Lectures on Theoretical Ecology

Stephen P. Ellner
Department of Ecology and Evolutionary Biology
Cornell University

May 29, 2009

Contents

Chapter 1

Foraging theory

1.1 Testing adaptationism

Darwinian natural selection produces organisms that do a better job at generating offspring, relative to others with whom they compete. According to ecologist Henry Horn (Horn 1971), we should think of plants as "crafty green strategists" who extend branches, grow leaves and sometimes discard them – and so on – in the way that gets them the largest possible share of the light and nutrients that they need for photosynthesis. Herbivores and predators, in this view, are machines for turning "your" offspring into "my" offspring, by eating the former and siring or bearing the latter. The species and genotypes that survive to this day are the ones that have done the best job of this.

The "adaptationist program" attempts to understand organisms' traits in terms of their adaptive value for survival, mating, and reproduction. It looks for the *ultimate* explanation for traits – what purpose are they designed to serve? – rather than the *proximate* means employed to create them, in terms of biochemistry, physiology, and so on. Why do some trees drop their leaves in winter, while others retain them? Why do some organisms typically reproduce several times during their lifespan (iteroparity) while in others reproduction is typically fatal (semelparity)? A hypothetical answer to this kind of "why" question invokes some costs and benefits for the various possibilities, relative to the ultimate goal of leaving the most descendants in future generations.

Models are the means by which such adaptive hypotheses can be tested. To support the hypothesis that trait X is designed to accomplish goal Y, we construct a model to let us predict what trait X *should be* in order to accomplish Y as best as possible. The components of such a model are

1. A *goal function* to be maximized: either Darwinian fitness, a proxy for fitness such as "number of grandchildren that survive to maturity", or some component of fitness such as survival, reproduction, energy intake, etc.
2. The *constraints* under which the organism is operating e.g., what information does it have? what

patterns of movement are possible in its habitat?

3. A set of *options* for the organism.

The goal function has to be computable for any option within the assumed constraints – either by deriving a formula for it, or by constructing a computational algorithm to compute it. Additional assumptions often become necessary at this point, in order to derive a formula or algorithm for the goal function.

Armed with such a model, adaptive hypotheses can then be tested by comparing predictions with observations. This is most often done in a comparative manner:

- By looking across species to see if trait X varies as predicted, as a consequence of differences between species in parameters or constraints, or
- By exposing individuals of one species to a range of conditions and observing how they respond.

Such tests are *qualitative*: we perturb some aspect of the situation, and see if the trait or behavior changes in the predicted direction. So one qualitative test doesn't mean much, but a series of them can add up to strong evidence for or against a theory. Making a quantitative prediction is much more difficult, and rarer.

From the adaptationist perspective, all aspects of ecological systems are shaped by natural selection, so we can seek their ultimate explanation. Our first example will be the structure of feeding relations in an ecological community, which is often summarized by a *food web* such as that shown in Figure 1.1. Like most empirical food webs this one is not completely resolved to the level of species, but it results from what each individual in each species chooses to eat. A community ecologist might take the web as "given" and ask how its structure affects the dynamics of the consituent populations. For example, the wider arrows indicate a loop of interactions (the arrows indicate who eats whom, but the resulting impacts on population growth are two-way), and such loops may have important implications for community stability. But from the adaptationist perspective, we try to understand *why* each species feeds on the ones that it does and not on others; how much time individuals put into feeding versus other activities; how different individuals will organize themselves in space as a result of competition for food; and so on. This is called *foraging theory*, and it is one of the best success stories for the adaptationist program.

1.2 The prey model

The "classical" phase of foraging theory, roughly 1966-1986, centered on a few basic models: the prey (1966), patch (1976), and Ideal Free Distribution (1970) models. The best reference for the classical phase is Stephens and Krebs (1986).

The prey model addresses the choice of which food items a forager should try to capture and eat, so it is often called the "optimal diet" model. As Stephens and Krebs (1986, Chapter 2) observe, essentially

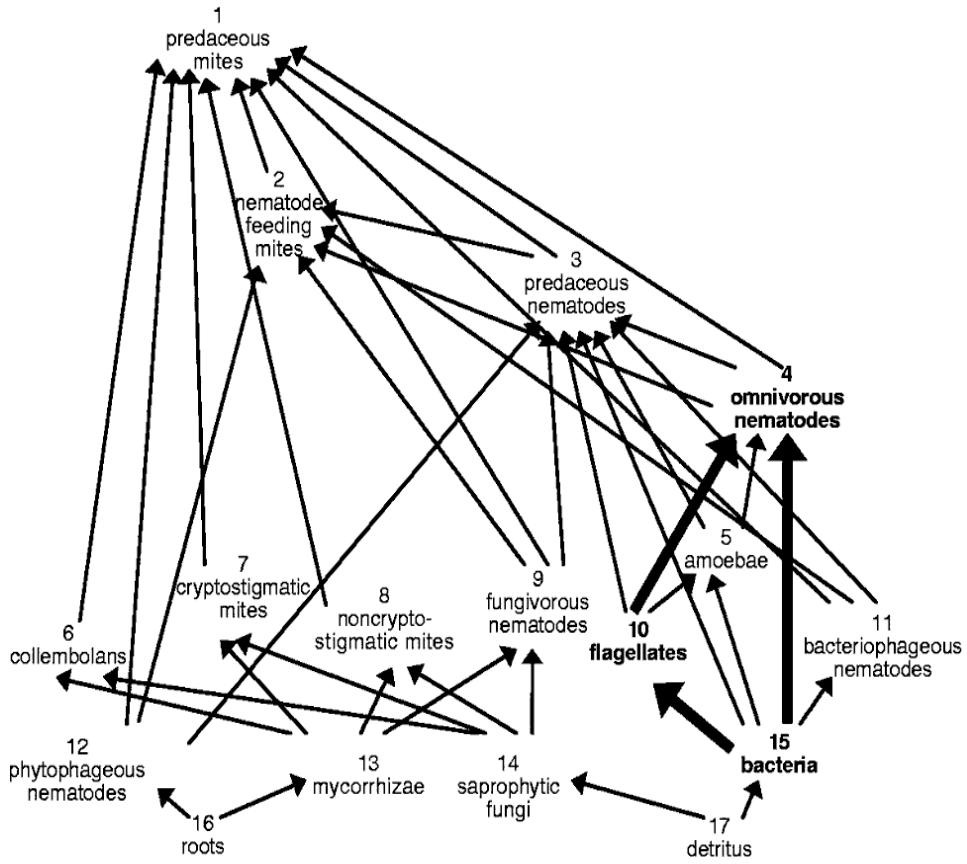


Figure 1.1: Food web summarizing feeding relations in a soil food web (native shortgrass prairie at the Central Plain Experimental Range, Nunn, Colorado) from Neutel et al. 2002.

the same model was proposed independently in half a dozen papers in the early 1970's, building on two similar models proposed independently by MacArthur and Pianka (1966) and Emlen (1966) in back-to-back papers. It was an idea whose time had come.

The model considers a single foraging individual, moving through space in search of food, and encountering prey items of several different types at random. We use "prey" to indicate any potential food item, which includes things like seeds or nectar-bearing flowers.

Model Notation

- i = index for prey type
- λ_i = rate of encounter with prey of type i (items/time = items/area $\times$ area/time)
- E_i = net energetic gain from catching and consuming a type-i prey item

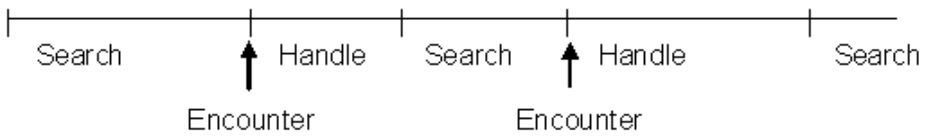


Figure 1.2: The search process assumed in the prey model.

- h_i = handling time for type i items (handle = catch + eat + resume search)
- c = energy cost per unit of time while searching

The classical prey model assumes that the forager knows all of these parameter values, and can identify which type of item it has encountered. If the forager decides to pursue an item, it switches to this activity, i.e. searching and handling are assumed to be exclusive.

Goal: Maximize the long-term rate of energy gain per unit time.

Constraints: The forager can only consume items that it has encountered. If it decides to pursue an item it switches to this activity, i.e. searching stops while the forager is handling an item that it wants to consume.

Options: when an item of type i is encountered, should the forager try to eat it, move on, or "coin toss" and pursue the item with some probability p between 0 and 1?

The forager's "decision variables" are therefore

$$p_i = \text{probability of attacking a type-i item, when one is encountered.}$$

To find the optimal diet, we need to compute the rate of energy gain as a function of the p_i. The forager alternates (by assumption) between searching for prey, and handling, as illustrated in Figure 1.2. S seconds (S large) of searching yields on average

$$S\lambda_i \text{ items of type } i$$

and requires

$$S\lambda_i p_i h_i \text{ seconds of handling time.}$$

Therefore,

$$\text{Total elapsed time is } T = S + S\sum_i \lambda_i p_i h_i$$

$$\text{Total net energy gain is } G = S\sum_i \lambda_i p_i E_i - Sc$$

The rate of energy gain (energy/time) is therefore

$$R = \frac{\sum_i \lambda_i p_i E_i - c}{1 + \sum_i \lambda_i p_i h_i} = \frac{\sum_i \lambda_i p_i E_i^*}{1 + \sum_i \lambda_i p_i h_i} - c \tag{1.1}$$

where $E_i^* = E_i + ch_i$. This informal calculation is rigorously justified by mathematical *renewal theory*. Let each search-and-handle segment of the foraging process be defined as a "cycle". The long term rate of gain is known to be given by the expected gain on a cycle, divided by the expected total time of a cycle. This gives the same result as our heuristic calculation above.

*From here on out we drop the *'s on E_j to simplify notation*, but remember that E has been adjusted in a way that lets us ignore search costs. We can also set $c = 0$ in equation (1.1), because the value of c has no effect on the optimal choice of p_i. An important principle is involved here: before trying to solve a problem, look for ways to simplify it or make it look simpler. In this case we've found a way to eliminate one parameter, c, by absorbing it into another. So instead of (1.1) we have

$$R = \frac{\sum\limits_i \lambda_i p_i E_i}{1 + \sum\limits_i \lambda_i p_i h_i}. \tag{1.2}$$

Having found R, our goal is to find the p_i that maximize it. We therefore compute $\partial R/\partial p_j$ and find that it has the form

$$\partial R/\partial p_j = [\text{positive}] \times \left(\frac{E_j}{h_j} - \frac{\sum\limits_{i \neq j} \lambda_i p_i E_i}{1 + \sum\limits_{i \neq j} \lambda_i p_i h_i} \right) \tag{1.3}$$

Here [positive] denotes an expression that is always positive, regardless of the values of the p_i. The term in parentheses does not depend on p_j so barring the exceptional situation where it is exactly 0, it is either positive or negative (mathematicians like to consider such exceptional situations, because they often lead to interesting mathematics, but for this class at least we're all biologists). We therefore have

Prediction 1. p_j is either 0 or 1 (no "partial preferences").

The two terms in the parentheses on the right-hand side of (1.3) can be interpreted as the rate of gain from eating a type-j item that has been encountered, and the average rate of gain from moving on with a policy of not eating type-j items. So if it is better to eat the item than move on, $p_j = 1$ is best. Otherwise, $p_j = 0$ is best. Put another way, the optimal diet consists of all prey items j where

$$\frac{E_j}{h_j} > \frac{\sum\limits_{i \neq j} \lambda_i E_i}{1 + \sum\limits_{i \neq j} \lambda_i h_i} \tag{1.4}$$

with the sum running over *all other prey types in the optimal diet.*

Equation (1.4) doesn't actually tell us what the optimal diet is, because you need to know the optimal diet to compute the right-hand side. But it does lead to the guess that foods with a higher value of E_j^*/h_j are better. In fact this is correct:

Prediction 2. Prey types should be ranked by profitability $\rho_j = E_j/h_j$, in the sense that the optimal diet consists of the items with the m highest values of ρ_j for some m.

To verify Prediction 2, we need to show the following: if $E_1/h_1 > E_2/h_2$ and type 2 is in the optimal diet, then type 1 must also be in the diet. We use proof by contradition. So suppose this is not true:

type 2 is in the optimal diet, and type 1 is not. The types in the optimal diet are then $j = 3, 4, \cdots, m$ for some m. Then from equation (1.4) we have

$$\text{type 2 in:} \qquad E_2/h_2 \quad > \quad \frac{\sum_{j=3}^{m} \lambda_j E_j}{(1 + \sum_{j=3}^{m} \lambda_j h_j)} \tag{1.5}$$

$$\text{type 1 out:} \qquad E_1/h_1 \quad < \quad \frac{\sum_{j=2}^{m} \lambda_j E_j}{(1 + \sum_{j=2}^{m} \lambda_j h_j)} = \frac{A}{B} \tag{1.6}$$

where A and B are the numerator and denominator in the last equation. We then have

$$\begin{aligned} \frac{E_2}{h_2} &> \frac{A - \lambda_2 E_2}{B - \lambda_2 h_2} \\ E_2 B - \lambda_2 h_2 E_2 &> h_2 A - \lambda_2 h_2 E_2 \\ \frac{E_2}{h_2} &> \frac{A}{B} \end{aligned} \tag{1.7}$$

But $\frac{E_2}{h_2} < \frac{E_1}{h_1} < \frac{A}{B}$, a contradiction proving Prediction 2.

So (once and for all) we can label prey types so that $\rho_1 > \rho_2 > \cdots$ and the optimal diet consists of types 1 through m for some m. To find m we can proceed sequentially using (1.4). Start with a "trial" value $\hat{m} = 1$ and ask "should type 2 be added to the diet?". That is, is

$$E_2/h_2 > \lambda_1 E_1/(1 + \lambda_1 h_1)?$$

If no: then m=1.
If yes: then set $\hat{m} = 2$ and ask if type 3 should be added:

$$E_3/h_3 > (\lambda_1 E_1 + \lambda_2 E_2)/(1 + \lambda_1 h_1 + \lambda_2 h_2)?$$

and so on.

This gives us a procedure for finding the optimal diet: it consists of items $1, 2, \cdots, m$ (ranked by profitability) where $m + 1$ is the smallest number such that

$$\frac{E_{m+1}}{h_{m+1}} > \frac{\sum_{i=1}^{m} \lambda_i E_i}{1 + \sum_{i=1}^{m} \lambda_i h_i} \tag{1.8}$$

is **not** true. That is, so long as (1.8) is satisfied by the highest-ranked item among those not already included in the sum, the new potential item should be added to the diet. If (1.8) fails, then the new item (and any of lower rank) should not be added.

Prediction 3. Whether or not a prey type is in the optimal diet depends on its profitability ρ but not its abundance, and on both the profitability and the abundance of more profitable prey items.

Assumptions of the Prey Model

- Search and handling time are exclusive
- Prey are encountered sequentially and at random in proportion to abundance
- Handling and recognition time are constants, unaffected by forager behavior or the abundance of food types in the habitat
- Encounter without attack takes no time or energy
- Perfect knowledge (forager knows the parameters and abundance of all food types)

Predictions

- 0-1 rule (no "partial preferences")
- Ranking by profitability (energy/handling time)
- More selective at higher encounter rates
- Selectivity independent of abundance of low-ranked (non-consumed) items
- Quantitative prediction of threshold profitability for inclusion in the diet

Results

- 51 qualitative tests, 21 quantitative
- 0-1 rule failed in **all** tests.
- Other predictions: 70% agree, 17% disagree, remainder inconclusive
- Mean fraction of assumptions satisfied: 84% for "Agree", 59% for "Disagree"

Table 1.1: **Tests of the Prey (optimal diet) model, based on Stephens and Krebs(1986)**

Prediction 4. An overall increase in abundance of all prey types will eventually increase selectivity (shrink the optimal diet).

Why? Because at each in/out test (1.8), if each of the λ's is increased, the value of the right-hand side becomes larger (verifying this is an Exercise below). That makes each test harder to pass, so m will eventually become smaller.

So does the model work? Stephens and Krebs (1986) reviewed over 70(!) laboratory and field tests of the prey model's predictions, 21 quantitative and 51 qualitative. Here "qualitative" means things like verifying that if you make all prey types more and more abundant, the optimal diet does eventually shrink. The results (Table 1.1) are amazingly good, except for the 0-1 rule.

There is a variant of the prey model in which it is assumed that an item has to be carried back to the

nest in order to be consumed, this is called "central place foraging". The handling time is then increased to $h_i + \tau$ where τ is the travel time back to the nest. Prey are then ranked by $E_j/(h_j + \tau)$, leading to the predictions that:

1. If τ is small, prey types are ranked by E_j/h_j
2. If τ is large, prey types are ranked by E_j
3. If $h_j \equiv h$ then increasing τ shrinks the optimal diet

For these, Stephens and Krebs report that 94% of tests were in agreement with the model predictions.

These analyses were recently updated and extended by Sih and Christensen (2001), who were then able to find 134 published tests of the optimal diet model. They again found very good agreement with predictions (Table 1.2, but identified different factors as determining agreement versus disagreement with the theory:

1. Qualitative versus quantitative tests, with better agreement for quantitative tests.
2. Prey mobility, with better agreement when prey were immobile or partially immobile.

In contrast to Stephens and Krebs, within each type of test (qualitative or quantitative) they found no relationship between the number of assumptions satisfied and how well model predictions were supported. As explanations for these patterns they suggested

1. Quantitative tests require greater knowledge of the study system, reducing the chance of making incorrect predictions due to mis-estimate parameters or some other misunderstanding of the study system.
2. With mobile prey, what a predator actually eats depends on prey behavior – how long and how successfully they can evade capture, for example – so the observed diet is not simply a reflection of the predator's decisions. Within-type variability in the ability of prey to defend themselves when attacked is likely to be large, but is ignored in the optimal diet model.

Predator attack is a strong selective force on the prey, favoring the evolution of defense traits and behaviors that can alter the predator's payoff for each prey type. Even very simple organisms can evolve defense traits – Yoshida et al. (2004) found that traits conferring partial defense against predation evolved in the unicellular green alga *Chlorella vulgaris*, when it was exposed to intense predation by the rotifer *Brachionus calyciflorus*. The impact of prey defense suggested by Sih and Christensen's analysis means that food webs and community structure have to be viewed from a *coevolutionary* perspective: asking what's optimal for the consumer is only half the story.

Exercise 1.1 For the following set of prey items, find the optimal diet according to the prey model.

Factor	**Quantitative tests**	**Qualitative tests**
Site of Study		
Laboratory	4.00 (NS)	2.31 (NS)
Field	3.89	2.52
Experiment?		
Yes	3.94 (NS)	2.31 ($P = 0.06$)
No	3.94	2.60
Forager type		
Invertebrate	3.67 (NS)	2.38 (NS)
Ectothermic vertebrate	3.63	2.37
Endothermic vertebrate	4.08	2.47
Prey type		
Immobile	4.22 ($P = 0.08$)	2.61 ($P < 0.005$)
Essentially immobile	4.25	2.40
Slightly mobile	4.00	2.44
Mobile	2.25	2.04

Table 1.2: Effects of different factors on the degree of fit between optimal diet theory and experimental tests, from Table 3 of Sih and Christensen (2001). For quantitative tests the level of agreement was scored from 1 to 6, with a value of 3 indicating qualitative agreement, and 6 indicating perfect quantitative agreement. The range of values for qualitative tests was therefore from 1 to 3. A study was considered to be an experiment if prey abundance was manipulated; all lab studies were classified as experiments. There were 35 quantitative tests and 99 qualitative tests included in these comparisons. Statistical significance of factors was assessed using Mann-Whitney U-test for two-level factors, Kendalls τ for multi-level factors; NS=not significant ($P > 0.05$). Year of study and number of assumptions met were not signficant for either type of test.

Item type	E	h	λ
1	2.6	1.2	4.8
2	2.6	0.7	6.7
3	4.3	1.4	7.3
4	5.3	1.4	7.0
5	7.8	1.9	1.5

Exercise 1.2 Verify the second half of equation 1.1, in which c has been eliminated by defining E_i^*.

Exercise 1.3 Verify that 1.3 is true, by finding the derivative and showing that it can be factored as claimed. If you need to brush up on differential calculus, see the Appendix at the end of this chapter.

Exercise 1.4 Verify that the right-hand side of (1.4) does indeed increase if the abundance of all prey types is increased, for any prey type j in the optimal diet. Hints: (1) the sums in (1.4) only involves prey types in the optimal diet, so we can ignore other prey types. (2) For each prey type k in the optimal diet, note that the right-hand side of (1.4) can be written in the form $(A + B\lambda_k)/(C + D\lambda_k)$. Use calculus to show that this last expression is an increasing function of λ_k for all prey types k in the optimal diet.

1.2.1 Risk-sensitivity

A second failure of the prey model, related to the 0-1 rule, is risk indifference. We assumed that all prey of a given type are identical. But the derivation of R remains the same if E_i and h_i are the average energy value and handling time for prey of type i. Consequently, the model predicts *risk indifference*: the amount of variance in type-specific parameters (with means held constant) should have no effect on the optimal diet.

There is much evidence that foragers are not risk-indifferent. Real (1981) allowed bumblebees and wasps to forage on blue or yellow artificial flowers, in arrays with

- Safe flower type: $2\mu L$ of nectar in each flower
- Risky flower type: either 0 or $6\mu L$ in each flower, with 2/3 of the flowers being empty

The bumblebees exhibited a strong preference for the safe flower type. To explain this Real et al. (1982) postulated that foragers have a nonlinear *utility function* U associated with energy gain. The constant flower yields $U(2)$, the variable one yields $(2U(0) + U(6))/3$. These need not need not be the same.

1. If $U'' < 0$, then $(2U(0) + U(6))/3 < U(2)$ – risk-averse favored
2. If $U'' > 0$, then $(2U(0) + U(6))/3 > U(2)$ – risk-prone favored

If the variance in energy gain X is small, then one can do a "small fluctuations" approximation which says that the long-term rate of energy gain is

$$E[U(X)] \doteq U(\bar{X}) + \frac{\sigma_X^2}{2} U''(\bar{X}) \tag{1.9}$$

where σ_X^2 is the variance in reward (we'll see later where this comes from). Real et al. (1982) showed that bumblebee behavior was consistent with some predictions from model (1.9). Harder and Real (1987) developed a mechanistic estimate of U for bumblebees, based on things like the time and energy required to extract different volumes of nectar, and found that it worked reasonably well to predict actual preferences for flower types in the lab.

Caraco et al. (1980) found both risk-prone and risk-averse types of behavior in foraging juncos: risk-prone if starved, but risk-averse if well fed before the experiment. This suggests that there is perhaps a *sigmoid* relationship between food reserves and fitness. Caraco et al. (1980) tested this idea by estimating the utility function of well-fed and hungry birds, and found that utility functions were indeed upward-curving for hungry birds, downward-curving for well-fed birds. (Here's one way to determine a utility function by observing behavior, assuming that the function exists and individuals behave so as to maximize their expected utility. Only relative utilities matter so we can take $U(0) = 0, U(6) = 1$ for 0 and 6 seeds obtained at a feeding station. Give birds a choice between a safe station with some constant number of seeds x, and a risky station having either 0 or 6 at frequencies $1 - p, p$. Then adjust p until

the bird is visits the safe and risky stations with equal frequency. This behavior is presumed to indicate that the safe and risky stations have the same utility for the forager. You can then estimate

$$U(x) = (1-p)U(0) + pU(6) = p.$$

Repeat this for $x = 1, 2, 3, 4, 5$ to fill in the utility function between 0 and 6. If you need to see more of the utility function, increase the reward at the safe station).

Initially much attention was given to simple models like (1.9) that combined the mean and variance of rewards in various ways into an overall reward function. However these have not stood up well against experimental data (Ellner and Real 1989), with one exception: models for "hunting by the hunted" that we will discuss below. The current approach to risk-sensitive foraging is discussed below in section 1.5

Exercise 1.5 For one of the Caraco et al. (1980) birds, the following utilities were estimated for 0 through 8 seeds: (0,0.05, 0.09, 0.18, 0.50, 0.55, 1.00, 1.05, 1.7). Given a choice between two feeder types offering
(1) either 2 or 7 seeds with equal odds
(2) either 1, 4 or 8 seeds with equal odds
which one should it prefer? What number of seeds x would be required for a safe feeder to be preferred over either of these risky feeders?

1.2.2 Predicting food web structure

We introduced foraging theory with the ambitious idea that the structure of real food webs, such as (Figure 1.1), could be predicted on the basis of foraging theory. Beckerman et al. (2006) have suggested that this actually works. Their "Diet Breadth Model" takes as given the number of species (for their analysis, the number of carnivores in real food webs). For each species in each web, they drew values of the parameters

- attack rate (A_{ij}) of species j on species i
- density (N_i)
- handling time (H_{ij})

from 123 published studies that estimated parameters of type-II functional responses across a wide range of taxa (data available from the authors upon request). The encounter rate λ_{ij} is assumed to be the product of attack rate and density and energetic content is assumed to be proportional to body mass:

$$\lambda_{ij} = A_{ij}N_i, \qquad E_i = eM_i$$

. They treat each species as a consumer, and predict which other species it should eat based on the prey model of optimal foraging theory.

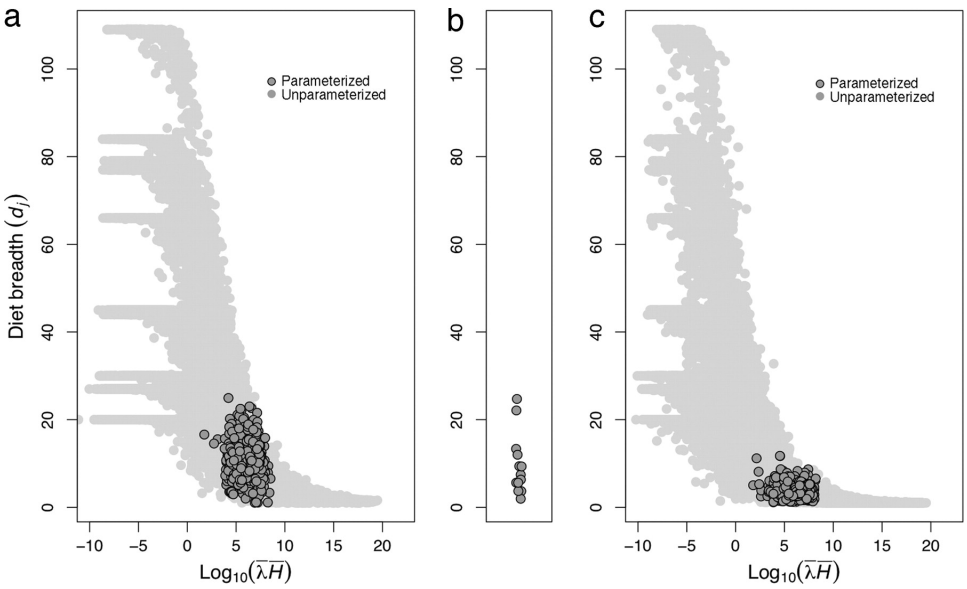


Figure 1.3: Predicted (a and c) and observed (b) diet breadths from Beckerman et al. (2006, Figure 1). (a) Predicted mean diet breadth from the parameterized Diet Breadth Model with 70% aggregation and 64 cumulative samples (dark circles). These values are not significantly different from the mean diet breadth of carnivores in 13 real food webs (b) (t test of predicted vs. observed diet breadths: mean observed = 9.7; mean predicted = 10.3; $t = 0.36$, $df = 12$, $P = 0.72$). Without parameterization (a; light gray circles), mean diet breadths can vary greatly. Diet breadth is influenced by the product λH. (b) Real food webs. (c) Identical to (a) but without aggregation and cumulative sampling. These diet breadths (black ringed circles) are are significantly lower than observed in real food webs. (mean observed = 9.7; mean predicted = 3.8; $t = 3.00$, $df = 12.0$, $P = 0.01$).

They then apply *aggregation and cumulative sampling* to model predictions. Aggregation means combining similar species into a "tropic species" and draw a link from i to j if any member of tropic species i eats some member of trophic species j. Cumulative sampling means that parameters are allowed to vary "over the year", and they count all links present at any time. The number of links defines the food web connectance $C = L/S^2$ (S=number of species, L=number of trophic links). The model leaves out quite a bit – for example, known effects of body size on foraging, and changes in foraging behavior as the density of prey species varies. Nonetheless it did a good job of predicting both the range of C values observed in real food webs(Figure 1.3), web-by-web connectances (Figure 1.4), and the relationship between C and S. Connectance is only one aspect of food web structure, but it is an important one for food web stability, and for robustness when a species is lost.

Brose et al. (2008) use the optimal diet model to make predictions about the strength of trophic links between predator-prey pairs: how the rate of energy flow along a link scales with predator body size. Models based on scaling of metabolism rates to body mass predict a monotonic relationship, whereas models based on foraging theory predict that the energy flow is maximized at intermediate predator body size. Laboratory experiments with predatory beetles and spiders supported the prediction from

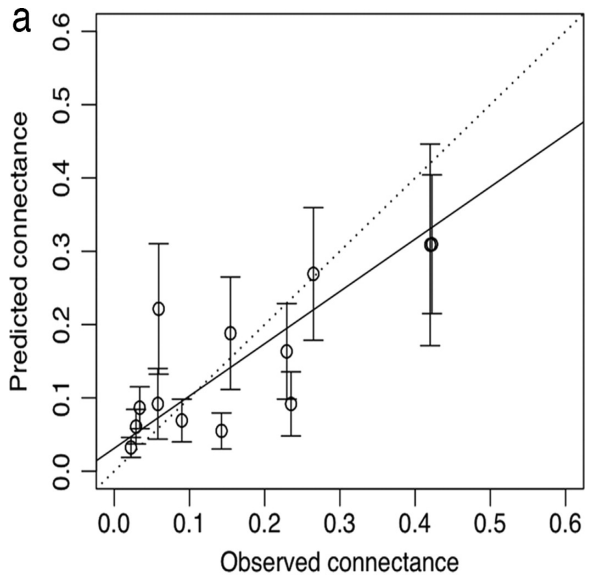


Figure 1.4: Comparison of predicted and observed connectance for 13 food webs from Beckerman et al. (2006, Figure 2). Dashed line is the 1:1 line, solid line is reduced major axis regression line. The slope of the regression line does not differ significantly from 1 ($P = 0.09$).

foraging theory.

These studies suggests that empirically parameterized models for individual adaptive decision-making may allow us to make explicit, mechanistic connections between individual-level choices and important ecosystem-level properties. The fundamental principle connecting individual and ecosystem scales (in this perspective) is natural selection: macroscopic properties are the result of individual properties shaped by current selective pressures, constraints, and evolutionary history.

1.3 Patch model

The patch model, originally due to Charnov (1976), operates on longer time and spatial scales. The optimal diet model pretends that prey abundances are constant, even while some prey are being eaten. We don't literally believe an assumption like this. It's a simplification that we adopt so that we can focus on one thing at a time, such as (in the diet model) an immediate decision between attacking and ignoring a single prey item that has just been discovered. But on a longer time scale, a forager or group of foragers may deplete the one "patch" of habitat and move on to another. The patch model addresses the question of when to abandon local search and start looking for a better patch somewhere else.

For now, we assume that all patches are exactly the same. The model is shown in Figure 1.5. The gain function $g(t)$, representing the totl net energy gain after time t in the patch, is assumed to be increasing but decelerating ($g'(t) > 0, g''(t) < 0$). The only other model parameter is the travel time T required

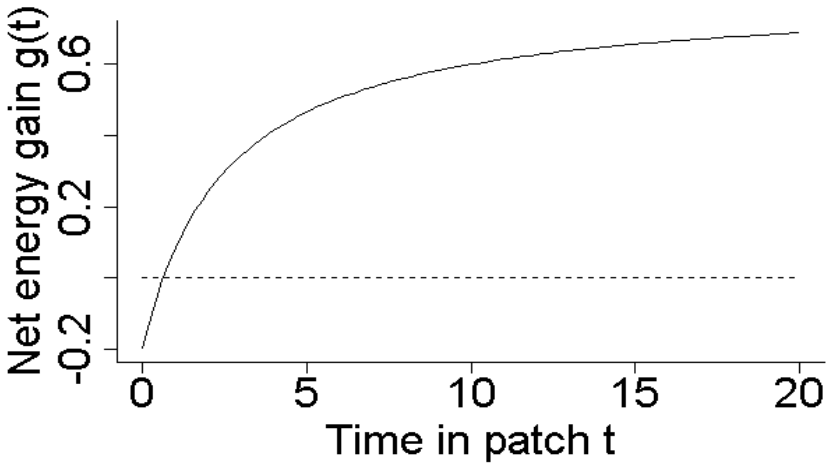


Figure 1.5: The patch model. Note that the energy gain on first arrival in the patch may be negative, to reflect the cost of getting to the patch.

for getting to the next patch.

