Contents

Chapter 1
Sensitivity Background

Abstract

1.1 History of SA

1.2 Simple Measures

1.2.1 Visual: e.g. Scatter Plots

1.2.2 Correlation Measures

1.2.3 One at a time – analytic versus other

1.3 Global Methods

1.3.1 Partial Rank Correlation Coefficient

1.3.2 Sobol' Measures

1.3.3 Importance Measures

1.4 To sample or not to sample

1.5 Monte Carlo

1.6 Rules of Thumb

1.7 Running list of methods

1. Scatter plots
2. regression
3. Derivative based
4. Pearsons correlation
5. Kendall?s correlation
6. Partial Rank
7. Importance
8. Gamma indices
9. Distributional tests: Smirnov, Von Mises

Chapter 2
Ecology

Abstract

2.1 Historical Background

Ecology is the study of organisms, their environment, and their interactions. It has been a distinct branch of science since at least the 1850's although studies of populations and their distributions were performed much earlier. In fact, one could argue that much of ancient Greek science revolved around understanding nature and how organisms and their environment fit together. However, the term 'ecology' was not coined until 1866 and it was only in common usage by the early 1900's when the British Ecological Society was formed. At that point some scientists referred to themselves as ecologists rather than the previously all encompassing term naturalists and by 1950 the idea of "the study of the structure and function of ecosystems," [**?**] was firmly established.

In some sense, ecology is a specialization of naturalism – that is the study of nature writ large. Naturalists grouped organisms in classes, developing methods for aggregating observations. There is evidence of the classification of plants during ancient Greece. Several hundred years later similar, more detailed classification of ant populations were performed. Both of these examples are focused on classification – how to group populations according to similarities and differences. One of the difficulties with this is that there was not a systematic way to rank similarities and differences. Even now it is well known that very small variations at the genetic levels can lead to large variations in phenotypes. It is also difficult to determine the organization of separate types of organisms. Without specific tests, comparing across types becomes problematic (consider the platypus!).

Major advances in ecology began in the 18th century, beginning with classification systems proposed by Carlus Linnaeus which is essentially the system still used today. However, in some sense this contribution reflects a continuation of the Greek method of classifying with the germ of the scientific revolutionary idea of un-

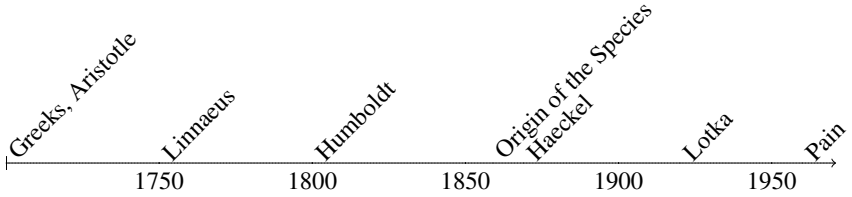
derstanding *why* the classifications are correct and *how* the defining characteristics emerge.

In the early 1800's Alexander Humboldt was one of the leading scientists of the day. He travelled extensively and collected prodigious data on species and, more importantly, their habitat. He was the first to describe the range of a species in conjunction with the classification and is often referred to as the Father of Ecology. In 1859 Darwin published the origin of the species which refined the methods of observation used by Humboldt, but most importantly Darwin proposed a general theory to explain the mechanisms behind the observations. In the 20th century, ecology turned to include human populations and in 1953 the first ecology textbook was written by Eugene and Howard Odum [**?**].

The idea of classification naturally turns to questions regarding the cause of distinct classes. Exploring causes naturally leads to the question of interrelationships – who eats whom? Who associates with whom? Who is related to whom? And then, inexorably to theoretical frameworks that can explain why these interrelationships exist. Hypotheses like the food chain that relates predators and prey hierarchically provide a framework to direct observations in a quest to discomfirm or support the hypothesis. As observations accumulate the simplistic linear view of relationships between organisms became untenable. Linear relationships between populations was eventually extended to a web of interconnected chains. In this, species interact directly and also indirectly.

The food web assigns a passive directionality in an ecosystem since energy/nutrient flows from prey to predator. It seemed apparent that it was the prey or herbivores that kept the populations in check. However, this theory was completely altered when Paine performed his classic experiment by creating a food web and removing a single predator species of starfish (*Pisastor ochraceus*) and showing that this drastically altered the community balance. This demonstrated directly that predators help maintain the ecosystem. Further demonstrations lead to the concept of keystone species, one that is more important that others in an ecosystem.

As ecology has become less of an observational science and more a quantitative science, specific questions can be posed. This is a standard development in a science. To answer the standard How? When? and Why? questions of science, ecology uses a mix of observations (weak inference), experimentation (strong inference) and theory. It is primarily the last of these that is most of interest to us.



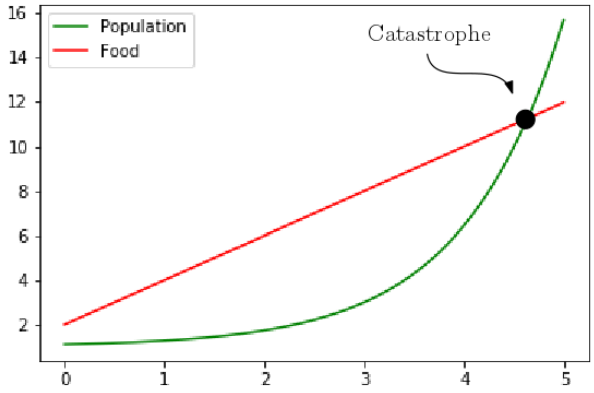


Fig. 2.1: Sketch of a Malthusian catastrophe.

2.2 Theoretical models

The theory of population dynamics began with observations of population growth within cities in the 18th century. During this period, Europe was becoming more industrial and less rural. As cities became larger, poverty and sanitation became rampant issues of the day. Thomas Malthus attempted to determine natural laws (e.g. theory) to explain the relationship between population growth and poverty. In his theory, population growth and food supply were governed by two different laws, with population accelerating and resources being produced at a constant rate. When the population outruns the food supply a catastrophe occurs (Figure 2.1).

2.2.1 The Logistic Model

One criticism of Malthus' theory is the observation that population growth causes resource scarcity. That is, the growth rate of the population should be directly tied to resource availability and similarly resources should be related to the population. In 1838, Pierre-François Verhulst argued that the growth rate should depend on the population level, which implicitly assumes dependence on resource. If the population is low the rate of growth is positive and if it is high the rate is negative. Verhulst proposed the differential equation, which he termed the *logistic* equation,

$$\begin{aligned} \frac{\partial P}{\partial t} &= r(1-\frac{P}{k})P \\ &= g(P;r,k)P. \end{aligned} \tag{2.1}$$

Where we think of the growth rate as $g(P;r,k)$ as a function of the populations rather than a constant as Malthus did.

2.2.2 Analysis

Qualitative analysis for the logistic equation is relatively straightforward. Recall that for autonomous differential equations the right-hand-side defines regions where P is increasing and decreasing – these are separated by special, constant solutions referred to as equilibria. In this case $g(P) = 0$ when $P = 0$ or k.

Equilibria solutions are in fact *bona fide* solutions. If $P_0 = 0$ the population remains zero for all time. If $P_0 = k$ then $\frac{dP}{dt} = 0$ and the population remains at the carrying capacity for all time.

But what happens if the population is not one of these special solutions? There are two answers to this – a local answer and a global answer. As is almost always the case with mathematics the principle of TANSTAFFL (There Ain't No Such Thing As a Free Lunch) means that the easiest answer, in this case the local behavior, is often not the most useful answer. While there is a lot of theory that helps bridge the gap, we will primarily use numerical methods. Once again, this requires care! Numerical methods are not (T)ruth... and should not be interpreted this way; however, in general we will lean on them.

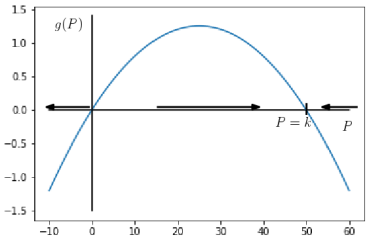
The answer to the behavior near the equilibria or the local answer rests on linearization. We write the solution $P = \bar{P} + \varepsilon\hat{P}(t) + \mathcal{O}(\varepsilon^2)$, where $\bar{P}$ is either 0 or k. Plugging this into Equation 2.1 and expanding the right-hand-side is,

$$\begin{aligned} r(1-\frac{P}{k})P &= r(1-\frac{\bar{P}+\varepsilon\hat{P}(t)+\mathcal{O}(\varepsilon^2)}{k})(\bar{P}+\varepsilon\hat{P}(t)+\mathcal{O}(\varepsilon^2)), \\ &= r(1-\frac{\bar{P}}{k})\bar{P}+\varepsilon r\hat{P}(1-\frac{2\bar{P}}{k}), \\ &= \begin{cases} r\hat{P} & \text{if } \bar{P}=0 \\ -r\hat{P} & \text{if } \bar{P}=k. \end{cases} \end{aligned}$$

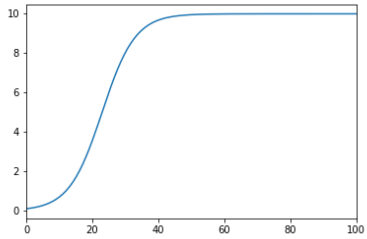
$\hat{P}$ increases near $P = 0$ and decreases near $P = k$ and population extinction is unstable and the population tends towards its carrying capacity. Plotting the phase-line along with g, we can read off the qualitative behavior (See Figure 2.2a).

2.2.3 Predator/Prey – Lotka-Volterra

As ecological hypotheses accumulated and more observations were mode of real ecosystems, the concept of a food web became the dominant paradigm. Here preda-



(a) Phase line behavior of the trajectories of the logistic equation. It is easy to see that $P = 0$ is unstable while $P = k$ is stable. Any nonzero initial condition will tend towards the carrying capacity, k.



(b) State-space trajectory for the logistic equation with $r = .1$ and $k = 10$.

tors eat prey and prey are eaten by predators (and may prey on creatures further down the food chain). A natural next step in modeling populations is to consider one link in the chain instead of treating the interaction between the population and the resource implicitly.

Working with the computer codes

We can do the same calculations using numerical approximations.

$$\frac{dP}{dt} = rP(1 - \tfrac{P}{k}) = f(P)$$

To find the steady-states, we have to solve $f = 0$. We call the Python/*R* code, `SS.py`/`SS.r`: `SS`(f,[a,b]) where $[a,b]$ is the search interval – some region where we think the steady-states might live. The search intervals matter a lot! Larger intervals lead to longer run-times and the possibility of jumping over steady-states but smaller search intervals restrict the domain. Unfortunately there is not a perfect method for this – one can spend much of calculus learning how to estimate intervals where functions have zeros... we *do not* want to do this here. Instead, we will augment our searching with curve plotting. Simple enough for this example (See Figure 2.2a). Running our scripts we find that , ...

Once we have steady-states, we need to determine stability by looking at the linearization. We run `LinearStability`(f,x_0) for each of the steady-states found in `SS` and everything matches up with our analytic treatment.

We can check whether trajectories near the equilibria agree with our calculations. We do do this by running the code `Traj`(f,x_0) with an initial population near the steady-states. This also agrees with our notion of stability.

Finally, we can see what happens far from equilibria – run `Traj`(f,x_0) with any old initial value you want!

