

# Mathematical and computational modeling in ecology

*Madeline Cowen*

**3 April 2019**

Steps to modeling a biological problem:

1. Formulate the problem (what do you want to know?)
2. Determine the basic ingredients (define the state variables, describe constraints on those variables (population density can't be 0, etc.), describe interactions among the variables, choose discrete or continuous time, and a time scale, units, define parameters and constraints on them)
3. Qualitatively describe the biological system (use a life cycle diagram, flow diagram, or a table for more complicated stuff)
4. Quantitatively describe the biological system (translate #3 to equations, perform checks to catch errors e.g., are units consistent, are constraints satisfied)
5. Analyze the equations (use the equations to simulate and graph the dynamics, perform appropriate analyses, (rarely) solve for general solution). Graphing: state variables vs. time, look for dependence on initial conditions, study dependence on parameter values (aka sensitivity analysis)
6. Checks and balances (check your model results against data or older analyses, determine how general your results are (based on parameter values and assumptions/constraints on variables), consider alternatives to the simplest model, extend or simplify (people usually add too much at the beginning; easier to extra stuff later and easier to understand role of new stuff you add because now you understand the base case) the model as needed and repeat 2-5)
7. Relate the results back to the question (do the results address your original question, are the results obvious or counter-intuitive, do you trust this, interpret the model verbally, describe potential experiments or observations that would test the model)

Formulating a model: population growth in discrete time, using the example of a sheep population on an island (closed population) with births in the spring and mortality during the winter. First question is how will the population change when they are first introduced, second question is how will it change when it gets crowded?

## Model 1

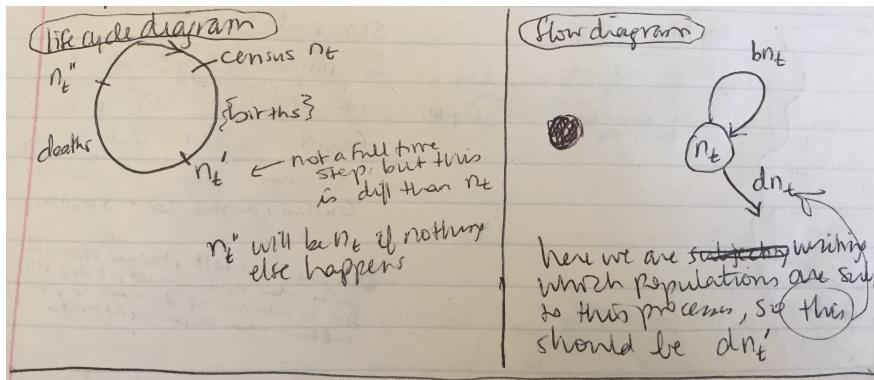
1. Let's specify females since they are often rate-limiting. How will the population size of females change when sheep are first introduced to the island? Assume resources are not limiting.
2. Basic ingredients:
  - Time is in discrete intervals of one year.
  - State variable:  $n_t$  = number of females in year  $t$
  - Parameters:  $b$  = per capita birth rate (expected # of female offspring per mother per year);  $d$  = fraction of the population that dies each year (when you multiply it by the population size), or the probability of dying for any individual (which sounds more stochastic and not deterministic, but this is a deterministic model so it's really more of an expectation of how many will die—if everyone has the same probability, that will manifest as a fraction of the population that will die).
  - Constraints: state variable must be  $\geq 0$ ,  $b \geq 0$  although  $b = 0$  is a trivial case,  $0 < d \leq 1$  (but if  $d = 1$  then everyone will die in the next year, as if they are annual plants). Another way to write this is:

$d \in (0, 1]$ .

- Units: Units of  $b$  is written as  $[b]$ .  $[b] = \# \text{ baby sheep} / \# \text{ mother sheep}$  and therefore is unitless. Similarly,  $d$  is a fraction or probability, which is also unitless. We don't have the "per year" as part of the units because this is discrete time, so time isn't flowing while you're in a particular time step.

### 3. Qualitative description:

- we'll count sheep just before breeding season.
- We can depict this in several ways (but these terms aren't used the same way by different authors).
- One is a "life cycle diagram". Another is the "flow diagram" (be careful whether you are writing per capita parameters or not)
- See image.



### 4. Quantitative description:

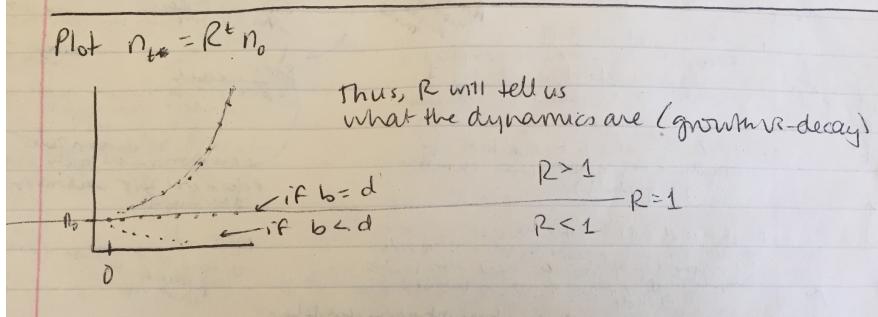
- for this, probably easiest to build from the life cycle diagram
- goal is to write down an equation for  $n_{t+1}$  in terms of  $n_t$ :  

$$n_{t+1} = n_t + bn_t - dn_t$$
 and  $n_{t'} = n_t + bn_t$  so we have:  $n_{t+1} = (n_t + bn_t)(1 - d) = n_t(1 + b)(1 - d)$
- In discrete time, we like to write that  $R = (1 + b)(1 - d)$ , which is sort of a discrete time growth rate, but let's avoid "rate" in discrete time and say it's a growth factor or growth multiplier. Thus,  $n_{t+1} = Rn_t$ . We also see this as:  $n_{t+1} = (1 + r_d)n_t$ , and we can call  $r_d$  is the per capita change in the number of sheep ("growth rate").
- Now let's perform some checks, starting with units.
- LHS:  $n_{t+1}$  is number of sheep. RHS:  $[R]$  is the number of future sheep divided by the number of present sheep, no units because  $b$  and  $d$  are unitless too.  $[n_t] = \text{number of sheep}$ . So  $[\text{LHS}] = [\text{RHS}]$ , yay.
- Sign check: check the state variables using constrained parameter values. We need  $n_t \geq 0, \forall t$  (aka for all time).  $R$  is always positive, because  $(1-d)$  and  $(1+b)$  will always be positive, so  $n_{t+1}$  will always be 0.

5. Analyze equations:

$n_{t+1} = Rn_t$  Call  $n_{t=0} = n_0$ . If we keep plugging in values for  $n_0$  and the next time step, we realize that there is a general solution:  $n_t = R^t n_0$

Then plot it! See image.



## 8 April 2019

(continuing with exercise from last class)

Usually you don't have a general solution. Most often you analyze the model qualitatively. This means that you consider the equilibrium points and their stability.

Definitions:

System is at equilibrium if the state variables don't change over time. In a discrete time model, this means you're looking for when  $n_{t+1} = n_t$ . Stability of an equilibrium describes whether it is attracting or repelling. Think about response to a perturbation from equilibrium (analogy—ball sitting on top of a hill or at the bottom of a valley, and you push it, how does it roll?)

Equilibrium point  $n^*$  is locally stable if the system near  $n^*$  approaches it, and is globally stable if the system will approach it from anywhere. It is unstable if the system near  $n^*$  moves away from it.

*Stability is a property of an equilibrium point in a model, and it will depend on parameter values!*

For geometric growth model,  $n_{t+1} = Rn_t$ .

Equilibrium point: Find values of  $n^*$  that satisfy  $n^* = Rn^*$ .

For  $R = 1$ , then any  $n^*$  works; this is the trivial case.

For  $R \neq 1$ , then  $n^* = 0$ , which is the only equilibrium.

Stability of  $n^*=0$  (based on the plot):

$n^* = 0$  is unstable when  $R > 1$

$n^* = 0$  is stable when  $0 < R < 1$  (when it's less than one, but remembering that it can't be negative). Looking at the model we know that it's globally stable, because the population density will always decrease when  $R$  is less than 1.

(trivial solution: when  $R = 1$ , that's neutral stability)

6. Checks and balances: Check our model against data. In this case, we can take advantage of the fact that we have a general solution, and transform our data to see if it matches our model:

$n_t = R^t n_0$  so let's log transform it:  $\log(n_t) = t * \log(R) + \log(n_0)$ . So we can log our data and plot it against time, and then we expect that the slope of the line should be equal to  $R$ . Ultimately, though, we'd expect that the model will break down; unconstrained growth can't go on forever on the island.

If you can get away with avoiding some nonlinearities, then do so, such as when there is early growth and things look exponential (or geometric if discrete time) at first. You don't have to bite off every possible complexity.

7. Relate back to biological question: what happens after introduction? Geometric growth or decay; it depends (on the values of the parameters, which represent the biology we're thinking about—here,

depends on demographic parameters—remember that  $R$  involves the birth and death rate). Note that right now we're talking about a deterministic world, but in the real world there's chance so there will be stochasticity, and there are models that deal with that (both environmental stochasticity, and the stochasticity that comes from being a discrete world where you can't have half of an individual).