The forager's decision is how long to stay in the patch, in order to maximize the rate of energy gain

$$R(t) = \frac{g(t)}{t+T}. \tag{1.10}$$

We have

$$\partial R/\partial t = \frac{(T+t)g'(t) - g(t)}{(t+T)^2}$$

so the maximum occurs when

$$g'(t) = \frac{g(t)}{t+T}. \tag{1.11}$$

This condition can be interpreted as follows. The right-hand side is equal to R, the long-term average rate of gain. The optimal "giving up time" is defined by the property that the instantaneous rate of gain within the patch, equals the overall rate of gain in the habitat as a whole. For this reason the solution of the patch model is called the Marginal Value Theorem (economists use "marginal" to mean "derivative of").

The condition defining the optimal departure time has a graphical solution shown in Figure 1.6. Mark the travel time T as a negative value on the time axis, and starting from that point $(-T, 0)$ swing a straight line down until it is just tangent to the graph of $g(t)$. Let $t*$ be point where the tangent line hits the graph of g. The slope of the line is then $g(t*)/(T+t*)$ (WHY?), which (because it is a tangent line) equals $g'(t*)$.

There is also an extension to multiple patch types. Suppose that the forager can tell what patch type it is in, but only after it has landed in the patch. Let P_i be the fraction of type-i patches and g_i the gain

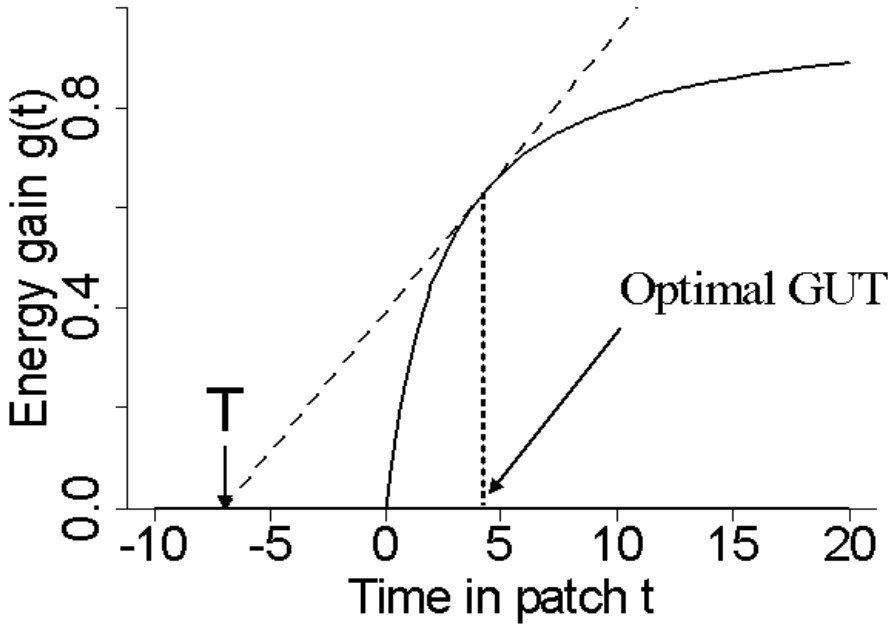


Figure 1.6: Graphical solution of the patch model. "GUT" is the forager's "giving up time", when it gives up on the current patch in order to go find another one

function for patches of type i. Then

$$R = \frac{\sum_i P_i g_i(t_i)}{T + \sum_i P_i(t_i)}$$

where t_i is the giving-up time for patches of type i. As in the prey model, R is given as the expected gain per cycle, divided by the expected time per cycle. With a bit of calculus the condition for optimal giving-up times turns out to be the same in this model as it was for a single patch type:

$$g_i'(t_i) = R.$$

Note that this model makes very specific assumptions about what the forager knows about the patch: "nothing" before it lands in it, and "everything" once it gets there. Under different assumptions, e.g. if there is assessment at a distance, things can be very different.

As with the Optimal Diet model, on the whole the Patch Model proved to be quite successful (Table 1.3 summarizes the review by Stephens and Krebs 1986). However, a subsequent review of 26 quantitative tests by Nonacs (2000) found that foragers consistently errer (when there were departures from Patch Model predictions) by staying in patches longer than predicted – cases where they left "too soon" were far rarer. This suggests that simplifying assumptions of the Patch Model lead to a systematic bias in

Assumptions of the Patch Model

- Foragers can recognize patch quality (if variable)
- Gain function in each patch type is known and constant

Predictions

- Giving-up time longer when travel time is higher, or mean quality of habitat is lower.
- All patches reduced to the same "marginal value" (rate of energy gain)
- More resources extracted from better patches
- Quantitative comparisons (ala Parker) of observed and predicted giving-up time.

Results

- 24 qualitative tests, 21 quantitative
- 70% agree, 8% disagree, remainder inconclusive.
- Mean fraction of assumptions satisfied=91% for "Agree", 75% for "Disagree"

Table 1.3: Tests of the Patch Model, based on Stephens and Krebs (1986)

its predictions. Nonacs showed that models incorporating other factors, such as starvation risk and predation risk, predicted longer patch residence times, and he suggested that these factors could explain the Patch Model's bias. Nonacs used Dynamic State Variable models, which are discussed below (section **??**). Carmel and Ben-Haim (2005) developed a very different type of model, based on the assumption that foragers are trying to achieve some required rate of energy gain G_{min}, and behave in the way that minimizes their risk of failure due to uncertainty about parameters. Instead of aiming for optimal and risking catastrophe, their foragers are trying to be good enough with certainty. Their key prediction, like Nonacs (2000), is that patch residence times should be longer than predicted by the Marginal Value Theorem.

The classical test of the patch model was Parker's (1978) study with dungflies, *Scatophaga stercoriara* (Figure 1.3. The "foragers" are males, the "patches" are females, and the "energy gain" is eggs fertilized. Males congregate at fresh cattle droppings, and attempt to copulate with arriving egg-bearing females. The longer a male remains in copulation, the higher (on average) fraction of her eggs will be fertilized using his sperm; females mate multiple times and an arriving female typically is carrying sperm from previous matings. After mating the male "guards" the female while she deposits eggs fertilized with his sperm, and this is included in the "travel time" (time until he mates with another female). Parker (1978) experimentally estimated the gain curve and the average "travel" (search+guard) time, and found close agreement between the predicted optimal copulation time (42.5 minutes) and the observed average (36.5 minutes).

Parker and Simmons (1994) looked more closely at variation in strategies between males of different sizes. Larger males have a faster rate of sperm displacement (a steeper gain curve) and a shorter search time, because they are able to displace a copulating smaller male and then fertilize most of the remaining eggs. Both of these favor shorter copulation times by larger males. Experimental results confirm this qualitative prediction, and there was also very good quantitative agreement between observed and predicted copula durations, except in the very smallest males. The model predicts very long durations (over 90 minutes) which are not observed. Parker and Simmons (1994) suggest that other factors or constraints might penalize such long durations, such as deterioration of the dung as a substrate for egg laying.

Exercise 1.6 Verify that (1.11) really specifies a maximum of $R(t)$ by computing $\partial^2 R/\partial t^2$, and using (1.11) and the fact that $g''(t) < 0$ to show that $\partial^2 R/\partial t^2 < 0$.

Exercise 1.7 Consider a habitat with patch-types having gain functions $g_i(t) = a_i t_i^{b_i}$, with $a_i > 0, 0 < b_i < 1$.
(a) For a habitat with only one patch type, having $a_1 = 2, b_1 = 0.5$ and travel time $T = 5$, find the optimal giving-up time according to the Marginal Value Theorem.
(b) For a habitat with $T = 3$, 70% patches of type 1 (as above) and the rest of type 2 with $a_2 = 1, b_2 = 0.7$, find the optimal giving up times in each patch.
You could do these by applying the Marginal Value Theorem, but it is more straightforward to do them numerically on the computer (for now, by evaluating the energy gain rate as a function of giving up times over a fine enough grid of values; later on we'll see how to do better).

Exercise 1.8 Best and Bierzychudek (1982) developed a patch-type model for bumblebees foraging for nectar on foxglove. Individual flowers are arrayed vertically on stalks, and (in their experiments) had 10 flowers per stalk labeled 1 to 10 from bottom to top. Bees were assumed to have the following strategy: on arrival to a stalk they checked the bottom flower for nectar. If it is empty they leave the stalk immediately, and if it contains nectar they visit $n \leq 10$ flowers on the stalk (in order from bottom to top) and then move on. The net energy intake rate, as function of n, is calculated as

$$R(n) = \frac{P(\sum_{i=1}^{n} E(i)) - P(n-1)E_w - PnE_f - (1-P)E_e - E_b}{T_b + P(n-1)T_w + PnT_f + (1-P)T_e}$$

where P=fraction nonempty bottom flowers = 0.375
$E(i)$ = mean nectar harvest from ith flower position = 20.19-1.7i cal
$E_b = 0.09$ cal, $T_b = 4.4$ sec are the energy loss and time flying between stalks
$E_w = 0.07$ cal, $T_w = 3.3$ sec are the energy loss and time flying between flowers within a stalk
$E_f = 0.02$ cal, $T_f = 14.7$ sec are the energy loss and time emptying a full flower
$E_e = 0.01$ cal, $T_e = 8.9$ sec are the energy loss and time checking an empty flower

(a) Justify the formula above for $R(n)$. Which of the following assumptions is it making: (1) Flowers on a stalk are independent (some are empty, some are full, with P being the fraction full) (2) Flowers on a stalk are either all empty or all full, and P is the fraction of full stalks.
(b) What is the optimal value of n? (answer this by writing a script file to compute R for any given

value of n).

1.4 Ideal Free Distribution

The third basic model is the Ideal Free Distribution (IFD), originally due to Fretwell and Lucas (1970). This model addresses the question of how a group of foragers, each trying to maximize their individual rate of energy gain, should distribute themselves among a group of patches.

Assumptions
1. Foragers are identical.
2. The reward (gain per time) in patch i is $w_i(n_i)$, where n_i is the number of foragers in patch i. This assumption means (tacitly) that resources are constantly being renewed in the patch – each flower makes more nectar, more females dungflies arriving at a dungpat, etc. – and the foragers in the patch are competing with each other for these new resources.
3. w_i is a decreasing function of n_i
4. Foragers are omniscient – in particular they know the reward functions w_i for each patch, and how many others are already in the patch – and are free to move among patches to achieve the highest possible reward, without constraint.

Prediction 1 All competitors experience equal gains, hence the average gain rate is the same in all patches.

To say any more, we need additional assumptions.

Prediction 2 If $w_i = q_i/n_i$ (where q_i is the resource supply rate in patch i) then n_i at steady-state will be proportional to q_i.

This is called the "input matching" or "habitat matching" rule, and has been repeatedly tested. Alternatively, we might suppose that there is interference between foragers. For example, in the prey model we expressed the prey capture rate as (prey/area) $\times$ (area searched)/time. Suppose the area searched per unit time is reduced by the presence of other foragers (giving each other a hard time) say $a_i = Qn_i^{-m}, m > 0$. The prey capture rate in patch i is then

$$\lambda_i = \frac{Qq_i n_i^{-m}}{1 + hQq_i n_i^{-m}}. \tag{1.12}$$

To equalize gains across patches, we therefore must have $q_i n_i^{-m}$ constant across patches, hence $n_i \propto q_i^{1/m}$.

The IFD has had the interesting property that its key prediction - average rate of gain per individual is the same in all patches - has been confirmed even when the model assumptions aren't met, in that there are consistent differences between individuals (Figure 1.8). To account for this, more recent models have incorporated differences between individuals, so that (e.g.) bigger ones make trouble for smaller ones. The trouble with this is that one can make up models *ad infinitum*, and predictions depend on assumptions for which there is still very limited experimental basis. In particular, once you drop the

assumption of forager equality, equality of average gain rate across patches is no longer a necessary outcome in the models.

1.5 Dynamic state-variable models

Current foraging theory is largely a response to the failures of the classical models: falsification of the 0-1 rule, sensitivity to risk, effects of variability and interactions among individuals, and the effects of hunger level on behavior .

Mangel and Clark (1986) said "let's bite the bullet" and admit that models including all these complications are too complex for analytic solution. Instead, they proposed using models that can be solved by a computational method called Dynamic Programming. These have the following ingredients

1. Individual state vector $X(t), t = 0, 1, 2, \ldots, T$ (T="final time").
2. The options available to an individual at each time t
3. Dynamic model for changes in $X(t)$ over time, and how these depend on behavioral decisions made at each time step.
4. An additive reward function of the form $W = \sum_{t=0}^{T} g_t(X(t))$

The optimal strategy can be found iteratively using Dynamic Programming. To introduce this method we have to talk about what it was really developed for: trucks (Figure 1.5).

Back to foraging. To see how Dynamic Programming can be used to find optimal state-dependent foraging decisions, we will build a deliberately oversimplified model for a forager having to decide between two patches, differing in food availability (λ_i = probability to find a food item) and the mortality risk (β_i = probability of being eaten). The forager's assumed goal is to remain alive until the final time T. The model components are:

1. State vector $X(t)$ = energy reserves at time t, $0 \leq X(t) \leq 5$.
When a forager falls to reserve level $X = 0$ it is dead of starvation, and $X = 5$ is full: even if more food is found, it can't be eaten.

2. Dynamic Model $X(t+1) = c(X(t) - 1 + Z(t))$
where $Z(t) = 0$ or 2 is the energy gain at time t, and $c(x)$ is the "chop-off" function

$$c(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 5 \\ 5 & x > 5 \end{cases}$$

Because $X = 0$ is death, we add the rule that $X(t+1) = 0$ if $X(t) = 0$. The dynamic model says that at each time step a still-living forager uses up one unit of energy. If it finds food in that time step $Z(t) = 2$ and then $X(t+1) = X(t) + 1$. If it doesn't find food, then $X(t+1) = X(t) - 1$.

3. Options At each time step the forager can choose to forage in habitat 1 or habitat 2, with parameters $\beta_1 = 0.01, \lambda_1 = 0.5$ or $\beta_2 = 0.05, \lambda_2 = 0.9$. Each of these is "risky" in its own way: Habitat 1 has a high risk of not finding food, Habitat 2 has a high risk of becoming somebody else's food.

4. Fitness (reward) function The assumed goal of surviving to time T means that the reward function is $W(X(T)) = 1$ if $X(T) > 0$, and $W(X(T)) = 0$ if $X(T) = 0$.

We can immediately see that this model leads to *partial preferences* – sometimes the risky patch will be used, and sometimes it won't be. For example, if a forager has $X(T-1) = 4$ then its best move is to go for the safe patch, since it doesn't need any more food to survive. But if $X(T-1) = 1$ it should go for the risky one, since the only way it will live is if it gets food, so it's best move is to maximize λ.

To find the optimal decision rule we define

$$J(x,t) = Pr\{\text{survive to time } T \text{ if } X(t) = x\} \text{ for a forager who behaves optimally between time } t \text{ and time } T. \tag{1.13}$$

$$P_i(x,t) = Pr\{\text{survive to time } T \text{ if } X(t) = x\} \text{ for a forager who choses patch type } i \text{ at time } t, \text{ and then behaves optimally between } t+1 \text{ and } T. \tag{1.14}$$

In general, $J(x,t)$ is the expected reward for an individual with state x at time t, assuming optimal behavior thereafter, i.e.

$$J(x,t) = E[W|X(t) = x].$$

J is analogous to "shortest route to the East Coast starting from where you are now" in the truck-routing example. For the reward function defined in item 4. above, the reward is $W = 1$ if the forager survives and $W = 0$ if the forager does not. So the expected reward is the probability that $W = 1$, which is the probability of survival up to time t. In other models, with different reward functions, the definition of J would be modified to correspond to the reward function. Similarly, the functions $P_i(x,t)$ are the expected reward for each of the choices available at time t.

As in the truck-routing example, we find values of J by starting at the end (time T) and moving backwards, alternating between computing values of J and values of P. The process starts by noting that the value of J at the final time T is determined by reward function W,

$$J(x,T) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \end{cases}$$

This holds because at time T there are no more decisions to make and no uncertainty as to the final outcome: the forager is either alive ($W = 1 if X(T) > 0$) or dead ($W = 0$ if $X(T) = 0$).

Then how do we move backwards? We know $J(0,t) = 0$ for all t so we only need to consider $x = 1, 2, 3, 4, 5$. We can then compute

$$P_i(x,t) = (1-\beta_i)[\lambda_i J(c(x+1), t+1) + (1-\lambda_i)J(c(x-1), t+1)] \tag{1.15}$$

or in words

(not eaten)[(eat)×J(new state if eat) + (not eat)×J(new state if not eat)]

For a forager with $X(t) = x$, the best decision on day t is to choose the patch type with the larger value of $P_i(x,t)$. We therefore have fitness function

$$J(x,t) = \max(P_1(x,t), P_2(x,t)) \tag{1.16}$$

and optimal patch choice

$$H(x,t) = 1 \text{ if } P_1(x,t) > P_2(x,t), H(x,t) = 2 \text{ if } P_2(x,t) > P_1(x,t). \tag{1.17}$$

So we know $J(x,T)$, and the steps above let us compute $J(x,T-1)$ for all x. Then we compute $J(x,T-2), J(x,T-3)$ and so on, all the way back to $J(x,0)$.

Typically, as you move backwards in time decisions eventually become a function of state, but not of time: $H(x,T-k) \to \tilde{H}(x)$ when k is large. $\tilde{H}(x)$ is sometimes called the *stationary solution*. In most applications, the model predictions that are reported are based on $\tilde{H}$, a time-invariant rule for how individual decisions depend on individual state.

The initial enthusiasm for these models is best seen in Mangel and Clark (1988), and the outcome in Clark and Mangel (2000). These models proved far harder to test than the classical foraging models: at most a handful of tests are reported in Clark and Mangel (2000), compared to the scores of tests of the basic foraging models in about the same amount of time. The reason for the difference is that there are no general predictions – everything is an inductive generalization from computed optimal strategies in a given model with a given set of parameters. So before you can test predictions you have to make them. That requires constructing the model and estimating the values of its parameters, including the state variable dynamics and fitness function under any possible sequence of decisions. That takes work and time.

For example, Clark and Mangel (2000, Chapter 1) summarize a study of prey choice in sticklebacks feeding on *Asellus* of different sizes. Experiments were conducted to estimate the energy content E_i, handling time h_i, and capture probability for prey of different sizes (3 to 9 *mm* in length), and to determine the energy requirements, stomach capacity, and energetic search costs of sticklebacks. Prey of intermediate sizes turned out to be the most profitable (largest E_i/h_i), while small prey were easiest to capture. A model was built, similar to our two-patch model, in which the fish's assumed objective was to remain above a critical level of food reserves $X(t)$, with death resulting from falling below the threshold. The model correctly predicted the observed partial preferences for the larger, riskier size classes: well-fed fish concentrated on the smaller prey that could reliably be caught, but fish with low energy reserves and thus a greater risk of starvation would take the gamble of pursuing larger prey when one was encountered.

Another issue is that the power and generality of dynamic programming make it tempting to build a complex model that's hard to understand. Mangel and Clark (1988) built a dynamic programming model based on Charnov and Skinner's (1984,1985) studies of wasps, *Trichogramma* and *Nasonia vitripennis*, encountering hosts for egg-laying of different types (size, species). When a female lays c eggs in a host of

type i, the contribution to her lifetime fitness can be measured by the resulting number of *grandchildren* (we count grandchildren rather than number of offspring, because offspring quality (size, vigor) is affected by host-type and egg number).

$$W_i(c) = \#\text{ of grandkids resulting from laying } c \text{ eggs on a type-i host.}$$

Experimental estimates of $W_i(c)$ have a maximum at some intermediate value of c, which is called the "Single Host Maximum" (SHM) because it is the number of eggs to lay if the female encounters only one host of type i. When actual egg-laying behavior is monitored
(1) females lay very variable clutch sizes, and
(2) clutches are almost all below the SHM.
This makes sense. A female should spread her eggs out over multiple hosts to reduce competition between her offspring.

Mangel and Clark (1988) model this by giving each female an initial reserve of R eggs. The model's state variable $X(t)$ is the number of eggs that she has left to lay, so that

$$X(t+1) = X(t) - c(t), X(0) = R. \tag{1.18}$$

The dynamic programming solution is based on $F(x,t)$=expected fitness increment from t to T, for a female with x eggs left to be layed at time t. Let

$$h_i(t) = \text{probability of encountering type-}i\text{ host at time } t.\ p_t = \text{probability of survival from } t \text{ to } t+1 \text{ (i.e. age-dependent)}$$

So T is defined as the age when p_{T-1} falls to 0. Then the dynamic programming iteration is

$$F^*(x,t) = \sum_i h_i(t) \max_{c \leqslant x} [W_i(c) + p_t F^*(x-c, t+1)] \tag{1.19}$$

and we know that $F^*(x,T) = 0$ (because she's dead).

Experimental data only let Mangel and Clark (1988) estimate W_i, so they generated numerical solutions for a range of plausible scenarios and looked for general properties. They predicted
(1) Older insects (near T) should lay larger clutches (because they are getting closer to the SHM situation).
(2) Due to different host encounter histories, females should have variable clutch sizes.
(3) An increase in mortality favors larger clutches.
(4) As h_i increases, the size of clutches on host-type i goes down.
(5) Females deprived of hosts for a while should then lay larger clutches than ones not deprived, and might use more host-types than a non-deprived female.
The last prediction matches experiments on apple maggot: after deprivation, females more readily used hosts marked by a pheromone showing that another female had already laid eggs there.

However, much of this was achieved with a simpler, analytic model (Parker and Courtney 1984). They assumed:

1. A female has R eggs that she can lay.

2. A female who finds a host has probability p of surviving to find another one.

3. The fitness payoff for laying c eggs on a host is $W(c)$.

From these assumptions, it was possible to derive the expected lifetime fitness as a function of the number of eggs laid per host. This model predicts that
(1) The optimal number of eggs per host is much smaller than the SHM.
(2) An increase in mortality favors larger clutches.
(3) A female who is running out of eggs should lay smaller and smaller clutches.

There will always be a tension between

- Simple broad-brush models that hope to make general predictions by including the most important bits, and hoping that the omitted details won't matter, and

- More complex models that take care to accurately represent case-specific details but therefore make case-specific predictions – generality coming only if general themes emerge from many specific cases.

The argument for the simpler model is that the analytic derivation of predictions lets you see exactly how predictions follow from assumptions. If a prediction is only there because of some dubious assumption that you made for the sake of simplicity, you'll know it and know not to put too much stake in the prediction. This may also lead to new predictions, because you'll see how predictions would differ for other situations where different assumptions would be appropriate. The argument for the more complex model is that you don't have to make so many dubious assumptions for the sake of simplicity, and you still get the conclusions that you need.

But of course you have a third choice – to build both kinds of model and get the benefits of each. Ideally the simpler models can be derived as approximations to the complex model, or as special cases of it, under some additional assumptions that don't hold exactly but might be close to the truth. The simpler model may then be a guide to the behavior of the more complicated one – providing hypotheses that can be checked computationally – while the more complicated one gives quantitative predictions that can be challenged with experimental data. We benefit from using both kinds of models, and properly used they can be mutually reinforcing.

Exercise 1.9 Write a program that uses Dynamic Programming to numerically solve the two-patch foraging problem with $\beta_1 = 0, \lambda_1 = 0.4$ or $\beta_2 = 0.05, \lambda_2 = 0.8$ for $x = 1 - 5$ from times $t = 99$ to 80, with $T = 100$. ("Solve" means to find, for each x and t, the optimal habitat for the forager to use). At what point (moving backwards from T) does the optimal strategy become time-independent?

Exercise 1.10 . The appeal of dynamic state variable models is ease with which complexities can be added. Patterned on the example above, develop a complete model for a forager deciding between two

patch types differing in prey availability λ and mortality risk β), which has both of the following two additional features:

(i) there is a third option "stay at home" with no chance to get food, or to get eaten ($\lambda_3 = \beta_3 = 0$).

(ii) the larger an individual is (as measured by their level of food reserves X), the higher their risk of death due to predation while actively foraging (perhaps they can't run as fast, or being larger are more easily detected)

Your description of the model should include a definition of any new variables or functions, all model components (state vector, options, dynamic model, reward function), and the backward-iteration equations analogous to (1.15)-(1.17) that could be used to solve the model and find the optimal strategy.

Exercise 1.11 Suggest another plausible feature that might be added to the basic two-patch foraging model, and predict what changes in optimal individual behavior might result from your modification of the model.

1.6 Hunting by the hunted

Another current aspect of risk-sensitive foraging is "hunting by the hunted". Most animals are at intermediate trophic levels: they are both predators and prey. Going out to hunt for food may place them at greater risk of being detected and eaten. There are a variety of models for different situations, and in some cases simple analytically solvable models have done well. The most successful of these is an approach originated by James Gilliam, based on the hypothesis that foragers minimize their mortality risk subject to getting enough food (to survive, or reach maturity, or whatever they are currently aiming to do).

Gilliam and Fraser (1987) considered a forager choosing among habitats $i = 1, 2, \cdots, n$ with associated energy gain rates $\mathbf{h} = (h_1, h_2, \cdots, h_n)$ and mortality rates $\mu = (\mu_1, \mu_2, \cdots, mu_n)$. The decision is what proportions of time to spend in each habitat, $\mathbf{p} = (p_1, p_2, \cdots, p_n)$. The choice of $\mathbf{p}$ implies

$$\text{Overall mortality rate } U = \langle \mu, \mathbf{p} \rangle$$
$$\text{Overall gain rate } H = \langle \mathbf{h}, \mathbf{p} \rangle$$

The forager's assumed **goal** is to minimize U subject to $H > H'$, where H' is some minimum requirement for energy gain – perhaps the lowest average rate for the day that will allow survival, or allow survival of the forager and its brood of young. The graphical solution is shown in Figure 1.10. The points represent the different habitats. The feasible region (corresponding to different choices of p) is the interior of the polygon (technically, the convex hull of the habitat points). The optimum is the lowest point in the polygon, to the right of the line $H = H'$. This is necessarily on the boundary, so the optimum involves at most two habitats.

Gilliam and Fraser (1987) tested this model experimentally with foraging creek chub minnow, given a choice of 3 habitats: refuge (no food or predators), safe (some food and predators), and risky (more

food and more predators). They used the model to predict the amount of food in the risky site that would lead the forager to switch from refuge+safe to refuge+risky as their two habitats of choice, and it worked (Figure 1.11).

Bednekoff (1997) developed a Dynamic Programming model for animals in a group making a decision between foraging and acting as sentinels. Each foragers decision is based on their own food reserve level (similar to the patch selection model) and on the number of others in the group acting as sentinels. A sentinel has a higher probability to detect an incoming predator (in which it alerts the flock and everyone escapes), but a higher risk of being the one attacked if a predator is not detected. A sentinel is performing a "good deed", but in this model its reason for doing so is selfish. The benefit for increased vigilance is assumed to outweigh the increased risk of being in an exposed position, so switching from forager to sentinel decreases an individual's risk of being eaten.

The solution to this model is a "switching curve": a line in the plane with axes x=Number of Sentinels and y=My Food Reserves (Figure 1.12). An individual above the curve (i.e., one with lots of food reserves) should be a sentinel, and one below the curve should forage. Finding that solution was hard because "my" decision depends on what everyone else's decision rule, so it takes more than just Dynamic Programming – but that's not something we can go into right now. After all this work, Bednekoff's model lets him make the following predictions:

1. Forage when you're hungry. When you aren't hungry, be a sentinel.

2. The higher the predation risk, the more hunger you'll tolerate before you quit being a sentinel and go back to foraging.

It's not clear that a full Dynamic Programming model was needed to reach these conclusions. On the other hand, the paper's predictions were soon put to the test experimentally, and in all cases the predictions were supported (e.g., Clutton-Brock et al. 1999). So there was a real benefit to having a "Swiss Army Knife" like Dynamic Programming that can handle all sorts of decision problems, even if it might sometimes be better to handle them in other ways.

Exercise 1.12 This exercise is inspired by (but not literally based on) Ydenberg and Clark's (1989) study of diving behavior in Western grebes. Imagine a sea bird diving in search of fish, and deciding how deep to go on each dive. The (hypothetical) tradeoffs are
(1) Fish are most prevalent in deeper waters, so the deeper a bird dives, the higher the probability that it will catch a fish on that dive (assume that the bird then has to surface to consume a fish).
(2) A deeper dive entails a higher energetic cost for the dive, and a higher probability of mortality.
Let's say that the energetic gain from catching and consuming a fish is 1 (all fish are equal, and we measure energy in fish units), the probability of catching a fish on a dive to depth y is $y^2/(50^2+y^2)$, the energetic cost of the dive is $(y/110)^2$, and the risk of mortality is $y/250$ with possible depths $y = 0$ to $y = 100$. For simplicity assume that the time required for a dive is independent of y – not realistic, but this is a homework problem, not a term project. A dive to "depth" $y = 0$ has no cost and no mortality risk – it can be interpreted as a decision to spend a time unit on shore rather than taking a dive.

(a) Build a complete Mangel-Clark style model for state-dependent depth choice during a series of dives at times $t = 0, 1, 2, \cdots, T$ during a day. Make any additional assumptions you need (for example, what is the reward function? – invent something plausible) and define any additional notation needed for your model. As in any such model, you will need to consider a discrete set of possible y values.

(b) Write out and explain the dynamic programming (backwards iteration) equations analogous to (1.15)-(1.17) that could be used to numerically solve your model.

(c) Build, and then find the solution for, a classical-type model that ignores mortality risk and assumes that the bird will dive so as to maximize its long term rate of energy gain.

(d) Build a Gilliam-Fraser style model, which assumes that the forager's goal is to minimize its mortality rate subject to the constraint that the expected rate of energy gain must be at least H. Specifically, give formulas for the expected energy gain and mortality risk as a function of dive depth y and use these to plot the model's graphical representation as in Figure 1.10 (note: you can build your plot by computing the energy gain and mortality risk for $y = 0, 1, 2, \cdots, 100$ and plotting one against the other).

(e) Describe how the value of H affects the optimal strategy in your model from part (d).

1.7 Reminder: the big picture

We have studied foraging theory as a paradigm of the adaptationist program, and it illustrates the structure of any adaptationist theory:

- The assumed goal (sometimes called the "currency", e.g. energy gain per time),
- The assumed set of constraints under which the organism is operating,
- The assumed set of options available (e.g., food types),
- The resulting policy which optimizes the goal subject to the options and constraints.

The last step is where models are essential, and make it possible to test the assumptions by comparison with experiments.

However, "testing optimality models is not an end in itself, but a means to gaining insight into the behaviour of individuals" (Krebs and Kacelnik, Ch. 4 in Krebs and Davies (1991)). We aren't testing whether animals are optimal. We are testing the adequacy of our assumptions about goals, options and constraints – which should include all the ways in which the organism is non-optimal, e.g. not omniscient, imperfect memory, and so on. It's an *iterative* approach to understanding the "why" of behavioral and other traits as as responses to ecological context, shaped by natural selection.

Even within animal behavior, adaptationist approaches have been applied to many things other than food-gathering:

1. fighting (e.g. males competing for females)
2. mate choice (e.g. males as "prey": mate with this one or search for a better mate?)
3. parental care (stay and raise this brood, or desert to sire another?)
4. conflicts between parents and offspring
5. territory acquisition and defense (how large a territory to defend? how to treat intruders?).