Motivated by this idea Lotka and Volterra developed a two-species interaction model. In fact, Lotka had been interested in biological models that were inspired by chemical models. Differential equation models in chemistry were relatively well grounded at the time and analogies between mass action and population interactions were being explored. At the same time, Volterra was interested in observations of predator fish shortly after WWI. There was a clear variation in the percentage of predator species caught before, during and after the war. Simultaneously there were variations in the prey species observed. During the war, Adriatic Sea had been fished extensively and presumably this played a role in the variation from a natural, base state. Volterra posed a two-species model consisting only of predators and prey. This model is equivalent to Lotka's model derived using mass action.

The model relates the change in population density of the predator species, H for hunters, with the prey species, P. We assume that the hunters require prey to survive and that otherwise their population decreases. Prey, on the other hand, have an abundance of food, so in the absence of hunters the prey grow at a constant rate. The two species interact with a rate proportional to the product of their densities (this is just mass action). When they meet, the predators convert prey into more predators while the prey are consumed. The equations for this are,

$$\frac{dH}{dt} = \alpha HP - \delta H \tag{2.2}$$

$$\frac{dP}{dt} = -\beta HP + \gamma P \tag{2.3}$$

We have to specify initial conditions, $H(0) = H_0$ and $P(0) = P_0$.

2.2.4 Analysis

While this is a fairly simple model it is surprisingly rich in behavior. There are several directions that we could explore at this point. One of the most fruitful is to linearize these equations near each equilibria. This provides information about the local (in phase-space) behavior and the long-term behavior. While this is quite useful and often the only 'analysis' that can be done, it does not reflect the behavior away from the steady-states – which is often the most useful information for an application. To estimate the behavior in other regions of phase-space, we can explore numerical approximations. Lotka-Volterra equations are sufficiently simple to allow us to extend this to other analytic techniques that do not always generalize.

We will proceed with both of these methods noting only that this motivates using Lotka-Volterra as a canonical example, since we can do this *and* further analysis to help illustrate our methods. We begin with linearization. We already saw this in the Section 2.2.2 in the case of a single equation. Just as we can use Taylors theorem to approximate any smooth function by a line (as long as we don't move the argument too much!), we can approximate the dynamics of a nonlinear system with that of a

linear system. Again, we have to stay in some neighborhood – this is a local result. In this case, the natural place to stay near is one of the equilibria, $(\bar{H},\bar{P})$. We will show shortly that there are two equilibria. We write the general solution for H and P as a perturbation from the equilibria,

$$H = \bar{H} + \varepsilon\hat{H} + \mathscr{O}(\varepsilon^2), \tag{2.4}$$
$$P = \bar{P} + \varepsilon\hat{P} + \mathscr{O}(\varepsilon^2), \tag{2.5}$$

where all of our errors coming from the approximation are lumped into the $\mathscr{O}(\varepsilon^2)$ term. Plugging these into Equations 2.2 and 2.3 we get,

$$\begin{aligned}\frac{d\left(\bar{H}+\varepsilon\hat{H}+\mathscr{O}(\varepsilon^2)\right)}{dt} &= \alpha\left(\bar{H}+\varepsilon\hat{H}+\mathscr{O}(\varepsilon^2)\right)\left(\bar{P}+\varepsilon\hat{P}+\mathscr{O}(\varepsilon^2)\right)\\ &\quad -\delta\left(\bar{H}+\varepsilon\hat{H}+\mathscr{O}(\varepsilon^2)\right), \end{aligned}\tag{2.6}$$
$$\begin{aligned}\frac{d\left(\bar{P}+\varepsilon\hat{P}+\mathscr{O}(\varepsilon^2)\right)}{dt} &= -\beta\left(\bar{H}+\varepsilon\hat{H}+\mathscr{O}(\varepsilon^2)\right)\left(\bar{P}+\varepsilon\hat{P}+\mathscr{O}(\varepsilon^2)\right)\\ &\quad +\gamma\left(\bar{P}+\varepsilon\hat{P}+\mathscr{O}(\varepsilon^2)\right). \end{aligned}\tag{2.7}$$

By expanding everything and collecting in terms of ε we find,

$$\frac{d\bar{H}}{dt}+\varepsilon\frac{d\hat{H}}{dt}+\mathscr{O}(\varepsilon^2) = \alpha\bar{H}\bar{P}-\delta\bar{H}+\varepsilon\left(\alpha(\hat{H}\bar{P}+\bar{H}\hat{P})-\delta\hat{H}\right)+\mathscr{O}(\varepsilon^2), \tag{2.8}$$
$$\frac{d\bar{P}}{dt}+\varepsilon\frac{d\hat{P}}{dt}+\mathscr{O}(\varepsilon^2) = -\beta\bar{H}\bar{P}+\gamma\bar{P}+\varepsilon\left(\beta(\hat{H}\bar{P}+\bar{H}\hat{P})+\gamma\hat{P}\right)+\mathscr{O}(\varepsilon^2). \tag{2.9}$$

Because we assume that H and P are smooth enough functions, we know that the coefficients of ε on the left must equal those on the right. This gives a hierarchy of equations beginning with the leading order (ε^0)

$$\frac{d\bar{H}}{dt} = 0 = \alpha\bar{H}\bar{P}-\delta\bar{H}, \tag{2.10}$$
$$\frac{d\bar{P}}{dt} = 0 = -\beta\bar{H}\bar{P}+\gamma\bar{P}, \tag{2.11}$$

since $\bar{H}$ and $\bar{P}$ are constants. Notice that these are equivalent to the equations we needed to solve to find the steady-states and we know there are two solutions.

The next order equations are,

$$\begin{aligned}\frac{d\hat{H}}{dt} &= \alpha(\hat{H}\bar{P}+\bar{H}\hat{P})-\delta\hat{H},\\ &= (\alpha\bar{P}-\delta)\hat{H}+\alpha\bar{H}\hat{P}, \end{aligned}\tag{2.12}$$
$$\begin{aligned}\frac{d\hat{P}}{dt} &= \beta(\hat{H}\bar{P}+\bar{H}\hat{P})+\gamma\hat{P},\\ &= \beta\bar{H}\hat{P}+(\beta\bar{P}+\gamma)\hat{P}, \end{aligned}\tag{2.13}$$

where we have dropped the ε from both sides.

These define the dynamics of the perturbations and are linear! Why does this matter? Because we can solve linear equations. So we can determine whether $\hat{H}$ and $\hat{P}$ go to zero or not. If they do, then we know that the steady-state is stable in the sense that if you start in a neighborhood of the steady-state solutions, you approach them in time. Otherwise, we call them unstable. You should look back at the phase-line for the logistic equation, this is the same idea.

The predator-prey model, Equations 2.13 and 2.13, can be written as a matrix (See the Appendix for details),

$$\begin{bmatrix} \hat{H} \\ \hat{P} \end{bmatrix} = J \begin{bmatrix} \hat{H} \\ \hat{P} \end{bmatrix} \tag{2.14}$$

where J is a 2×2 Jacobian matrix,

To recap, the process of linear stability requires us to,

1. Find steady-states by solving $f_1 = 0, f_2 = 0, ...f_n = 0$ for $(x_1, x_2, ..., x_n)$.
2. Write the solution as a perturbation of the steady-states
3. Expand and look at the order ε terms
4. Determine the eigenvalues of the Jacobian
5. If all eigenvalues are negative, the steady-state is stable

Let's follow this procedure with the Lotka-Volterra Equations. The right-hand-side functions are $f_1(H,P) = \alpha HP - \delta H$ and $f_2 = -\beta HP + \gamma P$. It is easy to see that there are two equilibria that satisfy,

$$\alpha HP - \delta H = H(\alpha P - \delta) = = 0, \tag{2.15}$$

$$-\beta HP + \gamma P = P(-\beta H + \gamma) = 0. \tag{2.16}$$

So that either $H = 0$ or $P = \frac{\delta}{\alpha}$ from Equation 2.16 while $P = 0$ or $H = \frac{\gamma}{\beta}$ from Equation 2.16. The two steady-states are $(\bar{P}, \bar{H}) = (0,0)$, $(\bar{P}, \bar{H}) = (\frac{\delta}{\alpha}, \frac{\gamma}{\beta})$.

The Jacobian is,

$$\begin{bmatrix} \frac{\partial f_1}{\partial H} & \frac{\partial f_1}{\partial P} \\ \frac{\partial f_2}{\partial H} & \frac{\partial f_2}{\partial P} \end{bmatrix} = \begin{bmatrix} \alpha P - \delta & \alpha H \\ -\beta P & -\beta H + \gamma \end{bmatrix} \tag{2.17}$$

At $(0,0)$ this is,

$$J(0,0) = \begin{bmatrix} -\delta & 0 \\ 0 & \gamma \end{bmatrix} \tag{2.18}$$

The eigenvalues of $J(0,0)$ are $-\delta \leq 0$ and $\gamma \geq 0$, which implies that the origin (or where both populations are zero) is not stable. In fact, we can see that if we start with only predators, the population goes to zero since there is no food source while starting with only prey, leads to unbounded growth.

At the other steady-state, $(\bar{P},\bar{H}) = (\frac{\delta}{\alpha},\frac{\gamma}{\beta})$, the Jacobian is,

$$J\left(\frac{\delta}{\alpha},\frac{\gamma}{\beta}\right) = \begin{bmatrix} 0 & \alpha\frac{\gamma}{\beta} \\ -\beta\frac{\delta}{\alpha} & 0 \end{bmatrix} \tag{2.19}$$

Working with the computer codes: Steady-states

Now we can run through the same process as before, but with the Lotka-Volterra model. We have differential equations of the form,

$$\frac{dx}{dt} = f_1(x,y)$$
$$\frac{dy}{dt} = f_2(x,y)$$

To find the steady-states, we have to solve three equations simultaneously: $f_1 = 0$ and $f_2 = 0$. In general, this is not simple. So we call the Python or *R* code, `SS.py` (`SS.r`). python `SS`$(f_1, f_2, [a,b])$ and we find numerical estimates steady-states. We can get some ideas of where the steady-states might live by just plotting $f_1(x,y)$. But this is a three-dimensional plot (two inputs and an output. It is relatively difficult to visualize this, but we can use a simple plotting technique. If we plot contours of the function, we get a topographical map of the function. It is a simple thing (once someone wrote a script for this!) to look only for contours at height zero. These are sometimes called the zero-level sets of a function. So we can look for zero level sets of f_1 and f_2. Where they intersect are the equilibria. Try running `contour`$(f_1, f_2, [a,b]$. Visually, you can see where the crossings are. This will give estimates to feed into `SS`. Why do we have to do both of these? Why not just run `SS` over a huge region? The answer is efficiency and accuracy. The crossings that show in contour plots only give rough estimates of the quantitative values of the intersections (so not very accurate). But you can sketch this over very large regions quickly (so reasonably efficiently). Finding the quantitative values using `SS` can be made very accurate, but it takes more time. Try running `SS` over larger and larger domains and time the results...