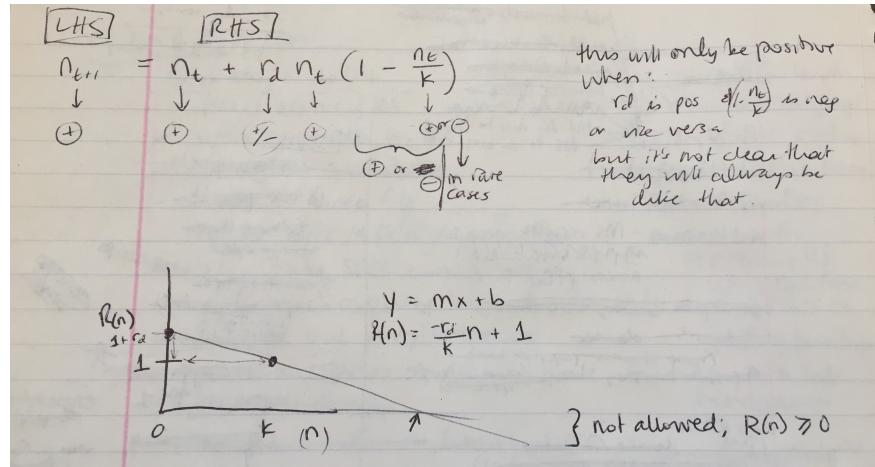
Okay, now let's move to question 2, what happens when it gets crowded?

## Model 2

- How does the population size change as resource become limiting?
- Ingredients: as before, but now let's say that  $R$  depends on the population size. We need an additional parameter, aka a carrying capacity (a quantity to level off, what is "crowded"). We'll call it  $k$ . This is the population size where each individual replaces itself each year. What values can  $k$  take? It must be larger than 0.
- Qualitative description:  $R$  depends on  $n$ , or mathematically we say  $R(n)$ . That's why we call this density dependence.  $R(n) \geq 0$ . Okay, now let's determine particular values:  
Assume (i)  $R(n=0)$ , this is the same  $R$  as in Model 1. We'll write it as follows:  $R(n=0) = 1 + r_d$ . At low population density, we know those dynamics. (ii)  $R(n=k) = 1$  by definition. How is  $R(n)$  changing between these points? There's no particular answer, but there is a convention: assume that  $R$  declines linearly with increasing values of  $n$ .
- Quantitative description:  $n_{t+1} = R(n_t)n_t$ . So, let's work with the rules to transform this into an equation. Think about it like an equation of a line between points  $(0, 1+r_d)$  and  $(k, 1)$  where  $R(n)$  is the y-axis and  $n$  is the x-axis, and slope is  $(1 - 1 - r_d)/(k - 0) = -r_d/k$ . So,  $R(n) = (-\frac{r_d}{k})n + (1 + r_d) = 1 + r_d(1 - \frac{n}{k})$ . Let's plug this back in to our equation:  $n_{t+1} = (1 + r_d(1 - \frac{n}{k}))n_t$  or in more classical form:  $n_{t+1} = n_t + r_d n_t (1 - \frac{n}{k})$ .

Let's check the units to make sure it's behaving: LHS:  $[n_{t+1}]$  = sheep; RHS:  $[n_t]$  = sheep,  $[r_d]$  is unitless (future sheep/current sheep),  $[1 - \frac{n}{k}]$  is also unitless (1-sheep/sheep). So yes,  $[\text{LHS}] = [\text{RHS}]$ .

Sign: what range of values can  $r_d$  take? If  $R(n)$  must be greater than or equal to 0, and is equal to 1 +  $r_d$ , then the lowest value  $r_d$  can take is -1.  $n_t + r_d n_t (1 - \frac{n}{k})$  has the following signs: (see image). In this case, it is not clear that the RHS will always be positive. Model pathology: red flag, because you need to make sure that you're working in the realm where you're not getting negative population sizes. It's the linear relationship between  $R(n=0)$  and  $R(n=k)$  that is suspect—if it gets extrapolated beyond  $k$ , then you get negative values or  $R(n)$ . see image



## 10 April 2019

continuing. . .

5. Analyze: Find equilibrium points by solving for values of  $n^*$  that satisfy  $n_{t+1} = n_t$ . From  $n_{t+1} = n_t + r_d n_t (1 - \frac{n}{K})$ , we have  $n^* = n^* + r_d n^* (1 - \frac{n^*}{K})$ , then  $0 = r_d n^* (1 - \frac{n^*}{K})$ . Solutions are therefore:  $n^* = 0$  and  $n^* = K$  ( $r_d = 0$  is trivial solution).

Next, look at stability: there are ways of doing this analytically (and he can help with that if we want to do this), but we can also look at model outputs to see what happens. He showed us two plots, one with positive values of  $r_d$  and one with negative values, and you can see that  $n^* = 0$  is unstable and  $n^* = K$  is stable when  $r_d$  is positive, and vice versa. Can look at sensitivity to initial conditions, for example by plotting population at time = 25 for different simulations with different initial population sizes against initial population size. Can look at sensitivity to  $r_d$  to see how stability of equilibrium points by plotting population at time = 5 against  $r_d$ . As you push  $r_d$ , you can get to the equilibrium point faster, or you can overshoot a little, and push the system into stable cycles, or into chaos.

7. Relate back to the question: how does the population grow as the island gets crowded? Answer: If  $r_d$  is not too high, population will approach K. If it is positive and large, you can get oscillations or even chaos.

### Continuous time models

Discrete time models make sense when your system has a kind of rhythmic clock to follow, but in some systems, the events can occur at any time and there isn't a seasonality to them. If your state variables can change at any time with no strong rhythm, you can let time be a continuous variable (world of calculus—infinitessimally small time steps).

Describe the rate of change of our system (state variable) using a differential equation model.  $\frac{dn(t)}{dt}$  = “some function of  $n(t)$ ”, where  $n(t)$  is the state variable and we are looking at the change in the state variable over time. This is called an ODE (ordinary differential equation, in other words, only looking at the rate of change with respect to one variable, here: time). FYI: This is in contrast to PDEs (partial differential equation), which looks at change with respect to more than one thing. For example:  $\frac{dn(t)}{dt} + \frac{dn(t)}{dx}$  where x is space or age.

Example: exponential growth model. Consider a closed population with per capita birth rate b per year, and per capita death rate d per year. See image for flow chart. Note that the model here updates instantly with any change, so need to worry about where the census happens or what the population size is that each b or d is affecting. Thus, we write:  $\frac{dn(t)}{dt} = bn(t) - dn(t) = (b - d)n(t) = rn(t)$  where  $r = b - d$  (so there is no limits on r) aka the per capita growth rate. You're just looking for changes in the population, so you don't have to add  $n(t)$ . This is a linear model, which means is it a simple linear function of the state variable (some constant times the state variable). Two consequences of this are that it's easy to model and work with, but it's also not capturing some of the more interesting biological interactions; for that you need nonlinearities. For now let's revel in this simple world and find a solution for this, by taking the integral!  $\frac{dn}{dt} = rn$  or  $\frac{1}{n}dn = rdt$  and take the integral of both sides, which yields  $\ln(n) = rt + C$ , which is  $n = e^{rt}e^C$  and this can be worked out to be:  $n(t) = n_0 e^{rt}$ . Usually we won't do this because it's too hard.

Equilibrium and stability: To find equilibrium points, you're looking for when  $\frac{dn}{dt} = 0$ . So, we're trying to solve:  $rn^* = 0$ . This is true when  $n^* = 0$  (for  $r \neq 0$ ).

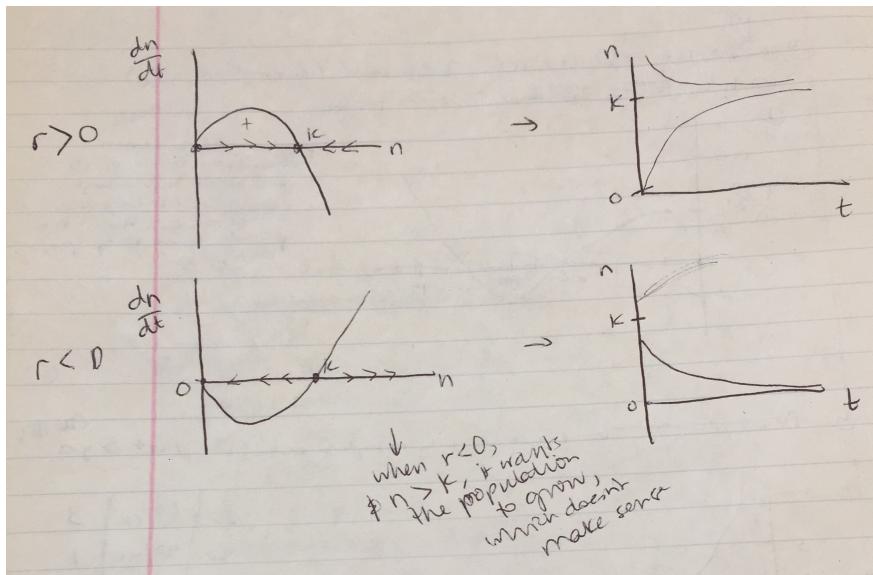
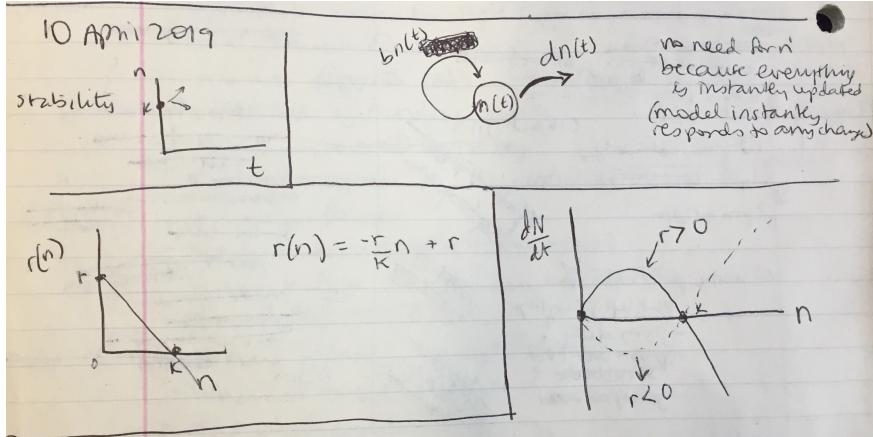
Stability: When will n approach  $n^* = 0$  or move away from it? We know that  $n(t) \geq 0$ , so this depends on r. If r is positive, it is unstable (exponential growth), and if r is negative, it is stable (exponential decay) (and it's globally stable).