A major focus of research now is on interactions between individuals, in particular on signaling, e.g. chicks begging to their parents for food, and whether showy male traits are reliable as signals of 'quality', and the evolution of cooperative behavior. Many of these require "game theory" models, which we will be studying in a later chapter. There is also an increased attention to the mechanisms whereby goals are achieved, in order to objectively specify options and constraints (i.e. to remove the possibility of getting the right answer by imposing constraints). The latest editions of the books by Krebs and Davies cited below will give a review of current research in adaptationist approaches to animal behavior.

1.8 References

Alonzo SH, P. Switzer P and M. Mangel. 2003. Ecological games in space and time: the distribution and abundance of Antarctic krill and penguins. Ecology 86:1598-1607.

Beckerman, A.P., O.L. Petchey, and P.L. Werren. 2006. Foraging biology predicts food web complexity. PNAS USA 103: 1374513749.

Brose, U. R. B. Ehnes, B. C. Rall, O. Vucic-Pestic, E. L. Berlow, and S. Scheu. 2008. Foraging theory predicts predatorprey energy fluxes. Journal of Animal Ecology 77: 1072-1078.

Bulmer, M. (1994) Theoretical Evolutionary Ecology (chapter 6). Sinauer Associates, Sunderland Mass.

Caraco, T., S. Martindale, and T.S. Whitham. 1980. An empirical demonstration of risk-sensitive foraging preferences. Animal Behaviour 28: 820-830.

Caraco, T. 1981. Energy budgets, risk, and foraging preferences in dark-eyed juncos (Junco hymelais). Behavioral Ecology and Sociobiology 8: 213-217.

Carmel, Y. and Y. Ben-Haim. 2005. Info-Gap Robust-Satisficing Model of Foraging Behavior: Do Foragers Optimize or Satisfice? American Naturalist 166: 633-641.

Charnov, E.L. 1976. Optimal foraging: the marginal value theorem. Theoretical Population Biology 9: 129-136.

Clark, C.W. and M.S. Mangel. 2000. Dynamic State Variable Models in Ecology. Oxford University Press, Oxford UK.

Clutton-Brock, T.H., M. J. ORiain, P. N. M. Brotherton, D. Gaynor, R. Kansky, A. S. Griffin, and M. Manser. 1999. Selfish sentinels in cooperative mammals. Science 284: 1640-1644.

Ellner, S. and L. A. Real. 1989. Optimal foraging models for stochastic environments: are we missing the point? Comments on Theoretical Biology 1: 129-158.

Emlen, J.M. 1966. The role of time and energy in food preference. American Naturalist 100: 611-617.

Fretwell, S.D. and H.L. Lucas (1970) On territorial behaviour and other factors influencing habitat distribution in birds. Acta Biotheoretica 19: 16-36.

Gilliam, J.F. and D. F. Fraser. 1987. Habitat selection under predation hazard: test of a model with foraging minnows. Ecology 68: 1856-1862.

Harder, L.D. and L.A. Real. 1987. Why are bumble bees risk averse? Ecology 68: 1104-1108.

Heimpel, G.E., J.A. Rosenheim, and M. Mangel. 1996. Egg limitation, host quality, and dynamic behavior by a parasitoid in the field. Ecology 77: 2410-2420.

Horn, H.R. 1971. The Adaptive Geometry of Trees. Princeton University Press, Princeton NJ.

J.R. Krebs and N.B. Davies, An Introduction to Behavioral Ecology. 3rd Edition, Blackwell (1993).

J.R. Krebs and N.B. Davies (eds.) Behavioral Ecology: an Evolutionary Approach. 2nd Edition (1984), 3rd Edition (1991), 4th edition (1997).

MacArthur, R. H. and E. R. Pianka. 1966. On optimal use of a patchy environment. American Naturalist 100: 603-609.

Mangel, M. and C.W. Clark. 1986. Towards a unified foraging theory. Ecology 67: 1127-1138.

Mangel, M. and C.W. Clark. 1988. Dynamic Modeling in Behavioral Ecology. Princeton University Press, Princeton NJ.

McNamara, J.M., and A.I. Houston. 1996. State-dependent life histories. Nature 380: 215-221.

Milinski, M. and G.A. Parker. 1991. Competition for resources. pp. 137-168 in J.R. Krebs and N.B. Davies (eds). Behavioral Ecology: An Evolutionary Approach. Blackwell, Oxford.

Neutel, A. M., J. A. P. Heesterbeek, and P. C. de Ruiter. 2002. Stability in real food webs: Weak links in long loops. Science 296:1120-1123.

Nonacs, P. 2000. State dependent behavior and the Marginal Value Theorem. Behavioral Ecology 12: 7183

Parker, G.A. 1978. Searching for mates. pp. 214-144 in: J.R. Krebs and N.B. Davies (eds.) Behavioural Ecology: an Evolutionary Approach. Blackwell Scientific, Oxford.

Parker, G.A. and L.W. Simmons. 1994. Evolution of phenotypic optima and copula duration in dungflies. Nature 370: 53-56.

Possingham, H. P., A.I. Houston, and J. M. McNamara. 1990. Risk-averse foraging in bees: a comment on the model of Harder and Real. Ecology 1622-1624.

Real, L.A. 1981. Uncertainty and Pollinator-Plant Interactions: The Foraging Behavior of Bees and Wasps on Artificial Flowers. Ecology 62: 20-26.

Real, L., J.R. Ott, and E. Silverfine. 1982. On the trade-off between the mean and variance in foraging: effect of spatial distribution and color preference. Ecology 63: 1617-1623.

Real, L., S. Ellner, and L.D. Harder. 1990. Short-term energy maximization and risk-aversion in bumblebees: a reply to Possingham et al. Ecology 71: 1625-1628.

Sih, A. and B. Cristensen. 2001. Optimal diet theory: when does it work and when and why does it fail? Animal Behaviour 61:379-390.

Stephens, D.W. and J.R. Krebs. 1986. Foraging Theory. Princeton University Press, Princeton NJ.

Verdolin, J.L. 2006. Meta-analysis of foraging and predation risk trade-offs in terrestrial systems. Behav. Ecol. Sociobiol. 60: 457464.

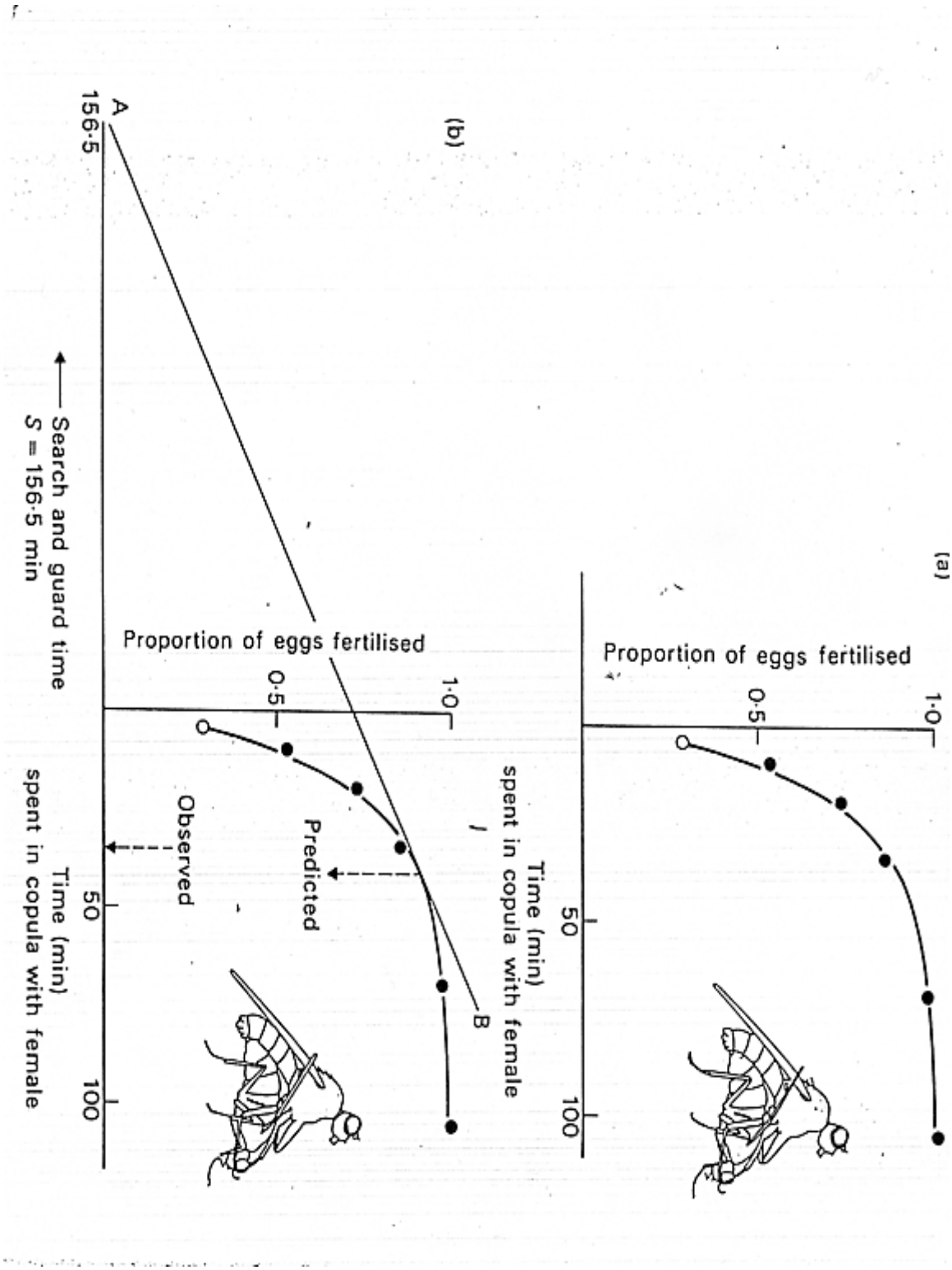


Figure 1.7: Results from Parker's (1978) test of the marginal value theorem using male "dungflies" foraging for egg fertilizations, with females as the "patches".

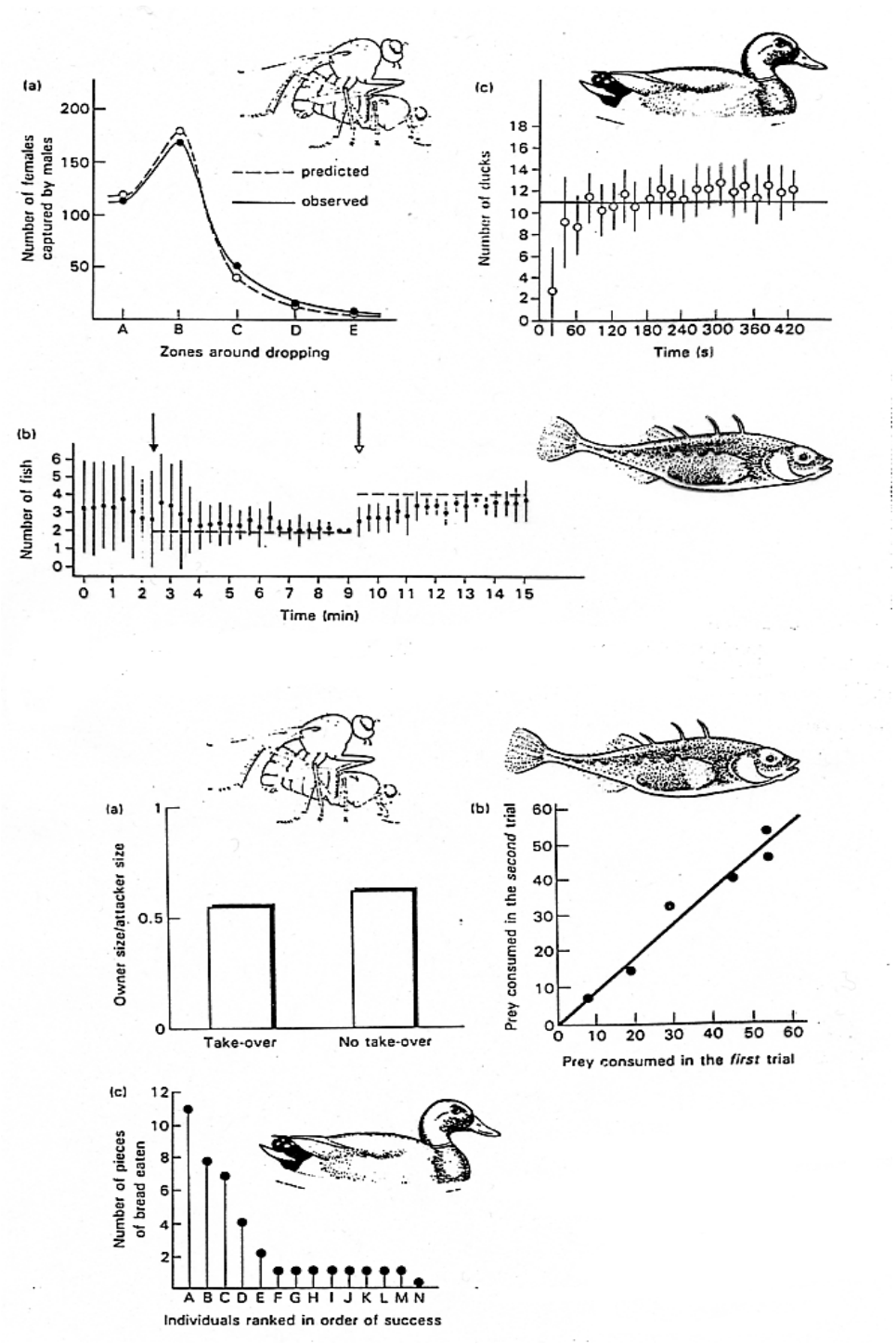


Figure 1.8: Experimental tests of the Ideal Free Distribution, from Milinski and Parker (1991). The upper 3 plots show experiments confirming the key prediction that individuals will distribute themselves so that the mean gain rate is the same in each patch. The lower 3 plots show that none of the studies actually conform to the assumption of forager equality.

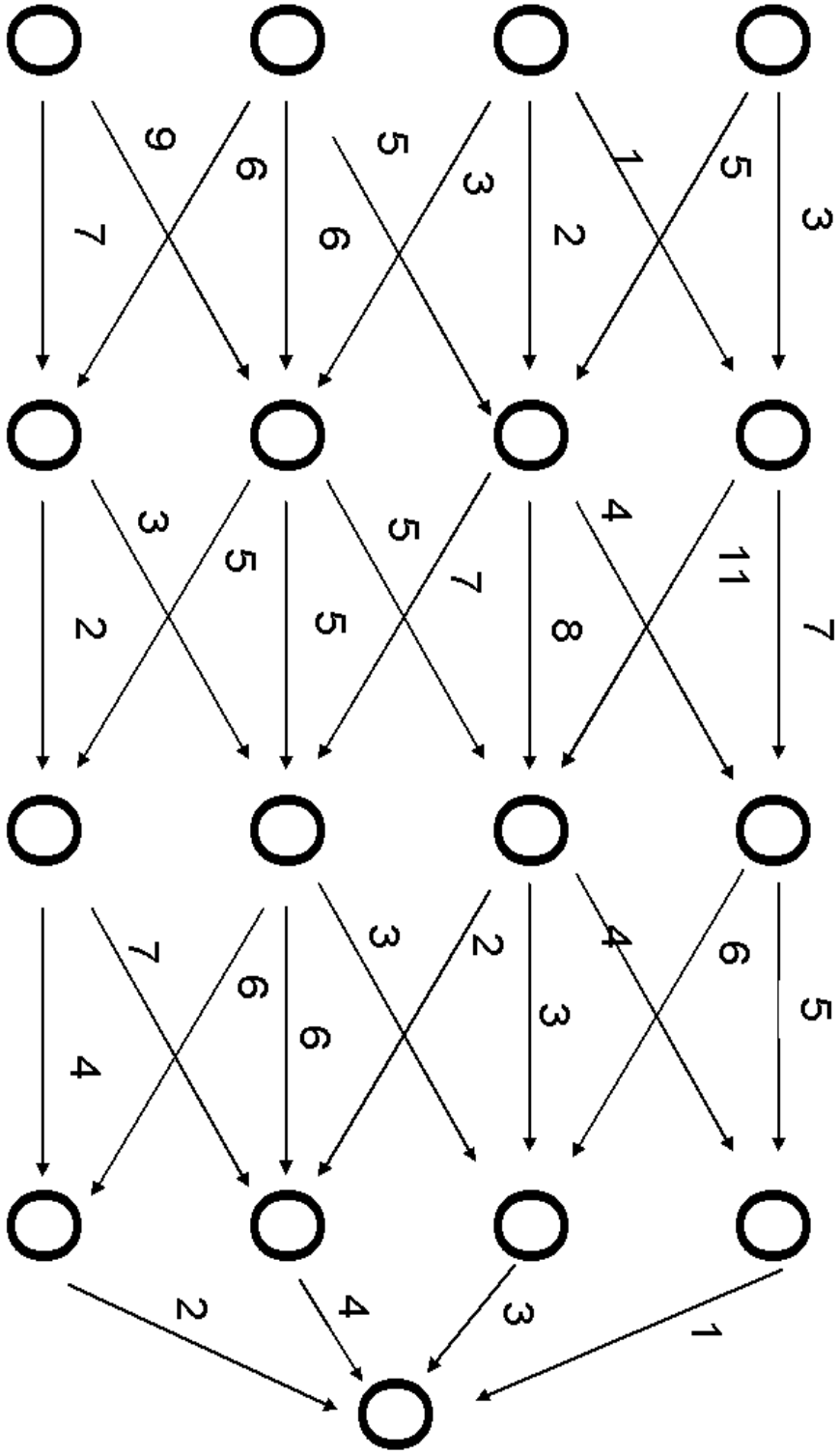


Figure 1.9: Problem: find the shortest route from a western node to the eastern node.

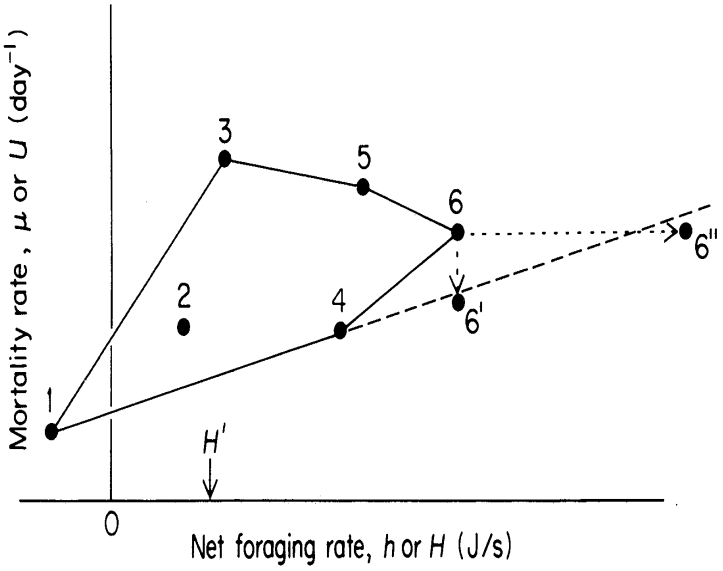


Figure 1.10: Graphical representation of the Gilliam and Fraser (1987) model of habitat selection for foraging under predation risk. Each possible habitat is characterized by its net foraging rate and predation risk. Lines connecting habitats represent foragers dividing their time between the two habitats. The optimal strategy is the lowest point of the polygon to the right of the required foraging rate H'. In this figure, the forager should divide its time between habitats 1 and 4. However, a sufficient improvement in habitat 6 – better foraging or lower risk – would lead to the forager abandoning habitat 4 in favor of habitat 6.

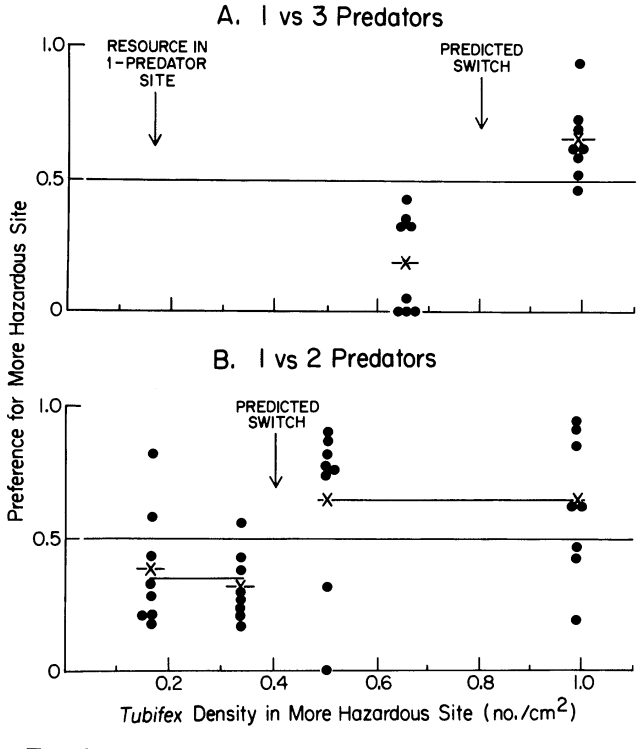


Figure 1.11: Test of the Gilliam and Fraser (1987) model using creek chubs foraging on *Tubifex* worms. The chubs were allowed to choose between 3 habitats: refuge, low-risk, and high-risk (like habitats 1, 4 and 6 in Figure 1.10). As predicted, an increase in the profitability of the high-risk habitat led to a switch from (refuge + low-risk) to (refuge + high-risk).

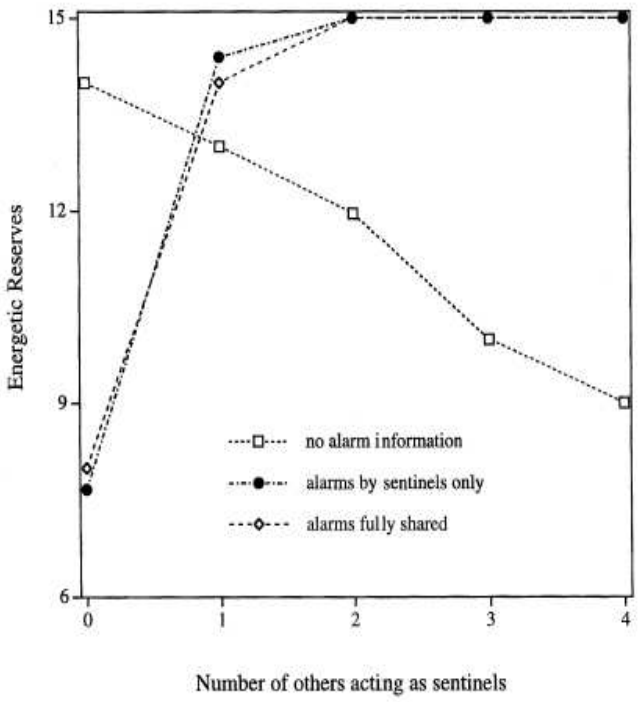


Figure 1.12: A numerical solution of optimal decision rules for the Bednekoff (1997) model of switching between foraging and sentinel behavior. The curves shown are for a group of 5 individuals, with individual food reserves restricted to 15 units or less. Individuals below the curve should forage, those above should be sentinels. The 3 curves correspond to different information about information sharing: no alarm by sentinel or forager who detects a predator (squares), alarms by sentinels only (solid circles) or by sentinels and foragers (diamonds).

Chapter 2

Bet hedging and Evolutionarily Stable Strategies

In the previous chapter we considered models for short-term decisions, making a choice among the available options for what to do "right now". However many important decisions involve a longer time scale, where the choice is between doing one thing now and doing another thing later – e.g., having more young this summer, versus being in better condition to survive the next winter and breed again next year. These longer term decisions are the subject of *life history theory*, which views the entire life cycle from the adaptationist perspective and seeks to answer questions such as:

- At what age should reproduction begin?
- Should reproduction be all-at-once (*semelparity*, meaning that there is a single reproductive event which is fatal to the parent) or spread out in time (*iteroparity*)?
- How many offspring should be produced each year?
- How should the parent allocate time and resources between maximizing offspring survival, and maximizing its own chances of living to breed again next year?

These questions are more difficult than short-term optimization, for at least two reasons. The first reason is the need to have some "common currency" to evaluate tradeoffs of survival against fecundity, or of fecundity now versus fecundity later. These different kinds of "payoff' must be combined into an overall measure of Darwinian fitness for the life cycle, using results from the theory of structured population models – so we'll come back to this in a later chapter.

The second reason – and the first subject of this chapter – is that there is enormous variation, from one year to the next, in the conditions under which decisions have to be made. – see Table 2.1. The world is very "noisy". For a forager trying to make it through the day, variation between one year and the

Long-lived adults	**Range of variation**	**Diapausing seeds/eggs**	**Range of variation**
Forest perennial plants	333	Chalk grassland annuals	1150
Desert perennial plants	4	Chapparal perennials	614
Marine inverebrates	591	Freshwater zooplankton	1150
Freshwater fish	706	Insects	31,600
Terrestrial vertebrates	38		
Birds	2200		

Table 2.1: Range of between-year variation in reproductive success (of long-lived adults or diapausing seeds or eggs), over years in which some reproduction occurred, based on Hairston et al. (1996). Hairston et al. (1996) compiled field studies that estimated per-capita reproductive success in a population, in several different years at the same location. They computed for each study the ratio between the highest and lowest annual per-capita reproductive success, omitting years when reproduction failed completely. The values in this table are the highest such ratio for species within each group listed.

next is not an issue. We can reasonably model foraging as if the current conditions were going to persist indefinitely, at least as a starting point. But when the structure of the entire life cycle is at issue, it is hard to justify models based on average conditions, such as the average payoff from a unit of time or energy invested in reproduction.

In this chapter we consider some simple models for individual decision making in the face of uncertainty, starting with a review based on Ellner (1997). The specific question is how reproduction should be spread across the life cycle, in situations where the payoff from trying to reproduce is highly unpredictable. So an attempt to breed this year may be fatal and unsuccessful, where waiting and trying next year might succeed – or the reverse might hold, and in advance one can't predict which.

2.1 It's not easy being a seed

The canonical example is Dan Cohen's (1966) theory for seed germination in a desert annual plant. A seed of a desert annual faces the problem of completing its life cycle and setting a seed crop within a short and unpredictable growing season that (in Israel, where Cohen worked) begins with rains in the winter and ends in a hot, dry summer. If a seed is going to germinate this year, it should do so at the first opportunity – the first good rain – but if that is not followed by enough rain later, the plant will die before setting seed.

Dan Cohen did his PhD on the physiology of seed germination, and then did a postdoc studying decision theory and applying it to germination and other biological "decisions" under uncertainty. It was a very good career move. As of this writing (2/7/07) his first resulting paper (Cohen 1966) has been cited 426 times, with 18 citations in 2006. The complete Introduction to that paper (whose brevity may not be unrelated to the paper's success) is as follows:

Figure 2.1: A seed-bearing dead skeleton of the Rose Of Jericho, *Anastatica hierochuntica*, in Algeria. Rainfall causes the skeleton to open partially, releasing some seeds to the soil where they can stay wet long enough to germinate. Lee Segel, one of the founders of modern mathematical biology, called this plant the ultimate Jewish Mother, who even after her death is keeping a tight grasp on her children and controlling their lives (*personal communication*). Image from *Botany Online: The Internet Hypertextbook* by Alice Bergfeld, Rolf Bergmann, and Peter von Sengbusch at `www.biologie.uni-hamburg.de/b-online`.

> Most living organisms are faced with considerable risk of failure when trying to reproduce. One obvious way to survive and reproduce in a risky environment is to spread the risk so that one failure will not be decisively harmful.

"Spreading the risk" of germination means that the seeds produced by one parent in one year will germinate at a range of times spread out over many years, even if they experience exactly the same conditions. This is accomplished by a variety of proximate mechanisms. A common one in desert plants involves germination inhibitors that must leach out of a seed before it can germinate. If seeds produced by a parent vary in the amount of germination inhibitor in their seed coat, two good things happen: germination is spread out over time, and a large rainfall causes more seeds to germinate than a small one. Another mechanism is differences in the sturdiness of a water-tight seed coat. Weaker ones will tend to crack sooner and allow the seed within to germinate, while stronger or thicker ones stay intact longer. Some desert plants retain seeds on their skeleton after they die, with mechanical processes causing some fraction to be released each time it rains. (Figure 2.1). If this piques your interest, see Evenari et al. (1971).

Figure 2.2 summarizes his model. At the start of growing season t, there are $n(t)$ buried seeds or eggs, of which a fraction H "decide" to hatch (seeds don't hatch, they germinate. But we'll eventually apply this model to copepod eggs so I'll call it the hatching fraction). Seeds that hatch each produce Y_t new

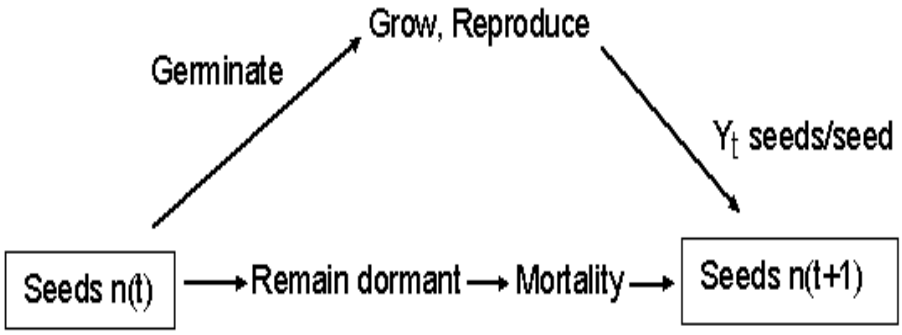


Figure 2.2: Cohen's (1966) model for optimal dormancy strategies in seeds of annual plants. At census times $t = 0, 1, 2, ...$ the population consists entirely of dormant seeds. Seeds that remain in dormancy have a constant risk of death, while seeds that germinate face a random and unpredictable set of conditions affecting their chances of completing the adult life cycle and producing a crop of new seeds.

seeds that re-enter the seed bank and survive until the start of growing season $(t+1)$. We consider Y_t to be highly variable: in some years there is a lot of rain and Y is large, in other years there is little or no rain and Y is small or 0. We also assume (for now) that Y_t is unpredictable: a seed has to germinate or not, without knowing if there will later be enough rain for it to grow, flower, and set seed. For seeds that don't hatch, life is predictable but unrewarding: a fraction $s < 1$ survive to time $t + 1$. The question is: should all seeds germinate, or should some "sit it out" until next year ($H < 1$)? And if so, how many?

We don't really believe that seeds are making a decision about whether or not to germinate. But to build an adaptationist model we begin by assuming that evolution will find a way to let individuals behave as they ought to. Then, if this model fails, we might ask if seeds are under some constraints that keep them from exhibiting the best possible decision rule.

Essentially the same model has recently been proposed (Stumpf et al. 2002) as an explanation for the *static latency* behavior of herpes viruses including herpes simplex HSV1 and HSV2, and varicella zoster virus (VZV) which is respondible for chickenpox and shingles. After an attack of chickenpox VZV lies dormant in nerve tissue. After many years it may re-emerge years to cause shingles. Shingles causes pain and discomfort that can last for months or even years, and an outbreak on the face may lead to permanently impaired hearing or sight. According to the US NIAID's FAQ page on shingles, roughly 20% of individuals in the US will get shingles at some time. Unlike the case of HIV, this is not an "active" latency resulting from host immune response – VZV has genes that actively maintain and regulate latency (Stumpf et al. 2002). Outbreak is more common in individuals with impaired immune function, but there is no known "trigger" for shingles to emerge from latency.