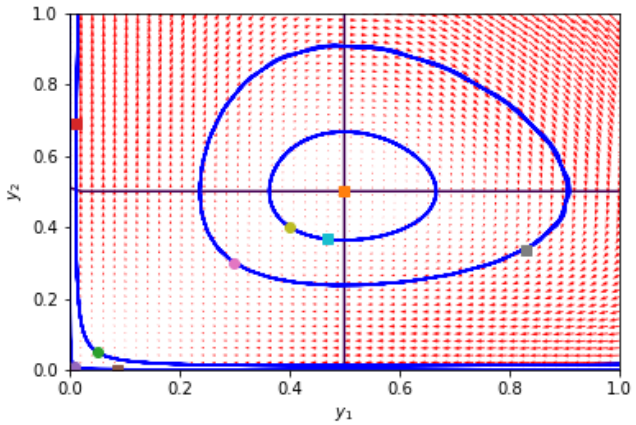


Fig. 2.3: Phase-plane– not consistent with the time course

This is pure rotation – since the eigenvalues are purely imaginary. All solutions near this steady-state are periodic.

The phase-plane analysis is also quite telling. Recall that for planar systems (two state variables), the behavior is organized by the nullclines ($f_1 = 0$, $f_2 = 0$). The trajectories (solutions) just follow the direction field. Starting with an initial population of predators and prey, the trajectories in the phase plane form closed curves (see Figure 2.3).

The trajectories are found by plotting the solution to the differential equations $H(t)$ and $P(t)$ as ordered pairs (H,P). The parametric curve in the phase-plane corresponds to the curves in state space as shown in Figure 2.4.

So which version provides the most detail? or the most intuition? As is the answer in most cases, the answer is 'It depends!'. It depends greatly on what your goals are.

Working with the computer codes: Jacobian

Once we have steady-states, we need to determine the sign of the eigenvalues of the Jacobian – recalling that if the real parts are positive the steady-state is unstable. We call $\texttt{Jac}(f_1, f_2, [\bar{x}, \bar{y}])|$. There are two equilibria to consider – the trivial solution, $(\bar{P}, \bar{H}) = (0,0)$ and the non-trivial one, $(\bar{P}, \bar{H}) = (\frac{\delta}{\alpha}, \frac{\gamma}{\beta})$, so we will have to run the scripts for both values. The script returns the eigenvalues and eigenvectors; however, we will not focus on the eigenvectors in this chapter.

In the context of sensitivity, it might be a bit easier to examine the state-space plot if one wanted to quantify the phase difference in the two curves. If one wanted to know the maximum or minimum either view is sufficient. One thing that is not as apparent in the phase-plane is the period since the trajectories repeat the same curve periodically. But it is important to understand that these are just different views of the solution to the equations, just obtained numerically.

2.2.5 Sensitivity

Based on the analysis in the previous sections, we can outline a few different questions that can be approached using sensitivity analysis. Obvious targets are things like the steady-states, maximum or minimum values, period etc. Thinking of this as an ecological model, one might be motivated by determining if the average popu-

Working with the computer codes: Trajectories/Phase-plane

To understand the behavior more (and to double-check our stability calculations), we use the script `Traj`(f_1, f_2, H_0, P_0) to explore the behavior. If we start 'close' to (H_0, P_0) we see behavior consistent with the linear theory – that is $(0,0)$ is unstable and there are rotations around $(\frac{\delta}{\alpha}, \frac{\gamma}{\beta})$.

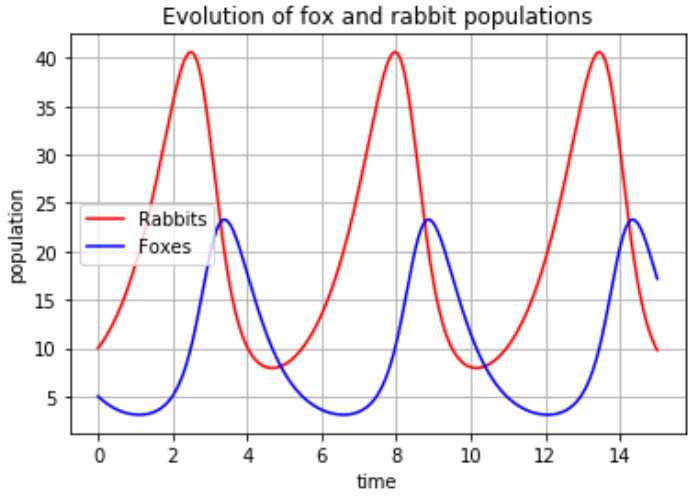


Fig. 2.4: A time course of the dynamics showing periodicity. For this simulation... description of the code used to generate this

lation is sensitive to variations in parameters. Each of these has some plusses and some minuses. For example, the steady-states can be calculated explicitly, so the sensitivity can be obtained explicitly. However, they have less biological meaning – in fact the population never actually reaches the steady-states since they are not attractors.

Focusing on the period is also instructive – this has some biological interpretation and might be important if one was comparing with observations since the observations must be structured in order to capture the oscillations. Therefore understanding the period helps with the design observational studies. One difficulty is that the periodicity is a local result – The eigenvalues of the Jacobian provide estimates of the period, but these estimates are not valid far from equilibrium. To capture the period outside the region where linearization fails (a region which is typically difficult to determine) requires numerical methods that can capture periodicity. This is not a simple task.

We will consider the maximum and minimum values within a window of time. One trick will be to let the window be a parameter so that we can determine whether we need to be careful about this. That is, we will let the sensitivity analysis help point the way to the natural period that needs to be considered. We do the simplest thing that we can think of – fix all parameters except for the parameter of interest. We 'roll the dice' (e.g. sample from a distribution) values of this parameter. This gives a full parameter set that we use to calculate the solution to the system of equations. We plot success pairs of the parameter value against the quantity of interest (QoI), here the maximum of the predator species. This scatter plot provides some inference into the how the output (QoI) depends on the input. This is some notion of sensitivity – if the scatterplot appears to be flat, there is little sensitivity. If the trend is increasing or decreasing there is some correlation (positive or negative) and the steepness of the trend is an estimate of sensitivity.

We will see later that there are at least two major issues with this sensitivity measure. The first is that scatterplots only make sense if one parameter is changed at a time. There is no way to examine cross-correlation or synergistic effects. The second difficulty is in quantifying the sensitivity. Calculating trendlines is reasonably simple if the relationship is almost linear, but if it is not, how should we assign a single number to the relationship? Regardless of this, scatterplots do provide some very nice insight in a computationally inexpensive way. This explains why scatterplots are ubiquitous in data analysis. It has been estimated that a large majority of graphs used in scientific publications are in the form of scatterplots – so we should not ignore them.

So, on to the problem at hand. How do we use scatterplots to develop insight into the problem of interacting species? As mentioned before, one trick that we will use first is to explore the effect of the length of time on the QoI. This is important, since we are estimating aspects of the population – here the maximum value. If we sample over too short a time we might miss the correct estimate. On the other hand, if we sample for a very long time to make sure we can capture the maximum, we are wasting computational resources. We first fix all parameters at some nominal level – this in itself is a complicated task and we *won't* worry about it here, but

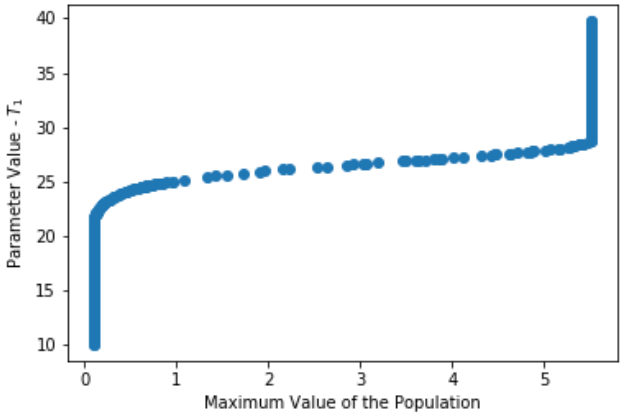
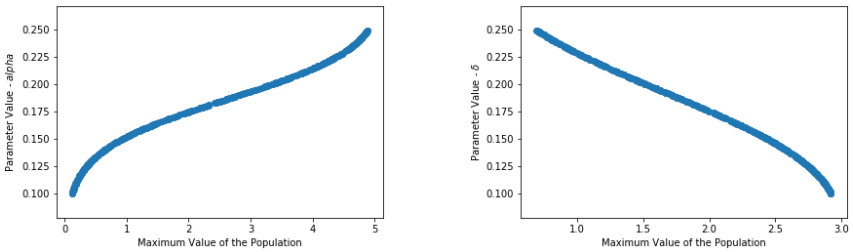


Fig. 2.5: Scatterplot of maximum value of the predator population as the end-time of the observation varies. For small times, the QoI is small, then increases up to some maximum.

keep in mind that this does matter. We will solve the equations on the time interval $t \in (0, T_1)$. Once we have selected a nominal value, we vary the end time, T_0.

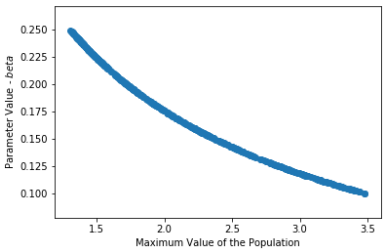
The results of 500 simulations are shown in Figure 2.5 and tell a simple, useful story. If the time is too short – say less than 22 hours, we vastly underestimated the maximum population value. Beyond some fixed time – about 26 hours, we don't see any changes. So, we will fix the time to be 30 (assuming that there is some variation in the time length as the other parameters vary. In fact, we have good reason to assume this. Two of the parameters are rates and help set the time scale! We can make sure that we have not missed anything once we determine how the QoI depends on the parameters. So, we fix all parameters but one and vary that one, moving through the remaining four parameters.

Looking at the scatter plots shown in Figure 2.6 it is relatively simple to see how each parameter affects the maximum value. Both α and γ are positively correlated with the maximal value of the predator population. These are growth rates, γ being the per capita rate of growth of the predators, so it seems reasonable that increasing this would lead to an increase in the maximum value. Increasing the growth rate of the prey also increases the predator maximum. It is interesting to note that there is an inflection in this correlation. After about 0.18, the rate of increase in maximum value slows down. This would be a place where sensitivity analysis can point towards interesting questions, but does not actually have explanatory power for this type of question. The other two parameters are decay terms. It seems intuitively reasonable for increases in these to lead to decreases in the maximum value of the population.

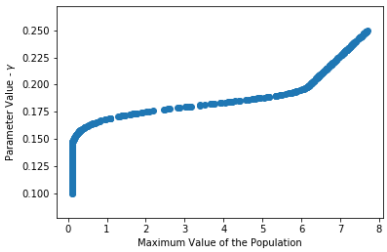


(a) α

(b) δ



(c) β



(d) γ

Fig. 2.6: All cases of competitive exclusion. There are two cases with nonzero steady-states, but only in Case 4 is this stable

2.3 Competitive Exclusion

We close this chapter with a small change in the model to include the concept of competitive exclusion. Rather than consider two species interacting in a linear chain, we can think of two species competing for the same resource. Why then, should only one species survive? Niche selection indicates that there are not typically similar species competing for the same resources. This is a fundamental question raised by Darwin, for example. There is evidence that Darwin's consideration of how species survive in niches, coupled with readings of Malthus helped him formulate his version of evolutionary theory.

