converting $\frac{1}{N(t)} \frac{dN(t)}{dt}$

into a discrete version of this equation with one endpoint of the derivative being evaluated at $t=0$ generations and the other being evaluated at $t=G$ generations. Your expression should involve a term like $N(t) = N(0)e^{g(r, \sigma, t)t}$, and you should expand the exponential to find a first-order asymptotic approximation $\tilde{r}_{asymp}(t)$ in the case that $g(r, \sigma, G)G \ll 1$. Does this expression match your answer in part g. for $\tilde{r}(t)$? Now exactly solve the equation

$$\frac{dN(t)}{dt} = \tilde{r}_{asymp}(t)N(t) \text{ to obtain } N(t) = N(0)e^{\int_0^t \tilde{r}_{asymp}(r, \sigma, t) dt}. \text{ Does } \tilde{r}(t) = g(r, \sigma, t)$$

from part g. match $g_{asymp}(r, \sigma, t)$ here? If it does not match, why is this the case and what might you have missed in your expansion of $\tilde{r}_{asymp}(t)$?

- i. A third way to derive an expression for $\tilde{r}(t)$ is by converting $\frac{1}{N(t)} \frac{dN(t)}{dt}$ into

a discrete version of this equation with one endpoint of the derivative being evaluated at $t=G$ generations and the other being evaluated at $t=G+1$ generations. Does this expression match your result from h. in any asymptotic

limits? You should now exactly solve $\frac{dN(t)}{dt} = \tilde{r}(t)N(t)$ for the $\tilde{r}(t)$

corresponding to this case. How does this answer compare with the results in parts g. and h.? Which choice is most meaningful in comparing to the original calculation in parts a. through d.? Comparing all three derivations for the case of constant growth rate, r , may help think about this.

- j. The spread of an infection and likelihood of an epidemic can be influenced by network structure and clustering. From class we talked about a measure of clustering that is a function of the expected number of triangle motifs (3 nodes for which each pair is connected by an edge) and the expected number of wedge motifs (3 nodes for which at least two of the three pairs are connected

by an edge). What was this function and what was the logic behind this being a measure of clustering? How does this clustering coefficient scale with network size (i.e., the total number of nodes, P) and the mean connectivity, λ , for an Erdos-Renyi random network? Do not worry about constant coefficients or other factors, just the proportionality and scaling. In terms of the scaling with the total number of nodes, P , and the mean connectivity, λ , how do these results compare with geometrical networks of the type in the paper by Itzkovitz and Alon in which the likelihood that two nodes would connect was a decreasing or decaying function of the distance, R , between the nodes?

- k. For some questions what matters most may be how many triangle or wedge motifs have at least one susceptible individual as one of the nodes or an infected individual as one of the nodes. For the scenario that you have (Itzkovitz and Alon)/geometrical networks and the number of susceptible individuals is a direct proportion of the total number of individuals (i.e., nodes), P , how would the number of wedge motifs, triangle motifs, and clustering coefficient of just "susceptible motifs" vary with total number of nodes, P ? What if the number of susceptible individuals is independent of total network size?