In the shingles model, the fluctuating environmental factor is the availability of susceptible hosts to infect, which they assume is unpredictable by the virus. In their model, an infected case can produce a

certain number c_1 of infections immediately, and another number c_2 after a latent period of duration τ unless the host dies beforehand, each modified by a random variable r_t proportional to the susceptible host density in year t:

$$N_{t+1} = r_t(c_1 N_t + c_2 m^\tau N_{t-\tau}).$$

Here m is the annual probability of host survival. Given enough variation in the abundance of susceptible hosts, Stumpf et al. (2002) show by simulation that large values of τ, similar to those observed for VZV, can be favored by natural selection – $\tau = 5$ years or higher for model parameters intended to represent preindustrial humans. This application illustrates the recent recognition that from the perspective of disease organisms **you are an ecosystem**. Ecological models are often relevant to understanding the cat-and-mouse game between the immune system and infections, and how each of these evolves in response to the other.

Now back to seeds: we assume for now that H is constant (and it can be proved that if Y_t is unpredictable, there is no advantage to having H vary between years). Then if H is the annual fraction that hatch, the population dynamics of the seedbank are

$$n(t+1) = [HY_t + (1-H)s]\, n(t). \tag{2.1}$$

This may seem like an odd way to specify an evolutionary model: where are the genes, Mendel's Laws, and the fitness differentials that drive evolution? But equation (2.1) has all the necessary information, if we adopt the simplifying assumption that "like begets like": all offspring have the same H as their parent. This can be made somewhat respectable by calling it a haploid asexual model: selection among competing clones. Each clone obeys (2.1) with its own value of H, and the changes in relative clone abundance translate into changes in genotype frequencies. So under like-begets-like, the winning H (which we denote H^*) is the one that results in the highest long-run population growth rate. Later in this chapter we discuss the limitations of this approach – for now we can view it as a useful simplifying assumption for a first attack on the problem.

So we have two jobs to do: we have to compute the long-run growth rate of a population obeying (2.1), and then figure out which value of H maximizes it, in order to identify the conditions in which be "winning" H is < 1. So first we have to review some probability theory.

Exercise 2.1 (a) Give one unstated simplifying biological assumption that is necessary for (2.1) to be valid as a description of population growth for our hypothetical desert annual plant with a seed bank ("simplifying" meaning, in particular, that the assumption is not likely to be exactly true in reality). (b) Propose a more realistic assumption, and derive the resulting replacement for (2.1) that applies under the more realistic assumption.

2.2 How to gamble if you're a seed

We can now return to our decision problem for dormant seeds. Denote $\lambda(t) = [HY_t + (1-H)s]$, so that

$$n(t+1) = \lambda(t) n(t). \tag{2.2}$$

Given $n(0) = n_0$ we then have

$$\begin{aligned} n(1) &= \lambda(0)n(0) \\ n(2) &= \lambda(1)n(1) = \lambda(1)\lambda(0)n(0) \\ &\vdots \\ n(t) &= \lambda(t-1)\lambda(t-2)\cdots\lambda(0)n(0) \end{aligned} \tag{2.3}$$

Since the Y's are independent the λ's are too. So if we define $\bar{\lambda} = E[\lambda]$ we have (by taking the mean of both sides)

$$E[n(t)] = \bar{\lambda}^t n_0 \tag{2.4}$$

This suggests that the key quantity for long-term population growth is the average annual growth rate $\bar{\lambda}$ – *but it isn't!* To see why we have to take logs in the last line of (2.3). The result is

$$\log n(t) = \log \lambda(t-1) + \log \lambda(t-2) \cdots + \log \lambda(0) + \log n(0) \tag{2.5}$$

Then dividing by t, we have

$$\frac{1}{t}\log(n(t)) = \frac{1}{t}\log(n(0)) + \frac{1}{t}\sum_{j=0}^{t-1}\log(\lambda(j)) \to E\log(\lambda(t)). \tag{2.6}$$

Thus $\log n(t) \sim \rho t$ where $\rho = E[\log \lambda(t)]$. A population will grow if $\rho > 0$, and decline to extinction if $\rho < 0$.

Note that if $E[\lambda] > 1$ but $E[\log \lambda] < 0$ (which is possible) the expected population size $E[n(t)]$ grows to infinity, despite the fact that the population converges to 0 since $\log n(t) \to -\infty$ (Lewontin and Cohen 1969). How is this possible? Think of "expected population size" as the average over "many" populations growing in separate patches with their own independent $\lambda_i(t)$. Eventually all of them go extinct ($n_i(t) \to 0$). But at any finite time, some of them are "lucky" and have very high population densities. The Weak Law of Large Numbers tells us that as time goes on, the fraction of "lucky" patches drops to zero. So the fact that the average keeps growing means that as time goes on, you have a smaller and smaller fraction of the populations doing weirder and weirder things, before crashing to 0 in the long run. Figure 2.3 shows an example, 250 independent runs where $\lambda(t)$ took the values 0.7 and 1.35 with equal probability, so $E[\lambda] = 1.025, E[\log \lambda] \doteq= -0.03$. The eventual tendency is downward, but at any given time a few oddball runs cause the mean to go up.

Exercise 2.2 (a) Use the Central Limit Theorem to show that the asymptotic (large-t) distribution of $n(t)$ is lognormal: that is, $\log n(t) \sim N(\mu_t, \sigma_t^2)$ where N is a Normal (Gaussian) random variable with mean μ_t and variance σ_t^2 (the subscripts indicating that the mean and variance depend on time t). What are the formulas for μ_t and σ_t in terms of the model parameters and the initial conditions? (b) How would you go about testing this prediction in the field? (be optimistic – assume you can do any population sampling etc. required to get the data that you want – but limit yourself to the typical 3-year time span of a research grant).

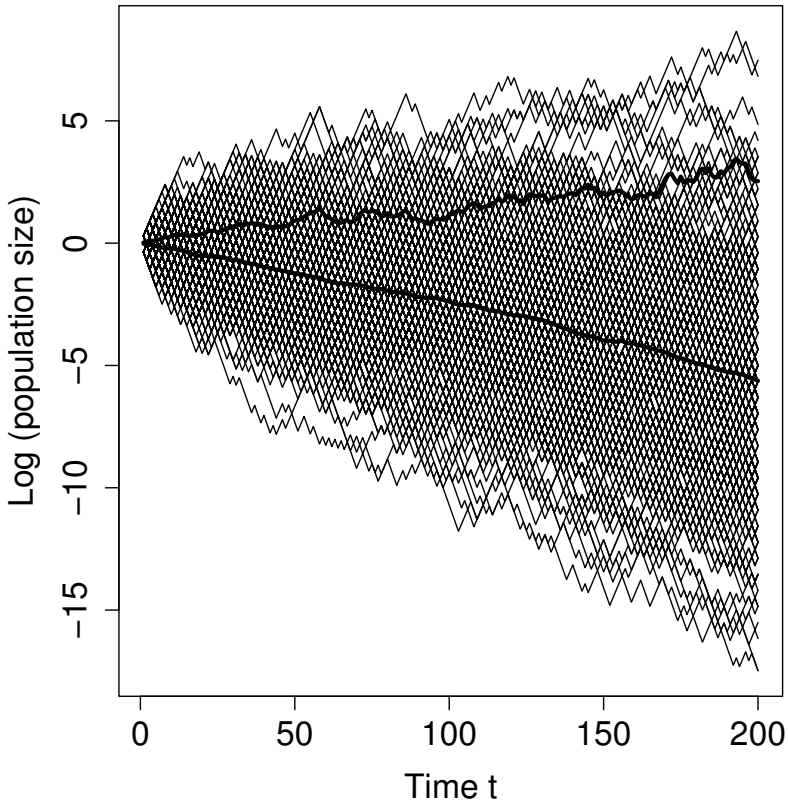


Figure 2.3: Random-environment model where $\lambda(t)$ took the values 0.7 and 1.35 with equal probability. The thin lines are $\log n(t)$ from 250 replicated simulations. The increasing thick line is the log of the average population size, the decreasing thick line is the average of the log population size.

2.2.1 Computing ρ

1. If $\lambda(t)$ is discrete: possible values $\lambda_1, \lambda_2, \cdots \lambda_n$ with probabilities $p_1, p_2, \cdots p_n$. Then

$$\begin{aligned} \rho &= E[\log \lambda] = \sum_i p_i \log \lambda_i \\ \text{hence } e^{\rho} &= e^{\sum_i p_i \log \lambda_i} = \prod_i e^{p_i \log \lambda_i} = \prod_i (e^{\log \lambda_i})^{p_i} = \prod_i \lambda_i^{p_i} \end{aligned} \tag{2.7}$$

This last quantity is known as the *geometric mean* of λ.

Example:

$$\lambda = \begin{array}{ll} \bar{\lambda} + \sigma & \text{with probability } 1/2 \\ \bar{\lambda} + \sigma & \text{with probability } 1/2 \end{array} . \tag{2.8}$$

In this case $E[\lambda] = \bar{\lambda}$, and $Var(\lambda) = \sigma^2$. Then

$$\begin{aligned} e^{\rho} &= (\bar{\lambda} + \sigma)^{1/2}(\bar{\lambda} - \sigma)^{1/2} = \sqrt{\lambda^2 - \sigma^2} \\ \rho &= (1/2)\log(\lambda^2 - \sigma^2) \end{aligned} \tag{2.9}$$

So a bad year hurts more than a good year helps, when their deviations from the mean are equal in magnitude. If the stock market goes up 10% one day and down 10% the next day (or vice versa), you have less money than you started with. You learned long ago that $(1+x)(1-x) = 1-x^2$, and now you know what it means for plants.

2. If $\lambda(t)$ is continuous with density $p(\lambda)$: Then by the Law of the Unconscious Statistician,

$$\rho = \int_0^{\infty} \log(\lambda) p(\lambda) d\lambda. \tag{2.10}$$

Equation (2.10) is typically useless since you can't evaluate the integral. So one can resort to the Small Variance Approximation with

$$f(x) = \log x, f'(x) = 1/x, f''(x) = 1/x^2$$

which gives

$$E[\log \lambda] \doteq log(\bar{\lambda}) - \frac{\sigma^2}{2\bar{\lambda}^2}. \tag{2.11}$$

where σ^2 is the variance of λ. This is pretty crude (because it makes no use of the distribution of λ) but it conveys the essential information: *variance hurts*.

"Bet hedging" is the general term for strategies that reduce the variance of rewards, at the price of a reduction in the mean reward. Phillipi and Seger (1989) identify two categories of bet-hedging: conservative and diversified. A conservative strategy for a shrub to deal with drought, for example, would be to store up water when it's available rather than using it all for immediate seed production, even if there is a cost to storage that reduced the number of seeds that can be produced per unit of water taken up. A diversified strategy involves some variability of behavior at the individual level; at the level of the genotype (present in multiple individuals) this results in "not putting all your eggs in one basket".

2.2.2 Optimal seeds

Our seedbank/eggbank model is a diversified strategy: some gamble on this year being good, others sit it out and try again later. The general theory above says that the winning H is the one which maximizes

$$\rho = E \log \left[HY_t + (1-H)s \right]. \tag{2.12}$$

To see what H this is, we differentiate within the expectation and do some curve-sketching of ρ as a function of H for $0 < H < 1$. [The expectation is either a sum or an integral, and in either case it is legitimate to move differentiation inside the operation of summing/integrating].

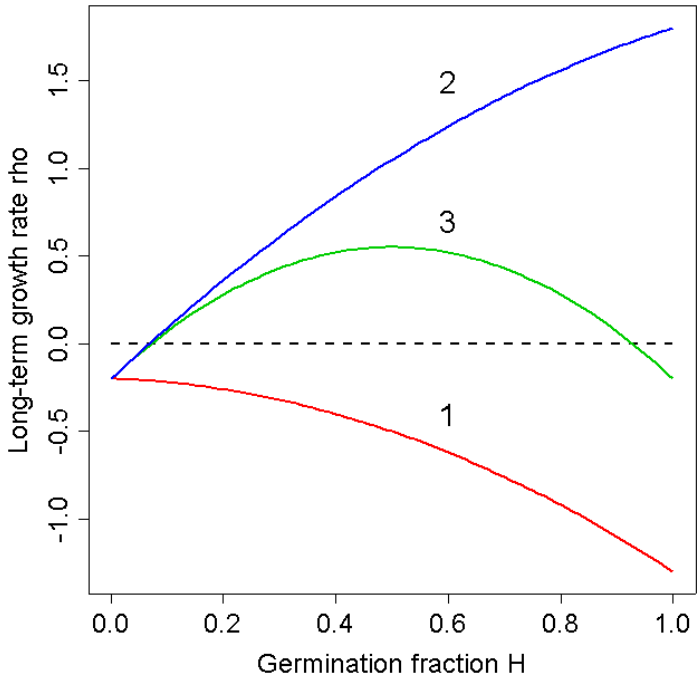


Figure 2.4: Curve-sketching $\rho(H)$ for Cohen's (1966) model. Because $\rho(0) = \log(s) < 0$ and $\rho''(H) < 0$, there are only three qualitative possibilities determined by the sign of ρ' at 0 and 1.

$$\rho'(H) = E\left[\frac{Y_t - s}{HY_t + (1-H)s}\right] \qquad \rho''(H) = E\left[\frac{-(Y_t - s)^2}{(HY_t + (1-H)s)^2}\right] < 0 \tag{2.13}$$

Because of the uniformly negative second derivative, there there are only 3 possibilities (see Figure 2.4) which can be distiguished based on the signs of $\rho'(0)$ and $\rho'(1)$:

1. $\rho(H)$ is decreasing on [0,1]. This occurs if $\rho'(0) < 0$, i.e. if $E(Y_t) < s$, and the optimum strategy is $H = 0$ giving population growth rate $\rho(0) = \log(s) < 0$, implying that all types decrease to zero abundance. So this possibility is actually not possible, and any real population must have $\rho'(0) > 0$.

2. $\rho(H)$ is increasing on [0,1], in which case the optimum strategy is $H = 1$, all hatch at the first opportunity.

3. $\rho(H)$ has a maximum at some H^* between 0 and 1, so the optimal strategy is to bet-hedge: some hatch each year and some wait until later.

We can tell which of the two possibilities holds by looking at the sign of $\rho'(1)$: The optimum is at $H = 1$ if $\rho'(1) = 1 - sE[1/Y_t] \geq 0$ and at some H^* between 0 and 1 if $\rho'(1) < 0$. Thus we found when some

seeds should "sit it out" each year:

$$H^* < 1 \text{ if and only if } sE[1/Y_t] > 1. \tag{2.14}$$

Having derived condition (2.14), we need to make sense of it. There are two special cases that usually help: *small variance*, and *good years/bad years*.

Exercise 2.3 Use the small variance approximation to derive that

$$E[1/Y_t] \doteq \frac{1}{\bar{Y}} + \frac{\sigma^2}{\bar{Y}^3} = (1/\bar{Y})[1 + (\sigma/\bar{Y})^2] \tag{2.15}$$

By substituting the result of the Exercise above into (2.14), we have the "small fluctuations" approximate condition for $H^* < 1$:

$$(s/\bar{Y})\left[1 + (\sigma/\bar{Y})^2\right] > 1. \tag{2.16}$$

$\bar{Y}$ must be > 1 (otherwise ρ would be negative for any H by Jensen's Inequality), so $(s/\bar{Y}) < 1$. The interpretation of (2.16) is therefore that all seeds should hatch each year if the variance in reproductive success is small, while if the variance is large some should sit it out. *Note the paradox here*: we derive the condition by assuming small variance, and interpret what it means when the variance is large. Don't let this bother you – we're going for qualitative insight, not mathematical rigor. As with any approximation, you check your insights on the computer to see if the model really acts like you expect it to.

The *good years/bad years* case is to suppose that $Y_t = M$ or m with probabilities p and $1-p$ respectively, where $m \ll 1 \ll M$. Then from (2.13) we calculate directly

$$\rho'(H) = p\left[\frac{M-s}{HM+(1-H)s}\right] + (1-p)\left[\frac{m-s}{Hm+(1-H)s}\right] \doteq \frac{p}{H} - \frac{1-p}{1-H} \tag{2.17}$$

using the assumption $m \ll s \ll M$ to simplify the fractions. Thus the optimum is approximately at $H = p$: the fraction of seeds that hatch should be approximately equal to the fraction of good years, p.

Since $\bar{Y} > 1$ but $s < 1$, the *expected* number of seeds next year (new plus survivors) is always maximized at $H = 1$. Thus the "decision" to have $H < 1$ when the variability in Y is large, really is a bet-hedging strategy: the mean payoff is reduced, but in the long run it pays off because it reduces the variance.

Exercise 2.4 Verify the statements above about the effect of H on the mean and variance of the population growth rate $\lambda(t) = HY(t) + (1-H)s$, i.e. show that the mean and variance of $\lambda(t)$ are both increasing functions of H [HINT: these can both be done using the rules on mean and variance listed in section 7.4].

2.3 Limits to Growth and Evolutionary Game Theory

One unrealistic aspect of Cohen's model is that population density increases without limit for many "losers" as well as the "winner" phenotype. This behavior comes from the simplifying assumption that

per-capita fecundity Y is unaffected by crowding. One has to start somewhere and often a density-independent model is a good place to start. But given that there always are limits to population growth imposed by crowding, it's important to see if a model with such limits makes different predictions.

The traditional way of modeling limits to growth is to assume that each individual's per-capita reproductive success is affected by the overall density of competitors, say $Y_t = K_t F(N_t)$ where K_t is random and $N_t = \sum_i H_i n_i(t)$ is the total number of seeds emerging from dormancy (i running over all competing types in the population). The key simplifying assumption is that the population is well-mixed in space, i.e. spatial variation in density is small enough to ignore.

Because strategies are now pitted against each other, instead of an "optimal" strategy we seek a "evolutionarily stable" strategy or **ESS**: a value H^* such that any competing type with a different H is at a disadvantage when type H^* dominates the population. Germination is then an *evolutionary game* in the sense of mathematical game theory: a contest where the payoff to each player depends not only on its own behavior, but also on what other players do. An ESS as defined above is closely related to the *Nash equilibrium* solution concept of mathematical game theory. A Nash equilibrium is a strategy (or set of strategies for all players) such that no player can do better by making a unilateral change of strategy.

Formally, let $\tilde{Y}_t$ denote the value of Y_t in a population consisting entirely of type-H individuals. Then the logic behind equations (2.2) and (2.6) still applies to give the growth rate of a rare type h invading an otherwise all-H population:

$$\rho(h|H) = E \log \left[h\tilde{Y}_t + (1-h)s \right] \tag{2.18}$$

This approach is a shortcut called an "invasibility analysis" that avoids having to analyze competition between two competing types. Instead we consider a "resident" tye H that has been established in the habitat for a long time, faced with a mutant "invader" h. We assume that the invader is at such low density that H goes on as if the mutant were not present: the resident affects the invader, but not vice-versa. If H is such that no invader can increase from its initial rare density, then H is called an **ESS** (evolutionarily stable strategy). This is much easier than analyzing competition, and usually gets the right answer(see Chesson and Warner 1981, Ellner 1985, Chesson and Ellner).

Note that $\rho(H|H) = 0$ (WHY? – no math needed, just a bit of insight). The condition for H to be an ESS is that

$$\rho(h|H) < 0 \text{ for } h \neq H,$$

hence H^* is an ESS if $\rho(h|H^*)$, considered as a function of h, has its maximum at $h = H^*$. That is: an ESS is defined by the property that it is the best invader of itself. As in Cohen's model we have

$$\frac{\partial \rho}{\partial h} = E\left[\frac{\tilde{Y}_t - s}{h\tilde{Y}_t + (1-h)s} \right], \quad \frac{\partial^2 \rho}{\partial h^2} < 0 \tag{2.19}$$

Because the second derivative is everywhere negative, for any H there is a single "best" invader. For a possible ESS $0 < H^* < 1$, the ESS property of being the best invader of one's self can be written as:

$$\partial\rho/\partial h = 0 \text{ at } h = H = H^*. \tag{2.20}$$

A more modest goal is to find when some dormancy is favored, i.e. under what conditions $H^* < 1$. As in the density-independent we do this by asking whether $H = 1$ can be invaded by $h < 1$, i.e., the condition for some dormancy to be favored is

$$\frac{\partial \rho}{\partial h} < 0 \text{ at } h = H = 1. \tag{2.21}$$

The formal calculations are the same as before and the result is equation (2.14) with Y_t replaced by the $\tilde{Y}_t$ for $H = 1$.

To proceed we will choose a convenient form of density dependence. In Cohen's model we assumed that the number of offspring in a given year was unlimited: each seed that germinates produces Y_t new seeds. Twice as many parents will produce twice as many offspring. As an opposite extreme, we can assume instead that the number of offspring in each year is strictly limited, regardless of the number of parents trying to reproduce.

Let K_t be the number of seeds produced in year t. If the population consists of a single resident with $H = 1$, then the number of seeds germinating in year t is exactly the number produced last year, i.e. K_{t-1} and the number of offspring per germinating seed is then $\tilde{Y}_t = K_t/K_{t-1}$.

Substituting into (2.19) and setting $h = H = 1$, the condition becomes for dormancy to be advantageous becomes

$$H^* < 1 \text{ if and only if } sE[K_t]E[1/K_t] > 1. \tag{2.22}$$

Exercise 2.5 Verify that (2.22) is correct.

Exercise 2.6 Derive the small-variance approximation for (2.22) in order to express the condition in terms of the mean and variance of K, and interpret what it predicts qualitatively about the conditions in which dormancy is favored ($H^* < 1$).

Exercise 2.7 Derive a good years/bad years approximation for for (2.22) along the lines of what we did above for Cohen's model. That is, assume $K_t = M$ or m with $m \ll M$, and keep only the dominant term(s) in the resulting expression for $sE[K_t]E[1/K_t]$.

Exercise 2.8 Explain why the good years/bad years approximation leads to the prediction that $H^* = 1$ if good years are either frequent or very infrequent, while some dormancy is favored if good and bad years are both frequent. Why does this make sense? – give an *intuitive, nonmathematical* explanation from the viewpoint of a seed deciding whether to germinate "now" versus waiting for next year.

2.3.1 Proceed with caution

We know that evolution involves

- Gene frequency changes governed by Mendelian genetics in a finite population.
- Competition among a suite of genotypes for different trait values.

To do an ESS analysis we *pretend* that it involves

- "Like begets like": uniparental clonal reproduction.
- Pairwise competition between an established "resident" and a rare "invader", ignoring effects of the invader on the resident.

These gambits don't always succeed.

- "Like begets like" is especially dubious in situations involving conflicts among offspring or between parent and offspring, such as allocation of resources within the family. Any time the interests of relatives don't coincide, it is essential to explicitly model the power structure.
- An ESS can't be dislodged once it is established, but that doesn't guarantee that evolution will move the population to the ESS (e.g. Eshel and Motro 1981, Takada and Kegami 1991). A strategy x^* with the latter property is called a **CSS** (continuously stable strategy). For a single strategy parameter, a CSS is defined by the property that if x is near x^* and y is between x and x^*, then y can invade x. The CSS and other related stability concepts are reviewed by Levin and Mueller-Landau (2000).
- The focus on the expected long-term growth rate of an invading strategy can be misleading when populations are small. The "coin tossing" randomness of individual survival and breeding then becomes important, and models based on the probability of a rare allele becoming fixed in the population then do a better job of predicting evolutionary dynamics (Proulx and Day 2001).

Because of these possible difficulties, an essential step in an ESS analysis is to check that all the tacit shortcuts have not led you into error. The available checks are general theory, special cases and simulation. "General theory" is a set of results giving conditions under which an ESS analysis agrees with the outcome of evolution in a mechanistic genetic model (e.g., Taylor 1989, Charlesworth 1990, Day and Taylor 1996, 1998). Almost all of these rely on weak selection approximations, so they offer comfort but not certainty. Convenient special cases of the genetics can often be used to check the "like begets like" gambit by raising the likely complications in the simplest possible setting (e.g., a diploid model with a few loci, a few alleles, but strong selection). However the pairwise competition gambit usually has to be checked by simulation, because multi-type competition models are high-dimensional and there is no general theory to help you with that.

Project Exercise 2.9 Check your conclusions for the good-bad years case by writing a program to simulate competition among multiple (10-25) types with different H values; run it first at parameter values where $H = 1$ should win, and then at parameters where $H = 1$ should lose, according to the criterion under the good-bad years approximation.

2.4 Empirical tests

2.4.1 Desert seeds

Phillipi (1993a) tested 3 predictions from Cohen's model on 6 desert annual species in the Southwestern US:

1. Of the seeds that do not germinate in their first year (under normal field conditions with adequate water), some should germinate in their second year under identical conditions.
2. The fraction germinating the first and second years should be equal
3. If first-year germination is prevented by bad conditions (a "no water" treatment in his experiments), this should not affect the germination fraction in the next year.

To be more precise, in the following experimental design

	Year 1	**Year 2**
Treatment 1	Good	Good
Treatment 2	Bad	Good

the observed value of H should be the same in all three "Good" trials.

Results:

Prediction 1: yes in 6 of 6 species
Prediction 2: no (either lower or higher
Prediction 3: no (either lower or higher)

Phillipi (1993b) looked at one species *Lepidium lasiocarpum* across multiple sites, to assess the response of H to environmental conditions.

1. H should be higher where average rainfall is higher ("good" years more common): *Yes, but only a very weak effect.* Germination rate remained nearly constant over a wide range of average rainfall.
2. There is a stronger, negative correlation between H this year, and the rainfall in the year when the seeds were produced.
3. Seeds produced by large mother plants were less likely to germinate in their first year.

All of these tend to favor the "limits to growth" ESS models. The first one is consistent with those models, and not with the basic prediction $H \approx p$ in the density-independent model. Clauss and Venable (2000) note that in general, experimental studies have not uniformly supported the prediction H should be higher in seeds from populations with a higher fraction of "good" years. In fact, sometimes the trend is the reverse of the prediction.

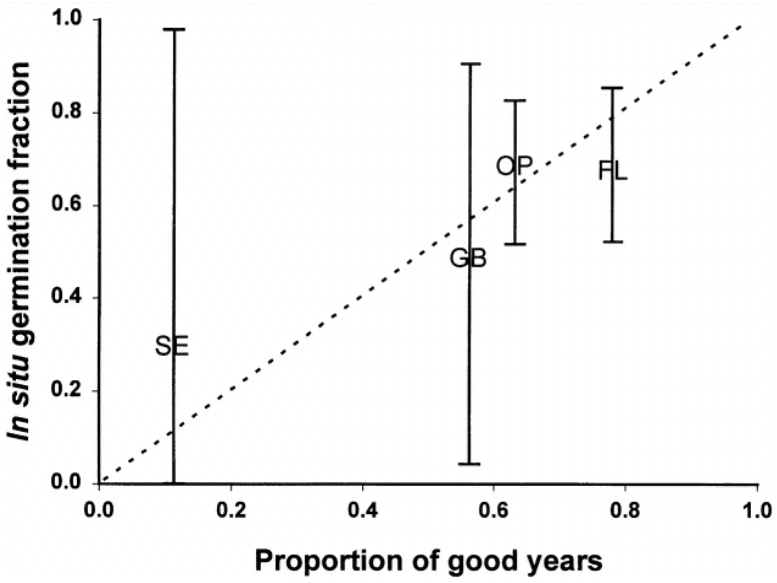


Figure 2.5: From Clauss and Venable (2000), comparing predicted values of H^* with actual mean germination rates in the field.

One interpretation of Phillipi's second and third results is that seeds are responding to "cues" about the density of competitors. If last year was good, or if "your" mother was large and therefore produced many seeds, "you" are probably surrounded by a lot of other seeds, so it's a good idea to stay dormant until the crowd thins out. Models bear this out (but we need to move on, so you'll have to take this on trust): in a density-dependent setting where seeds can estimate how crowded it is, Phillipi's observations are properties of the ESS.

Clauss and Venable (2000) focused on the response of seeds to conditions at the time of germination. They extended Cohen's model by assuming that years come in 3 types: *No Rain, Good, Bad*. In a No Rain year, seeds just sit it out. In a Good year, early rains that allow germination are followed by sufficient late rain to allow seed set. In a Bad year, early rains are not followed by adequate late rain. The prediction that they derive for this model is

$$H^* = \frac{P(G|r)Y - s}{Y - s} \approx P(G|r)$$

where $P(G|r)$ is the probability that the year is good, given that some early rains occur, and Y is the per-capita reproductive success in a good year.

Comparing this prediction with field-germination rates in the desert annual *Plantago insularis*, the results were quite good (Figure 2.5). Clauss and Venable (2000) reported that a density-dependent version of their model gave similar predictions to the density-independent model whose predictions are plotted, but don't go into details.

On the other hand, they found that under "common garden" conditions in the lab, there was **no** tendency for seeds from better (more good year) sites to have a higher annual germination rate. In fact, seeds

Figure 2.6: A female *Diaptomus sanguineus* carrying a clutch of eggs. This individual is about 1.5mm long. Picture courtesy of N.G. Hairston, Jr.

from drier sites germinated under less restrictive conditions than seeds from wetter sites. So the pattern in Figure (2.5) resulted from the interaction of intrinsic seed properties with conditions in the field – laboratory tests of seeds' propensity to germinate may be totally misleading about germination rates in the field. Clauss and Venable (2000) suggest that this might account for the inconsistent results in lab-experiment tests of bet-hedging models.

2.4.2 Freshwater copepod eggs

There are (at least) two decision problems in the life-cycle of the freshwater copepod *Diaptomus sanguineus* (Figure 2.6). The life-cycle of *Diaptomus* in Bullhead Pond, RI, is summarized in Figure 2.7, based on experimental work over nearly two decades by Nelson Hairston, Jr. and collaborators (summarized in Ellner et al. 1999).

Eggs hatch in the Fall, mature, and lay eggs that will hatch immediately and produce a second adult cohort. Second-cohort females begin by doing the same, but switch in the Spring to making diapausing eggs that will remain dormant at least until the next Fall. Diapausing eggs are safe from predation by fish, which quickly intensifies when the pond warms up in Spring. It is not too inaccurate to imagine that fish suddenly becoming active on a single "catastrophe date" each year. The catastrophe date varies widely from year to year and there is no apparent way to predict it in advance.

So for *Diaptomus* females, a very early switch date is sure to yield at least some diapausing eggs before fish become active. But a late switch date is a gamble on when the catastrophe date will come this year. If it comes late, she keeps on pumping out eggs that hatch and produce many daughters and

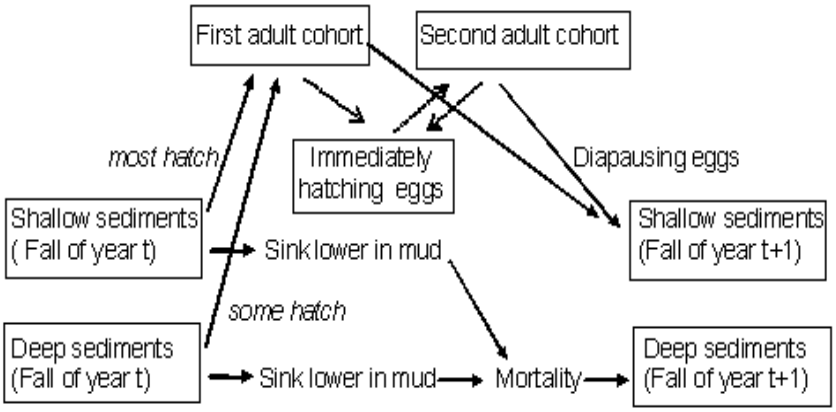


Figure 2.7: Schematic life cycle of Diaptomus sanguineus in Bullhead Pond, RI. Eggs hatch in late fall and there are typically two generations completed before predation wipes out the active population in the Spring. The population persists through the summer as diapausing eggs. Eggs that do not hatch the next Fall sink deeper into pond sediments and can only hatch in later years if they are mixed upwards to the surface of the sediment layer, e.g. by fish laying nests.

grand-daughters before predation intensifies. But if it comes early, the female and her (still active) offspring suffer high mortality and leave few descendents. Selection against those who guess wrong is strong enough to cause year-to-year changes in mean switch date – it shifts earlier after a year of low predation, and later after a high predation year.