Suppose we select a region of space and insert two populations that have similar foraging strategies. Suppose the consume roughly the same nutrients, have roughly the same motility, reproductive rates etc. Weak inference via niche observations indicates that these populations will not live happily together. Instead, one will have some advantage and, by exploiting it, will outcompete the other population in a process referred to as exclusion. Beginning in the 1920's controlled experiments began to be performed. In part, this was an effort to reconcile one underlying ecological argument that had been brewing for some time. To what extent does the density of the population affect the behavior? Ecological theories were essentially split between

camps that argued that most parameters were density dependent and those that disagreed. To help distinguish this controlled experiments were performed to see to what extent behaviors depended on density. When the two species were similar and competing for limited resources, it was shown that one species emerged victorious. George Francis Gause [**?**] formulated this precisely "Two species competing for limited resources can only co-exist if they inhibit the growth of the competing species less than their own growth." It is not a stretch to argue that the principle of competitive exclusion helped organize modern ecology.

2.3.1 Model

By combining a Lotka-Volterra type model of population interaction with a Pearl-Verhulst type model for population reproduction, a simple model of competition between species with population N_1 and N_2 is,

$$\frac{dN_1}{dt} = r_1 N_1 \frac{\kappa_1 - N_1 - \alpha_{12} N_2}{\kappa_1}, \tag{2.20}$$

$$\frac{dN_2}{dt} = r_2 N_2 \frac{\kappa_2 - N_2 - \alpha_{21} N_1}{\kappa_1}. \tag{2.21}$$

In his classic paper [**?**], Gause termed α_{12} and α_{21} the *coefficients of the struggle for existence*. The imagery evoked by Guase is apt. As species struggle for existence, they can do this actively as in predator versus prey or more passively by competing for resources. By considering different species of yeast, Gause studied how the populations grew in isolation and in competition for resources. Later he performed similar experiments on other organisms including paramecium (*Paramecium caudatum* and *P. aureli*).

Using different species of yeast as experimental subjects, Gause studied how the populations grew in isolation and compared this to growth with both species present and in competition for resources. Later he performed similar experiments on other organisms including paramecium (*Paramecium caudatum* and *P. aureli*). Gause also provided estimates of the parameters that reflected his data. Using these parameters and starting with two orders of magnitude more of species N_1, eventually N_2 outcompetes the other population. That is, the 'struggle for existence' is won by N_2 (see Figure 2.7).

2.3.2 Analysis

How do we analyze this model? We can proceed as before:

1. Find steady-states by solving $f_1 = 0, f_2 = 0, \ldots f_n = 0$ for $(x_1, x_2, \ldots, x_n)$.

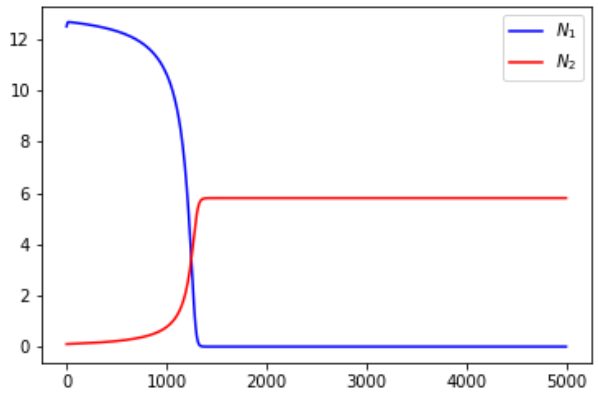


Fig. 2.7: An example of exclusion using parameters from Gauses paper [**?**]. We start with an order of magnitude more of N_1 than N_2, but N_2 exerts more pressure on N_1 than the other way around. In other words, N_2 is more *fit* and eventually N_1 is forced to zero and the competition is won by N_2.

2. Write the solution as a perturbation of the steady-states
3. Expand and look at the order ε terms
4. Determine the eigenvalues of the Jacobian
5. If all eigenvalues are negative, the steady-state is stable

Here $f_1(N_1,N_2) = r_1 N_1 \frac{\kappa_1 - N_1 - \alpha_{12} N_2}{\kappa_1}$ and $f_2(N_1,N_2) = r_2 N_2 \frac{\kappa_2 - N_2 - \alpha_{21} N_1}{\kappa_2}$. The steady-states satisfy,

$$r_1 N_1 \frac{\kappa_1 - N_1 - \alpha_{12} N_2}{\kappa_1} = 0, \tag{2.22}$$

$$r_2 N_2 \frac{\kappa_2 - N_2 - \alpha_{21} N_1}{\kappa_2} = 0. \tag{2.23}$$

The nullclines are reasonably simple. The axis of the phase plane are nullclines, so if $N_1 = 0$ the N_1 population cannot change and the same is true for the N_2 population. On these axes, we can understand the behavior. In the absence of one species we have the standard logistic behavior. So the individual populations tend towards their carrying capacity $\kappa_{1,2}$. The other parts of the nullclines come from,

$$\kappa_1 - N_1 - \alpha_{12} N_2 = 0,$$

and

$$\kappa_2 - N_2 - \alpha_{21} N_1 = 0,$$

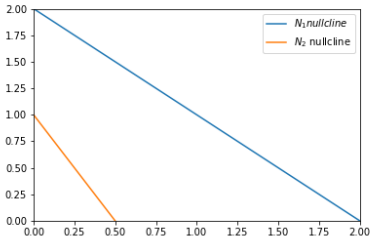
These are lines in the phase-plane (See Homework). The behavior depends on specifically on the parameters – for example, there are parameter sets for which the steady-states are not in the interior of the first quadrant, that is, one of the species is zero! It is a bit tedious, but not difficult to explore all possibilities. There are four cases, sketched in Figure 2.8. Note that in two of the cases (a and b) there are no other steady-states than either $N_1 = 0$ or $N_2 = 0$, so depending on the initial conditions the trajectories tend towards elimination of one species. When there is a coexistence steady-state, there is one case where the equilibrium is unstable (case c) in which case you again go to one of the elimination steady-states. For the final case, there is a nontrivial steady-state (e.g. a state where neither population is eliminated.

For this case, the equilibria is stable. But what does this mean biologically? If competitive exclusion is a guiding principle, how can we interpret this observation? One way to get insight it o examine a simplified case. What if both populations were exactly the same *except* for their coefficient of struggle? That is, suppose each population had the same carrying capacity and growth rates? Then the condition that has to be satisfied to be in case 4 is that $\alpha_{12} < 1$ and $\alpha_{21} < 1$. This means that the competition terms are small. This begs the question: How small? or Small with respect to what? Now that we have some idea that the magnitude of the competition terms matter, if we go back to the general case (where the populations do not have the same parameters), the inequalities that must be satisfied are $\alpha_{21} < \frac{\kappa_2}{\kappa_1}$ and $\alpha_{12} < \frac{\kappa_1}{\kappa_2}$. So the interaction pressure must be less than the ratio of carrying capacities.

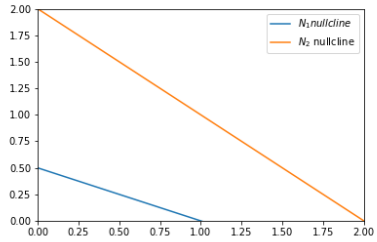
In Figure 2.9 we show results to compare with the null-cline images in Figure 2.8. We see that all cases but case 4 lead to exclusion of one species.

Working with the computer codes: Steady-States

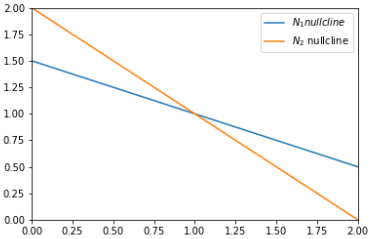
Just as we did for the logistic equation and the predator-prey model, we will use numerical approximations of the solutions in two ways. First we can confirm that our local analysis is correct – does the behavior near the equilibria match our eigenvalue interpretation? Running the script `Traj(f,g,H0,P0)` for initial populations near the equilibria show that our interpretation is correct. This can get quite taxing in terms of organizing the information – we have 4 cases to handle with multiple equilibria. To organize this we can use the script `analysis(f,g)` to make a table for stability and trajectories (This would be a nice wrapper – some code that took rhs, found steady-states, stability, and produced phase-plane plots).
Second, we can look far from the equilibria. We will not go into the theory, but remember that trajectories are restricted in their behavior – at least for planar systems. The converge to equilibria, convert to periodic solutions or run off to infinity.



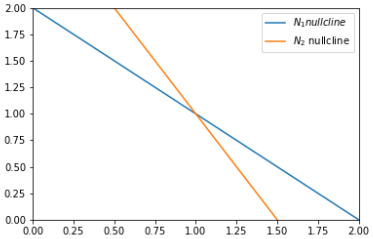
(a) Case 1: $\frac{\kappa_2}{\alpha 21} > \kappa 1$ and $\kappa_2 > \frac{\kappa_1}{\alpha 12}$



(b) Case 2: $\frac{\kappa_2}{\alpha 21} < \kappa 1$ and $\kappa_2 < \frac{\kappa_1}{\alpha 12}$



(c) Case 3: $\frac{\kappa_2}{\alpha 21} < \kappa 1$ and $\kappa_2 > \frac{\kappa_1}{\alpha 12}$



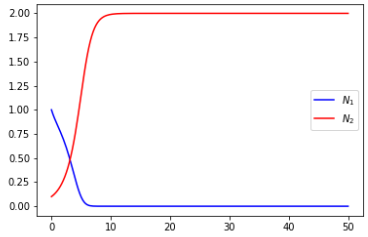
(d) Case 4: $\frac{\kappa_2}{\alpha 21} > \kappa 1$ and $\kappa_2 < \frac{\kappa_1}{\alpha 12}$

Fig. 2.8: All cases of competitive exclusion. There are two cases with nonzero steady-states, but only in Case 4 is this stable

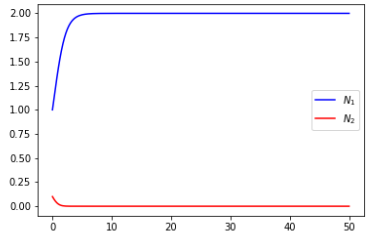
2.3.3 Sensitivity

In this section, we will introduce a sensitivity method related to one-at-a-time scatter plots discussed in Section 2.2.5. It is difficult to rank the parameters in their order of sensitivity by examining the scatterplots directly. This requires some quantification to compare the parameters. Regression is in keeping with the idea of correlation that is visually apparent in the scatterplots but adding a quantitative aspect. Regression means a lot of different things but we will typically talk about it in the informal *curve fitting* approach. The idea is that you have some data and you want to connect them with some function. If you had only two data points you can form a line through the two points. This line could be used to extrapolate to the an independent value outside the data domain (See Figure 2.11 blue star) or interpolate two a value within the domain (See Figure 2.11 green star).