Now let's add *density dependence* to this model. As a start, let's follow the same set of logic as in discrete time. We want to capture the influence of crowding. Qualitatively, let r be a function of n:  $\frac{dn}{dt} = r(n)n$ . Let K be the carrying capacity. Three assumptions, as before, but they look a little different: (i)  $r(0) = r$  (aka at low density, we assume it follows exponential growth rules), (ii)  $r(K) = 0$  (continuous time is all

about rates of change instead of replacing individuals in discrete time), and (iii) we assume that there is a linear decline between these two rates. See image for the plot. Slope is  $\frac{-r}{K}$  and intercept is  $r$ , gives us  $r(n) = \frac{-r}{K}n + r = r(1 - \frac{n}{K})$ . Let's plug this back in and that gives us the *classic continuous time logistic growth model*:  $\frac{dn}{dt} = r(1 - \frac{n}{K})n$ .

Equilibrium points: want  $n^*$  such that  $r(1 - \frac{n}{K})n = 0$  when  $n^* = 0$  or  $n^* = K$ .

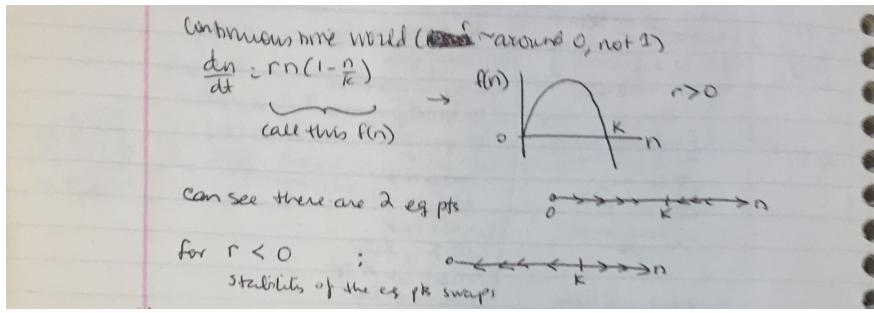
Stability: for a given set of parameter values, are the points stable? Let's plot how  $\frac{dn}{dt}$  depends on the population size.  $\frac{dn}{dt} = r(1 - \frac{n}{K})n = rn - \frac{rn^2}{K}$ . We know when  $\frac{dn}{dt} = 0$ , and let's draw this from knowing this is a parabola. But, it's going to depend on the sign of  $r$ : if  $r$  is positive, then the coefficient is negative, so it's going to go downward. See image. Can consider those  $\frac{dn}{dt}$  plots to determine stability, but can also write them out as a plot of  $n$  by  $t$ . See image.



## 15 April 2019

### Review

In the continuous time world: ( $r$  leads to different dynamics if it is greater than or equal to 0, not 1). If we call  $\frac{dn}{dt} = rn\left(1 - \frac{n}{K}\right) = f(n)$  and we plot  $f(n)$  by  $n$ , you can see that there are two equilibrium points,  $n^* = 0$  and  $n^* = K$ . You can plot the phase line and see the stability of these points (depending on  $r$ ; stability swaps when  $r>0$  vs.  $r<0$ ). See image.



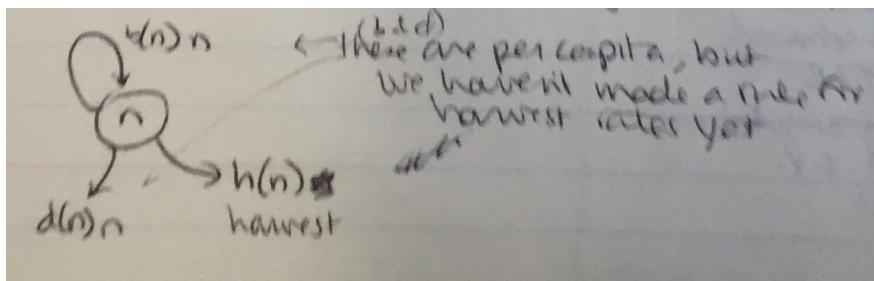
As you can see, knowing whether the slope of the function at the equilibrium point is positive or negative is enough to know the stability of the curve. Thus we can evaluate the derivative of the function at the equilibrium point to see the slope of the line. See image.

$\left  \frac{df}{dn} \right _{n=n^*}$	$< 0$ stable (stable) $> 0$ unstable
--	---

Plot of  $f(n)$  versus  $n$  shows you density dependence, aka how the growth rate changes with density.

### Adding complexity: models with “harvesting”

Developing models to know how much to harvest is one of the ways this field developed, especially in fisheries. Let's build off of continuous time logistic and add more stuff, namely that some of the population is lost to harvesting: See image.



Questions: can we define a sustainable harvest rate? Can we maximize yield without risk of collapse? Are some policies better than others for this?

General model:

$$\frac{dn}{dt} = f(n) - h(n)$$

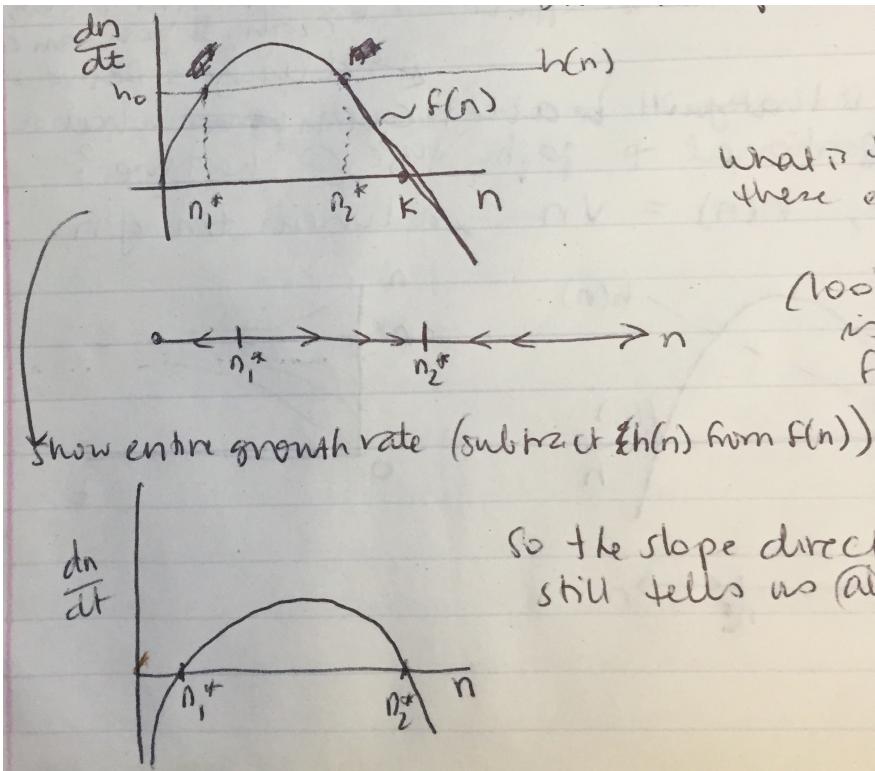
where  $f(n)$  is the “natural growth dynamics” ( $b(n)n - d(n)n$ ) and  $h(n)$  is the harvest rate.

Assume logistic growth applies: (and assume  $r > 0$ )  $f(n) = rn\left(1 - \frac{n}{k}\right)$

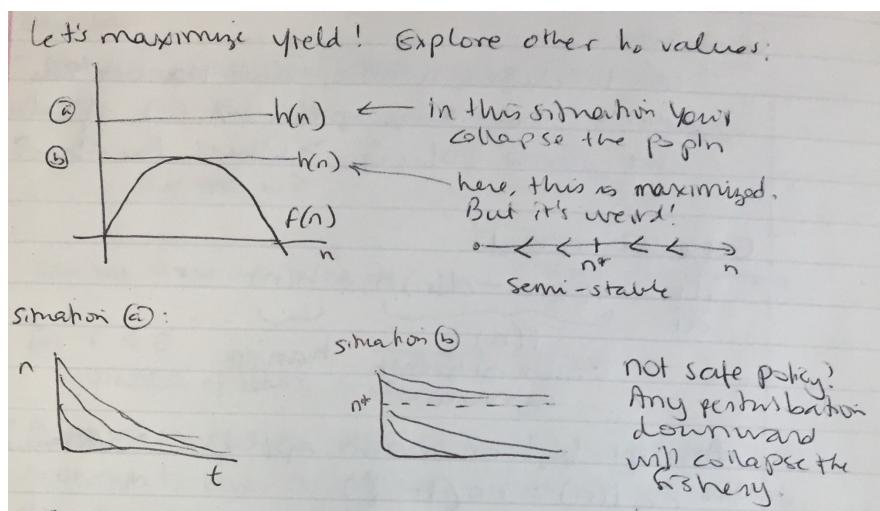
Where are the equilibria? Answer where  $f(n) = h(n)$ , aka at the intersection of the two curves.

*Policy 1:* constant yield (aka  $h(n)$  doesn't depend on  $n$ )

$h(n) = h_o$  where  $h_o$  is some constant. Draw the two components of  $\frac{dn}{dt}$  separately. You can determine the stability by looking at what  $\frac{dn}{dt}$  will be by subtracting  $h(n)$  from  $f(n)$ . See image. You can also draw what you would get by subtracting the two (just one function on the plot; see image), and you can see that the slope of the line rule still applies.