We initially studied the system in order to understand the high levels of heritable (i.e. genetically determined) variation in switch date among females (Figure 2.8), which could not be explained by the then-current theory from population genetics (Hedrick 1986, Bull 1987, Karlin 1988, Turelli 1988, Barton and Turelli 1989, Frank and Slatkin 1990, Gillespie 1991). We developed models that could explain this (which is another story), and then to test our models made some predictions about differences in switch-date distribution between descendants of active females in the water column, and descendants of eggs buried in the pond sediments. Diapausing eggs that remain dormant are eventually covered by the accumulation of sediment at the bottom of the pond, and cannot hatch unless they are mixed up to the surface, e.g. by fish building nests or a tree limb falling into the pond and disturbing the sediments. Such buried eggs can survive for centuries: in the lab, 50% hatching success was obtained on eggs buried for 300 years, according to lead-210 dating of the sediments from which they were extracted (Figure 2.9).

Our models (leaving out some early ones that didn't work) are conceptually similar to the seedbank model, but consider two different decision variables: switch date (z) and first-year hatching rate H.

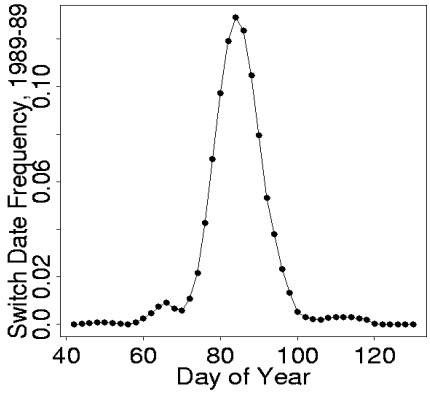


Figure 2.8: Average distribution of switch dates over 10 years in Bullhead Pond, RI, based on cohort sample data of Nelson G. Hairston, Jr.

Conceptually, the model is

$$n(t+1) = n(t)[HY(z,t) + (1-H)s] \tag{2.23}$$

which is the basic seedbank model except that reproductive success Y is partially determined by switch date z. In particular, we assumed that Y is determined jointly by z and the (unpredictable, random) "catastrophe date" $\theta(t)$) when predation commences in year t.

$$Y(z,t) \propto e^{(-c(z-\theta(t))^2)}$$

with the total number of eggs being higher in years when predation is late ($K(t) = Ke^{a\theta(t)}, \quad a > 0$). This fitness function is shown in Figure 2.10. We mostly used a "limits to growth" version of the model (with a cap on total population size), and for comparisons with data added other features (such as there being two generations per year) – but (2.23) is the gist of all the models.

With these assumptions we got the following predictions:

1. When Var(θ) is low, the ESS is to have z at the average optimum switch date.
2. When Var(θ) is high, there is no ESS. Instead, the population should consist of a discrete set of types $(z_1, H_1), (z_2, H_2), \ldots, (z_n, H_n)$ having distinctly different switch dates and first-year hatching rates.
3. Unless Var(θ) is really, really high, two types coexist with a negative correlation between switch date and hatch rate.

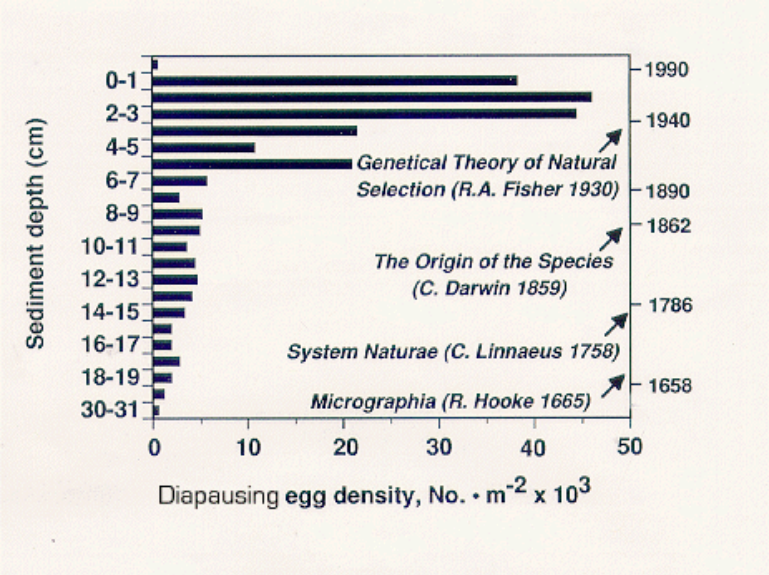


Figure 2.9: Distribution of diapausing eggs buried in the sediments of Bullhead Pond, RI and estimated ages based on lead-210 dating, compared to some landmarks in the history of biology. Figure courtesy of Nelson G. Hairston, Jr.

The predicted negative correlation is easy to understand: types with a early early switch date never encounter predation, so every year is the same and they should have a high H. Late-switching types experience a highly variable environment and should have a low H.

We estimated (Ellner et al. 1999) that Var(θ) was high but not really, really high, so the prediction to test was the third one above. Figure 2.11 shows evidence for the existence of two egg types, based on the amount of incubation time required for hatching to occur in the lab, in conditions that would evoke hatching in the field. While suggestive, this is not quite enough, because the difference could simply be a bet-hedging strategy of variable diapause timing: we also need to check whether it lines up with differences in switch date. To do that, offspring of those eggs were assayed for what kinds of clutches they produce, under lab conditions mimicking those in the pond at the average switch date. The results (Figure 2.12) confirm the association: quick hatchers produce a mixture of clutch types, late hatchers are overwhelmingly producing diapausing eggs (these figures use data from the experiments reported by Hairston et al. 1996b)

The overall picture is that the copepod population was a mix of genotypes that used different ways of coping with the risk of reproduction in the face of unpredictable predation, one "conservative" and one "diversified". Early-switch genotypes are conservative: as adults they accept the reliable payoff for producing diapausing eggs before things get risky, so they have no reason to bet-hedge as eggs and therefore hatch readily. Late-switch genotypes gamble as adults on the big payoff that comes in years of low predation risk, and therefore play it safe by bet-hedging as eggs, spreading their hatching over many years. A larger lesson from this is that how safe or risky the world is, is not necessarily an intrinsic

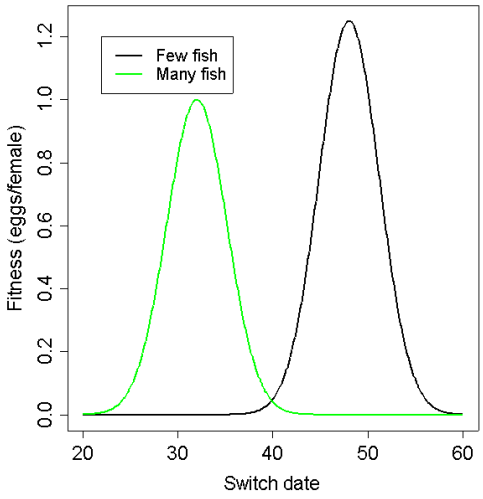


Figure 2.10: Fitness function (eggs per female as a function of switch date) in model for bet-hedging by *Diaptomus sanguineus*. Under intense predation, it pays to have an early switch date; under weak predation it pays to have a late switch date, and the potential reproduction is greater.

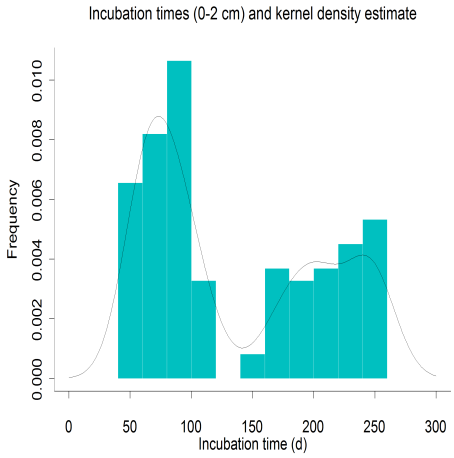


Figure 2.11: Distribution of incubation time (days) required for hatching to occur.

property: a lot can depends on whether the organism avoids or seeks risky opportunities.

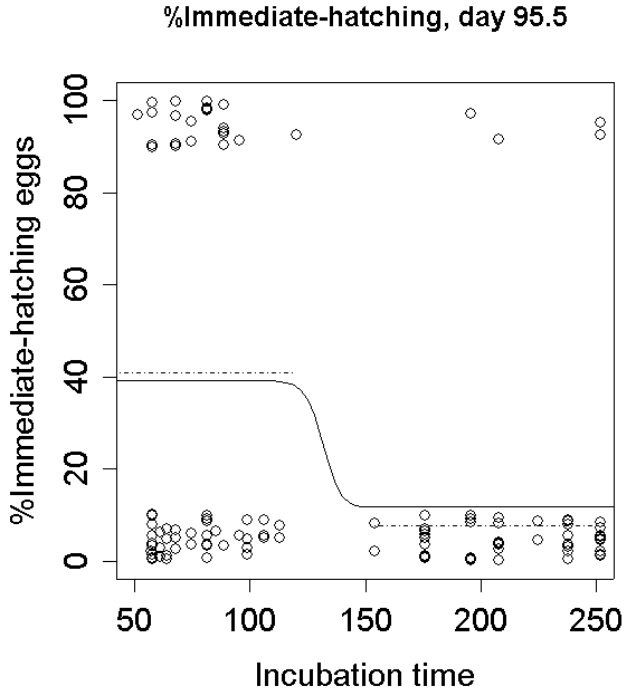


Figure 2.12: Association between incubation time and the fraction of immediately-hatching egg clutches, under lab conditions mimicking those in the field on Julian day 95.5

2.5 Reminder: the big picture

As we leave the topic of individual-level decision making, it's again a good time to expand our focus. Foraging theory introduced classical models that ignore variability. The assumed goals involve long-term average rates, and ignore short-term accidents. Dynamic state variable models incorporate variability during the lifetime of an individual, or on the time-scale of daily decision-making. Bet-hedging theory takes account of variability on longer time scales, such that successive generations, or one individual over the course of its lifetime, face conditions that vary unpredictably. To develop the theory we had to learn about population growth under temporally varying conditions, mathematical techniques for approximating stochastic growth models. Taking account of limits to growth forced us to learn about invasibility analysis and evolutionarily stable strategies (ESS).

All of these general ideas and methods have much broader application. In particular, the ESS approach of evolutionary game theory is the fundamental tool for models of animal behavior whenever interactions between individuals are involved. Examples include: contests for food, territory or mates; signaling; coordination of group activities (e.g. sentinel/forage decisions); division of labor in social insects and allocation of reproductive output among individuals; cooperation among unrelated individuals. The book edited by Dugatkin and Reeve (1998) reviews the wide scope of game theory models in the study

of animal behavior.

Another major application area is life history theory, as mentioned at the start of this chapter:

- Body size: when to stop growing.
- Age at maturity: when to start allocating resources to reproduction rather than just growth and survival.
- Reproductive lifespan: continuous breeding over many years, versus a single suicidal "spawning".
- Offspring size and number: many small or few large offspring at each breeding?
- Aging: patterns of change in fecundity and survival over time.

As in our models for short-term decision making, these are viewed as solutions that optimize some assumed goal, that is a proxy for long-term representation in the gene pool. As always, optimization is subject to constraints and tradeoffs, e.g., putting more resources into breeding decreases growth or survival. In a historical development similar to that of foraging theory, life history theory has progressed from classical models that ignore variability, to state-dependent models that take account of within- and between-generation variability in conditions. So although the applications are different, the tools and concepts are the ones that you've seen in this and the previous chapter. Stearns (1992, 2000) and Roff (2001) are good recent reviews of life theory and its applications.

2.6 References

Barton, N. H. and M. Turelli. 1989. Evolutionary quantitative genetics: how little do we know? Annual Review of Genetics 23: 337-370.

Bull, J. J. 1987. Evolution of phenotypic variance. Evolution 41: 313-315.

Bulmer, 1995. Theoretical Evolutionary Ecology. Sinauer Associates, Inc., Sunderland Mass.

Caswell, H. 2001. Matrix Population Models (2nd edition). Sinauer Associates, Inc., Sunderland Mass.

Charlesworth, B. 1980. Evolution in Age-Structured Populations. Cambridge University Press, Cambridge, UK.

Charnov, E. L. 1993. Life History Invariants. Oxford University Press, Oxford.

Chesson, P. L. 1994. Multispecies competition in varying environments. Theoretical Population Biology 45: 227-276.

Chesson, P. L., and R. R. Warner. 1981. Environmental variability promotes coexistence in lottery competitive systems. American Naturalist 117: 923-943.

Chesson, P.L. and S. Ellner. 1989. Invasibility and stochastic boundedness in monotonic competition models. J. Mathematical Biology 27:117-138.

Clauss, M.J. and D.L. Venable. 2000. Seed germination in desert annuals: an empirical test of adaptive bet hedging. American Naturalist 155: 168-186.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. J. Theoretical Biology 12: 119-129.

Day, T. and P. D. Taylor. 1996. Evolutionarily stable versus fitness maximizing life histories under frequency-dependent selection. Proceedings of the Royal Society of London B 263:333-338

Day, T. and P. D. Taylor. 1998. Unifying genetic and game theoretic models of kin selection on continuous traits. Journal of Theoretical Biology 194:391-407

De Stasio, B. T. 1989. The seed bank of a freshwater crustacean: copepodology for the plant ecologist. Ecology 70:1377-1389.

Dugatkin, L.A. and H.K. Reeve. 1998. Game Theory and Animal Behavior. Oxford University Press, Oxford UK.

Ellner, S. 1985. ESS germination strategies in randomly varying environments. Theoretical Population Biology 28: 50-116.

Ellner, S. and N. G. Hairston Jr. 1994. Role of overlapping generations in maintaining genetic variation in a fluctuating environment. American Naturalist 143: 403-417.

Ellner, S. 1997. You bet your life: life-history strategies in fluctuating environments. pp. 3-24 in: H.G. Othmer, F. R. Adler, M.A. Lewis and J.C. Dallon (eds.) Case Studies in Mathematical Modeling: Ecology, Physiology, and Cell Biology. Prentice-Hall, NJ.

Ellner, S.P., N.G. Hairston, Jr., C.M. Kearns, and D. Babai. 1999. The roles of fluctuating selection and long-term diapause in microevolution of diapause timing in a freshwater copepod. Evolution 53: 111-122.

Eshel, I. and U. Motro. 1981. Kin selection and strong evolutionary stability of mutual help. Theoretical Population Biology 19: 420-433.

Evenari, M., L.Shanan, and N. Tadmor. The Negev: The Challenge of a Desert. Harvard University Press, Cambridge Mass.

Falconer, D. S. 1981. Introduction to Quantitative Genetics (2nd edition). Longman Scientific & Technical, Harlow, Essex, England.

Frank, S. A., and M. Slatkin. 1990. Evolution in a variable environment. American Naturalist 136:244-260.

Gillespie, J. H. 1991. The Causes of Molecular Evolution. Oxford University Press, Oxford UK.

Hairston, N. G., Jr. 1988. Interannual variation in seasonal predation: its origin and ecological importance. Limnology and Oceanography 33:1245-1253.

Hairston, N. G., Jr., and T. A. Dillon. 1990. Fluctuating selection and response in a population of freshwater copepods. Evolution 44:1796-1805.

Hairston, N.G. Jr., R.A. van Brunt, C.M. Kearns, and D.R. Engstrom. 1995. Age and survivorship of diapausing eggs in a sediment egg bank. Ecology 76:1706-1711.

Hairston, N. G., Jr., S. Ellner, and C. M. Kearns. 1996. Overlapping generations: the storage effect and the maintenance of biotic diversity. in: O. E. Rhodes, R. K. Chesser, and M. H. Smith (eds.). Population Dynamics in Ecological Space and Time. University of Chicago Press.

Hairston, N.G. Jr, S. Ellner, and C.M. Kearns. 1996b. Phenotypic variation in a zooplankton egg bank. Ecology 77: 2382-2392.

Haldane, J. B. S. and S. D. Jayakar. 1963. Polymorphism due to selection in varying directions. Journal of Genetics 58: 237-242.

Hedrick, P. W. 1986. Genetic polymorphism in heterogeneous environments: a decade later. Annual Review of Ecology and Systematics 17:535-566.

Karlin, S. 1988. Non-Gaussian phenotypic models of quantitative traits. pp. 123-144 in:E. J. Eisen, M. M. Goodman, G. Namkoong, and B. S. Weir, eds. The Second International Conference on Quantitative Genetics. Sinauer, Boston.

Levin, S. A. and H. C. Muller-Landau. 2000. The evolution of dispersal and seed size in plant communities. Evolutionary Ecology Research 2: 409-435.

Lewontin, R.C. and D. Cohen. On population growth in a randomly varying environment. PNAS USA 62: 1056-1060.

Metz, J. A. J., R. M. Nisbet, and S. A. H. Geritz. 1992. How should we define 'fitness' for general ecological scenarios? Trends in Ecology and Evolution 7: 198-202.

Nagylaki, T. 1993. Evolution of multilocus systems under weak selection. Genetics 134: 627-647.

Nowak M.A. and K. Sigmund. 199. Primer-Evolutionary Game Theory. Current Biology 9: R503-R505.

Orzack, S. H. 1993. Life history evolution and population dynamics in variable environments: some insights from stochastic demography. pp. 63-104 in: J. Yoshimura and C. W. Clark (eds.) Adaptation in Stochastic Environments. Lecture Notes in Biomathematics Vol. 98. Springer-Verlag.

Phillipi, T. and J. Seger. 1989. Hedging one's evolutionary bets, revisited. Trends in Ecology and Evolution 4: 41-44.

Phillipi, T. 1993a. Bet-hedging germination of desert annuals: beyond the first year. American Naturalist 142: 474-487.

Phillipi, T. 1993b. Bet-hedging germination of desert annuals: variation among populations and maternal effects in *Lepidium lasiocarpum*. American Naturalist 142: 488-507.

Proulx, S.R. and T. Day. 2001. What can invasion analyses tell us about evolution under stochasticity? Selection 2:1-16

Roff, D.A. 2001. Life History Evolution. Sinauer Associates, Sunderland MA.

Seger, J. and H.J. Brockmann. 1987. What is beg-hedging? Oxford Surveys in Evolutionary Biology 4: 182-211.

Sasaki, A. and S. Ellner. 1995. The evolutionarily stable phenotype distribution in a random environment. Evolution.

Stearns, S. 1992. The Evolution of Life Histories. Oxford University Press, Oxford.

Stearns, S.C. 2000. Life history theory: successes, limitations, and prospects. Naturwissenschaften 87: 476-486.

Stumpf, M.P.H., Z. Laidlaw, and V. A. A. Jansen. 2002. Herpes viruses hedge their bets. PNAS USA 99: 15234-15237.

Takada, T. and J. Kigami. 1991. The dynamical attainability of ESS in evolutionary games. Journal of Mathematical Biology 29: 513-529.

Taylor, P.D. 1989. Evolutionary stability in one-parameter models under weak selection. Theoretical Population Biology 36A: 125-143.

Turelli, M. 1988. Population genetic models for polygenic variation and evolution. Pages 601-618 in E. J. Eisen, M. M. Goodman, G. Namkoong, and B. S. Weir, eds. The Second International Conference on Quantitative Genetics. Sinauer, Boston.

2.7 ESS in two easy lessons

The purpose of this section is to review some classical examples of evolutionary game theory models for behavior and life-history traits, and give an indication of how they can be analyzed using the idea of an Evolutionarily Stable Strategy. Evolutionary game theory is now a vast subdiscipline of theoretical ecology and this chapter is an ultra-minimal introduction. For a broader view of the many current applications of evolutionary game models, a good starting point is the volume edited by Dugatkin and Reeve (1998).

First some motivation. The reason for needing a different approach to individual adaptation is *frequency dependence*, meaning that the costs and benefits of a particular action or trait depend in part on the actions or traits of other individuals in the population. A consequence of frequency dependence is that evolution is not expected to produce optimality in any absolute sense. As a simple example to show why this is true, consider a simple 1-locus, 2-allele model for gene frequency change in a randomly mating diploid population with constant population size. Let A and a denote the 2 alleles, $p(n)$ the frequency of allele A in generation n, and W_{AA}, W_{Aa}, W_{aa} the genotype fitnesses. We don't need to analyze (or even to fully write down) this model. All we need is one result about the model: if $W_{AA} > W_{Aa} > W_{aa}$ in generation n, then $p(n+1)$ will be higher than $p(n)$.

Now suppose that the genotype fitnesses have the following frequency dependence:

$$W_{AA} = 5 - 5p, \quad W_{Aa} = 4 - 4p, W_{aa} = 3 - 3p. \tag{2.24}$$

Then the frequency of allele A always increase from generation to generation (unless it is already at frequency 1), so the population converges onto $p = 1, W = 0$ – which is the lowest possible fitness. Things aren't always this bad. But this example shows that when there is frequency dependent fitness for a trait or behavior, we can't use optimality principles to predict the outcome of evolution.

2.7.1 Lesson 1: Hawks and Doves

The canonical example of an evolutionary game model is Hawk-Dove model for pairwise conflicts between individuals. This was motivated originally by the question of why animal conflicts are often settled without full-scale battle between the individuals, e.g. males resolving contests over mates without either being greatly harmed by the other.

Maynard Smith and Price (1973) proposed a simple model for pairwise contests in the framework of noncooperative game theory. They assume that pairs of individuals are drawn at random from the population, and contest for some item of value V. Each individual has the choice of two behaviors: they can Display (a show of strength without actual attack) or Escalate (a real attack on the other individual). If both individuals Display, then the item goes to one of them chosen at random. If one displays and the other Escalates, the latter gets the item. If both Escalate there is a fight (with cost c to both participants), and at the end the item goes to one of them chosen at random.

Maynard Smith and Price (1973) considered a variety of strategies in which the decision to Escalate or

Display depended on past interactions between the two individuals (for example, an individual could Escalate as a form of retaliation for what their current opponent had done in the past). But the essential ideas can all be seen in a stripped-down model where past history is ignored, as if each interaction involved individuals who had never met before. This has come to be known as the Hawk-Dove game: a Hawk escalates, a Dove displays. The two individuals are assumed to choose their behaviors simultaneously – meaning that each must decide without knowing what the other will do. Most models in evolutionary game theory have made this assumption.

We can summarize the model by writing down the expected "payoff" to an individual its own behavior and that of the other member of the pair. Let $W(X, Y)$ denote the payoff to an individual with behavior X when playing against an individual with behavior Y. The model is then summarized by the payoffs:

$$W(D, D) = V/2, \quad W(D, E) = 0, \quad W(E, D) = V, \quad W(E, E) = V/2 - c.$$

These can be represented in the *payoff matrix*

	Display	Escalate
Display	$V/2$	0
Escalate	V	$V/2 - c$

(2.25)

Entries in the matrix are the payoff to the individual using the behavior on the left, if its opponent uses the behavior at the top. All decisions depend on relative payoffs – differences between entries in the matrix – so nothing changes if a constant is added to each entry in the matrix. So don't be worried about the possibility of negative fitness (less than zero offspring?) – the entries in the matrix should be thought of as modifications to some baseline fitness W_0, where W_0+(any matrix entry) is positive. For example, males may be contending for high-quality breeding territories, and the losers still get to use lower-quality habitat.

The language being used to describe the model reflects the roots of game theory in economics. In applying these models to behavior, a tacit assumption is that the "reward" or "payoff" translates into individual Darwinian fitness. Individuals are assumed to behave so as to maximize their expected payoff.

Frequency-dependent fitnesses result from the fact that the expected payoff for a given behavior depends on what others in the population are doing. For example, suppose first that the population is dominated by Hawks, individuals who alway escalate. Then an individual who Displays gets payoff $W(D, E) = 0$, and one who Escalates gets payoff $W(E, E) = V/2 - c$. If $V/2 - c > 0$, then the best option for the focal individual is to Escalate. The Escalate strategy is then an *Evolutionarily Stable Strategy*, or ESS – given that everyone else does it, you ought to do the same. But if fights are very costly so that $V/2 - c < 0$, the always-Escalate strategy is not an ESS, because an invading mutant who chooses to Display gets higher fitness.

In contrast, the always-Display strategy cannot be an ESS. If the population is dominated by Doves who always Display, then an individual who Displays gets $W(D, D) = V/2$ while one who Escalates gets $W(E, D) = V$, which is greater.

So the interesting situation is when fights are costly, $V/2 < c$, so neither Dove (always Display) or Hawk

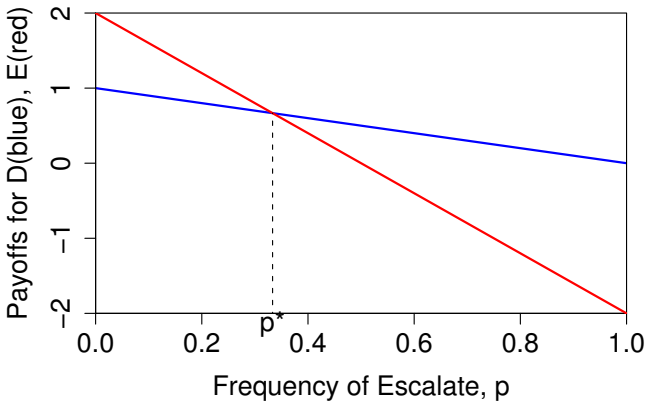


Figure 2.13: Graphical representation of the finding the ESS mixture for the Hawk-Dove game when $V/2 - c < 0$. Note that this is qualitatively like the example (2.24), in that a higher frequency of Escalation decreases the mean payoff from all strategies. The ESS mixture p^* does not optimize fitness at either the individual or population levels.

(always Escalate) is an ESS. What then?

Well, it depends. We have to decide whether individuals are stuck playing one of the *pure strategies* Hawk, Dove – meaning that each individual uses the same behavior on each encounter with another individuals – or if they can play a *mixed strategy*, doing different things on different occasions.

In either case, however, the prediction is a mixture of strategies at the population level. Specifically, an evolutionarily stable population state must have the property that in any one play of the game, the expected payoffs for Display and Escalate must be equal. If the payoffs were unequal, say if there was a higher payoff for Display, the best choice for an individual would be to always Display. So an invader that always used the Display strategy would have higher fitness than the population average, and Displayers would increase in frequency.

To find the stable mix, let p be the population frequency of the Escalate behavior. That means that an individual's opponent has probability p of Escalating, probability $1 - p$ of Displaying. The expected payoffs for the individual's options are then

$$\begin{aligned} W(D, pE + (1-p)D) &= pW(D,E) + (1-p)W(D,D) = (1-p)(V/2) \\ W(E, pE + (1-p)D) &= pW(E,E) + (1-p)W(E,D) = p(V/2 - c) + (1-p)V \end{aligned} \tag{2.26}$$

The stable mix p^* is then found by setting these equal,

$$(1-p^*)(V/2) = p^*(V/2 - c) + (1-p^*)V.$$

This gives the nice prediction

$$p^* = \frac{V}{2c}. \tag{2.27}$$

(see Figure 2.13). That is, if the cost of a fight is high relative to the value of the item being contested, your frequency of escalating to a fight should be low – just high enough to deter others from assuming that you won't ever fight, so they might as well Escalate and take the item at no cost to themselves.

The attractive feature of the ESS concept is that you don't have to say anything about the genetic basis for the trait or about the dynamics of gene-frequency change. Everything is strictly in terms of the payoff function. An ESS strategy is defined by the property that "cheaters don't prosper" – if everyone else in the population is behaving according to the ESS, then you can't get a higher payoff than the others by moving to some other strategy. That is,

$$W(p, p^*) \leq W(p^*, p^*) \tag{2.28}$$

if p^* is the ESS. Equation (2.7.1) is actually not quite enough to define an ESS, because the ESS often has the property that $W(p, p^*) = W(p^*, p^*)$ for all p. This is true in the Hawk-Dove gam, and many others where the strategic choice is the relative frequency of several different behaviors, because the ESS results in all behavioral options having the same payoff. The equality of payoffs for all options means that there is not necessarily any force tending to keep the population at the ESS.

To cover those situations, (2.7.1) has to be expanded, as follows: for all $p \neq p^*$,

$$\begin{aligned} &\text{either } W(p, p^*) < W(p^*, p^*) \\ &\quad \text{or } W(p, p^*) = W(p^*, p^*) \text{ and } W(p, p) < W(p^*, p). \end{aligned} \tag{2.29}$$

The *either* condition means that alternatives to the ESS are strict losers once the ESS is established in the population. The *or* condition means that an alternative strategy may be as good as the ESS if it's so rare that it always plays against the ESS, but if it rises in frequency (so that sometimes its opponent is also playing the alternative strategy) this gives an advantage to the ESS.

Excercise 2.10 Consider a general payoff matrix with two strategies I, J,

	I	J
I	a	b
J	c	d

(2.30)

(a) Show that there is a mixed ESS if $a < c$ and $b > d$, and interpret this condition in words.
(b) Find the ESS frequency p^*.

Exercise 2.11 A parental care game (based on a book in preparation by R.A. Johnstone and D.J.D. Earn). Each "round" of this game is played by two parents who simultaneously face the same decision: whether to remain and **Care** for the young they have produced together, or to **Desert** the brood and seeking a new mate. Assume that there are n offspring in a brood. If both parents Care, a proportion s_2 of these offspring will survive. If only one parent stays and the other Deserts, the proportion of young surviving will be $s_1 \leq s_2$. If both parents Desert, offspring survival will be $s_0 \leq s_1$. A parent

that Deserts has a probability m of finding a second mate, with whom it can produce a second brood of n offspring, of which a proportion s_L survive.
(a) Let $W(x, y)$ denote the payoff to a parent with behavior x when the other parent has behavior y, $x, y \in \{C, D\}$ standing for **Care,Desert**. Explain in words why the model's assumptions lead to the following expressions for the payoffs:

$$W(C,C) = ns_2 \quad W(C,D) = ns_1 \quad W(D,C) = ns_1 + mns_L \quad W(D,D) = ns_0 + mns_L \tag{2.31}$$

(b) Under what conditions on the parameters is **Care** an ESS?
(c) Under what conditions on the parameters is **Desert** and ESS?
(d) Can **Care** and **Desert** both be ESS's, and if so, when?
(e) Under what conditions will neither **Care** nor **Desert** be an ESS, and what is the predicted evolutionary outcome then?

Exercise 2.12 This problem involves ESS analysis of the Rock-Paper-Scissors game. Rock-Paper-Scissors is a good model for male mating strategies in the side-blotched lizard *Uta Stansburiana*; see Sinervo and Lively (1996), The three possible actions are *R*ock, *P*aper, and *S*cissors, and the payoff matrix is

		Opponent's action		
		Rock	Paper	Scissors
	Rock	0	-1	1
My action	Paper	1	0	-1
	Scissors	-1	1	0

(2.32)

and the matrix entries are the payoffs to me.

(a) Explain why none of the pure strategies (always R, always P, or always S) is an ESS.

(b) Write expressions for the expected payoffs W_R, W_P, W_S for playing a pure strategy, when the resident population plays the three possible actions with probabilities f_R, f_P, and $f_S = 1 - f_R - f_P$.

(c) Show that $f_R = f_P = f_S = 1/3$ is an ESS for this game.

(d) Show that $f_R = f_P = f_S = 1/3$ is the only ESS for this game.

2.7.2 Lesson 2: sex allocation in hermaphrodites

One of the earliest applications of "ESS thinking" – long before ESS's were named, and possibly known to Darwin – was to the evolution of sex ratios: what proportion of offspring should be male versus female? A more recent extension, and one that has made considerable contact with data (see the inspiring book by Charnov (1982)), is sex allocation to male versus female "function" (modes of reproduction) by hermaphroditic species, such as plants that produce both pollen and ovules. To find the ESS allocation, we consider a population dominated by a resident type with a particular allocation to male and female function, and compute the fitness of a rare invader with a different allocation. The simplest model goes like this:

- Nonoverlapping generations, with adults diploid and gametes haploid, and a constant number of adults N in each generation.
- Resident adults produce F gametes through female function, and M gametes through male function.
- The next generation is formed by random mating of gametes.

We can compare different allocation strategies based on the number of gametes that each individual contributes, on average, to the next generation. We do this first for the resident, and then for a rare invader with a different allocation to male versus female function.