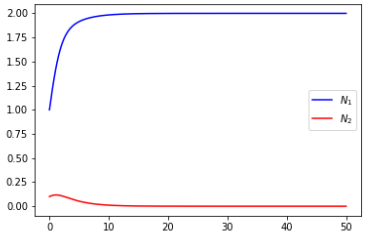
This begs the question about what to do with more observations? It is impossible to think of their relationship being exactly linear as there so many sources of error and uncertainty. So, what does one do? The easiest quantification follows by *assuming* a linear relationship and asking what is the line that best fits the data. Of course this begs an entirely different question: What constitutes the 'best fit'? We will use the relatively robust concept of minimum error, where the error is measured by the



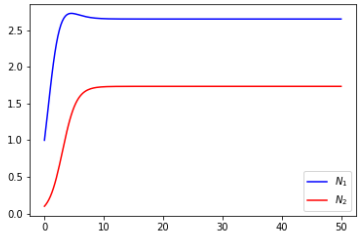
(a) Case 1: $\frac{\kappa_2}{\alpha 21} > \kappa 1$ and $\kappa_2 > \frac{\kappa_1}{\alpha 12}$



(b) Case 2: $\frac{\kappa_2}{\alpha 21} < \kappa 1$ and $\kappa_2 < \frac{\kappa_1}{\alpha 12}$



(c) Case 3: $\frac{\kappa_2}{\alpha 21} < \kappa 1$ and $\kappa_2 > \frac{\kappa_1}{\alpha 12}$



(d) Case 4: $\frac{\kappa_2}{\alpha 21} > \kappa 1$ and $\kappa_2 < \frac{\kappa_1}{\alpha 12}$

Fig. 2.9: All cases of competitive exclusion. There are two cases with nonzero steady-states, but only in Case 4 is this stable.

sum of the square distances between the model and the data. This is often referred to as the least-squares fitting. This topic is a deep topic with details that we don't need to understand right now. There are many packages out that will do linear and nonlinear regression. That is they will determine the polynomial (including a linear one) or other nonlinear function that best fits the data. Most sensitivity methods that use regression use linear regression only, since comparing the slopes of regression lines provides robust quantitative comparisons. The major caveat here is that the relationship *must* be linear, or near linear (whatever that means). Remembering the scatterplots Figure 2.8, we know this is not typically true. So we might have to restrict our parameter range, which may weaken our conclusions (TANSTAAFL again!).

We proceed with quantitative estimates for the competitive exclusion model. But what should the QoI be? Perhaps one of the questions is which parameters are most important for determining which population wins the competition. We define outputs $s_1 = \frac{N_1}{N_1+N_2}$ and $s_2 = \frac{N_2}{N_1+N_2}$. Both s_1 and s_2 are between 0 and 1 (why?). If N_1 is close to zero, s_1 is also. If N_2 is close to zero s_1 will be close to 1. The problem is symmetric, so this is likely to be redundant, but it is simple enough to see what happens. We start as we did with the scatterplots and define intervals for the parameters to be sampled from. We then fix all but one parameter at their nominal

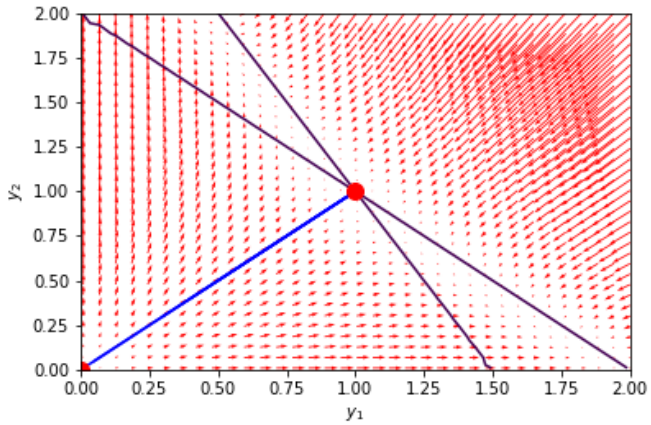


Fig. 2.10: Example phase-plane for competitive exclusion for the special case where there is co-existence.

value and vary the last one. The ranges of the parameters selected here are list them hereWe put all the selected parameter values into a vector-array, p and the calculated s_1's into a vector-array, S. Running `RegressionLinear(p,S)` finds the best fit line, that is the line with the minimum error. The slope of this line is a measure of the sensitivity. Looking at Figure 2.12, we see very distinct trends. Some of these

Working with the computer codes: Trajectories and Phase-Plane

this is a placeholder – the code is currently not structured this way Just as we did for the logistic equation, we will use numerical approximations of the solutions in two ways. First we can confirm that our local analysis is correct – does the behavior near the equilibria match our eigenvalue interpretation? Running the script `Traj(f,g,H0,P0)` for initial populations near the equilibria show that our interpretation is correct. This can get quite taxing in terms of organizing the information – we have 4 cases to handle with multiple equilibria. To organize this we can use the script `analysis(f,g)` to make a table for stability and trajectories (This would be a nice wrapper – some code that took rhs, found steady-states, stability, and produced phase-plane plots).
Second, we can look far from the equilibria. We will not go into the theory, but remember that trajectories are restricted in their behavior – at least for planar systems. The converge to equilibria, convert to periodic solutions or run off to infinity.

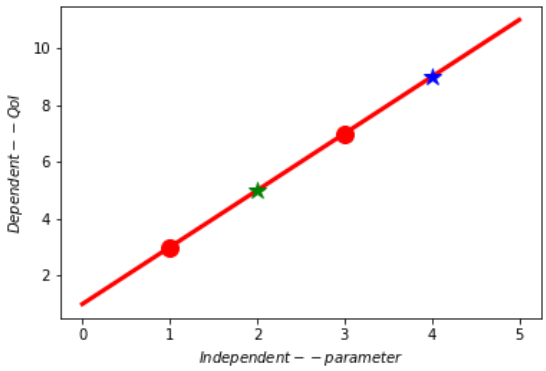
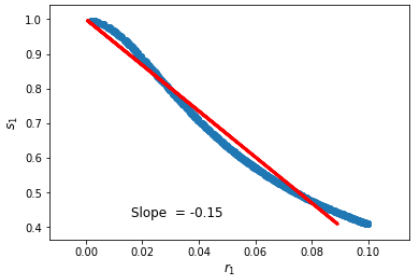


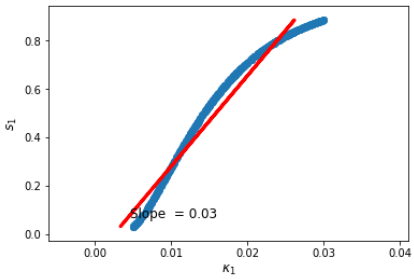
Fig. 2.11: Linear extrapolation and interpolation. Given two observations (indicated as red circles), if we assume a linear relationship we can extrapolate (blue star) and interpolate (red star) to get estimates of the QoI away from the observations.

are easily understandable. For example, as the carrying capacity for population N_1 increases, so does s_1. As the carrying capacity for N_2 increases, s_1 decreases. Similarly the competition terms agree with our intuition. As the pressure on population N_1 due to N_2, α_{12} increases, N_1 decreases. Less clear is the affect of growth rates which appear to be counter intuitive (better check this). But even more interesting the scale of the affects. Looking at the estimated slopes, we see that in our parameter regime the competition terms change the output qualitatively more. Since the slopes are several orders of magnitude larger than those for the other parameters, we can conclude that in our parameter regime it is the competition parameters that matter the most – or are the most sensitive.

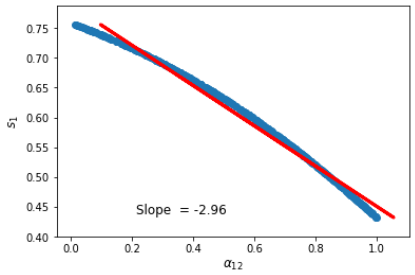
There is a major caveat here. One can cheat by restricting the parameter range by choosing a box that is too large and the variance in the output outpaces the variance in the input parameter selections. This can exacerbate the differences in regression lines. One can also unfairly restrict the parameter ranges which can reduce the appearance of sensitivity. The problem is that there is no global method for determining parameter ranges to search over, mainly rules of thumb. It is very useful if these can be restricted by the experimental/biological/physical system. For example, bacterial growth varies but in planktonic settings this is relatively well controlled and can be estimated. If this is not possible, one should take care to, at the very least, test different parameter sets and this should be clearly identified.



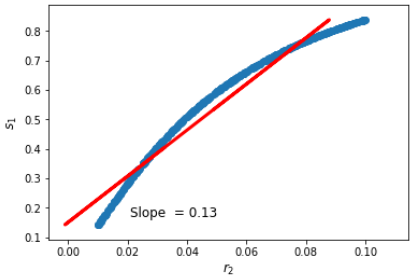
(a) Case 1:



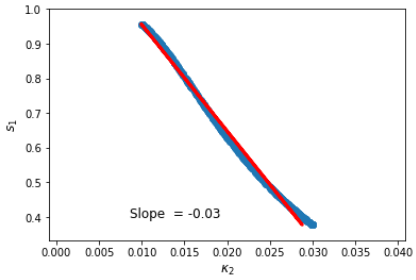
(b) Case 2:



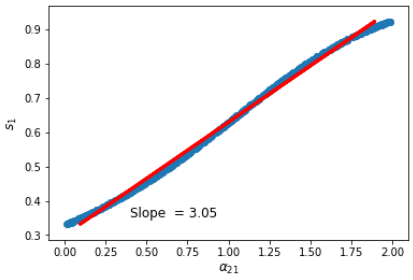
(c) Case 3:



(d) Case 4:



(e) Case 5:



(f) Case 6:

Fig. 2.12: All cases of competitive exclusion. There are two cases with nonzero steady-states, but only in Case 4 is this stable

2.4 State of the Art and Caveats

2.5 Problems

2.1. Use the script `Traj`(f_1, f_2, H_0, P_0) to how that if the initial population of prey, $P_0 = 0$ in the predator/prey equations (Section 2.2.3) leads to extinction of the pop-

ulation. Also show that starting with only prey leads to unbounded growth. Interpret this using the phase-plane.

2.2. Problem 3.4 indicates a difficulty with the simplifying assumptions of the Lotka-Volterra model, namely unbounded growth. Often logistic growth is used when there is evidence that the population is self-limiting. Analyze the logistic Lotka-Volterra model:

$$\frac{dH}{dt} = \alpha HP - \delta H \tag{2.24}$$

$$\frac{dP}{dt} = -\beta HP + \gamma rP(1 - \frac{p}{k}), \tag{2.25}$$

with parameters $\alpha =$, $\beta =$, $\delta =$, $\gamma =$, $r =$, and $k =$.

1. Find steady-states by solving $f_1 = 0, f_2 = 0, ...f_n = 0$ for $(x_1, x_2, ..., x_n)$.
2. Write the solution as a perturbation of the steady-states
3. Expand and look at the order ε terms
4. Determine the eigenvalues of the Jacobian
5. If all eigenvalues are negative, the steady-state is stable

2.3. Use scatter plots to determine the sensitivity of the following QoI for the Lotka Volterra equations:

1. The difference in predators and prey
2. Minimum Prey
3. The Period of oscillations. Note that this requires checking if the solutions have converged to a periodic solution and then estimating the period

2.4. Use linear regression to determine the rankings of the parameters in the competitive exclusion model, when the QoI is the maximum of species N_1. Verify that the behavior is symmetric in species. That is if you compare the rankings with respect to the maximum of N_2 they should be opposite. What is a reasonable interpretation of the sensitivity rankings?

2.5. Demonstrate that the solutions to Lotka-Volterra are periodic – this can be done numerically (See previous problems). Use the period of the output as the QoI and include the initial condition as a parameter. What do you find about the sensitivity of the IC? Why does this make sense?