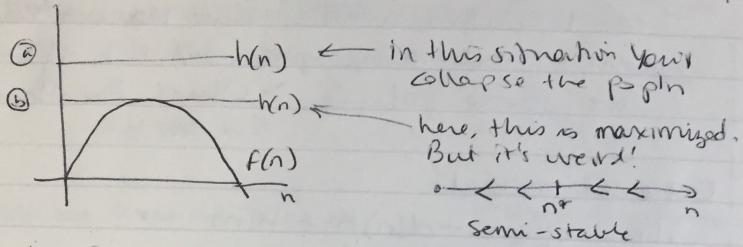
Let's maximize yield! Explore other values of  $h_o$ .



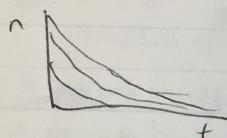
In some cases, you get a semi-stable equilibrium point, but this isn't safe because any perturbation downward will collapse the system. So what do we do? Let's try other policies. Why don't we try to adapt harvest rate based on the current population size.

*Policy 2:* constant effort (instead of constant number of fish, say catch all you can during a period of time). The idea is that you will be able to catch what is proportional to population size at the time. So, now  $h(n) = vn$ , a linear function of  $n$ . See image.

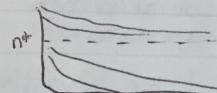
Let's maximize yield! Explore other  $h_0$  values:



situation (a):



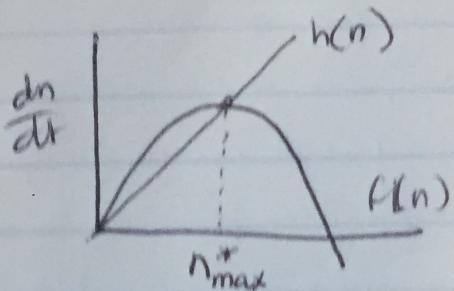
situation (b)



not safe policy!  
Any perturbation  
downward  
will collapse the  
fishery.

How do you maximize the yield here? Choose a  $v$  where  $h(n)$  intersects  $f(n)$  at its peak. See image.

Still looks stable:



$$\frac{dn}{dt} = rn(1 - \frac{n}{K}) - vn$$

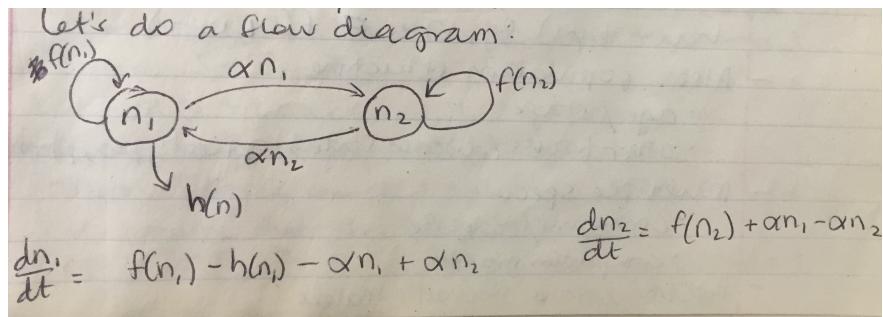
Still looks stable! This is a much better policy because it's self-correcting. But, if  $vn$  is too steep (i.e., bigger than  $r$ ) then you are harvesting too much and the population will crash:

$\frac{dn}{dt} = rn\left(1 - \frac{n}{K}\right) - vn$ . When  $n \ll K$ ,  $\frac{n}{K} \approx 0$ , so  $\frac{dn}{dt} = rn - vn$  (at low densities) and will crash if  $v > r$ . This happens when humans "cheat" with technology and more effort, but also when other fluctuations lower  $r$ .

17 April 2019

Marine Reserve Model: qualitative description: now assume two habitat patches, one that is harvested and one that is protected. Assume both are subject to logistic growth, same parameters  $r$  and  $K$ . Also assume they are connected by migration where each fish can move to the other patch with a constant per capita rate  $\alpha$ . Let's do a flow diagram:

See image.



$$\frac{dn_1}{dt} = f(n_1) - h(n_1) - \alpha n_1 + \alpha n_2$$

$$\frac{dn_2}{dt} = f(n_2) - h(n_2) - \alpha n_2 + \alpha n_1$$

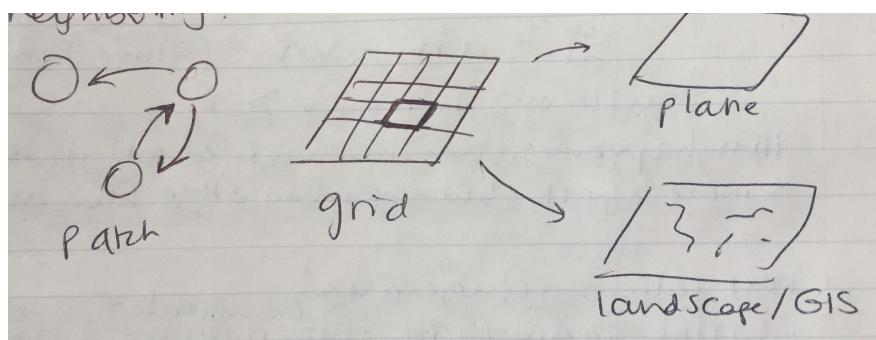
With logistic growth:

$$\frac{dn_1}{dt} = r n_1 \left(1 - \frac{n_1}{k}\right) - h(n_1) - \alpha n_1 + \alpha n_2$$

$$\frac{dn_2}{dt} = r n_2 \left(1 - \frac{n_2}{k}\right) - h(n_2) - \alpha n_2 + \alpha n_1$$

### Model elaborations

Models with patches with connections (a “patch model”) can be elaborated into a network model. Can also go into a grid system where you divide space into plots and pay attention to which plots are neighboring. Can look at just existence of space  $\rightarrow$  plane, or take into account landscape detail. See image.



- think about number of dimensions (going down a river is 1 dimension, down and across or down and into different water depths are 2 dimensions)
- can add time delays into the system (responses have lags) (e.g.,  $n_{t+1} = f(n_t, n_{t-1})$ )
- also: population structure (age/stage, other traits such as disease states, genotypes, phenotypes)
- multiple species (consumer/resource models, competition...)
- deterministic vs. stochastic models

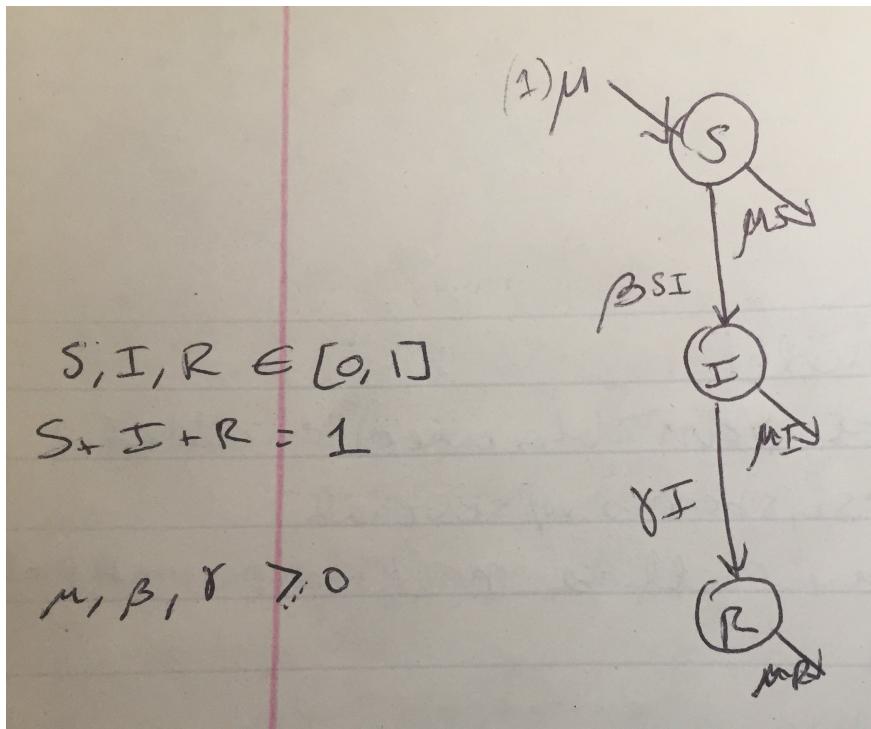
We'll talk about a kind of multicompartmental model (the bubbles of state variables are often called compartments; systems with several compartments are called multicompartmental models) called SIR epidemic models. SIR models: consider an infectious disease like measles. It's acute and immunizing (once you've

gotten it, you can't get it again), so they're called "childhood diseases", very amenable to vaccination programs.

Questions:

- When infection is introduced to a population, what determines whether you get an outbreak?
- How much of the population do you need to vaccinate to prevent an epidemic? "Herd immunity"  $\rightarrow$  once vaccination level is beyond a certain point, everyone is protected, even those who can't get vaccinated.

Divide the population into groups based on their status with respect to disease: Susceptible (never been infected); Infectious (infected and can infect others); Recovered (previously infected, no longer infectious) (hence S-I-R). Let state variable S, I, and R represent the proportion of the population in each group. Assume closed, constant population size, allow for birth and death but assume they balance out. Since S, I, R are proportions,  $S, I, R \in [0, 1]$  and  $S + I + R = 1$ .



Births and deaths:  $\mu$  is the per capita birth rate into S group, and also the per capita death rate from all groups.

New infections: rate of new infections depends on I individuals contacting S individuals, so we assume it depends on abundance of S and I individualshosts. Assume random mixing of the population, which gives us a rate of about  $\beta SI$ .  $\beta$  is the transmission rate.

Recovery: assume infectious people recover to immune state at some fixed per capita rate  $\gamma$ .