Figure 2.14 summarizes the calculations. The N resident adults produce NF gametes through female function. Each of these mates with a male gamete produced by *by some resident*, because the population is dominated by the resident type and we assume that males are common enough that all female gametes get mated. We note for future reference that the mating success of a male gamete is $NF/NM = F/M$. Mating yields NF zygotes containing a total of $2NF$ resident gametes. To maintain a constant population size of N, the survival probability of zygotes must be $1/F$. The N resident adults therefore put $2NF/F = 2N$ gametes into the next generation, for an average fitness of 2 gametes per resident adult.

Now consider a rare invader with allocations F_i, M_i. By female function the invaders produce N_iF_i gametes. Each of these mates with a resident's male gamete, yielding N_iF_i invader gametes in new offspring. The zygote survival is $1/F$, so invaders put N_iF_i/F gametes into the next generation through female function. By male function invaders produces N_iM_i gametes. These have the same mating success as a resident-type male gamete, so the number of gametes in offspring is $(N_iM_i) \times (F/M)$, hence the number of gametes in the next generation is N_iM_i/M. The N_i invader adults therefore put $N_i\,(F_i/F + M_i/M)$ gametes into the next generation, for an average fitness of $(F_i/F + M_i/M)$ per invader adult. The invader increases if this is higher than the resident fitness (which is 2), and decreases otherwise.

Therefore, an ESS allocation M^*, F^* must have the property that

$$\frac{M_i}{M^*} + \frac{F_i}{F^*} < 2 \text{ for all } (M_i, F_i) \neq (M^*, F^*). \tag{2.33}$$

This is known as the *Shaw-Mohler* equation. To give it a more intuitive interpretation, let p denote the fraction of resources put into male function, so that M and F are functions of p. The Shaw-Mohler condition is that

$$W(p, p^*) = \frac{M(p)}{M(p^*)} + \frac{F(p)}{F(p^*)} < 2 \text{ for all } p \neq p^*. \tag{2.34}$$

Because $W(p^*, p^*) = 2$, the ESS p^* must have the property that

$$\frac{\partial W}{\partial p}(p, p^*) = 0 \text{ when } p = p^*.$$

That is,

$$\frac{M'}{M} + \frac{F'}{F} = 0 \Rightarrow M'F + F'M = 0 \Rightarrow (MF)' = 0 \tag{2.35}$$

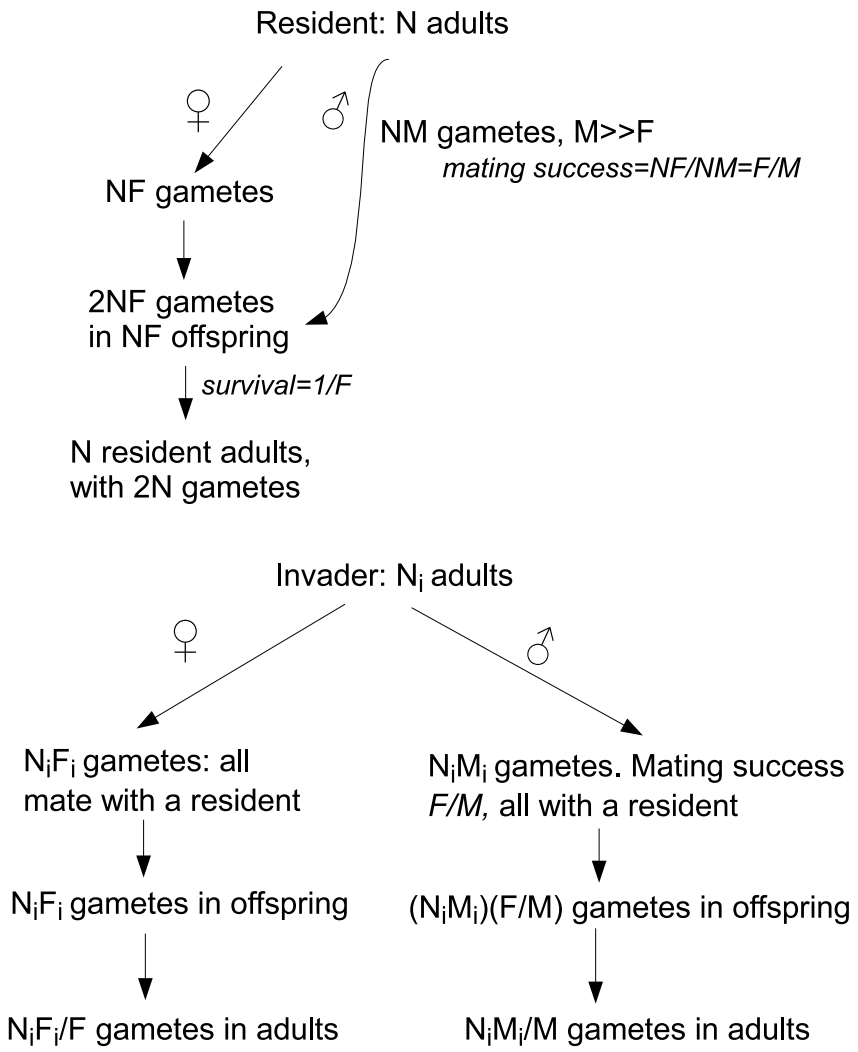


Figure 2.14: Calculations leading to the Shaw-Mohler equation

where $'$ indicates the derivative with respect to p. The ESS allocation to male versus female reproductive functions therefore *maximizes the product of the numbers of viable gametes produced through male and female functions.*

Charnov's (1982) book develops sex-allocation theory, including the Shaw-Mohler equation and corresponding theory for other situations (such as non-hermaphrodites) and is full of inspiring applications.

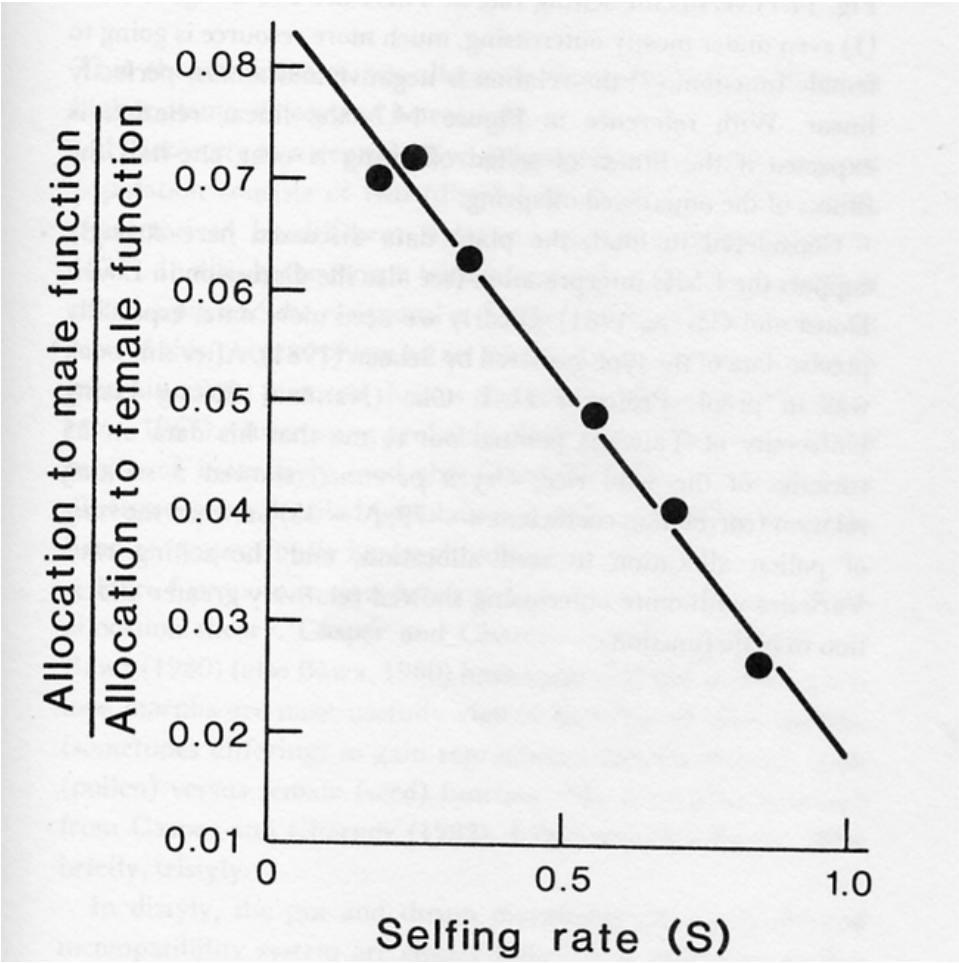


Figure 2.15: Allocation of resources to male versus female function in relation to the selfing rate S for several populations of *Gilia achilleifolia*.

Figure 2.15 shows one example, taken from that book. This is Figure 16.3 in Charnov (1982), and in the caption there Charnov notes "The observed data are very close to the theoretically expected results", which is a nearly linear relationship between selfing rate and the ratio of male- to female-function allocation. It should be noted, however, that the theoretical prediction being tested is that the relationship should be a line with negative slope, *not* the particular line drawn in the Figure. Nonetheless, it is far from trivial to predict that the *ratio* of male to female function should be a linear function of selfing rate, and to find that the data match it so well.

Exercise 2.13 Use the Shaw Mohler equation to show that if male and female function gametes are equally costly, the ESS is $M = F$. "Equally costly" means that $M + F = R$ where R is the total resource available for gamete production. What happens if there are unequal costs, so that $c_M M + c_F F = R$, where c_F, c_M are the costs per gamete?

Exercise 2.14 Modify the derivation of the Shaw-Mohler equation so that it applies to single-sex individuals (each individual is either male or female), in order to derive the ESS sex ratio of offspring as a function of the cost for producing viable male versus female offspring. The "map" from the hermaphrodite model to sex-ratio calculations is as follows: Instead of male-function and female-function gametes, consider male and female offspring. Instead of counting gametes in the next generation, count grandchildren. Assume for simplicity that Mom's genotype determines the sex ratio of her offspring. Your goal is to find a Shaw-Mohler type equation that applies to the production of Male versus Female offspring. As a special case, show that a 50:50 sex ratio is optimal if male and female viable offspring are equally costly.

2.7.3 Proceed with caution II

As noted earlier in this chapter, the ESS approach is an attempt to shortcut many biological details and get straight to predictions. Our examples all feature one particular shortcut: ignoring the mechanics of trait inheritance and the dynamics of gradual trait evolution due to natural selection. The stability concept for an ESS is that once the population is at the ESS, there is no benefit for a change in strategy. This says nothing about whether or not the population would ever get to the ESS: if the population is close but not quite (e.g., a mix of Hawks and Doves in some proportion $p \approx p^*$) will evolution move the population closer to the ESS? An ESS with this property is said to be *convergence stable*. Unfortunately, not all ESS's are convergence stable. For some payoff functions, a population near an ESS can be invaded by individuals further from the ESS, but not by individuals closer to it. Convergence stability may also depend on which model is adopted for the evolutionary dynamics of strategies – exactly how does a difference in current payoffs translate into the change in behavioral strategies between this generation and the next?

Finally, one really needs to be careful about checking second-order (i.e. second-derivative) conditions for a maximum. A purported ESS identified by a first-derivative condition such as (2.26) or (2.35) might turn out to be a fitness *minimum* rather than a maximum. And it can even turn out – if you rig things just right – that a population can converge to a strategy that (once the population gets there) is a fitness minimum rather than a fitness maximum. This is called an *evolutionary branch point*, and can result in a population splitting into sub-populations with different strategies.

These are still murky and very contentious waters, as can be seen in (for example) Waxman and Gavrilets (2005) and the responding articles in that issue of *Journal of Evolutionary Biology*. As Waxman and Gavrilets (2005) review, the first-order conditions on an appropriate payoff function define an *Evolutionarily Singular Strategy*; calling this an ESS has become established usage, but it's such a *really* bad idea that I will call it an ESIS instead. Any ESS has to be an ESIS, but an ESIS might not be an ESS. Many of the questions raised in the two preceding paragraphs can be settled by looking at the second derivatives of the payoff function at an ESIS. For example, let $r(v,u) = W(v,u) - W(u,u)$ be the relative payoff in Hawk-Dove for an individual who escalates with probability v, in a population where the resident strategy is to escalate with probability u. A v-type invader will increase in frequency

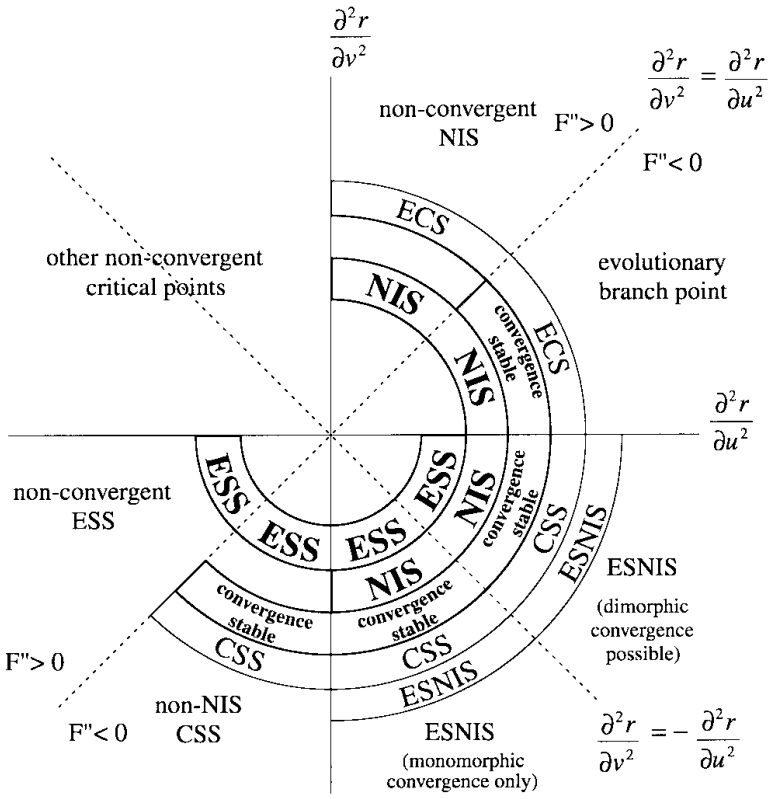


Fig. 1. Classifications of the properties of critical points of the fitness function, r, in terms of their pure second partial derivatives with respect to the type of the resident, u, and of the invader, v. ESS = evolutionarily stable strategy; NIS = neighbourhood invader strategy; CSS = continuously stable strategy; ECS = evolutionarily compatible strategy; ESNIS = evolutionarily stable neighbourhood invader strategy. Note that the points to the right and below the line where the two derivatives are equal (convergence stable points) are ones for which the function F has a maximum.

if $r(v,u) > 0$. An ESIS p^* in this setup is defined by the property that $\frac{\partial r}{\partial v} = 0$ at $v = u = p^*$. The predicted evolutionary outcome depends on the second derivatives of r at $v = u = p^*$.

Figure 2.7.3, lifted from Levin and Mueller-Landau (2000) is a nice summary of the many possibilities. The key thing to notice is that non-invasibility (the property defining an ESS) does not exactly coincide with convergence stability. So a population may fail to converge to an ESS, or it may converge to an ESIS that is not an ESS. The exciting possibility is that evolutionary branch points could be an important mode of speciation, and so may have been a major force in the history of life. The unexciting possibility is that evolutionary branch points are rare and special things, easy to create on the computer but hard to find in nature.

2.8 References

Charnov, E.L. 1982. The Theory of Sex Allocation (Monographs in Population Biology 18). Princeton University Press, Princeton NJ.

Dugatkin, L. and H.K. Reeve, eds. 1998. Game Theory and Animal Behavior. Oxford University Press, Oxford UK.

Levin, S.A. and H. C. Mueller-Landau. 2000. The evolution of dispersal and seed size in plant communities. Evolutionary Ecology Research 2: 409-435.

Maynard Smith, J. and G.R. Price. 1973. The logic of animal conflict. Nature 246: 15-18.

Maynard Smith, J. 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

B. Sinervo and C.M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380: 240-243

B. Sinervo, E. Svensson, and T. Comendant. 2000. Density cycles and an offspring quantity and quality game driven by natural selection. Nature 406: 985-988.

Waxman, D. and S. Gavrilets. 2005. 20 questions on Adaptive Dynamics. Journal of Evolutionary Biology 18(5), 11391154.

Chapter 3

Single-species Population Dynamics

In this chapter we move up to the level of the population. The models we consider here attempt to explain and predict patterns of change over time in population *density*, the number of individuals per unit area or volume.

Population dynamics has always been a core topic in theoretical ecology. One of the original – and still strong – motivations for developing mathematical models is to understand the cause of cycles in particular populations, such as those shown in Figure 3.1. Is the famous lynx cycle the result of the trophic link between lynx and hare, or is it a cycle driven by the interaction between hares and their food resources with the lnyx just "along for the ride" as their food source waxes and wanes? We could imagine experimental approaches to this question, but their practicality is serious issue, and even when such experiments have been attempted their interpretation is not always clear (Turchin 2003). Models are therefore play two important roles: they identify classes of population interactions that could lead to the observed type of dynamics, and they may be able to identify unique predictions of each competing hypothesis that can be tested using observational data.

We are also interested in understanding patterns in cross-species comparisons – how do macroscopic properties such as the cycle period depend on properties of the individual organisms? For example, Figure 3.2 shows that when the period of population cycles is scaled relative to the individual maturation time (the time between birth and sexual maturity), some scaled periods are not observed (the INT category), and there is a clear difference in scaled period between specialists and generalists (a specialist feeds almost exclusively on just one species, a generalist feeds on many species). At the end of a long road that we are now starting, we will understand how this pattern was predicted *a priori* using models that account for individual growth and development, and their interactions with competition for resources.

Population modeling is also important for species management: managing fisheries for the highest possible sustainable yield, developing recovery plans for species threatened by extinction, or trying to contain or prevent the spread of invasive species. For these questions, we are often interested in how a population, or set of interacting populations, responds to a change in system parameters. For example: how

does a change in harvesting effort affect an exploited population, or how does a reduction in some cause of juvenile mortality affect a species at risk of extinction? How sure can we be about our predictions, or (more importantly) about the degree of uncertainty in our predictions?

3.1 Unstructured populations in continuous time

"Unstructured" means that we ignore differences between individuals, and pretend that a total headcount – irrespective of age, sex, breeding status, disease state, etc. – provides all the necessary information for predicting future population changes. You have to start somewhere, this is the easiest place to start, and easy is where you always start when you're trying to develop something new.

The starting point for modeling population change is the fundamental "BIDE" Balance Law for total population size $N(t)$:

$$N(t+h) = N(t) + \text{ Births } + \text{ Immigration } - \text{ Deaths } - \text{ Emigration} \tag{3.1}$$

Equation (3.1) is always true, but it is vacuous until we specify the values of B,I,D, and E over the time interval $(t, t+h)$.

3.1.1 First simple models

Exponential growth results from the simplest possible assumptions, constancy of the per-individual birth and death rates:
Births = (# parents) × (births/parent/time) × (length of time interval)

$$B = N(t) \times b \times h$$

and similarly

$$D = N(t) \times d \times h$$

and $I = E = 0$ (closed population). Then

$$\begin{aligned} N(t+h) &= N(t) + N(t)bh - N(t)dh \\ \frac{N(t+h) - N(t)}{h} &= (b-d)N(t) \end{aligned} \tag{3.2}$$

Now we let $h \to 0$ to get

$$\frac{dN}{dt} = (b-d)N(t) = rN(t) \quad \text{where } r = (b-d). \tag{3.3}$$

This is the differential equation every ecologist can solve,

$$N(t) = N(0)e^{rt}. \tag{3.4}$$

The above is a paradigm of how a model is created.

1. The *state variable(s)* of interest are specified: here just $N(t)$, total population size, based on the "unstructured" assumption.
2. The *processes* affecting the state variable are specified, here by the general BIDE equation.
3. The *process rates* (how many births per time interval?, etc.) are specified as functions of the state variables to give a closed system of equations.
4. The final result is a *dynamic equation* specifying how the state variables change over time, that can be solved (analytically or numerically) to determine what the model predicts

It often helps the modeling process to build up a compartment ("box and arrow") diagram of the model as you proceed, with boxes representing state variables and arrows the processes that cause a state variable to increase or decrease. For example, Figure 3.3 shows the diagram for the exponential growth model:

The solution (3.4) says that if $r > 0$ (i.e. if $b > d$) the population grows exponentially without limit, and if $r < 0$ it decreases to 0. The latter is all too credible nowadays for nonhuman species, but even for humans the former is not (regardless of what Julian Simon would like us to believe). A more plausible model has to put limits on growth.

The *logistic model* assumes that the net birth-death rate decreases linearly with N. The traditional form is

$$dN/dt = rN\left(1 - \frac{N}{K}\right), \qquad r, K > 0. \tag{3.5}$$

This has a long history and remains a favorite representation of limited growth. But despite early (1930's) claims that it was a general quantitative law, it really is just a convenient, simple model: linear decrease is sometimes seen, sometimes not. Note that the model makes no sense biologically if $r < 0$, since then a population with $N(0) > K$ would increase without limits.

The behavior of solutions to the logistic model can be found by a simple graphical analysis.

1. Graphing dN/dt versus N we have a parabola, with zeros at $N = 0$ and $N = K$ and maximum at $N = K/2$.
2. Any N where $dN/dt = 0$ is an *equilibrium = steady state = fixed point = critical point.* The equilibria for the logistic model are therefore at $N = 0$ and $N = K$.
3. We can also determine the *stability* of fixed points: 0 is unstable, K is stable (N^* stable means: if $N(0)$ is near N^*, then $N(t) \to N^*$ as $t \to \infty$. Unstable means: not stable).
4. The graph of dN/dt also allows us to sketch the qualitative form of solutions $N(t)$.
    - If a solution starts at $0 < N(0) < K/2$, then $N(t)$ is an increasing function of t because $dN/dt > 0$. Moreover, as $N(t)$ increases, so does dN/dt until $N(t)$ reaches $K/2$; after that dN/dt decreases over time while $N(t)$ continues to increase up to K. The graph of $N(t)$ is therefore sigmoid, with an inflection point when $N(t) = K/2$.

- If a solution starts at $K/2 < N(0) < K$, then dN/dt is always decreasing while $N(t)$ increases to K, so the solution curve is concave down.
- What is the shape of solutions when $N(0) > K$?

5. The fixed points and their stability can also be found by graphing the net (birth rate - death rate) $r(N) \equiv (1/N)dN/dt = r(1 - N/K)$, so long as we remember that there is an equilibrium at $N = 0$ even though $r(0) > 0$. Positive values of $r(N)$ imply that the population will grow, negative values imply that it will shrink, and any nonzero N^* where $r(N^*) = 0$ will be an equilibrium.

Exercise 3.1 Analyze graphically the dynamics of the following equations. In each case, plot dx/dt versus x, find all the equilibria, determine their stability, and sketch the qualitative shape of $x(t)$ versus t for various starting values, as we did for the logistic equation in lecture. You may find it useful to use **R** for plotting dx/dt versus x.
(a) $dx/dt = 1 - x^2$ (consider also $x < 0$)
(b) $dx/dt = x(x - 0.5)(2 - x)$ (consider only $x \geq 0$)

3.1.2 Scaling out parameters

Numerical solutions of a model depend on the values of all its parameters. However, qualitative features of model behavior are usually controlled by a smaller number of parameter combinations. One way to find these useful quantities is by re-scaling the model into dimensionless form.

For the logistic model, we first define a new state variable $x = N/K$. We rescale in this way because N and K are in the same units, so this makes x a dimensionless quantity, meaning that its value does not depend on the units in which N and K are measured. Then

$$\frac{dx}{dt} = \frac{1}{K}\frac{dN}{dt} = r\frac{N}{K}\left(1 - \frac{N}{K}\right) = rx(1 - x).$$

Next we rescale time. r is in units 1/time (WHY?) so rt is dimensionless. Let $\tau = rt$. Then

$$\frac{dx}{d\tau} = \frac{dx}{d(rt)} = \frac{1}{r}\frac{dx}{dt} = x(1 - x).$$

(this calculation can be justified formally using the Chain Rule: $\frac{dx}{d\tau}\frac{d\tau}{dt} = \frac{dx}{dt}$).

So apart from the scaling of the axes, the behavior of the logistic model is independent of the values of r and K (though note that we use the assumptions of the logistic model that r and K are both positive – otherwise the rescaled model would have time running backwards rather than slower or faster, and negative numbers of individuals).

A general – though not universal – rule is that rescaling can be used to "eliminate" one parameter per state variable, and one more for the time variable in a continuous time model. In a discrete time model

you can knock out one per state variable and that's it. Once all the variables (including time) in a model are dimensionless (like $x = N/K$), you can't get any more simplifications by rescaling. However there may be several different ways of rescaling a model into dimensionless form, and some may be more helpful than others for the tasks at hand.

Exercise 3.2 Show how the following model for a harvested population

$$dN/dt = rN\left(1 - \frac{N}{K}\right) - H\frac{N}{A+N}, \qquad r, K, A, H > 0.$$

can be re-scaled into the dimensionless form

$$\frac{dx}{d\tau} = x(1-x) - h\frac{x}{a+x}$$

and give the expressions for τ, x, h, a in terms of the variables and parameters of the original model. In the second term of the model H is a measure of fishing effort (e.g. proportional to the number of boat-days per fishing season), and A is the fish population density at which the capture rate per unit of fishing effort reaches half its maximum possible value.

3.1.3 Spruce Budworm

The graphical analysis of equilibrium stability suggests that single-species models $dN/dt = f(N)$ are totally boring. That's not true once you consider how the population can respond to changes in parameter values. As an example, we consider now a classical model for the dynamics of spruce budworm (Ludwig et al. 1978). The goal of the model was to explain the qualitative pattern that budworm exhibits sudden outbreaks from low to high density, which causes defoliation of the forest, and then a sudden collapse back to very low numbers, rather than either gradual oscillations or convergence to a steady balance between bugs and trees. Our presentation draws on Strogatz (1994) and the original article by Ludwig et al.

The model for the insect (budworm) assumes that the population is regulated by two things: competition for resources (foliage) and predation. Competition is modeled by the logistic equation (as usual, because it's the easiest thing to use). Predation is modeled by a loss term depending on budworm density N,

$$dN/dt = RN\left(1 - \frac{N}{K}\right) - \frac{BN^2}{A^2+N^2}.$$

The predation term is a so-called "type III" functional response representing a generalist predator at constant density that switches among potential prey items as they change in abundance. Despite optimal foraging theory, this is often observed. $\dfrac{N^2}{A^2+N^2}$ represents how much "attention" the predator is paying to budworm.

In the logistic growth term, we think of K as depending on the foliage level in the trees. A more complete model, developed by Ludwig et al. (1978), includes a dynamic equation for changes in foliage. Here we will treat foliage level as a parameter, and think about how gradual changes in foliage level will affect the insect population.

First we need to non-dimensionalize; there are many options. In this case it is convenient to let $x = N/A$ (we'll see why later) so

$$\frac{N^2}{A^2 + N^2} = \frac{x^2}{1 + x^2}.$$

This results in

$$\frac{dx}{dt} = Rx\left(1 - \frac{x}{k}\right) - \frac{B}{A}\frac{x^2}{1 + x^2} \tag{3.6}$$

where $k = K/A$. Now we want to re-scale time. We have a choice: "kill R" or "kill B/A". It is convenient (for the same reason) to kill B/A. So we let

$$\tau = (B/A)t$$

and get

$$\frac{dx}{d\tau} = \frac{dx}{(B/A)dt} = \frac{A}{B}\frac{dx}{dt} = \frac{RA}{B}x\left(1 - \frac{x}{k}\right) - \frac{x^2}{1 + x^2} \tag{3.7}$$

$$\Rightarrow \qquad \frac{dx}{d\tau} = rx\left(1 - \frac{x}{k}\right) - \frac{x^2}{1 + x^2} \tag{3.8}$$

where $r = RA/B$.

To analyze the model, we start by looking for fixed points and analyzing their stability.

1. There is always a fixed point of (3.8) at $x = 0$, and it is always unstable (because the linear term dominates near x=0, giving $\dot{x} > 0$ (**Warning:** $\dot{x}$ denotes the derivative of x with respect to time. This notation is convenient because it lets us gloss over whether time is t or τ; but at the end you do have to remember how you scaled things).

2. Positive fixed points ($x > 0$) occur where $\dot{x}/x = 0$, i.e. where

$$r\left(1 - \frac{x}{k}\right) = \frac{x}{1 + x^2} \tag{3.9}$$

   This is equivalent to a cubic equation, so there can be at most 3 solutions. There has to be at least one (since a cubic tends to opposite signs as $x \to \pm\infty$ but there can be 1,2 or 3 depending on the values of r and k (see Figure 3.4).

We can study fixed points and their stability graphically by graphing the two sides of equation (3.9) versus x, on the same graph; see Figure 3.4. Fixed points then occur where the graphs of the two sides intersect. In Figure 3.4 k is constant and r varies, but the same happens as k varies with r held fixed: starting from the bottom dashed line where there are 3 fixed points, as k decreases we go to 2 fixed points and then to 1.

We can determine the stability of fixed points by graphing the flow on the x axis, as usual. From the expression for $\dot{x}$ we see that $\dot{x}$ is positive where the line (the left hand side in (3.9)) is above the curve (the right hand side), and negative where the line is below the curve. Doing this for the possible cases in Figure 3.4, we find that

1. When there are 3 fixed points $a < b < c$, we have bistability: the largest and smallest are stable, the intermediate fixed point is unstable.

2. When there is only one fixed point, it is stable.

We get from 3 fixed points to 1 when the fixed points b collides with one of the others, and they mutually self-destruct. This is called a *saddle-node bifurcation* – because in two-dimensional systems the colliding fixed points are a (stable) node and (unstable) saddle. A *bifurcation* is the term for a change in the qualitative nature of the solutions to a dynamical system. So long as all 3 fixed points are present, the qualitative picture is the same (the system tends to fixed points a or c depending on where it starts); after the bifurcation things are different.

It is useful to study bifurcations because they let us map out where in "parameter space" (the (k, r) plane) we have 1 versus 2 versus 3 fixed points. The model goes from 1 to 3 fixed points along the curve in parameter space where there are exactly 2 fixed points. This occurs at parameters where the two rate functions are tangent (see Figure 3.4). We find these tangencies by requiring that the curves be equal in both value and derivative:

$$r\left(1 - \frac{x}{k}\right) = \frac{x}{1+x^2} \tag{3.10}$$

$$-r/k = \frac{1-x^2}{(1+x^2)^2} \tag{3.11}$$

The best way to express this is by solving for r and k as a function of x, the (scaled) budworm density at which the tangency occurs. It goes like this: from the condition (3.10) we have

$$r = \frac{r}{k}x + \frac{x}{1+x^2}.$$

The condition (3.11) gives us r/k as a function of x. Plug that into the expression above for r and simplify:

$$r = \frac{2x^3}{(1+x^2)^2}. \tag{3.12}$$

Then take this expression for r as a function of x and plug it into (3.11), and simplify to get

$$k = 2x^3/(x^2-1). \tag{3.13}$$

Equations (3.12) and (3.13) are graphed in Figure 3.5. The cusp in the bifurcation diagram occurs because the maximum of $r(x)$ and the minimum of $k(x)$ occur at the same point, $x = \sqrt{3}$ (proof: $r'(x) = k'(x) = 0$ at $x = \sqrt{(3)}$). But the slope of r versus k (given by $\dfrac{dr}{dk} = \dfrac{dr/dx}{dk/dx}$) is a smooth function at this point. That means that the two branches of the cusp are tangent – they come into the cusp at the same slope.

It has taken a while to understand this model, but the all we're using is high school algebra, freshman calculus and (crucially) the computer. *The computer changes how we figure out the behavior of models.* We could try to proceed deductively from first principles using pencil and paper. But it's faster to first

see what happens on the computer screen (e.g., there is a cusp in the bifurcation curve, and it seems to be a tangency), and then use math to understand why those things happen.