2.6 Appendix

We can generalize this idea. We write our system of n ODEs,

$$\begin{aligned}
\frac{dx_1}{dt} &= f_1(x_1, x_2, ..., x_n), \\
\frac{dx_2}{dt} &= f_2(x_1, x_2, ..., x_n), \\
... \quad &..., \\
\frac{dx_n}{dt} &= f_n(x_1, x_2, ..., x_n).
\end{aligned}$$

Just as our single equation (in the case of the logistic model) and in the system of two equations (as in the case of the Lotka-Volterra model), we can linearize the behavior around the steady-states. For a general system, we define the steady-state as a vector $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_n)$ and the expansion $\mathbf{x} = \bar{\mathbf{x}} + \varepsilon\hat{\mathbf{x}}$ leads to a linear system of ODEs,

$$\frac{d\hat{\mathbf{x}}}{dt} = J\hat{\mathbf{x}}, \tag{2.26}$$

where , $\mathbf{x} = (x_1, x_2, ..., x_n)$. We use the notation J since the matrix we have obtained is referred to as the Jacobian. The Jacobian is a useful concept, and for us one of the uses is to shorten some of the calculations. It turns out that we can generalize the above calculations. The matrix Jacobian,

$$J = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{x_n} \\
\frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{x_n}
\end{bmatrix} \tag{2.27}$$

Just as the solution to the simplest linear ODE $\frac{dx}{dt} = ax$ is $x = x_0 e^{at}$, we can write the solution to Equation 2.26 as $\mathbf{x} = \mathbf{x_0} e^{Jt}$. Well, sort of. We might not really understand what it means to exponentiate a matrix. If we replace J with the diagonalized version,

$$J = UDU^{-1}, \tag{2.28}$$

where D is a diagonal matrix with the eigenvalues of J along the diagonal,

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{2.29}$$

We find that the solution to Equation 2.26 is $\mathbf{x} = \mathbf{x_0}\left(U \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix} U^{-1}\right)$. If the eigenvalues of J are negative, $\mathbf{x} \to 0$. Otherwise the perturbation does not decay and we can conclude that the steady-state solution is not stable.

Chapter 3
Within Host Disease Models

Abstract

As was described in Chapter 4, mathematics has been used to study disease progression within a population since at least the 1600s when John Graunt used death records from London to develop a method to estimate the risk of dying for different diseases [**?**]. However, t

3.1 Historical Background

To be able to model the spread of a disease within a population we have to know the disease progression *within* the host and modes of transmission. For example, there are multiple diseases such as the common cold, flu's and gastroenteritis have different progression timing and are transmissible at different stages of the disease. Also, differences between diseases spread by insect vectors, body fluids, via the air etc. clearly alter the terms in a between-host model. Finally, understanding the dynamics of diseases within the host plays an important role in developing treatments since when to treat, how long to treat and how aggressively to treat the disease depends on the hosts response to the disease and treatment.

It is interesting that the development of within-host models is far more recent that epidemiological models. This is primarily do to the dearth of understanding of the immune system. Immunology was essentially an experimental subject during the development of vaccines for diseases such as smallpox and cholera. By the 19th century several competing theories had been developed to understand the nature of diseases. Pasteur developed germ theory by exploring the process of fermentation. He noted that liquids that easily fermented did not behave the same after they were boiled. Since he had already connected the process of fermentation with living organisms, it was a short step to identifying specific bacteria that caused diseases. These were further refined by numerous contemporaries including Koch who studied anthrax.

At the same time, while exploring vaccination, Ehrlich focused on soluble serum or anti-toxins (now known as antibodies). Ehrlich followed pioneering work on blood serum by Von Behring and identified chemical mechanisms that formed antibodies against bacterial challenges. Antibodies work by identifying components of invaders (antigens). Antibodies interact with antigens and tag them so that specialized cells can attack and kill them as well as trigger the immune system to train cells to identify and kill the invading cells. Along with studies focusing on specialized components of the blood (phagocytes), Ehrlich and Metchnikoff received the 1908 Nobel Prize in Physiology.

So bacteria were identified as causal agents of disease and the several mechanisms bodies used to fight infections were identified. However, this did not explain the variation in immune protection. In other words, if antibodies were the bodies method of protecting against disease challenge, what causes or mediates the diversity of antibodies? It was not until the 1950s that the immune system was compartmentalized into innate and active with immunological memory mediated by specific regulatory machinery including humoral response, active cellular response, and messaging apparatus.

The messaging apparatus has undergone massive revisions from simple chemical activation to lymphocytic interactions and antigen identification, presentation and production dynamics. The most fundamental advance in immunology was the refinement of the concept of clonal adaptation. Specialized cells travel throughout the body some producing antibodies in response to their environment (B-cells) and some identifying and responding to antigens presented on the surface of cells. The upregulation of production of lymphocytes trained against specific invaders in response to recognition of presented antigens is referred to as clonal selection. Basically there is a small population of immune cells that have a library of antigens that they recognize as hallmarks of invaders. If one of the immune cells finds this antigen in the body, it begins to proliferate and lead the charge to attack the invading pathogen.

Until the hypothesis of clonal expansion was solidified in the 1970s there was little identifiable mathematical modeling of within-host dynamics. But this hypothesis leads to quantifiable questions such as how large of an antigen repertoire is required to provide immunity? How do the cells recognize self and non-self antigens? What are differences between the immune response for viral and bacterial pathogens? How can vaccines be developed and refined? The mathematics relevant to these questions leapt to the forefront during the advent of HIV in the 1980s. We will wait for a bit to focus on HIV given the magnitude of the literature on this disease. Instead, we will focus on general within host models.

3.2 Within Host Models

Within host dynamics focus on how a disease progresses within a single host. These are among the most detailed models that have been introduced. For a simple intro-

duction, we will focus on a model proposed for the interaction between a tumor and the immune system. Now, both tumor biology and immune system biology is complicated, so we will simplify this immensely. The drawbacks of this simplification is that we cannot provide detailed insight into any quantitative dynamics. In fact, we likely can't even associated our state variables with biologically measurable processes. The benefit is that the model is quite general and can be adjusted to model many different situations. The basic units of this model represent the disease (in this case a tumor) and the immune response. Recalling that the immune response can be separated into two processes, innate and adaptive (or lots of different terminology). But basically, there is a piece of the immune system that just hangs around and kills invaders and a part that responds to the invader.

3.2.1 Model

This model considers the adaptive immune system. Therefore, we require a few things. First, if there is no tumor the adaptive response should return to some sort of minimal level. This return to a basal state is referred to as homeostasis and ensures that the host does not become over-run by an immune response without a target. We also require the immune response to increase if there is an invader. In the model below, the rate of increase is bounded for all invader levels. For simplicity, we will also expect that the invaders are bounded for all time.

We define T as the tumor cells and E as the effector cells – the aggregate of all immune response cells. Tumor cells grow and are killed by effector cells. Effector cells are produced by the body at a constant rat, decay, are recruited by tumor cells (one could lump this into the rate of production by the body, but it is more typical to separate the processes) and die when they remove tumor cells. The word model then looks like,

$$\begin{aligned}\text{rate of change of E} &= \text{production} - \text{decay} + \text{recruitment} - \text{death},\\ \text{rate of change of T} &= \text{growth} - \text{death}.\end{aligned}$$

We just have to describe each of the terms. It is typically a good idea to start simple unless one knows specific observations that need to be accounted for. For example, we have a choice for the tumor growth compartment. We could just assume exponential growth ($\frac{dT}{dt} = kT$). But we have already seen some issues with this and we do know that tumors can't grow forever. We could include nutrient as a limiting factor. That is we could assume $\frac{dT}{dt} = kTf(N)$ for some function of the nutrient. But then we would need an explicit description of the nutrient. This might be useful, but certainly makes the model more complicated. In between these is logistic growth. It leads to bounded behavior but does not require explicit treatment of the nutrient. This is a great place to start, as long as we keep in mind a few things. Logistic growth forces a stable, nontrivial equilibrium. This may actually be inappropriate

for some situations. Fortunately, we can always go back and refine the model and this is a place where we know that we are 'cheating'.

We will follow previous work done by Kuznetsov and Perelson for the other terms,

$$\frac{dE}{dt} = s - dE + pE\frac{T}{g+T} - mET, \tag{3.1}$$

$$\frac{dT}{dt} = rT(1-\frac{T}{k}) - ET. \tag{3.2}$$

This implies that production of effector cells by the body occurs at a constant rate (probably not a perfect assumption!) and that the reaction between effector and tumor cells reduces each population. This is mass action and the parameters indicate the yield (how many tumor cells are removed by a single effector cell, which has been scaled to one). We will see this in more detail in Chapter 6. Effector decay is at a constant rate.

Recruitment is the most complicated term. First, it takes effector cells to be recruited and they respond to tumor cells. We could have assumed the simplest interaction, pET, but biological observations indicate that there is a maximum rate that effector cells can be recruited so that this rate should be bounded as $T \to \infty$. Again, we could try the model without this and see what happens (See the problems).

We will use the parameter estimates provided as nominal values:

Parameter	Meaning	Value
s	effector production rate	.1181 (units!)
d	Effector decay rate	0.3743
p	Effector recruitment rate	1.131
r	Tumor growth rate	1.636
k	Tumor carrying capacity	0.5×10^3
g	Effector response scaling	20.19
m	Effector death from tumor reaction	3.11×10^{-3}

3.2.2 Analysis

This is a system of nonlinear equations, which means it is likely to be difficult to solve them analytically. But we can proceed as we did in the previous chapter. Recall the basic steps:

1. Find steady-states by solving $f_1 = 0, f_2 = 0, ...f_n = 0$ for $(x_1, x_2, ..., x_n)$.
2. Write the solution as a perturbation of the steady-states
3. Expand and look at the order ε terms
4. Determine the eigenvalues of the Jacobian
5. If all eigenvalues are negative, the steady-state is stable

To find the steady-states, we just need to solve,

$$s - dE + pE\frac{T}{g+T} - mET = 0, \tag{3.3}$$

$$rT(1-\frac{T}{k}) - ET = 0. \tag{3.4}$$

The first equation can be solved for E,

$$E = \frac{s}{d - p\frac{T}{g+T} + mT}$$

and by substituting this into 3.3, we find the ungainly expression,

$$rT(1-\frac{T}{k}) - \frac{s}{d - p\frac{T}{g+T} + mT}T = 0. \tag{3.5}$$

Clearly $T = 0$ is one solution. That makes sense since we are not considering how the tumor is initiated. If we do a bit of algebra, we can see how many other solutions there are (note that one could just plot this, but this comes with a warning that without a decent understanding of how many solutions there are it is easy to miss one). First, we remove the $T = 0$ (corresponding to the steady-state $(E,T) = (0.3155, 0)$) and then eliminate all the fractions,

$$\begin{aligned} 0 &= r(1-\frac{T}{k}) - \frac{s}{d - p\frac{T}{g+T} + mT}, \\ &= r(1-\frac{T}{k})(d - p\frac{T}{g+T} + mT) - s, \\ &= r(1-\frac{T}{k})\left(d(g+T) - pT + mT(g+T)\right) - s(g+T) = SS(T). \end{aligned} \tag{3.6}$$

If you look long enough at Equation 3.6, you can see that it is a cubic in T which means that, in general, there are three nontrivial solutions. In Figure **??**, we plot this just to get a visual understanding of the behavior. . In addition the null clines intersect in three points B = (1.6093,8.158), C = (0.7172, 280.8), and D = (0.1825, 442.2). To determine the stability of the steady-states we can either compute the eigenvalues of the Jacobian matrix or use the numerical methods to estimate this (See the shaded box on page...).