## 22 April 2019

SIR models continued:  $\mu$ ,  $\beta$ , and  $\gamma$  cannot be negative. They are rates, and if they were negative it'd be as if you could reverse the direction of a state. All must be equal to or greater than 0.

Aside on residence time: residence time refers to how long an individual stays in a given state. In an ODE model, the expected residence time in a state is the reciprocal of the total outflow rate. For example, if you have a state X and individuals leave at a per capita rate  $a$ , then the expected residence time is  $\frac{1}{a}$ . If state Y has outflow  $cY$  and  $bY$ , then the expected residence time  $E(T_Y) = \frac{1}{c+b}$ .

The  $\gamma$  term is the recovery rate. We can come up with an estimate of  $\gamma$  from observable biology, but it's hard to estimate the recovery rate. Instead, you can look at how long individuals stay in the infected class and use

that to estimate the rate at which they leave the class: estimate parameter values using data/observations of residence times in the real world.

Back to SIR: let's say that the infectious period for measles = 8 days. Therefore, we can estimate  $\gamma = \frac{1}{8\text{days}} = 0.125\text{days}^{-1}$ . Similarly, if we say "residence time" of being alive is 80 years, then  $\mu = \frac{1}{80\text{years}} = 0.0125\text{years}^{-1}$ . If you want to use these two models, you need to convert these so you're either in days or years. As you can see, these are on very different orders of magnitude, so it allows us to focus on the effect of measles–death by measles is more likely than average death rate).

Adding death from measles: if you wanted to, you could have  $\mu' = \mu + \alpha$  where  $\mu$  is natural death and  $\alpha$  is death from measles.

Quantitative model:

$$\frac{dS}{dt} = \mu - \mu S - \beta SI$$

$$\frac{dI}{dt} = \beta SI - \mu I - \gamma I$$

$$\frac{dR}{dt} = \gamma I - \mu R$$

First question: what determines whether an introduction of disease leads to an epidemic? Epidemic means that you have transmission of infection and an upswing of infections (and ultimately a downswing when you run out of people to infect).

First, initial conditions:  $I > 0$ , but small:  $I \ll 1$ , naive population  $R = 0$ , so  $S = 1 - I$  but since  $I$  is small,  $S = 1$ . This is essentially asking if, when we add a few  $I$  individuals, will  $I$  grow?

Two ways to translate this to math. First way ("instantaneous flow"): is  $\frac{dI}{dt} > 0$ ? This is true when  $\beta SI - \mu I - \gamma I > 0$ . Simplifying this and taking into account initial conditions aka  $S(0) = 1$  (and that  $I$  is not 0, so we don't have to worry about dividing by 0), we can see that:  $\beta > \mu + \gamma$  or  $\frac{\beta}{\gamma+\mu} > 1$ . Aka, is this equilibrium stable? Does a small perturbation cause  $I$  to grow or not?

Second way ("generational approach"): are infected individuals infecting more than just one other person before they recover, aka is the expected number of new infections caused by each case greater than 1. The "expected number of new infections caused by each case" in epidemiology is called  $R_0$  or the basic reproductive number (we call it this when the population is naive and susceptible). The condition we're looking for is  $R_0 > 1$ . This is the dominant paradigm for how we think about disease spread, borrowed from population demography (are genotypes replacing themselves–absolute fitness).

$R_0 = (\text{rate of production of new cases by 1 infectious individual in susceptible population}) \times (\text{expected duration of infectious period})$  which is  $(\beta S)$  times  $(\frac{1}{\gamma+\mu})$ , but  $S(0) = 1$  so:

$$R_0 = \beta * \frac{1}{\gamma+\mu} \text{ and we want } R_0 > 1.$$

These two approaches give the same formula, so you can think about it two ways and get the same condition.

Second question: what is the critical vaccination threshold to reach the state of "herd immunity" and prevent an epidemic? Goal is to shift an unstable equilibrium point (where a perturbation would lead to epidemic) to a stable one.

Assumptions: a perfect vaccine (vaccination changes  $S$  to  $R$ ). Let  $v$  = proportion of population that is vaccinated when the pathogen is introduced. This means that we are going to deal with vaccination in the initial conditions instead of as a flow in the model:  $S(0) = 1-v$ ,  $R(0) = v$ ,  $I(0) = \text{small but positive}$ . Also assume that  $R_0 > 1$ . How big does  $v$  need to be to prevent epidemic?

## 24 April 2019

Effective population number:  $R_{eff}$  or  $R$  is the expected number of cases caused by an infected individual in a partially immune population. The  $R_0$  world is best case scenario for the pathogen (completely susceptible population). Now we're trying to find the conditions that give us  $R_{eff} > 0$ .

$\frac{dI}{dt} > 0$  when  $\frac{dI}{dt} = \beta S - \gamma - \mu > 0$ . With initial conditions  $S = 1 - v$ , then:  $\beta - \beta v > \gamma + \mu$  and we need  $v_{critical} < 1 - \frac{\gamma+\mu}{\beta} = 1 - \frac{1}{R_0}$ , so now we can just plug in  $R_0$ . For measles,  $R_0$  is thought to be 20 (very infectious!), so the critical vaccination level is 0.95. You could even plot the critical vaccination level needed for different values of  $R_0$ .

From the approach of  $R_{eff}$ , you get  $R_{eff} = R_0(1 - v)$  aka  $R_{eff} = R_0S$  and since we want  $R_{eff} > 1$  then we find the same critical vaccination level.

What is this missing? Assumes random mixing, doesn't take into account heterogeneity in the population.

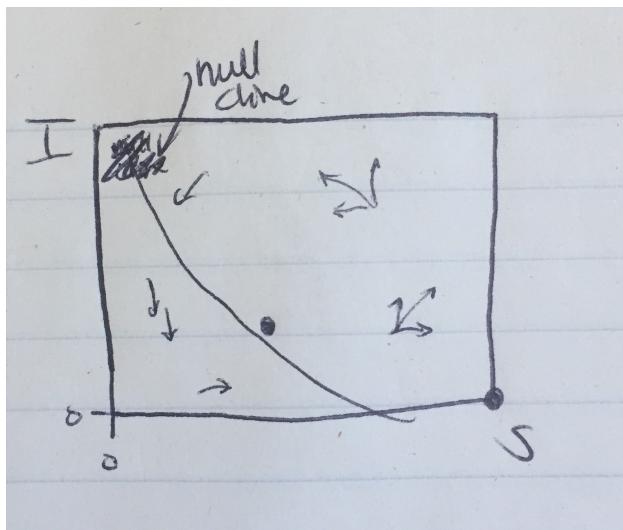
Qualitative analysis: equilibria and stability. You can find equilibrium points in a multicompartmental model by setting all equations equal to 0 and solving a system of equations. Also, because we are working with proportions, we only need to work with two of the equations, because we can figure out the other state since  $S + I + R = 1$ . Do some math to find values of  $S^*$ ,  $I^*$ , and  $R^*$  to satisfy equilibrium assumptions.

Disease-free equilibrium:  $S^* = 1$ ,  $I^* = 0$ , and  $R^* = 0$

Endemic equilibrium (disease exists in the population):  $S^* = \frac{\gamma+\mu}{\beta}$ ,  $I^* = \frac{\mu}{\beta}(\frac{\beta}{\gamma+\mu} - 1)$ , and  $R^* = 1 - S^* - I^*$ .

This makes sense because at equilibrium,  $R_{eff} = 1 = R_0S^*$  so it makes sense that  $S^* = \frac{1}{R_0}$ .

Stability: instead of phase line, we can look at a phase plane. In this case, let's look at  $S$  and  $I$ . The system has to be in the box somewhere, and if it doesn't add up to 1 then you know the remainder is in the  $R$  group. You can basically map out the vector field (what direction will the system go in), or also do null clines (draw boundaries). See image.



Slides on HIV. First question: why does the viral load peak and then drop off? (either immune systems figures it out, or the virus runs out of CD4 cells to infect) Someone tracked infected/uninfected CD4 cells and saw that depletion of CD4 cells leads to that viral load peak pattern without invoking immune response. Second question: why the long stable "set-point" phase before viral load increases? (could be that virus-infected cells live a long time, turnover very slowly) Ho et al 1995 treated HIV patients (as a way of cutting off the births) and looked at viral load response to determine death rate. Found that those cells are actually dying really quickly, so to hold this set-point the virus is constantly rapidly re-generating itself (major conceptual breakthrough, knowing that it can evolve so quickly).

Model 3: how will antiretroviral drugs impact the HIV epidemic? There was a fear of "behavioral disinhibition", that people would go back to risky behavior if drug therapy lowered fears of contracting HIV. Blower et al 2000 model that allowed for change in behavior and for evolution of resistance, used the model

to explore uncertainty.

Model 4: estimating the rate of new infections, and who is at risk. Williams (2000) saw that HIV is still happening among the women with the most HIV, and that there was greater risk for younger women than you could tell just by looking at prevalence.

Model 5: impact of male circumcision on HIV risk.

## 29 April 2019

Pseudo-code: informal way to plan out the structure and logic of your coding programs. See Jamie's slides.

Practice pseudo-code:

```
Assign 6 to x and 2 to y
Calculate x+y and x*y
Do ifelse statement to check if x+y > x*y
If yes, then print "the sum wins"
If no, print "boo" etc.
```

Could also write it out as something closer to code:

```
x <- 6
y <- 2
sum <- x + y
product <- x*y
If sum > product, print("the sum wins")
Else print("boo")
```

Good practice: write out your pseudo-code, and then turn those into commented headings with your code underneath, so you have nice commented code that's easy for your future self or collaborators to understand.