So how was this model used to explain budworm outbreak and collapse? In the full model of Ludwig et al. (1978) the buworm population model was linked to a set of two equations for variables characterizing the state of the forest – the size and energy reserves of the trees – in such a way that parallel increases in tree size and health caused both $r =$ and k to increase. Both r and k are dimensionless ratios of parameters, and the full Ludwig et al. (1978) model – equations (20)-(22) in that paper – had the property that forest recovery after a budworm outbreak caused both r and k to increase. The referring to figure (3.6) we can have the following scenario:

1. Following an outbreak, the forest has collapsed to a state of low r and k such that there is only a single fixed point a at which budworm density is very low.

2. As the forest matures, r and k increase (more foliage, greater potential for budworm increase). However the budworm remain at a until the saddle-node bifurcation occurs where a and b collide and annihilate each other. The budworm density then suddenly increases to the fixed point c.

3. The same now occurs in reverse: as the forest is defoliated – decreasing r and k – the budworm remain at c until the saddle-node bifurcation occurs at which c is eliminated, and they suddenly collapse down to a.

This is an example of what is called *hysteresis*: as parameters are changed to new values, the system changes; but as parameters change back to old values, the system does not retrace its steps in reverse. Instead it follows a different path, because the change in parameters resulted in it jumping from one stable fixed point to another.

Exercise 3.3 We return here to the (re-scaled) model for a harvested fish population

$$\dot{x} = x(1-x) - h\frac{x}{a+x}, \qquad x > 0 \tag{3.14}$$

with parameters $a, h > 0$. h is a measure of fishing effort (relative to the intrinsic rate of growth of the fish population) and a is the value of (rescaled) fish density at which the harvest per unit of fishing effort reaches half its maximum possible value. Consequently a is "given" by the biology of the system, but h as subject to regulatory control.

(a) As in the budworm model, there is always a fixed point at $x = 0$. Is it always unstable in model (3.14)? [HINT: show that for $x > 0$, $\dot{x}$ has the same sign as $g(x) \equiv (a-h) + (1-a)x - x^2$. Positive fixed points and their stability can therefore be studied as if $\dot{x}$ were equal to $g(x)$.

(b) Graphically show how (3.14) can have either 0,1 or 2 positive fixed points. HINT: use $g(x)$ again. It is very useful to observe that $g(x)$ can be written as $[a + (1-a)x - x^2] - h$, the difference between a parabola and a horizontal line. Feel free to draw graphs freehand rather than using the computer, so long as you get things qualitatively right.

(c) Graphically show how the following scenario can occur: As the fishing rate h is increased, the fish population suddenly collapses and seems to be decreasing exponentially to extinction. When h is returned to its initial, lower value, the population does not recover, but instead continues to decrease. HINT: same hint as for (b).

(d) Find the expression for the curve in the (a, h) parameter plane where the saddle-node bifurcation occurs (going from 0 to 2 positive fixed points). HINT: use g again, and recall that the solutions of a quadratic equation $Ax^2 + Bx + C = 0$ are

$$\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

3.1.4 Local stability analysis of fixed points

This section is based on Chapter 2 of Bulmer's book, which can be consulted for additional details.

For the model

$$\dot{n} = f(n)$$

we want to determine the local stability of a fixed point $\hat{n}$. To study the dynamics near $\hat{n}$ we define $p(t) = n(t) - \hat{n}$ – the deviation from the fixed point. Note that $f(\hat{n}) = 0$ since it is a fixed point. Then

$$\begin{aligned} \dot{p} = \dot{n} &= f(\hat{n} + p) \\ &= f(\hat{n}) + f'(\hat{n})p + \cdots \text{ [we drop higher order terms in } p\text{]} \\ &\doteq f'(\hat{n})p \end{aligned} \tag{3.15}$$

So our equation for the local dynamics is $\dot{p} = rp$ with $r = f'(\hat{n})$. Consequently $\hat{n}$ is

- Stable if $f'(\hat{n}) < 0$
- Unstable if $f'(\hat{n}) > 0$

If $f'(\hat{n}) = 0$ it could be stable or unstable – the linear analysis doesn't tell us.

Example: Spruce budworm model at $x = 0$,

$$\dot{x} = rx\left(1 - \frac{x}{k}\right) - \frac{x^2}{1 + x^2}.$$

We could compute $f'(x)$ and set $x = 0$, but there's an easier way for $x = 0$. Quadratic and higher-order terms won't affect the slope at $x = 0$ so we can eliminate them *before* taking the derivative. That leaves us with

$$\dot{x} = rx$$

so $x = 0$ is always unstable (under the model's assumption that $r > 0$).

Example: $dn/dt = rn(1 - n), \quad r > 0$. We can find the fixed points and determine their stability by computing the derivatives of $f(n) = rn(1 - n)$ at the fixed points. But graphing $f(n)$ versus n it is easy

to see that $f'(0) > 0$ (unstable) and $f'(1) < 0$ (stable) – local linearization gives us the same answer as our previous graphical analysis.

Exercise 3.4 (based on Hess 1996). Levins (1966) introduced a simple model for species persistence in a *metapopulation* – a collection of local populations spread across a set of *habitat patches*, and linked through migration of individuals from between patches. The model tracks the fraction of patches that are *occupied* – meaning that some individual of the species is present – but does not pay attention to the variation in population density among occupied patches. The standard form of the model is

$$\dot{p} = mp(1-p) - xp \tag{3.16}$$

where p is the fraction of habitat patches occupied, m is the migration rate, and x is the extinction rate.
(a) The term $mp(1-p)$ represents recolonization of empty patches, through the arrival of a migrant. What biological assumptions lead to this expression for the recolonization rate?
(b) The term xp represents extinction of occupied patches. What biological assumptions lead to this expression for the extinction rate?
(c) Find the conditions under which the model (3.16) has a positive equilibrium p^*. Show that there is at most one positive equilibrium, and use local stability analysis to show that that the equilibrium is stable whenever it exists.
(d) What happens to the population when there is no positive equlibrium?
(e) Based on the Levins model (and others), many conservation biologists have suggested that increased connectivity between habitat patches will help the persistence of species whose habitat has become fragmented (a Web search on "conservation corridor" will find many examples where this idea is now being implemented). Hess (1996) pointed out that increased connectivity might also affect the rate of extinctions, by facilitating the spread of predators, pathogens, exotic organisms, and disturbances (fire, toxic waste) that could have negative impacts on the target species. He therefore modified the Levins model to

$$\dot{p} = r(c)p(1-p) - x(c)p \tag{3.17}$$

where c is the connectivity between patches, and r is the recolonization rate. Suppose that $r(0) = 0$ and r is an increasing function of c. Show that, under some assumptions about how the extinction rate x depends on c, increased connectivity could be beneficial up to a point but then harmful: the species goes extinct if c is too small or too large, and persists at some intermediate range of c values.

3.2 Single unstructured population, discrete time

We continue to follow Chapter 2 of Bulmer's book. The basic model in discrete time is

$$n_{t+1} = F(n_t) = n_t C(n_t) \tag{3.18}$$

Example: the Ricker map (called the "discrete logistic" in Bulmer),

$$n_{t+1} = n_t e^{r(1-n_t/K)}$$

As usual we rescale the model by setting $x = n/K$, which gives us

$$x_{t+1} = x_t e^{r(1-x_t)}.$$

But note that we can't "kill" r like we did in the continous time logistic. In discrete-time models you can typically scale out one parameter per state variable, by making each state variable dimensionless. In continuous-time models you can generally scale out one more by making time nondimensional. But in discrete-time models time already is nondimensional – the model describes what happens in one "time step" whether that step is an hour, day, or year.

For a difference equation like this, fixed points occur where $F(x) = x$, intersections between the graph of $y = F(x)$ and the 45-degree line $y = x$. **(Draw the graph)** Note the potential for overshoot of a fixed point, impossible in $dn/dt = f(n)$. That is, it is possible to have n_t below a fixed point and n_{t+1} above it.

3.2.1 Local stability analysis

As usual we let $n_t = \hat{n} + p_t$ where $\hat{n}$ is a fixed point, meaning that

$$F(\hat{n}) = \hat{n}.$$

Then

$$\begin{aligned} n_{t+1} = F(n_t) &= F(\hat{n} + p_t) \\ &\doteq F(\hat{n}) + F'(\hat{n})p_t \\ &= \hat{n} + F'(\hat{n})p_t \end{aligned} \tag{3.19}$$

and therefore

$$p_{t+1} \doteq F'(\hat{n})p_t.$$

Consequently, $\hat{n}$ is

- Stable if $|F'(\hat{n})| < 1$, and unstable if $|F'(\hat{n})| > 1$. If $F'(\hat{n}) = \pm 1$ we don't know: the local linear analysis is inconclusive.
- Monotonic if $F'(\hat{n}) > 0$, and
- oscillatory if $F'(\hat{n}) < 0$ (oscillatory means that successive values of n_t are on opposite sides of the equilibrium).

Note: all 4 combinations of (Stable, Unstable) × (Monotonic, Oscillatory) are possible.

Example: what most people call the discrete logistic model,

$$n_{t+1} = rn_t(1 - n_t/K), \quad r > 0.$$

What are the equilibria, and how does their stability depend on parameter values?

The first step is to get rid of one parameter; letting $x_t = n_t/K$ we have

$$x_{t+1} = rx_t(1 - x_t), \quad r > 0. \tag{3.20}$$

Equilibria of (3.20) satisfy the equation

$$\hat{x} = r\hat{x}(1 - \hat{x}).$$

One solution is $\hat{x}_1 = 0$. To find others we divide through by $\hat{x}$ and get

$$1 = r(1 - \hat{x}),$$

so the second equilibrium is

$$\hat{x}_2 = 1 - \frac{1}{r},$$

which is only meaningful when $r > 1$ so that $\hat{x}_2 > 0$.

To examine the stability of these equilibria we need to look at the derivatives of $F(x) = rx(1 - x)$. Expanding the quadratic and differentiating, we have

$$F'(x) = r - 2rx = r(1 - 2x).$$

- Stability of $\hat{x} = 0 : F'(0) = r$, so this equilibrium is stable when $r < 1$ and unstable for $r > 1$.
- Stability of $\hat{x} = 1 - 1/r : F'(1 - 1/r) = r(1 - 2(1 - 1/r)) = r(2/r - 1) = 2 - r$, so this equilibrium is stable whenever

$$\begin{gathered} -1 < 2 - r < 1 \\ -3 < -r < -1 \\ 1 < r < 3 \end{gathered} \tag{3.21}$$

  Convergence to $\hat{x}$ will be monotonic if $0 < 2 - r < 1$, which is $1 < r < 2$, and oscillatory for $2 < r < 3$.

This leads us to make the following reasonable guesses:

1. For $0 < r < 1$ all solutions tend to $\hat{x} = 0$, so the population dies out.
2. For $1 < r < 3$ all solutions tend to $\hat{x} = 1 - 1/r$, so the population settles down to a constant value.
3. for $r > 3$ the population doesn't die out and it doesn't settle down, so it must constantly fluctuate.

Exercise 3.5 Do some computer experiments to see how much of our "reasonable" guesswork is correct (answer: almost all of it, but not exactly 100%).

Exercise 3.6 Pulliam (1988) developed a model for populations divided between *source* and *sink* habitat. In source habitat the species is self-sustaining and exports migrants. In sink habitat the species cannot maintain itself unless the local population is augmented by immigrants. Pulliam's (1988) simple didactic model is as follows:

• The source habitat contains $\hat{n}_1$ breeding sites, which are always fully occupied. Each year the source habitat population grows by a factor $\lambda_1 > 1$, with $\hat{n}_1$ individuals remaining in the source habitat, and the remaining $(\lambda_1 \hat{n} - \hat{n})$ moving to the sink habitat.
•The sink habitat population $n_2(t)$ shrinks by a factor $\lambda_2 < 1$ each year, followed by arrival of the migrants from source habitat.
These give the following equation for the sink habitat population dynamics:

$$n_2(t+1) = \lambda_2 n_2(t) + \hat{n}_1(\lambda_1 - 1) \tag{3.22}$$

Note that the only state variable is n_2; $\hat{n}_1 and \lambda_1, \lambda_2$ are parameters whose value doesn't change over time.

(a) Find the nonzero fixed point for the sink habitat population n_2 (call it $\hat{n}_2$) and show that $\hat{n}_2$ is locally stable by a linearized stability analysis.

(b) For what values of the model parameters will $\hat{n}_2$ be greater than $\hat{n}_1$?

(c) Show that for suitable parameter values, the number of immigrants to the sink population may be far smaller than the number of individuals in the sink population.

The conclusion of this paper – and the reason it got so much attention – are the ways for things to be other than they seem:

- most individuals of a species may be in *unsuitable* habitat where it cannot sustain itself.
- A seemingly trivial number of immigrants, coming from a small fraction of the occupied habitats, may be crucial for a population to persist.

A lot of conservation planning is still based on range maps and the idea that you can preserve a species by preserving the places where it is found. But if source-sink is credible (as may believe), a static picture of "who lives where" may be very misleading. Even if a species and its habitat are protected across most of its range, it may go extinct if most of the protected area is sink habitat.

Exercise 3.7 Here we explore the possible role of *lattice effects* in the dynamics of these simple population models. "Lattice effects" are the consequences of the fact that actual number of individuals in a population must be an integer. For this exercise, look at lattice effects in the unscaled logistic model

$$N_{t+1} = rN_t(1 - N_t/K)$$

with $r = 3.7, K = 100$.
(a) Do some computer experiments to compare the dynamics of the logistic model to that of the lattice logistic model

$$N_{t+1} = \text{round}\,(rN_t(1 - N_t/K))$$

(here round(x) is the integer closest to x), and report what you discover. It will be informative to start by comparing what happens if N_0 is close to the fixed point of the non-lattice model, but don't stop

with that.
(b) For the lattice model, draw a plot of N_{t+1} versus N_t, and add the line $N_{t+1} = N_t$ to the plot. What does that tell you about stability of the fixed point for the lattice model? What else does this explain about your discoveries in part (a)?
(c) How much difference do you think it would make if you rounded down in the lattice model, instead of rounding to the nearest integer?
For more on lattice effects, see Henson et al. (2001) and subsequent work citing that paper.

3.3 Fitting the models to data

We now have two different models for the changes in abundance of a single unstructured population

$$\begin{aligned} \dot{n} &= f(n), \\ n_{t+1} &= F(n_t) \end{aligned} \tag{3.23}$$

Both of these are ways to specify "what happens next" – given the current state (population density n at time t), the first gives the instantaneous rate of change dn/dt, the second specifies the finite rate of change $n_{t+1} - n_t$.

Biologically, we think of these as corresponding to overlapping versus discrete generations. In the former case, the ongoing stream of births and deaths is approximated by an equation for the rate of ongoing population change (net (birth rate - death rate) as a function of current population size). In the latter case, the population really jumps from one value (this generation) to another (next generation).

However, any differential equation implicitly deterimines a difference equation: $n(t+1) =$ where you wind up after one time unit, starting from $n(t)$. So if a continuously growing population can be described by a model of the form $dn/dt = f(n)$, we can also fit a model of the form $n_{t+1} = F(n_t)$, and both models should make the same predictions.

Figures 3.7 and 3.8 show an example. The data are from a study by Veilleux (1976, 1979) on a protozoan predator-prey system *Paramecium aurelia* and *Didinium nasutum*, which is a classic model system for population studies. *Paramecium* was grown on Cerophyl medium (which provides nutrients for the bacterial populations upon which the *Paramecium* feed). Varying the Cerophyl concentration corresponds to varying the prey carrying capacity. Methyl cellulose was added to thicken the medium and slow down predator and prey movements. Persistent predator-prey cycles were obtained in this medium for Cerophyl concentrations ranging from $0.675 - 0.9 g/l$. The data shown in the figures are *Paramecium* growing in the absence of the predators, at $0.9 g/l$, digitized from Figure 2b in Veilleux (1976). Jost and Ellner (2000) analyzed these data as a step towards estimating a model for the predator-prey dynamics.

The differential equation model requires us to find a smooth function representing $\dot{x}$ as a function of x. A simple approach is to use the $x(t)$ data to estimate $\dot{x}$; fitting the model is then a statistical regression problem. The data are at regularly spaced times t_i, so a crude estimate of the derivative at time t_i is

the centered difference quotient

$$\dot{x}(t_i) \doteq \frac{x(t_{i+1}) - x(t_{i+1})}{t_{i-1} - t_{i-1}}.$$

We can then plot these estimated derivative values versus the corresponding x value, and fit a curve (bottom panel in Figure 3.7). A logistic model (quadratic without intercept) fits the $\dot{x}$ data reasonably well, and solutions of the model approximate the data reasonably well.

The difference equation is easier: just plot "next generation" versus "this generation" and fit a curve to it. Again a logistic model is a reasonable fit (the discrete logistic map in the usual sense, $n(t+1) = an(t) - bn(t)^2$), and iterating this model gives a good fit to the data (Figure 3.8).

The **R** code for doing these fits is shown below. In reality (Jost and Ellner 2000) one would go beyond this crude fit of the differential equation model. First, the derivative can be estimated more accurately by fitting a smooth curve through the data, and taking the derivative of the fitted curve. Second, there seems to be overshoot: the population gets above K and then converges with oscillations (this is more pronounced in other data sets, where the simple logistic model consequently is not such a good fit). Jost and Ellner (2000) found that the data from this and other experiments could be fitted better by a logistic model with time delay,

$$\dot{n} = rn(t)(1 - n(t-\delta)/K).$$

Mechanistically this corresponds to individuals alive at time t giving birth and dieing at a rate determined by population density (a surrogate for prey availability?) at time $t - \delta$. If so, we should see in the data that $\dot{\log}(n(t))$ is a linear function of $n(t-\delta)$. This seems to be true, more or less, for $\delta = 12h$, which was interpreted as corresponding to the time time between cell divisions estimated by Vielleux (9-15h); see Figure 3.9.

3.4 Period doubling bifurcation

We have seen that in the logistic map there is a single fixed point $\hat{n}$ for $r > 1$, which becomes unstable at $r = 3$. Specifically, a r increases past 3, $F'(\hat{n})$ goes from $-1 + \varepsilon$ (stable) to $-1 - \varepsilon$ (unstable).

What happens for $r > 3$, when no fixed point is stable? Look at the second iterate map

$$F_2(x) = F(F(x))$$

[Note: this is also notated as F^2 or $F^{(2)}$. $F_2(x)$ has two humps instead of 1. This is not a special property of the logistic map. It happens because the maximum value of $F(x)$ – the top of the hump – is above 0.5, the value of x at which the hump occurs. So as x increases from 0 to 1, $F(x)$ increases from 0 to something above 0.5, and then decreases back to 0. So $F(F(x))$ climbs up the hump in F and partway down it, but then reverses direction and goes back over the hump.

Note that $F_2(\hat{n}) = F(F(\hat{n})) = F(\hat{n}) = \hat{n}$, so $\hat{n}$ is also a fixed point of F^2. *Is it stable or unstable?* To answer this question we look at the derivative of F^2 using the chain rule:

$$F_2'(n) = (d/dn)F(F(n)) = F'(F(n))F'(n)$$

```
######### Fitting the logistic ODE by gradient matching
require(odesolve);
# read in the data
x=as.matrix(read.table(file="whatever"));
tvals=x[,1]; xvals=x[,2];  nt=length(xvals);

# Estimate derivatives by centered differences, fit logistic model
xt=xvals[2:(nt-1)];
dxt=(xvals[3:nt]-xvals[1:(nt-2)])/(tvals[3:nt]-tvals[1:(nt-2)]);
fit=lm(dxt~xt+I(xt^2)-1); # "-1" means to omit the intercept

#solve the fitted logistic equation
logist=function(t,x,parms) {
    r=parms[1]; b=parms[2];
    return(list(xdot=r*x+b*x^2));
}
times=(0:65)/10;
parms=fit$coef; x0=xvals[1];
out=lsoda(0.75*x0,times,logist,parms);

# plot the results
par(mfrow=c(2,1),cex.axis=1.25,cex.lab=1.25,cex.main=1.25,mar=c(5,6,4,4));
plot(tvals,xvals,type="o",pch=16,xlab="time t (d)",ylab="Population x(t)")
points(out[,1],out[,2],type="l",lty=2,col="red",lwd=2);
title("Data and fitted solution");

px=0:550; py=fit$coef[1]*px+fit$coef[2]*px^2;
plot(xt,dxt,pch=16,xlim=c(0,550),xlab="Population x(t)",ylab="Estimated dx/dt");
points(px,py,type="l"); title("Fitting logistic to dx/dt vs x")
```

Table 3.1: Fitting the logistic differential equation by gradient matching.

For $n = \hat{n}$ we have $F(\hat{n}) = \hat{n}$ so

$$F_2'(\hat{n}) = [F'(\hat{n})]^2$$

Thus $F_2'(\hat{n})$ goes from (1 - a bit) (stable) to (1+ a bit) (unstable) as r increases past 3, and at $r = 3$ the graph of F_2 is exactly tangent to the 45-degree line (slope=1).

So when $\hat{n}$ becomes unstable through the slope of F_2 becoming slightly bigger than 1, this creates two new fixed points of F_2, say $\hat{n}_1$ and $\hat{n}_2$. The new fixed points are both stable because the slope of F_2 at the fixed points is positive but less than 1. Again, these are not special properties of the logistic map.

```
x=as.matrix(read.table(file="whatever",skip=1));
tvals=x[,1]; xvals=x[,2];  nt=length(xvals);

# fit difference equation
xt0=xvals[1:(nt-1)]; xt1=xvals[2:nt];
fit=lm(xt1~xt0+I(xt0^2)-1);

# solve fitted difference equation
r=fit$coef[1]; b=fit$coef[2];
xhat=rep(0,nt); xhat[1]=xvals[1];
for(j in 2:nt) {
  xhat[j]=r*xhat[j-1]+b*xhat[j-1]^2
}

# plot the results
par(mfrow=c(2,1),cex.axis=1.25,cex.lab=1.25,cex.main=1.25,mar=c(5,6,4,4));
plot(tvals,xvals,type="o",pch=16,xlab="time t (d)",ylab="Population x(t)")
points(tvals,xhat,type="o",lty=2,col="red",lwd=2);
title("Data and fitted solution");

px=0:550; py=fit$coef[1]*px+fit$coef[2]*px^2;
plot(xt0,xt1,pch=16,xlim=c(0,550),ylim=c(0,550), xlab="Population x(n)",
ylab="Population x(n+1)");
points(px,py,type="l");
points(px,px,type="l",lty=2,col="red");
title("Fitting quadratic map to x(n+1) vs x(n)")
```

Table 3.2: Fitting the logistic map by ordinary least squares regression.

The two fixed new fixed points of F_2 are initially close to $\hat{n}$, so

$$F_2'(\hat{n}_i) \approx F_2'(\hat{n}) \approx 1 > 0.$$

The fact that $F_2'(\hat{n}_i) < 1$ is a consequence of the double-hump shape of F_2. At both fixed points the graph of F_2 crosses the 45-degree line from above to below, implying that the slope of F_2 is < 1. Thus $0 < F_2'(\hat{n}_i) < 1$, so the new fixed points are locally stable.

The new fixed points of F_2 are then a two-point cycle or period-2 oscillation in F, i.e.

$$F(\hat{n}_1) = \hat{n}_2, \qquad F(\hat{n}_2) = \hat{n}_1$$

The population oscillates up and down, alternating between values above and below the unstable fixed point $\hat{n}$.

GRAPH stable 2-point cycle, n_t versus t, oscillating around $\hat{n}$.

As r increases further, this process repeats. $\hat{n}_1$ goes from being a stable fixed point of F_2 (with slope near 1) to unstable (slope < -1), creating a pair of fixed points for $F_2(F_2(n)) = F_4(n)$, i.e. a four-point cycle. Further increases in r give 8,16,32, etc. point cycles, until eventually you reach deterministic chaos at which the oscillations become aperiodic. Quoting Marc Mangel, it is hard to be awake in the 21^{st} Century without being aware of chaos and the interest that it has aroused in many areas of science.

3.5 What really happens?

Even in our simplest population models, we have seen a range of qualitative dynamics: stability, cycles and chaos. So what really happens?

A short answer is that we see stability:cycles:chaos in proportions 2:1:0.

A longer answer is: **Chaos** The appeal to (at least some) ecologists of chaos was the promise that it would provide a simple explanation for complicated population dynamics. That is, a few strong feedbacks of simple form might be all it takes to explain any patterns whatsoever in the dynamics of populations. Figure 3.11 is the kind of thing that got people excited: all the complexities of real data coming from a 1-parameter, one-dimensional map. Maybe life was simple after all.

This has not panned out (Zimmer 1999), but it took a while not to. Early reviews of the available long-term data were essentially negative (Hassell et al. 1976). But the approach used in that study was soon rendered obsolete by developments in "nonlinear time series analysis" as practised by physicists. Those approaches, applied to ecological data, began to produce abundant evidence of chaos. Unfortunately, equally stong evidence could also be obtained using nonchaotic artificial "data" that had been generated on the computer using random numbers to simulate effects of environmental variation superimposed on other factors (seasonality, age structure, etc.). The methods from physics proved to be unreliable when the data did not come from controlled experiments in the laboratory. With more appropriate methods, suitable for data on populations in the wild, one finds a few apparent examples but not many (Ellner and Turchin 1995, Rees et al. 2002, Freckleton and Watkinson 2002). An "optimist" might now think that there are 3–6 solid examples out of the data sets examined for signs of chaos, which would extrapolate to saying that $\frac{1}{2}$–1% of populations are chaotic. But a "pessimist" could say that there are no solid examples outside of laboratory experiments.

Cycles These are common. The best empirical estimate (Kendall et al. 1998, see Table 3.3 is that about 1/3 of all animal populations show evidence for periodicity in their dynamics. These are not necessarily periodic orbits – they could be damped oscillations that are continually "kicked" into action by exogenous factors that perturb populations away from equilibrium. Nonetheless, the evidence suggests that ecologists' fascination with population cycles was not misplaced: cycles happen often enough to merit study. There is also a latitudinal trend, with populations near the poles being more likely to show cycles. In particular, the fitted trend line (Kendall et al. 1998, Figure 1) estimates that the majority

Taxon	# populations	% Periodic*
Birds	139	13
Mammals	328	33
Fish	129	43
Insects	79	16
Aquatic inverts	18	44
All	693	29

*Significant peak in the power spectrum.

Table 3.3: Fraction of long-term population records showing statistical evidence of cyclic fluctuations, classified by taxonomic group. Retabulated from Table 1 of Kendall et al. (1998)

of mammal populations at 60^o lattitude or higher exhibit cyclicity. The data base for these estimates suffers from observation bias: interesting data (and cycles are interesting!) are more likely to be collected for long enough to count as a long-term data set (Kendall et al. required 25 or more years of data for inclusion in their survey). On the other hand, given such short series and the high sampling error in population counts, one would expect that many cyclic cases would be missed because the signs were not strong enough to achieve statistical significance. So perhaps these balance out, and 1/3 is a good estimate for the frequency of cycling in natural animal populations.

Stability This is really a catch-all for non-cyclic patterns of population variation, that don't show any signs of being chaotic. Stability sounds boring, but in fact is the most problematic pattern for theory, and is still intensely studied despite decades of attention to it already. Exploiter-victim systems are prone to cycling; some of them (host-parasitoid) are, in simple models, prone to extreme cycles. All organisms are either exploiters or victims, or both. So why do we see so much stability in the real world? Maybe the simple models in this chapter are leaving out something essential for understanding real populations.

3.6 References

Ellner, S. and P. Turchin, 1995. Chaos in a noisy world: new methods and evidence from time series analysis. American Naturalist 145: 343-375.

Freckleton, R.P. and A.R. Watkinson. 2002. Are weed population dynamics chaotic? Journal of Applied Ecology 39: 699-707.

Hassell, M.P., J.H. Lawton, and R.M. May. 1976. Patterns of dynamical behavior in single-species populations. Journal of Animal Ecology 45: 471-486.

Hastings, A. 1997. Population Biology: Concepts and Methods. Springer, NY.

Henson, S.M., R. F. Costantino,J. M. Cushing, R. A. Desharnais, B. Dennis, A. A. King. Lattice effects

observed in chaotic dynamics of experimental populations. Science 294: 602 - 605.

Hess, G.R. 1996. Linking extinction to connectivity and habitat destruction in metapopulation models. American Naturalist 148: 226-236.

Kendall, B.E., J. Prendergast, and O.N. Bj**o**rnstad. 1998. The macroecology of population cycles: taxonomic and biogeographic patterns in population cycles. Ecology Letters 1: 160-164.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America 15: 237-240.

Ludwig, D., D.D. Jones, and C.S. Holling. 1978. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. Journal of Animal Ecology 47: 315-332.

Murdoch, W.W., B.E. Kendall, R.M. Nisbet, C.J. Briggs, E. McCauley, and R. Bolser. 2002. Single-species models for many-species food webs. Nature 417: 541-543.

Pulliam, H. R. 1988. Sources, sinks and population regulation. American Naturalist 132: 652-661.

Rees, M. et al. 2002. Snow tussocks, chaos and the evolution of mast seeding. American Naturalist 160: 4459.

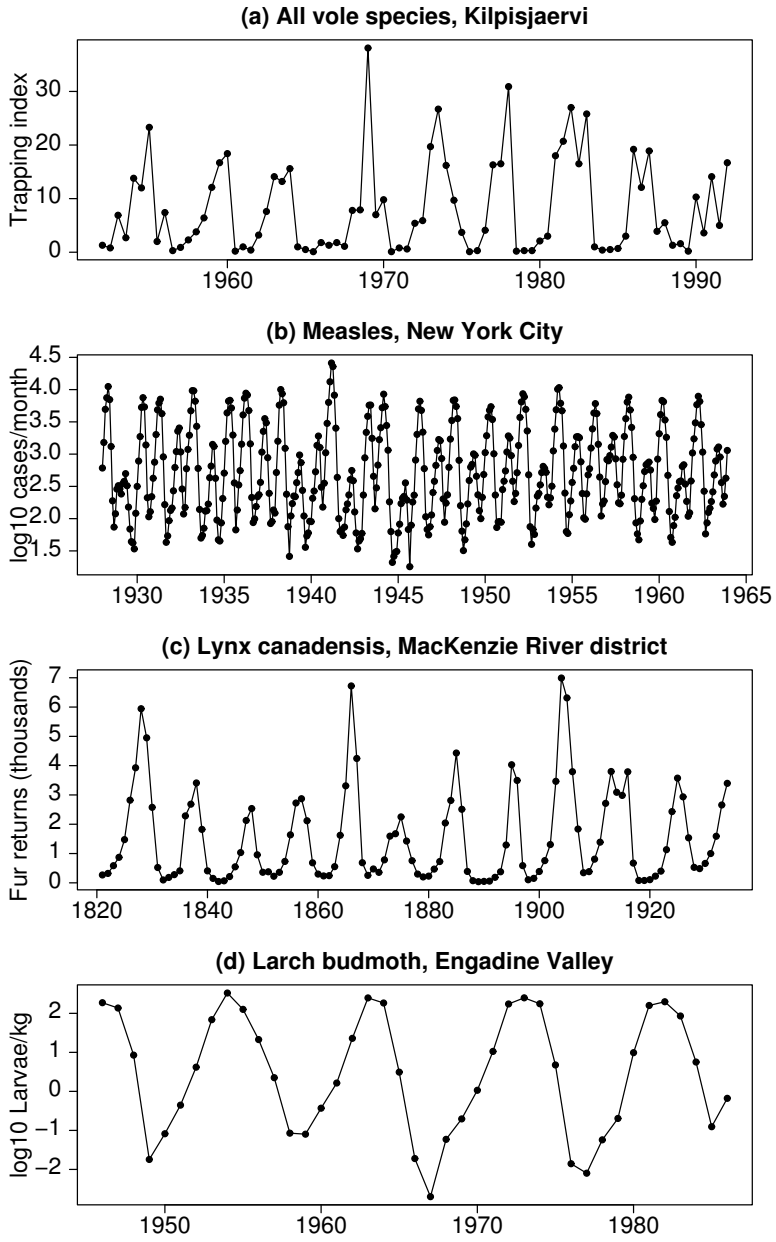
Strogatz, S. 1994. Nonlinear Dynamics and Chaos. Perseus Books, Reading Mass.