So, we have determined that there are four equilibria. The tumor free state is unstable and there are two stable equilibria, a low tumor/high effector cell state and a high tumor/low effector state. At this point, there are some options for what to do next. One could note that the model is oversimplified and begin adding complexity, with the aim of incorporating important details to complement the biological investigations. That is, experiments show that compartmentalizing the effector cells this way misses an important feedback loop. Namely, some classes of effector cells (CD 4+) are used to recruit more effector cells. These 'helper' cells have the job of interacting with the tumor cells, which activates them and induces them to produce

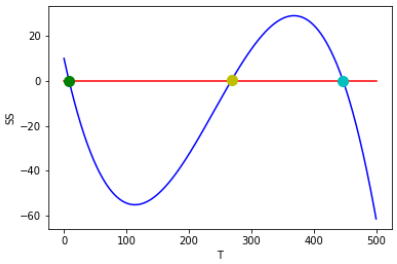


Fig. 3.1: Simplified nullclines showing steady-state values of T.

numerous signaling molecules. One of the most important of these is the specific signal interleukin-2 so some authors include a compartment for the signal that autoinduces recruitment of effector cells[].

A different tack could be to use the simplified model to develop and explore treatment methods. One method is to add effector cells (immune therapy). With a fairly simple model such as this one, one can consider adding a bolus of effector cells and try to force the dynamics into the low tumor state (See Homework). This is again the principle of TANSTAAFL – the two variable model is not (T)ruth, but is much more tractable. Often, analyzing the simple models provides a road-map for a more complicated model.

Working with the computer codes: Steady-states

The equations are again, two dimensional:

$$\frac{dE}{dt} = s - dE + pE\frac{T}{g+T} - mET, \tag{3.7}$$

$$\frac{dT}{dt} = rT(1-\frac{T}{k}) - ET. \tag{3.8}$$

We call the Python or *R* code, python `SS`$(f_1, f_2, [a,b])$ and find numerical estimates steady-states. The code `TE_ss` has a lot of parts including a numerical Jacobian calculator. We find that ....It would be much better to be able to make this general for arbitrary sized systems. Currently only set-up for two variables.

3.2.3 Sensitivity

We now turn back to sensitivity analysis of our model. In this chapter, we focus on differential methods. One could argue that all SA methods stem from this method. The question of how much does the output change as the input (parameter) changes is certainly phrased like 'rate-of-change' questions in calculus. In practice and in historical context, there are plenty of classical examples where analytic methods are used but we will avoid those since it is far more typical for the models to be sufficiently complicated to preclude this type of analysis. This is the context where the confusion of local versus global might arise. Differential methods are by their nature local since differentiation is a local process (think derivative at a point, or the tangent to a curve at a point). It is important to note that almost all methods of SA are local to some extent. One has to define a region in parameter space. The difference is that the conclusions for differential methods are all evaluated at a point (or a nominal parameter value) and are extended to a neighborhood with caution.

A word of caution here as well. This rapidly becomes computationally difficult for many parameters. What we describe below will be for a single parameter, but it is extendible to a vector of parameters. We should have code for this as well. A useful homework would be to show examples were it gets more time-intensive as the number of parameters increases.

A useful method for performing local sensitivity using differential thinking to derive a dynamic equation for the sensitivity index of a parameter. Assume that you are given a dynamic model (like Kuznetsov and Perelson) that consists of several ODEs that depend on a parameter p,

$$\frac{dy_i}{dt} = f_i(y_i, t; p), \tag{3.9}$$

with initial conditions,

$$y_i(0) = y_0. \tag{3.10}$$

We define a sensitivity measure $S_i = \frac{\partial y_i}{\partial p}$. This describes how variations in p affect the output y_i. If the exact solution was completely known, this is a useful tool. For example, the logistic equation introduced in Chapter 2, so we can do this explicitly (See the appendix). But typically this is not possible or may be very complicated. Instead, we can augment our ODE system. We define new variables, $S_i = \frac{\partial y_i}{\partial p}$ that are the sensitivity measures of each state variable, y_i with respect to a given parameter p. The 'trick' to this method is to consider the dynamics of the index in time, $\frac{dS_i}{dt}$. We can then introduce the dynamics of the state variables (which are known from the ODEs). We take the time derivative of the indices and interchange the time derivative and derivate with respect to the parameter of interest,

$$\begin{aligned}\frac{dS_i}{dt} &= \frac{d}{dt}\left(\frac{\partial y_i}{\partial p}\right),\\ &= \frac{\partial}{\partial p}\left(\frac{dy_i}{dt}\right).\end{aligned} \tag{3.11}$$

Now, interchanging derivatives can be tricky business so we will just acknowledge that sometimes this might be a problem, but if the state variables are well behaved and there aren't discontinuities, we should be fine.

Since $\frac{dy_i}{dt} = f_i(y_i,t;p)$ the chain rule implies,

$$\begin{aligned}\frac{dS_i}{dt} &= \frac{\partial}{\partial p}\left(f_i(y_i,t;p)\right),\\ &= \sum_{j=1}^{n}\frac{\partial f_i}{\partial y_j}\frac{\partial y_j}{\partial p} + \frac{\partial f_i}{\partial p}\end{aligned} \tag{3.12}$$

We just derived a dynamical system (ODE system) that defines how the sensitivity index changes in time. There are two contributions. The first is from how the model depends explicitly on the parameters: $\frac{\partial f_i}{\partial p}$. The other has to do with how the right-hand-sides change implicitly as the dynamics of the state variables are affected by variations in the parameters. This is encapsulated in $\sum_{j=1}^{n}\frac{\partial f_i}{\partial y_j}\frac{\partial y_j}{\partial p}$ Since $\mathbf{S} = (S_1,S_2,...,S_n) = (\frac{\partial y_1}{\partial p},\frac{\partial y_2}{\partial p},...,\frac{\partial y_n}{\partial p})$, we can write the implicit dependence in terms of the Jacobian,

$$\frac{d\mathbf{S}}{dt} = J\mathbf{S} + \mathbf{f_p},$$

where $\mathbf{f_p} = \left(\frac{\partial f_1}{\partial p},\frac{\partial f_2}{\partial p},...,\frac{\partial f_n}{\partial p}\right)$.

If we have a differential equation for each of the indices, we have to define the initial conditions. We will not show this here, but claim that if no parameters appear in the initial condition of the state equations (Equations 3.10) zero. If the parameter appears explicitly in the initial condition for some state variable y_k, we define $S_i(0) = 0$ unless $i = k$ when $S_k(0) = 1$.

How does this help? Did we just make things harder? It helps because we have tools to estimate the indices. We can just augment our system of ODEs given in Equations 3.9 with those defined in Equations 3.13,

$$\begin{bmatrix}\frac{d\mathbf{y}}{dt}\\ \\ \frac{d\mathbf{S}}{dt}\end{bmatrix} = \begin{bmatrix}\mathbf{f}\\ \\ J\mathbf{S}+\mathbf{f_p}.\end{bmatrix} \tag{3.13}$$

This doubles the size of the system of equations we are solving since $\mathbf{S}$ is the same length as $\mathbf{y}$. We also have to compute the Jacobian matrix at every time-step unless we can precompute this. For systems that are too large this can be computationally difficult (see Problem 3.2). But given those issues, we have a direct method

for calculating the sensitivities. This is a good thing! One downside is that it is not as simple to look for other QoI, at least without thinking it through a bit more. There are methods for estimating 'feature' sensitivities from direct sensitivities, but these are mainly *ad hoc*.

One additional thing that this indicates is that the sensitivities are dynamic. In retrospect this is an obvious observation, but it is very apparent in this method. One could ask a question that is more detailed than 'What is the most sensitive parameter with respect to the QoI y_i?' Instead, one could ask whether the ranking of the sensitivities change?

Working with the computer codes: Direct SA

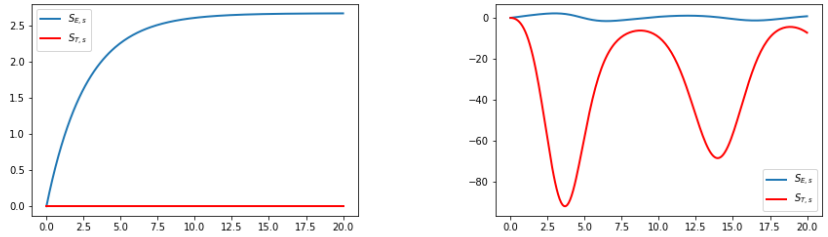
We will use our scripts to study direct sensitivity by augmenting the original dynamical system,

$$\frac{dE}{dt} = s - dE + pE\frac{T}{g+T} - mET = F,$$
$$\frac{dT}{dt} = rT(1-\frac{T}{k}) - ET = G,$$

s is probably wretched notation! and there is something wrong in the subscripts with the following system for the index of parameter s,

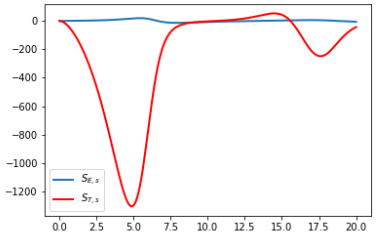
$$\frac{dS_{E,s}}{dt} = \frac{\partial F}{\partial E}S_s + \frac{\partial F}{\partial T}S_s + \frac{\partial F}{\partial s}$$
$$\frac{dS_{T,s}}{dt} = \frac{\partial G}{\partial E}S_s + \frac{\partial G}{\partial T}S_s + \frac{\partial G}{\partial s}.$$

We can just change our script that we use to solve the differential equations to estimate the solution of the system of four variables. There is a bit of notational difficulty as well – there are indices for each parameter and each QoI (state variable). So we use notation $S_{i,j}$ where i is the QoI (in this case $i = (E,T)$ and j denotes the parameter in question. Here $j = (s,d,p,g,m,r,k)$ In Figure ...there are too many different views: 7 parameters, 4 SS. The SA's in time are interesting. There might be issues with my barplots. Which should I show? Right now I show one QoI and a the two different steady-states, a couple of parameters.
In Figure 3.2, we show the sensitivities in time for the outputs E and T with respect to the single parameter s. There are caveats here – this does not rank which parameters are most sensitive, but indicates for which output is s the most sensitive parameter. What is interesting is the comparison between this near different steady-states.

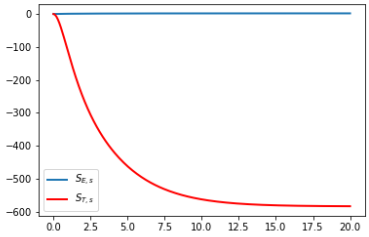


(a) Smallest steady-state

(b) next steady-state



(c) β



(d) γ

Fig. 3.2: Sensitivity indices near each steady-state for the parameter s.

3.3 Acute Infection

In this section we consider the case of an acute infection. There is a distinction between this and a chronic disease, even if the models are similar. In particular, acute infections such as influenza, dengue and Zika have finite duration and are typically resolved within a few weeks whereas chronic infections, such as HIV, persist. The length of time where the disease dynamics plays a role allows for some aspects of the biology such as the production of uninfected cells can be neglected. Additionally, the models are required to exhibit clearance steady-states to be realistic. A generic model is separates the host cells into target cells, T (those that are not infected and possible targets), infected cells, I, and virus, V [**?**]. When target cells and virus molecules interact the target cells may transition to infected cells. Infected cells can die and produce virus particles which the decay. Using mass action kinetics we can describe the dynamics of each population,

$$\frac{dT}{dt} = -\beta TV, \tag{3.14}$$

$$\frac{dI}{dt} = \beta TV - \delta I, \tag{3.15}$$

$$\frac{dV}{dt} = pI - cV. \tag{3.16}$$

The parameters define the infectivity per virus particle (β), death rate of I (δ), virus production per infected cell (p) and viral decay rate (c). The system has initial conditions T_0, I_0, and V_0. This ignores any latency, where the target cells have to incubate the virus particles. Additionally, this model does not include the immune system. A simple model for the immune response, A, is,

$$\frac{dA}{dt} = kAV + S_A - dA. \tag{3.17}$$

This is a simple model for the immune response to antibodies produced by virus particles since A is produced by the interaction between A and V.