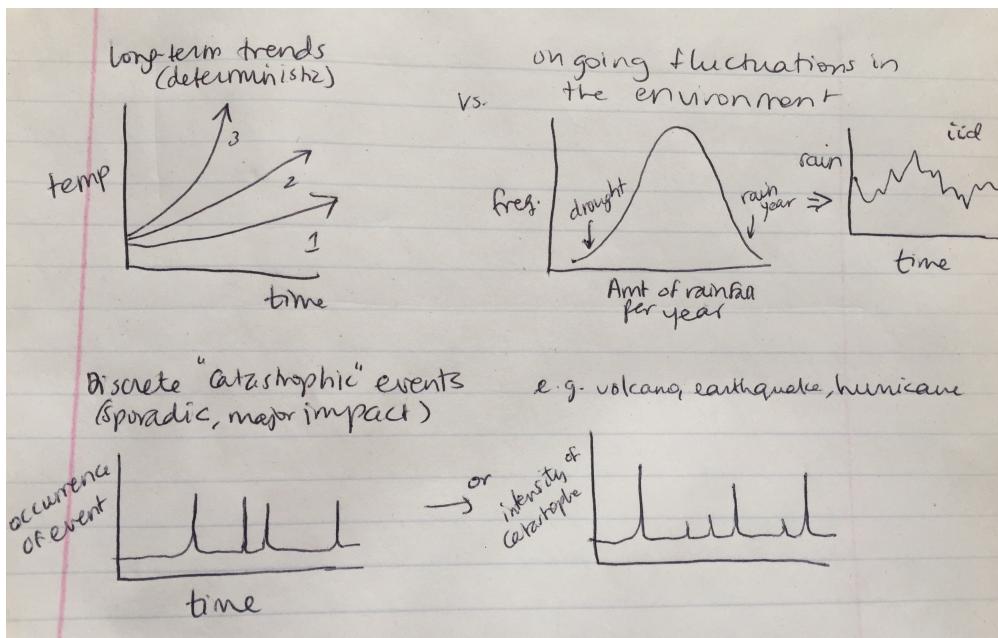
## Population dynamics in a changing environment

Until now, we've assumed that nothing changes except our state variables. We've assumed our parameter values are all constant. This is obviously not true; many examples of things that can affect parameters: climate change, habitat loss/degradation/fragmentation, overfishing/harvesting increases (including pet trade), hurricanes, geological disasters, behavior, weather, and more.

We can model this! Two broad classes of change:

(1) Deterministic changes: these are predictable; they follow some rule, so you can predict *how* and *when* they will act. These can be long-term trends (such as temperature increasing over time due to climate change, modeling several possible climate change scenarios). Can also be discrete events, such as a policy change.

(2) Stochastic changes: these are unpredictable, because they involve chance. These are random events, or very complicated things that are not included in the model. We can describe these in terms of probabilities and distributions, but can't predict exactly when or how they will act. We draw from a distribution (independent and identically distributed, or "iid"). These can be distributions of intensity of ongoing fluctuation, or distributions of when certain discrete events will happen. See image.



Four-part model script:

1. Setup stuff
2. Parameter assignments, initial conditions, initialize variables
3. Calculations and simulations
4. Output

When including parameters that will change, you should describe the distribution/changing environment in step 2, and then sample from that in step 3. For example, in step two you say the mean and standard deviation of the distribution, and then you sample from that distribution during the simulation. (every step of the for loop in step 3 is what happens every year or every time interval, so you want to sample from the distribution each year/time interval). In pseudo-code this will look like:

```
for(t in 1:100){
  figure out environmental conditions and update parameter values
  n(t+1) <- fn of n(t), given current parameters
}
```

Example 1: Retrospective study. Assume you know how parameters changed through time. e.g., you have  $R[t]$ , a vector of parameter values through time. If you have all of these parameter values through time, then in step 2, you'd assign to the vector all of those parameters, and then in step 3, in each step of your for loop, you'd assign a value from the vector.

(...)

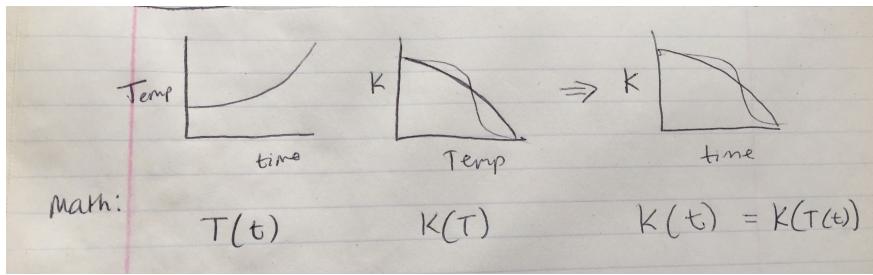
```
Rvec <- assign vector of parameter values
```

```
for(t in 1:100){
  Rnow <- Rvec[t]
  n[t+1] <- Rnow * n[t]
}
```

(etc.)

Example 2: More complicated scenario: Polar bears and climate change. Temps are rising in the Arctic. What is the effect on polar bears? Base model: discrete logistic model. How will rising T impact polar bear population ecology? Which parameters could be impacted?

Version #1: Rising temperature melts sea ice, so carrying capacity declines due to habitat loss. Assume no other impacts. Logic: we need an equation for how temperature will change through time, and we also need a relationship between carrying capacity and temperature. These two things will tell us how carrying capacity will change through time. See image.



```

for(t in 1:100){

  # update parameter values
  Tnow <- get it from T(t) equation
  Know <- plut Tnow into K(T) equation, use it below!

  # update state variables (using updated parameters)
  n[t+1] <- n[t] + rn[t](1 - n[t]/Know)

}

```

## 1 May 2019

Deterministic vs. stochastic models, continued... see slides.

Stochastic effects are especially important when numbers are small, such as during invasion or near extinction. Demographic stochasticity (variation arising because individual outcomes are not certain) vs. Environmental stochasticity (variation arising from fluctuations in the environment or rare unpredictable events).

You can write stochastic models where you write analytic stochastic models and explore their behavior, such as with Monte Carlo approaches, but in this class we'll just explore them with simulations.

“Flipping a coin” in R: see Jamie’s slides for more detail. For example, simulate “heads” occurring with probability  $p = 0.6$ :

```

coinflip <- runif(1) # generates a random number from [0,1]
if(coinflip > p) {
  # do something
} else {
  # do something
}

```

In the pseudo-code above, `runif()` simulates the coin flip, and then the if statement checks to see whether the flip is head or tails by incorporating the probability. You can simulate a multi-sided coin by checking the cumulative sum of the probabilities for each side (see slides) using the function `cumsum()`.

Instead of simulating a two-sided coin flip each time, you can draw from the binomial distribution (discrete distribution that describes the probability of getting a particular number of successes (k) when you conduct N independent trials (which succeed with probability p and fail with probability 1-p)).

$X \sim \text{Binomial}(N, p)$  then the random variable X can take values from 0, 1, 2, ..., N and the expected value (aka mean) is  $E[X] = Np$ . In R, you can use the function `rbinom()` to draw a random number from this distribution. See slides for syntax.

### Approaches to including stochasticity in your models (add from Jamie's slides)

#### *Environmental stochasticity*

- if environment can be described as a small number of states, flip a coin each year to decide what state the system is in and then use the parameter values that correspond to that state - if environment is more complicated, for example if there is a continuous range of environmental conditions, then draw parameter values from a distribution that describes how the parameters vary - if environment has occasional catastrophes, flip a coin to decide when the catastrophe will occur, and then use a binomial distribution to determine survival from the catastrophe. S

Example 1: discrete states. 45% chance of “good year” vs bad year. Randomly draw from a distribution to see what the year is, and then use the respective parameters.

Example 2: continuous states: draw parameters from two different distributions.

Example 3: every year there is a 5% chance of a catastrophe that affects survival with varying probability. See slides for pseudo-code.

## 6 May 2019

What to do with stochastic models:

- calculate mean value of some outcome
- calculate variance of some outcome
- calculate the probability of some outcome

### Managing complexity in models

We can always build a set of equations. What about BIG systems? For example, let's say you want to build a patch model for an archipelago with 30 islands. That's a lot of compartments and arrows to write down manually! Another example: sea lions live many years, so you could be keeping track of numerous age classes and sexes for up to 40 different groups if you want to keep track of every year. This would be a lot of equations to write by hand, and it could be easy to make mistakes. If you want to use lots of equations, use *matrix models* to keep track.

We'll see how to use matrix models to study age/stage structured populations (stage could mean developmental stage, which do/don't map consistently on to age), as well as spatially structured populations.

### Lightning intro/review of matrices

A matrix is an array of numbers. We typically use a large capital letter to denote it, and large brackets to define it. Matrices have rows and columns.

$$\mathbf{A} = \begin{bmatrix} 4 & 2 \\ 0 & -1 \end{bmatrix}$$

In the example above, A is a  $2 \times 2$  matrix.  $A_{ij}$  is entry in row i, column j. For example,  $A_{2,1} = 0$ .

Matrix math:

- Can only add/subtract matrices of the same size. Do it element by element. For example, if  $A + B = C$ , then  $C_{ij} = A_{ij} + B_{ij}$
- Can only multiply matrices A and B if their inner dimensions match. For example, if A is a  $2 \times 2$  matrix and B is a  $2 \times 3$  matrix, we can multiply them, because there are 2 columns in A and two rows in B. Their product will have the dimensions of the outer dimensions; i.e.  $2 \times 3$ . However, you couldn't multiply B and A.
- The transpose of the matrix is a matrix where you have flipped the rows and columns. For example:

$$\mathbf{D} = [ \begin{array}{ccc} 1 & 4 & 2 \end{array} ]$$

$$\mathbf{D}^T = \left[ \begin{array}{c} 1 \\ 4 \\ 2 \end{array} \right]$$

- So, you couldn't multiply  $B$  times  $A$  (matrix multiplication is not commutable), and you couldn't multiply  $A$  times  $B^T$ .

Doing matrix multiplication: Given a  $2 \times 2$  matrix A and a “column vector” (aka  $1 \times \_$  matrix) B:

$$\mathbf{A} = \left[ \begin{array}{cc} a & b \\ c & d \end{array} \right]$$

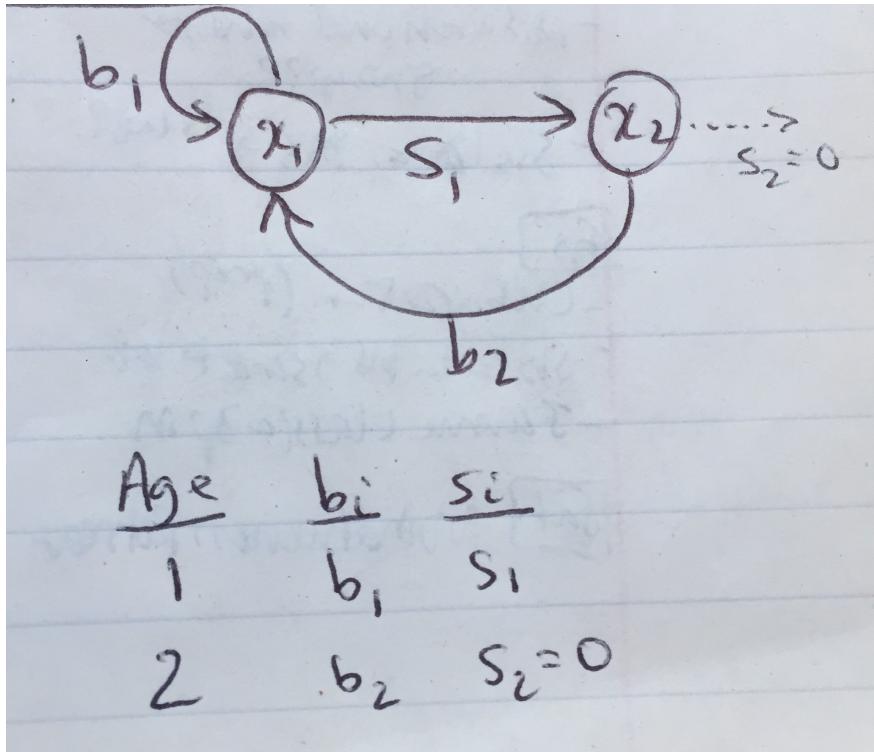
$$\mathbf{B} = \left[ \begin{array}{c} m \\ n \end{array} \right]$$

Multiply the first item in the first row of A times the first item in the first column of B and add that to product of the second item of the first row of A and the second item of the first column of B (and continue until you run out of items in the first row of A/first column of B). This sum of products is in the first row and first column of the output matrix. Continue with the second row of A:

$$\left[ \begin{array}{cc} a & b \\ c & d \end{array} \right] \left[ \begin{array}{c} m \\ n \end{array} \right] = \left[ \begin{array}{c} am + bn \\ cm + dn \end{array} \right]$$

Age-structured models: Consider a population whose individuals live up to 2 years and have offspring both years. Let  $x_i(t)$  = number of individuals of age i at time t (discrete time).

Flow diagram/“life cycle graph”; can convert this into a life table: See image.



Then, write these as equations:

$$\begin{aligned} x_1(t+1) &= b_1 x_1(t) + b_2 x_2(t) \\ x_2(t+1) &= s_1 x_1(t) \end{aligned}$$

Note that first years either die or move on to year 2, so our equation accounts for this. For example, we could say  $x_1(t+1) = x_1(t) - dx_1(t) - s_1 x_1(t) + b_1 x_1(t) + b_2 x_2(t)$  but  $x_1(t) = dx_1(t) + s_1 x_1(t)$  so these terms all cancel each other out and we get the equation with just births.

Let's express these equations in matrix form: we have some vector  $\bar{x}(t+1) = L\bar{x}(t)$  where

$$\mathbf{L} = \begin{bmatrix} b_1 & b_2 \\ s_1 & 0 \end{bmatrix}$$

and

$$\bar{\mathbf{x}}(t) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

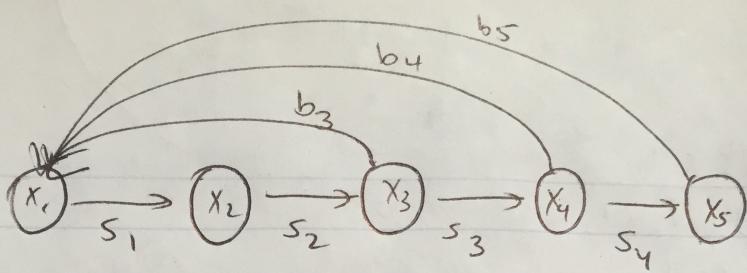
$L$  is the Leslie matrix. The Leslie matrix is a projection matrix; it projects the age-structured population from  $t$  to  $t+1$ . Each element of the Leslie matrix  $L_{ij}$  is the per capita contribution of stage  $j$  to stage  $i$ .

## 8 May 2019

In the Leslie matrix, the rows are where the contribution is going “to”; the columns are where the contribution is coming “from”. Again,  $L_{ij}$  is the per capita contribution of stage  $j$  to stage  $i$ . Usually we write the stages in the matrix as youngest to oldest, so usually the first row is the birth rate from each of the stages.

Generalize: another example: Animal that lives up to 5 years, starts reproducing in year 3. Age-specific parameters  $s$  and  $b$ .

See image.



$$x_1(t+1) = b_3 x_3(t) + b_4 x_4(t) + b_5 x_5(t)$$

$$x_2(t+1) = s_1 x_1(t)$$

$$x_3(t+1) = s_2 x_2(t)$$

$$x_4(t+1) = s_3 x_3(t)$$

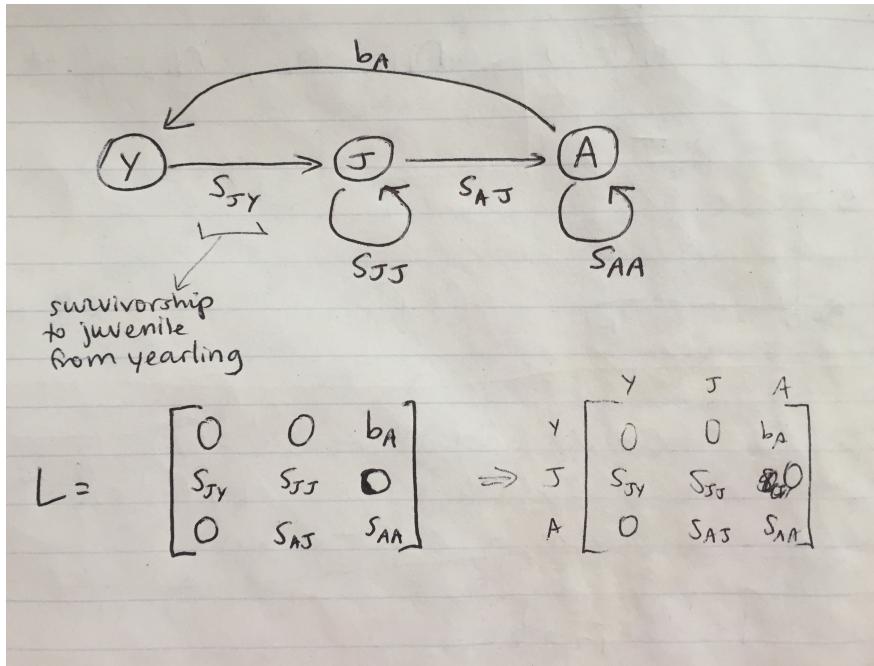
$$x_5(t+1) = s_4 x_4(t)$$

$$L = \begin{bmatrix} 0 & 0 & b_3 & b_4 & b_5 \\ s_1 & 0 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 & 0 \\ 0 & 0 & s_3 & 0 & 0 \\ 0 & 0 & 0 & s_4 & 0 \end{bmatrix}$$

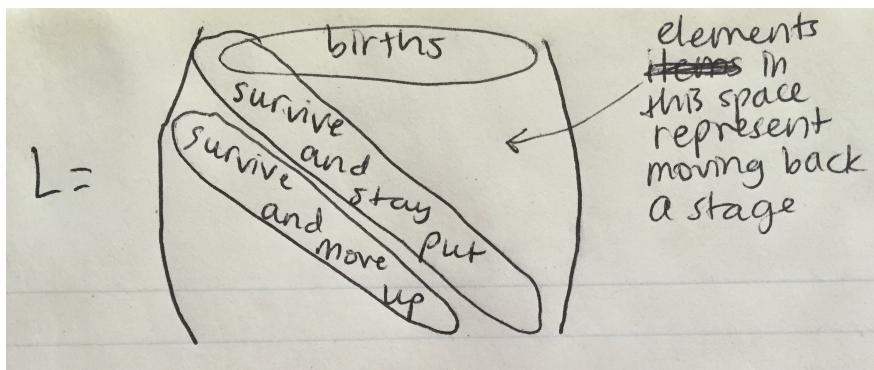
### Stage-structured models

Sometimes it makes more sense to think about the system in stages instead of years. For example, yearling, juvenile, adult. One reason is that it's simpler and appropriate for the system. Another reason is that sometimes movement between stages depends on, for example, the environment, and having a stage-structured model allows you to model that.

See image.



Some people in the literature only call age-based models Leslie models, and stage-structured models Leslie-Lefkovitch models. See image for description of what elements in the Leslie matrix mean. Note that it usually doesn't make sense for a stage to contribute to a previous stage (besides births), but this can make sense in some developmental instances, and also if different stages are patches in space.



Also note that if the yearlings can give birth too, and can remain in the yearlings class, then  $L_{11} = b_Y + s_{YY}$ .