But of course, if we do not include the effects of the immune response on the virus the model has only one-directional effects. That is, the virus signals the immune system without any reciprocal effects. We can couple these by altering 3.16,

$$\frac{dV}{dt} = pI - cV - c_A VA. \tag{3.18}$$

3.3.1 Analysis

We will study the system,

$$\frac{dT}{dt} = -\beta TV, \tag{3.19}$$

$$\frac{dI}{dt} = \beta TV - \delta I, \tag{3.20}$$

$$\frac{dV}{dt} = pI - cV - c_A VA, \tag{3.21}$$

$$\frac{dA}{dt} = kAV + S_A - dA. \tag{3.22}$$

This is a system of nonlinear equations, which means it is likely to be difficult to solve them analytically. But we can proceed as we did in the previous chapter. Recall the basic steps:

1. Find steady-states by solving $f_1 = 0, f_2 = 0, \ldots f_n = 0$ for $(x_1, x_2, \ldots, x_n)$.
2. Write the solution as a perturbation of the steady-states
3. Expand and look at the order ε terms
4. Determine the eigenvalues of the Jacobian

5. If all eigenvalues are negative, the steady-state is stable

Often one hears that algebra is "easier than calculus". This does not mean all parts of algebra are equally straightforward. It would be more accurate to say that calculus requires algebra, while algebra does not require calculus so there is some sort of ordering. But it is notoriously difficult to find roots of algebraic equations when the system becomes large – there is no closed form solution like the quadratic formula if the degree of the polynomial is 4 or higher. This does not mean nobody knows what it is, but that it is not possible to find one! This is a good lesson in using numerical methods wisely. If you look at the algebraic system,

$$-\beta TV = 0, \tag{3.23}$$
$$\beta TV - \delta I = 0, \tag{3.24}$$
$$pI - cV - c_A VA = 0, \tag{3.25}$$
$$kAV + S_A - dA = 0, \tag{3.26}$$

the first says that steady-states occur only when either $T = 0$ or $V = 0$ – no target cells left of no virus particles. Remember that we are considering acute infections, so the time-scale led us to neglect cell growth. From a purely mathematical standpoint we can reconcile this by arguing in detail about what we mean by long-term. This is a relative description rather than a quantitive description. If we kept going in this direction we would develop the theory of approximations which is often called assymptotics or perturbation theory. While these are very important concepts, they do fall a bit outside the scope of this book. So we will just note that the $T = 0$ case is not relevant for the diseases that we are focusing on. So the steady-state requires $V = 0$. It is simple to see that 3.24 means that $I = 0$. Plugging these into 3.25, we see that it is satisfied and Equation 3.26 gives $A = \frac{S_A}{d}$. That is we expect that there is a time-independent solution where there are are a given number of target cells, no infected cells, no virus particles and the immune response is at a basal level that is the ratio of production to decay.

3.3.2 Sensitivity Analysis

Thinking about it for a minute, this is a practically useless result other than noting that once an infection is introduced it will eventually run its course. Instead, it might be useful to ask some more practical questions that a person who is sick might ask. How severe is the sickness going to be? How long will it last? When can I expect to start feeling better? All of these provide good QoI targets! Examining Figure **??**, you can see that the basic disease time course from a small infection, is monotonically increasing virus load until a maximum is reached and then monotonic decrease until the virus is cleared. We could just ask which parameters are most important for increasing the rate of change in the viral load. This would tell us what to target during the ramping up of the infection (say to reduce the viral load) as

well as during the recovery (to increase the rate of reduction). Mathematically we are interested in $\frac{\partial}{\partial p}\frac{dV}{dt}$. If we optimistically interchange the derivatives we find the sensitivity can be determined from $\frac{d}{dt}\frac{\partial V}{\partial p}$. But this is just the time derivative of S_V! Which we have already decided how to calculate. One way to address this, that leans on the computer codes, is to determine an ODE for this so that we can solve it simultaneously. Or we can post-process of solution curves to determine this as well. Need to show the solution curves for reasonable values and try derivatives of that.

3.4 Problems

3.1. Consider the within-host model:

$$\frac{dE}{dt} = s - dE + pE\frac{T}{g+T} - mET + \alpha\delta(t_0), \tag{3.31}$$

$$\frac{dT}{dt} = rT(1-\frac{T}{k}) - ET. \tag{3.32}$$

Where we have included a source of effector cells at time t_0. Beginning with high tumor, low effector cells $(.1, 500)$, consider how the outcome depends on α and t_0.

1. For a fixed t_0, systematically vary α and demonstrate that above a certain level, little changes

Working with the computer codes: Steady-state Analysis

Estimates of the steady-states of

$$\frac{dT}{dt} = -\beta TV = f_1, \tag{3.27}$$

$$\frac{dI}{dt} = \beta TV - \delta I = f_2, \tag{3.28}$$

$$\frac{dV}{dt} = pI - cV - c_A VA = f_3, \tag{3.29}$$

$$\frac{dA}{dt} = kAV + S_A - dA = f_4. \tag{3.30}$$

can be found from the Python or *R* code `SS`$(f_1, f_2, f_3, f_4[a,b])$. We already noted that the clearance state is the most relevant. We can ask the stability of this steady-state using the code `TE_ss`. We find that the eigenvalues of the Jacobian matrix are The code `TE_ss` has a lot of parts including a numerical Jacobian calculator. We find that ....It would be much better to be able to make this general for arbitrary sized systems. Currently only set-up for two variables.

2. Determine whether the time, t_0 plays a role in the affect of α.
3. Use the differential method to determine the sensitivity of the tumor volume on the two additional parameters α and t_0.

3.2. We mentioned that calculating the indices by solving the augmented system might become computationally difficulty. In this problem, we will show that as the size of the system increases, the numerical methods slow down. I think we can do this with a linear system that grows: $\frac{dy_i}{dt} = \sum \alpha_{ij} y_j$. We can do this analytically since it is linear. We can do it numerically with the Jacobian explicit, and numerically calculating the Jacobian.

3.3. What does SA in time imply about the other parameters in our model? Do the ranking of most sensitive change in time?

3.4. Consider the overly simplified model of within host dynamics:

$$\frac{dE}{dt} = s - dE + pET - mET, \tag{3.33}$$

$$\frac{dT}{dt} = aT - nET. \tag{3.34}$$

1. Find the steady-states
2. Use numerical methods to determine if the steady-states are stable or not.
3. Discuss what biological observations this might support or contradict.

3.5. In the appendix, we derive explicit estimates of differential sensitivities for the logistic equation. Compare these estimates with those obtained by scatterplots and augmented SA

3.6. At the end of the chapter, we showed that the direct sensitivity method could be used for certain 'feature' sensitivity. In particular, we showed how to estimate which parameters played a dominant role in controlling the rate of clearance of the disease but differentiating the sensitivity measure numerically.

1. Adjust the numerical scripts to calculate the feature sensitivity by augmenting a differential equation for the rate of change of the derivative S_V with respect to time. Do you get similar estimates?
2. Use either method to estimate the sensitivity of the time to maximum virus concentration and the maximum concentration – discuss how this provides targets to shorten the disease course.

3.5 Appendix

$$\frac{dy}{dt} = ry(1 - \frac{y}{k}) \tag{3.35}$$

has the exact solution,

$$y = \frac{Ce^{rt}}{\left(1 + \frac{C}{k}e^{rt}\right)}. \tag{3.36}$$

where $y(0) = C$. Interestingly we see that the initial condition can be viewed as a parameter of the system. Homework problem here!

Estimating the S_i's is "only" a problem of calculus. It is not terribly difficult to show that

$$\begin{aligned} S_C &= \frac{k^2 e^{rt}}{(k + Ce^{rt})^2}, \\ S_r &= \frac{Ck^2 e^{rt} t}{(k + Ce^{rt})^2}, \\ S_k &= \frac{C^2 e^{2rt}}{(k + Ce^{rt})^2}. \end{aligned} \tag{3.37}$$

Some things to observe from this are that the sensitivities depend explicitly on time. This is something that we won't discuss here, but sensitivity in time is a useful concept to keep in mind for later. But all sensitivities actually go to zero as $t \to \infty$. This makes sense since we know that for any positive initial condition, the logistic equation tends to a steady-state. If $C < 0$ stranger things can happen! But they are not physical.

But typically it is not possible to find the analytic solution. Instead we turn to approximate techniques. If we do not know the solution the simplest way to estimate S_i is using a difference approximation. $S_i \approx \frac{y_i(p+\Delta p)}{\Delta p}$. This is exact at p and approximates the rate of change/slope of the tangent line etc. For a single parameter, you can take to realizations of the model (say from a numerical solution) at two different parameter values. Looking at the difference in the output, scaled by the difference in parameters gives an approximation of the indices.

Difference approximation

We can compare the results of the discrete estimate of the sensitivity and the analytic for the logistic equation.
We consider,

$$\frac{dx}{dt} = rx(1-\tfrac{x}{k}),$$

with the nominal parameter values $r_0 = .1$, $k = 10$ and initial condition, $C_0 = 1$. For simplicity, we will evaluate the sensitivity at some point before the solution has equilibrated, based on our simulations, we use $T = 30$. The exact indices from Equations 3.37. We find that $S_C \approx 2.2191$, $S_r \approx 66.5716$, $S_k \approx 0.4457$. That seems to imply that the most sensitive parameter, at that time, is the growth rate – by a lot!
Now we can try the discrete estimate. We have to run $\texttt{Traj}(f, x_0, p)$ four different times. clearly Traj has to be organized differently The first calculates the value of the output at the relevent time, $T = 30$. We will call that Y_30. We then have to vary the parameters, one at a time. To make them comparable, we will vary each parameter by 10% of the nominal value and $S_i \approx \frac{y_i(p+\Delta p)}{\Delta p}$ for the three parameters $p = r, k, C$. We find that the discrete approximation yields $S_C \approx 2.374$, $S_r \approx 64.1062$, $S_k \approx 0.4531$ which is pretty reasonable agreement.

Chapter 4
Between Host Disease Models

Abstract

4.1 Historical Background

Mathematics has been used to study disease progression within a population since at least the 1600s when John Graunt used death records from London to develop a method to estimate the risk of dying for different diseases [**?**]. In fact, it is arguable that alongside ecological models, epidemiological models have been among the earliest models developed as well as those with some of the most lasting affect on peoples daily lives. Imagine, your insurance costs depends on factors that are, in part, determined by these models. The development of drugs from antibiotics, to antivirals, to flu vaccines (both in the development of the vaccine and in the use of the vaccine to prevent the spread of the flu). Where the government allocates resources such as HIV prevention (needle exchange? education?) and outbreak diseases like Ebola, West Nile virus – all of the policy decision making is influenced by epidemiological models.

daksflaksjdfaskljdf

4.1. What does SA in time imply about the other parameters in our model? Do the ranking of most sensitive change in time?