Can we make this density-dependent? One way is to make the birth rate or survivorship density-dependent in the Leslie matrix, but this makes the model less straightforward. You'd have to do this in simulation mode, where you change the values of  $L_{ij}$  every run of a for loop.  $\bar{x}(t+1) = L\bar{x}(t)$  is fundamentally linear, so you can't express this as a density-dependent expression. It's still a powerful tool, but it's specific, where you can describe the system where every parameter is a static per capita rate, and if you want to do density-dependence then you need to simulate.

### What do we do with this model?

*Projecting population dynamics:*

$$\begin{aligned}
\bar{x}(t+1) &= L\bar{x}(t) \\
\bar{x}(t+2) &= L\bar{x}(t+1) \\
&= L * L\bar{x}(t) = L^2\bar{x}(t) \\
\bar{x}(t+N) &= L^N\bar{x}(t)
\end{aligned}$$

You can then plot different things:

- plot  $N[t+1]/N[t]$  over time - plot proportion in each age class/time - plot total population/time - plot number of one age class/time

Next time: we can predict what these dynamics look like

## 13 May 2019

Matrix Models ~ Eigen Analysis

Observations from our simulations:

1. Growth rate of the total population size,  $\frac{N(t+1)}{N(t)}$  approaches a constant value. This is called the long-term growth rate, called  $\lambda$ .
2. The fraction of the population in each age class also approaches a constant value. This is called the stable age distribution. We're going to write this as a vector  $\bar{w}_1$  (less universal notation than  $\lambda$ ).

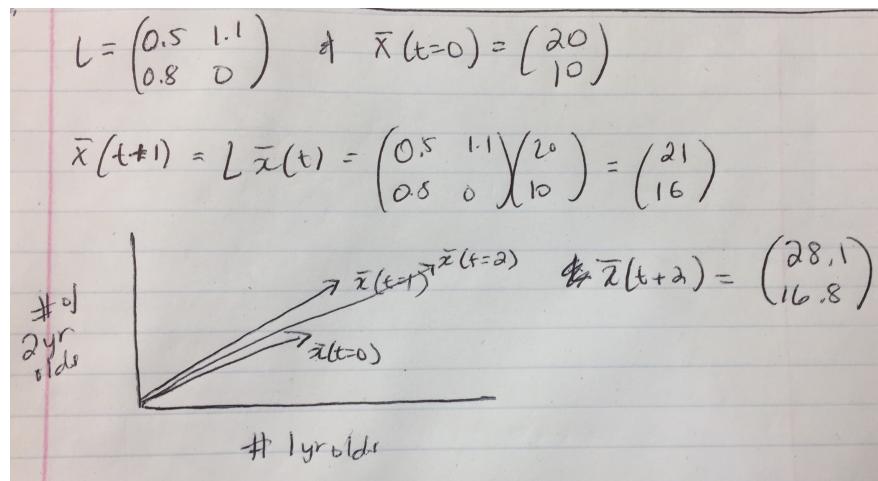
Aside: Eigen-analysis is a mathematical technique to summarize important information/properties of square matrices. We saw that whenever you use projection matrices in an equation where you are multiplying a state vector, you end up with square matrices. Find all of the solutions for  $\lambda_i$  and  $\bar{w}_i$  that solve this equation (the eigenvalue equation):

$$L\bar{w}_i = \lambda_i\bar{w}_i$$

.  $\lambda_i$  is a number, aka a 1x1. The  $\lambda_i$  are eigenvalues, and the  $\bar{w}_i$  are eigenvectors. "Eigen" is a German word that means "self"; something about this is capturing "selfness" as it relates to the projection matrix and these vectors.

In general, if you multiply a vector by a matrix, it change direction (here, that tells us that the proportion of age classes has changed) and also magnitude. A matrix is what's called an operator, and changes things about the vector it's operating on. Typically, you get a rotation, a stretch (or shrinkage).

See image.



If we just keep multiplying the state vector by the matrix, we'll eventually get to a point where we multiply the vector by the projection matrix, and we'll get the same vector (same direction), but just a little longer

or shorter. For the system we have in this example, all of the vectors will be pointing in the same direction, but they will keep getting longer. This means that the proportion of individuals in the different age classes will settle down to something constant, and once the system settles down to a constant proportion of age classes,  $\lambda$  tells you how the total population size is changing over time.

For an  $n \times n$  matrix, you will get  $n$  solutions for  $\lambda_i$ . Each  $\lambda_i$  has a corresponding  $\bar{w}_i$ . The Leslie matrix has special properties. In the one that you start with, there are some non-zero and some zero elements of the matrix—ie.,  $L_{ij} \geq 0$ . Usually you raise Leslie matrix to a high power, all of the elements will be greater than 0. This means that all states/classes are in contact with each other. In matrix world, it means that L is “irreducible and primitive”. For an irreducible and primitive matrix (aka almost any Leslie matrix), that means that you get one eigenvalue that is bigger than all of the others, and is always positive. This is called the “dominant” eigenvalue,  $\lambda_1$ , which has a corresponding dominant eigenvector,  $\bar{w}_1$ . This comes from the Perron-Frobenius Theorem“.

$\lambda_1$  and  $\bar{w}_1$  characterize the information in L in a way that corresponds to high powers of L, which basically means that they describe the long-term dynamics of the population. And so,  $\lambda_1$  or just  $\lambda$  is the long-term growth rate, and  $\bar{w}_1$  is the stable age distribution (make sure you normalize it—aka that the entries in that vector need to add up to 1). Note that R doesn’t normalize the vector, so you will need to add up all of the entries and divide by the number of them, and use that resulting vector.

## 15 May 2019

Review of Leslie matrices so far...

$$\bar{x}(t+1) = L\bar{x}(t) \text{ where } L \text{ is nxn.}$$

Solve  $L\bar{w}_i = \lambda_i \bar{w}_i$ . L has eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  with corresponding eigenvectors  $\bar{w}_1, \bar{w}_2, \dots, \bar{w}_n$ .

They form a “basis”, which means that you write down any state vector  $\bar{x}(t)$  as a linear combination of the eigenvectors (can add them together in some combination). Linear combination means that there is some set of constants that exist that add up the eigenvectors that we can add up and reach any current state of the system.

This means that each time you multiply by L, yes, you are raising L to some power and multiplying it by  $\bar{x}(t)$ , but it’s also as if you are adding up different eigenvalues x eigenvectors, so it’s basically you are writing the state of the system in terms of eigenvalues raised to the number of steps of the system.

$$\bar{x}(t=0) = c_1 \bar{w}_1 + c_2 \bar{w}_2 + \dots + c_n \bar{w}_n$$

$$\text{General solution: } \bar{x}(t) = c_1 \lambda_1^t \bar{w}_1 + c_2 \lambda_2^t \bar{w}_2 + \dots + c_n \lambda_n^t \bar{w}_n$$

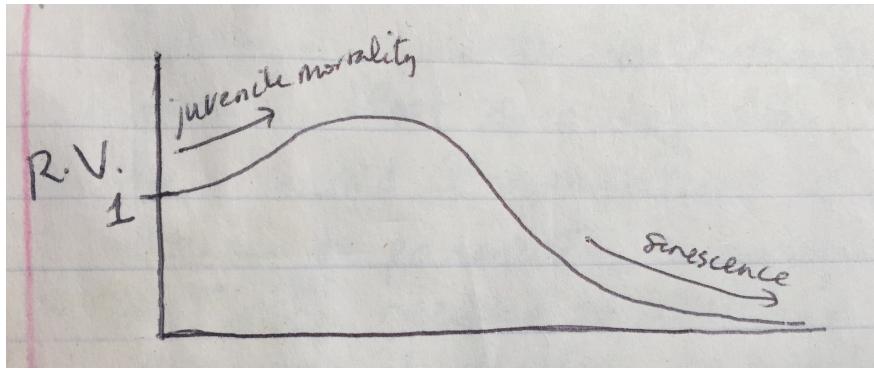
So, as you can see, over time, these  $\lambda_i$  are being raised to higher and higher powers, so the  $\lambda_i$  that is the biggest, is having the biggest contribution to the current state of the system, gets way bigger than all the other  $\lambda_i$  to the point where their contribution is negligible.

As t gets large,  $|\lambda_1^t| >> |\lambda_{i \neq 1}^t|$ , so for large t,  $\bar{x}(t) = c_1 \lambda_1^t \bar{w}_1$ .

Take-homes:

1. Longterm growth rate given by dominant eigenvalue  $\lambda_1$  or  $\lambda$ .  $\lambda > 1$  means growth, and  $0 < \lambda < 1$  means longterm decline.
2. Stable age distribution given by dominant eigenvector, appropriately normalized.
3. Other eigenvalues/eigenvectors tell you about transient dynamics (these can matter a lot, because if a system is experience perturbations, it might be very influenced by transient dynamics).
4. Reproductive values.
5. Sensitivity and elasticity analysis.

*Reproductive value* is the expected contribution of a female of age i to present and future reproduction. Typically RV assumed to start at 1. It goes up (instead of starting at the highest possible value) because it first factors in juvenile survival, and goes down at the end due to senescence. Calculate this using the left dominant eigenvectors of L.



Solutions to  $L\bar{w}_i = \lambda_i \bar{w}_i$  give the right eigenvectors  $\bar{w}_i$ . But, we can also solve:  $\bar{v}_i L = \lambda_i \bar{v}_i$  which gives the left eigenvectors  $\bar{v}_i$ . Note that these are the SAME eigenvalues. Also note that the dimensions of  $\bar{v}_i$  are 1xn (row vector now).

Trick: in R, you can calculate  $\bar{v}_i$  as the right eigenvectors of  $L^T$ . In R, eigen(L) gives you all of the eigenvalues and the right eigenvectors of L.

Once you have these, pick the dominant left eigenvector, normalize it so the first entry is 1, and then you have the reproductive values!