Tilman, D. 1982. Resource Competition and Community Structure. Princeton University Press, Princeton NJ.

Veilleux, B. G. 1976. The analysis of a predatory interaction between *Didinium* and *Paramecium*. Master's thesis, University of Alberta.

Veilleux, B. G. 1979. An analysis of the predatory interaction between *Paramecium* and *Didinium* Journal of Animal Ecology 4: 787-803.

Zimmer, C. 1999. Life after chaos. Science 284: 83-86.



.

Figure 3.1: Examples of cyclic fluctuations in natural populations. (a) All vole species at Kilpisjärvi, Finnish Lapland. The data are an index of vole abundance based on the number caught in a standardized trapping protocol each Spring and Fall. (b) Monthly case reports of measles in New York City, prior to vaccination. (c) Abundance of *Lynx canadensis* in the MacKenzie River district, Canada, based on annual fur capture records of the Hudson's Bay Company. (d) Number of larvae per kg of foliage of larch budmoth *Zeiraphera didiniana* in the Engadine Valley, Switzerland. A one-stop source for long term population data is the Global Population Dynamics Database [NERC Centre for Population Biology, Imperial College (1999). The Global Population Dynamics Database. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html].

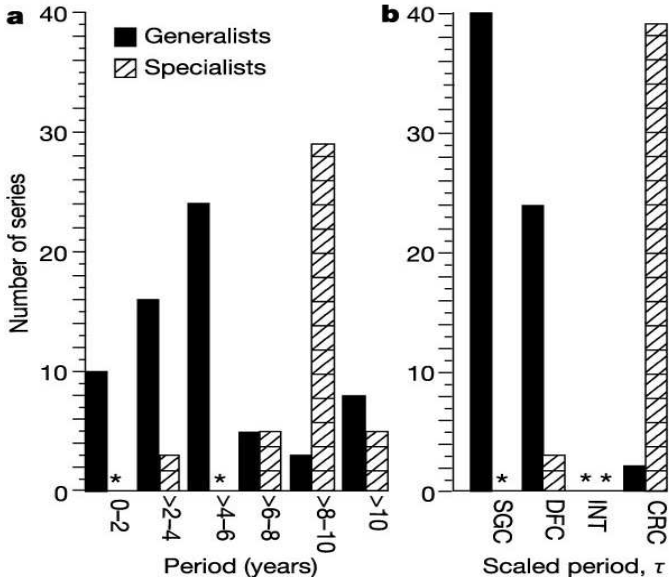


Figure 3.2: Population cycles classified by period, from Murdoch et al. (2002). Asterisk indicates zero in the class. (a) Number of cyclic populations with various periods in years. (b) Distribution of cycles among classes defined by scaled cycle period τ=(cycle period)/(maturation time). SGC, single generation cycles (τ= 1); DFC, delayed-feedback cycles ($2 \leq \tau \leq 4$); CRC, consumer-resource cycles (period in years $\geq 4T_C + 2T_R$ wherer T_C and T_R are the maturation times of consumer and resource species).

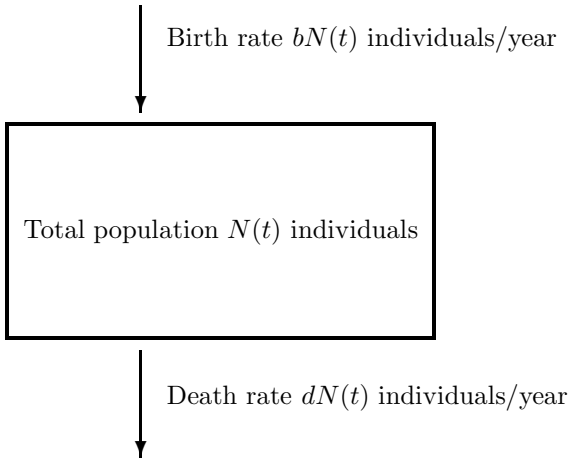


Figure 3.3: Compartment diagram of the exponential growth model. The rectangle denotes the state variable – the number of individuals in the population – and the arrows denote births and deaths.

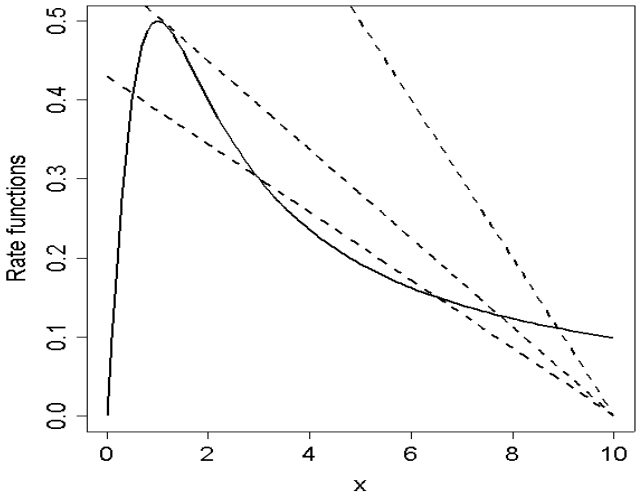


Figure 3.4: Examples of Budworm model with 1,2 or 3 nonzero fixed points for the insect population density. The solid curve is $\frac{x}{1+x^2}$, and the dashed lines are $r(1-x/k)$ for $k = 10$ and three different values of the $y-$intercept r.

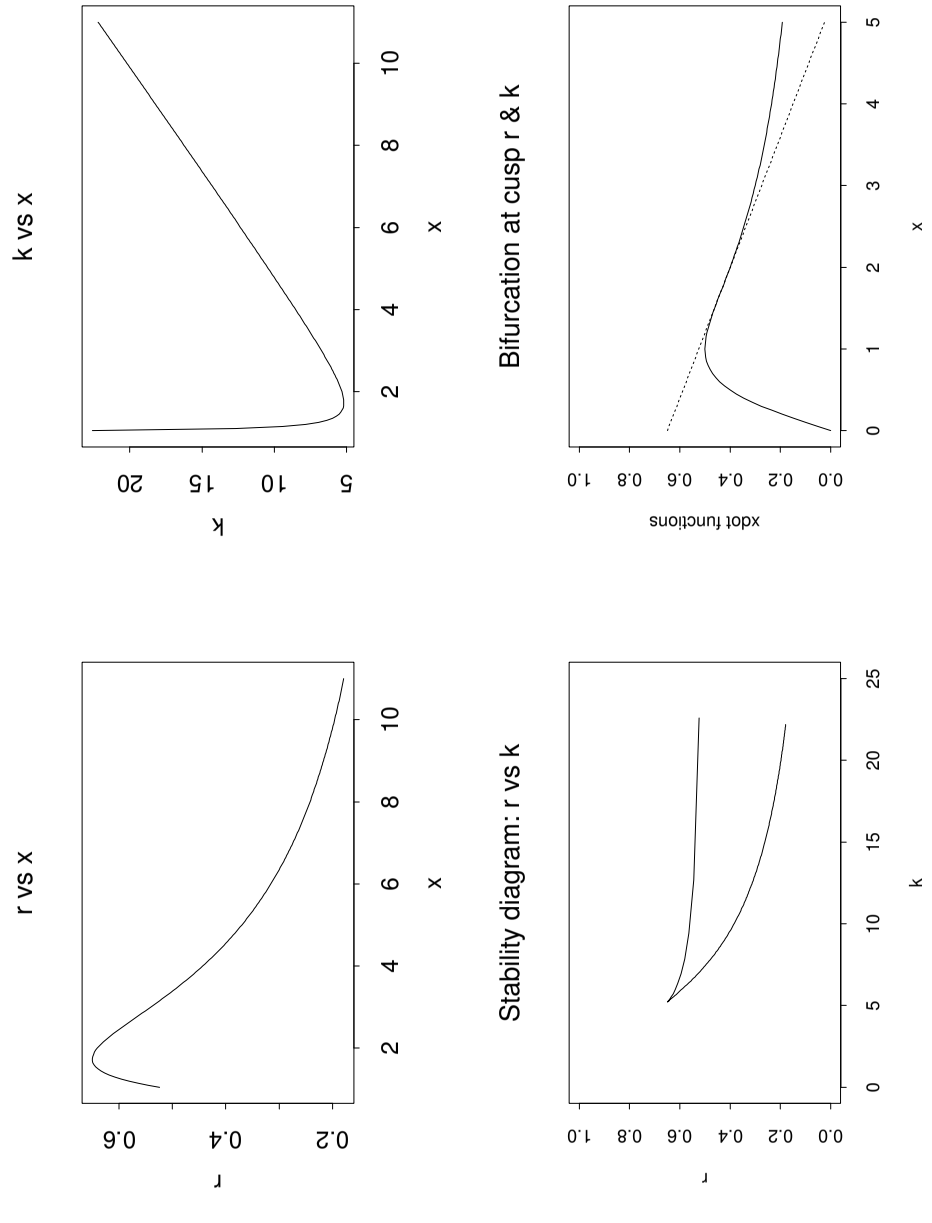


Figure 3.5: Bifurcation curve in budworm model

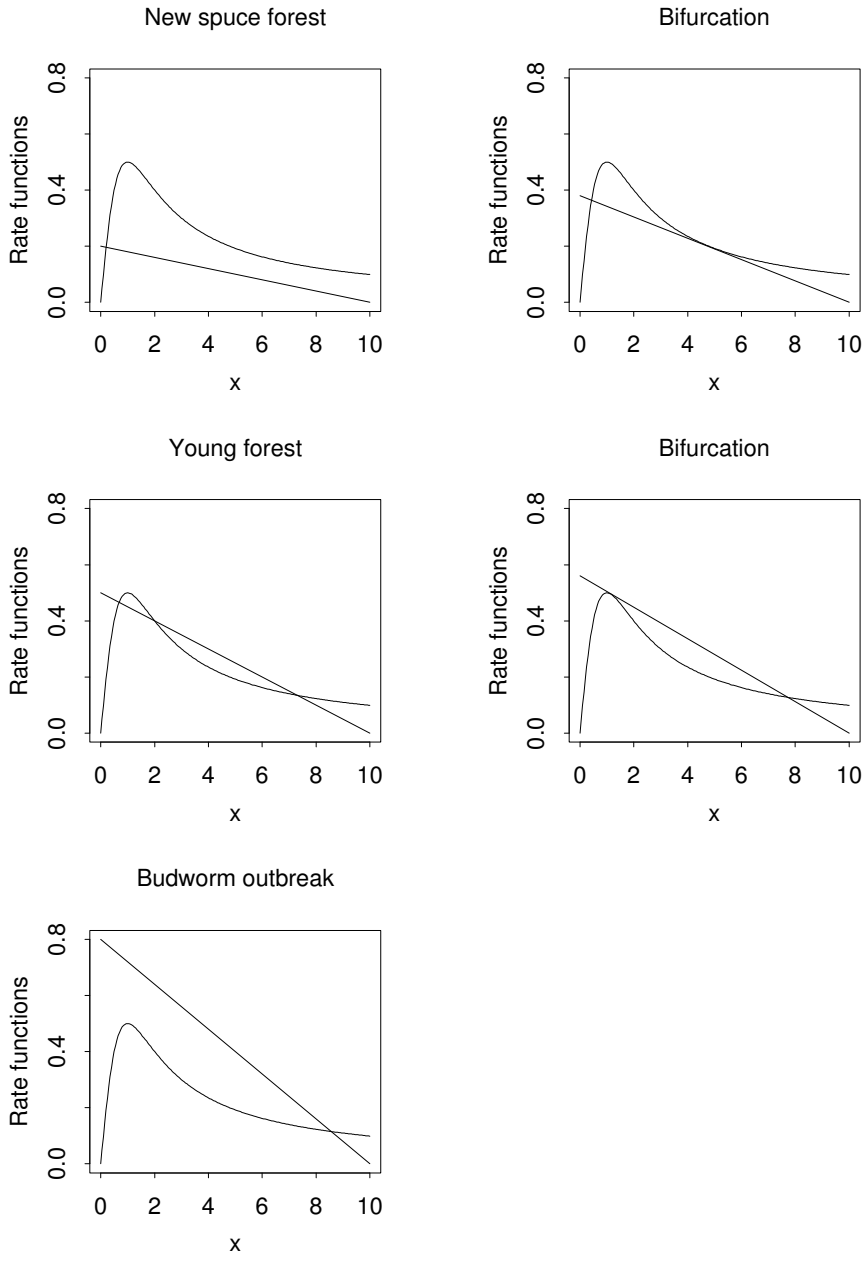


Figure 3.6: Scenario for budworm sudden outbreak and collapse through saddle-node bifurcations, resulting from forest growth and then defoliation by budworm

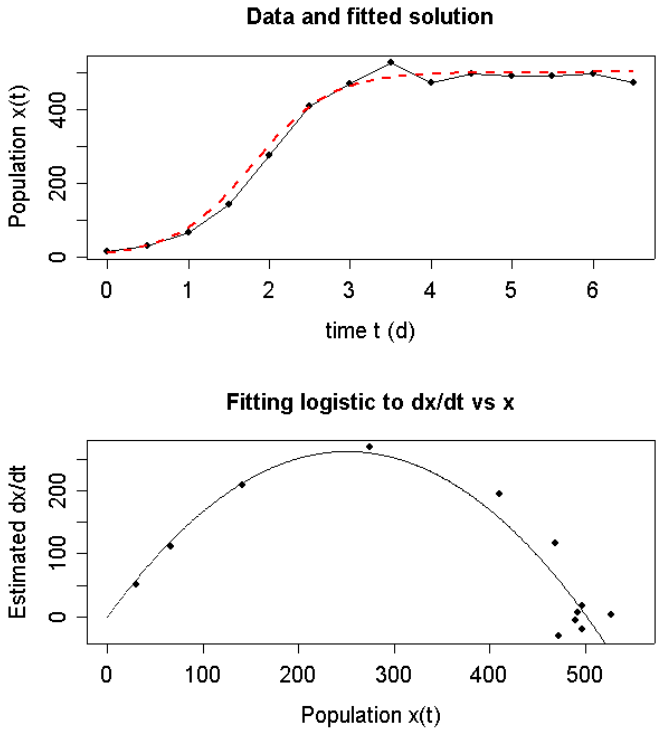


Figure 3.7: Fitting $\dot{x} = f(x)$ to Veilleux (1976) data on growth of *Paramecium* on Cerophyl medium. Top shows the data and a solution trajectory of the fitted model. Bottom shows the plot of estimated $\dot{x}$ versus x on which the model is based, and the fitted logistic model.

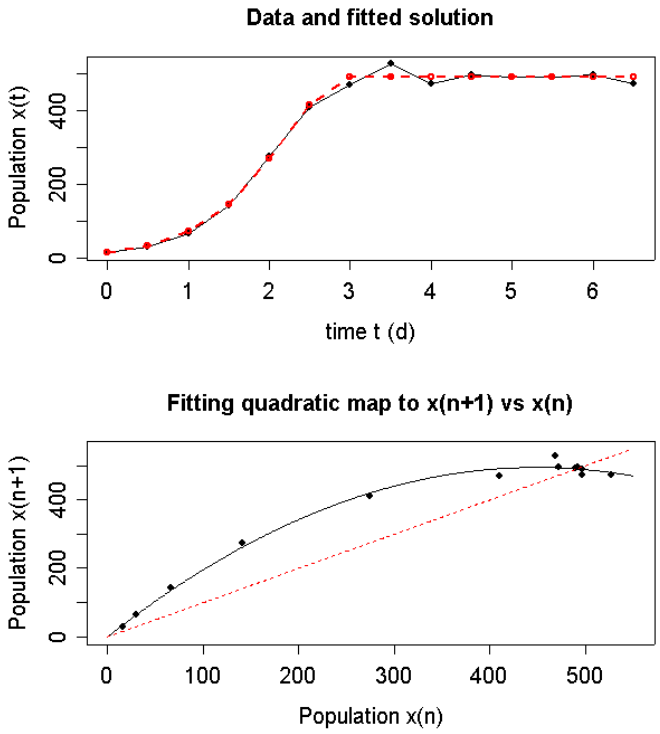


Figure 3.8: Fitting $n_{t+1} = F(n_t)$ to Veilleux (1976) data on growth of *Paramecium* on Cerophyl medium. Top shows the data and a solution trajectory of the fitted model. Bottom shows the plot on which the model is based, and the fitted quadratic (discrete logistic) model.

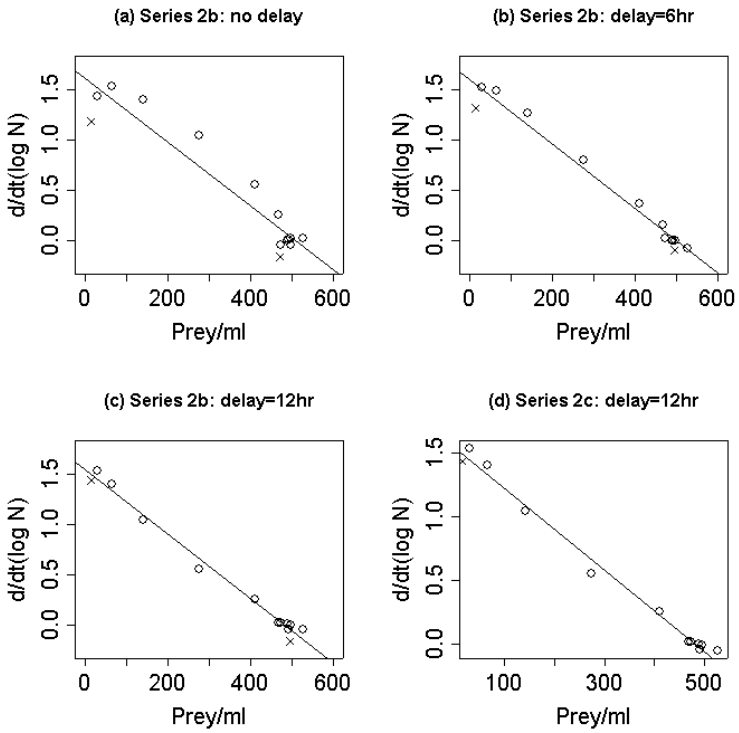


Figure 3.9: Fitting the time-delay logistic $\dot{n}(t) = rn(t)(1-n(t-\delta)/K)$ by regressing estimated derivative of $\log n$ versus $n(t-\delta)$. Panels a,b,c show fits with 3 values of δ for the data set used above for the logistic without time delay; panel d shows the fit with $\delta = 12h$ for another data set. Points marked with $\times$ are the first and last in the data series, at which the derivative estimates are least accurate. This figure is a re-plot of part of Figure 2 in Jost and Ellner (2000).

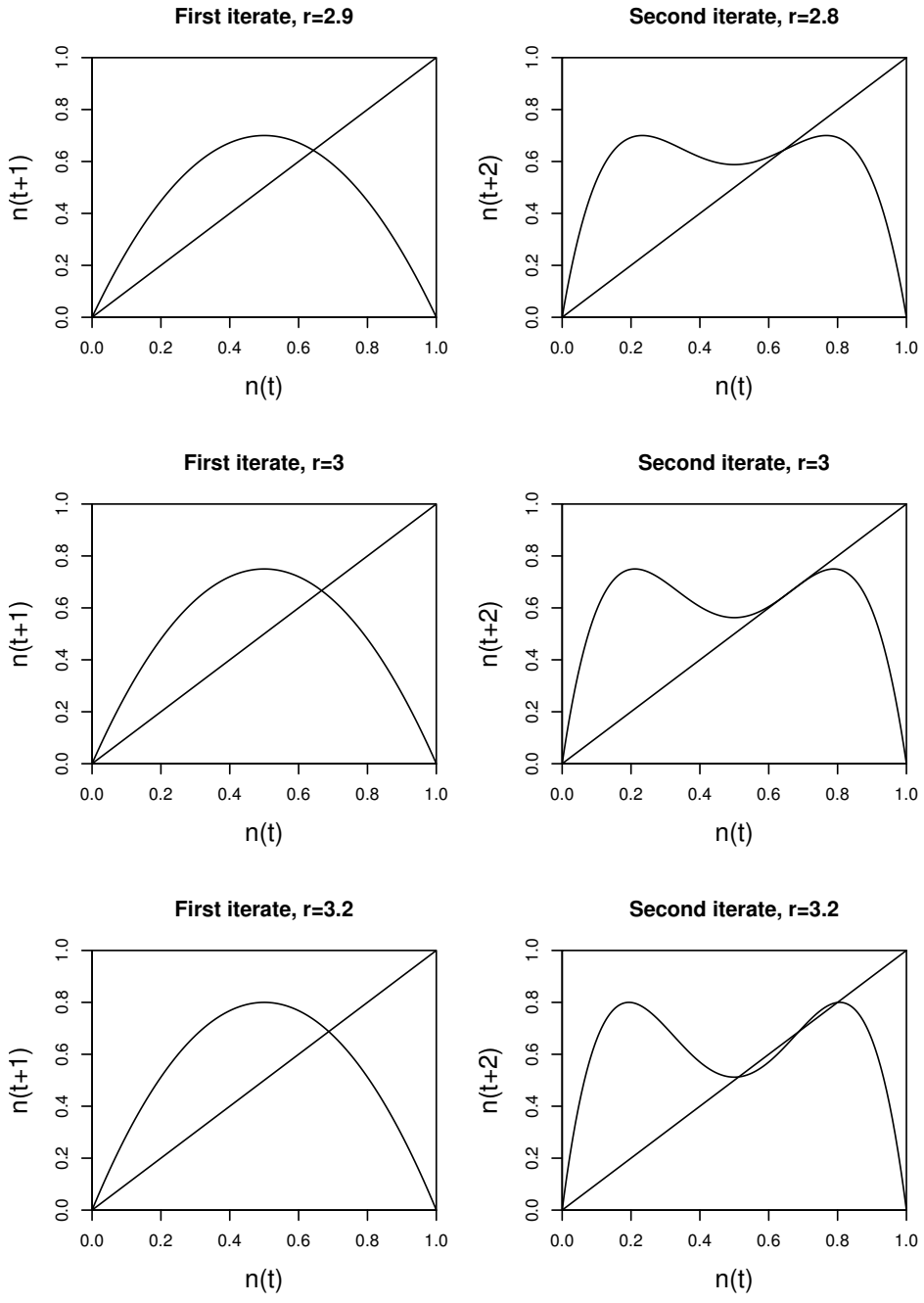


Figure 3.10: The first period-doubling bifurcation in the logistic map, which occurs at $r = 3$. The humped shape of $F(n)$, and the fact that the maximum value of $F(n)$ is > 0.5, implies that $F_2(n)$ has a double-hump shape.

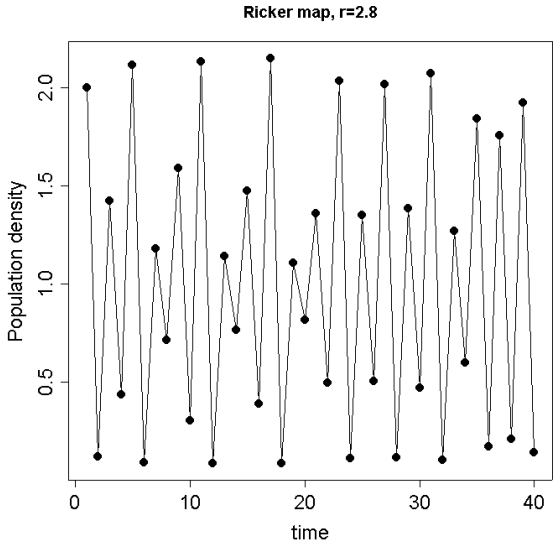


Figure 3.11: Chaotic dynamics in the Ricker map $n_{t+1} = n_t e^{r(1-n_t)}$ for $r = 2.8$

Chapter 4

Interacting Populations: Competition, Predation, and Parasitoids

Even in simple one-species unstructured models we have seen the potential for stability, cycles or chaos. In this chapter we consider another potential source of complex dynamics: interactions between different species. We will examine a series of models for different kinds of interactions, along the way learning the techniques for dealing with multi-species models. As in the last chapter, we will continue (for now) to use unstructured models that ignore within-species differences between individuals. We start with models in continuous time, and then consider discrete-time models.

4.1 Lotka-Volterra Competition Model

The Lotka-Volterra models are the extension to multiple species of the single-species logistic model. That is, the state variables are the total density of each species, and the per-individual effects are linear – the effect of species i on the per-capita growth rate of species j is proportional to the density of the species. In studying these models, we don't really believe the assumption of linearity – but it's the natural starting point, and it gives you the right "picture" for thinking about more general models.

The model for two competing species is

$$\begin{aligned} \dot{N}_1 &= r_1 N_1 \left(1 - \frac{N_1}{K_1} - \beta_{12}\frac{N_2}{K_1}\right) \\ \dot{N}_2 &= r_2 N_2 \left(1 - \frac{N_2}{K_2} - \beta_{21}\frac{N_1}{K_2}\right). \end{aligned} \tag{4.1}$$

Each β_{ij} is a positive parameter measuring the relative impact of one species j individual on the growth

of species i, relative to the impact of one species i individual.

The model is only interesting if each species could persist in the absence of the other, which means $r_1, r_2 > 0$. In that case we can't have both species going extinct – for if we did, then each species would eventually have $\dot{N}_i/N_i > 0$ and it would then be increasing rather than decreasing. The possible outcomes in the model are then persistence of both species, or persistence of one while the other goes to extinction.

We can get a complete picture of this model's behavior by examining its *nullclines*. The nullclines are the curves in the (N_1, N_2) plane where $\dot{N}_i = 0$. Nullclines have two important properties:

- Solution curves cross the N_1 nullcline going vertically up or down, because at that location only N_2 is changing. They cross the N_2 nullcline going horizontally to the left or the right.
- So long as $\dot{N}_1$ and $\dot{N}_2$ are continuous functions of (N_1, N_2) the sign of $\dot{N}_i(N_1, N_2)$ can only change when you cross the N_i nullcline.

For the Lotka-Volterra competition model the nullclines are

$$\begin{aligned} \dot{N}_1 = 0: &\quad N_1 = 0 \text{ or } \frac{N_1}{K_1} + \beta_{12}\frac{N_2}{K_1} = 1 \\ \dot{N}_2 = 0: &\quad N_2 = 0 \text{ or } \frac{N_2}{K_2} + \beta_{21}\frac{N_1}{K_2} = 1 \end{aligned} \tag{4.2}$$

The second condition in each nullcline defines a line with negative slope in the first quadrant of the (N_1, N_2) plane. Any intersection of the nullclines is a fixed point, so there are potentially four fixed points:

$$(0,0), (K_1, 0), (0, K_2), (N_1^*, N_2^*), \tag{4.3}$$

the last existing if the off-axis parts of the nullclines intersect.

The nullclines are lines, so they are determined by their points of intersection with the axes:

The N_1 nullcline runs from $\left(0, \dfrac{K_1}{\beta_{12}}\right)$ to $(K_1, 0)$.

The N_2 nullcline runs from $(0, K_2)$ to $\left(\dfrac{K_2}{\beta_{21}}, 0\right)$.

There are four possible configurations of the nullclines, depending on which of them has the higher intersection with the N_2 axis, and which of them has the rightmost intersection with the N_1 axis.

Consider first the case where the N_2 nullcline lies entirely above the N_1 nullcline. From the differential equations, we see that each $\dot{N}_i$ is negative above the nullcline for that species, and positive below it. That tells us the direction of flow for solution curves in each region of the positive quadrant (Figure 4.1).

Looking at the directions of flow, we should suspect that all nonzero solutions tend to $(0, K_2)$ – species 2 wins – and we can prove it. Consider first a solution starting about the N_2 nullcline. If the solution never

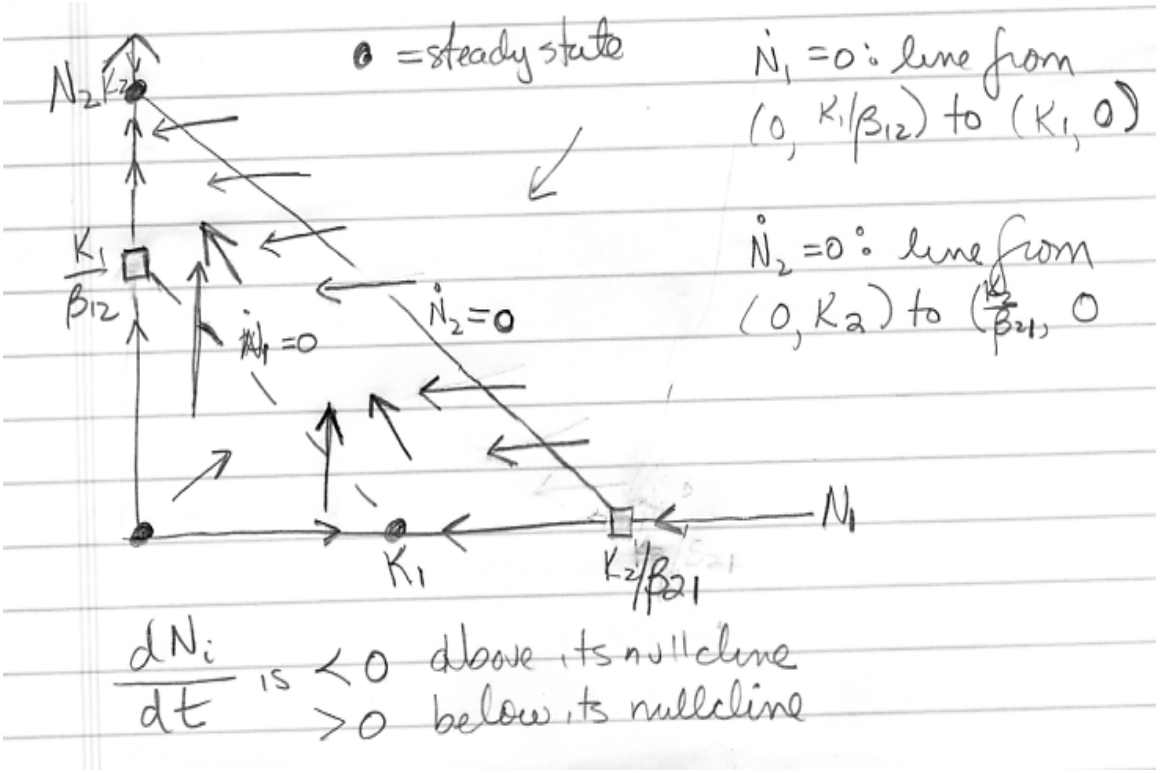


Figure 4.1: Nullclines in the Lotka-Volterra competition model, when the N_2 nullcline lies entirely above the N_1 nullcline. Arrows indicate the direction of flow implied by the nullclines.

crosses the N_2 nullcline, then N_1 and N_2 are monotonically decreasing and therefore converge to limiting values. Convergence to a limit means that $\dot{N}_i \to 0$, so the limit must be a fixed point. The only possible limit point is therefore $(0, K_2)$. If the solution does cross the N_2 nullcline, it can never escape the region between the nullclines. Therefore, once it enters that region N_1 is monotonically decreasing and N_2 is monotonically increasing. So again the solution must approach a limit that must be an equilibrium, which must be $(0, K_2)$.

We've already shown that a solution starting between the nullclines converges to $(0, K_2)$. A solution starting below the nullclines (but not at $(0, 0)$) must either remain below the nullclines (and therefore increase monotonically to a limit that must be $(0, K_2)$) or else cross into the region between the nullclines.

If the N_1 nullcline lies entirely above the N_2 nullcline – just swap $1 \leftrightarrow 2$ and we have the species 1 wins, and $N_2 \to 0$. The cases where the nullclines cross are slightly different, but the style of argument is the same: so long as solutions don't cross a nullcline they are monotonic increasing or decreasing, so solutions curves must converge to a limit, and the limiting point must be an equilibrium. The nullcline configurations and the conclusions are shown in Figure 4.2.

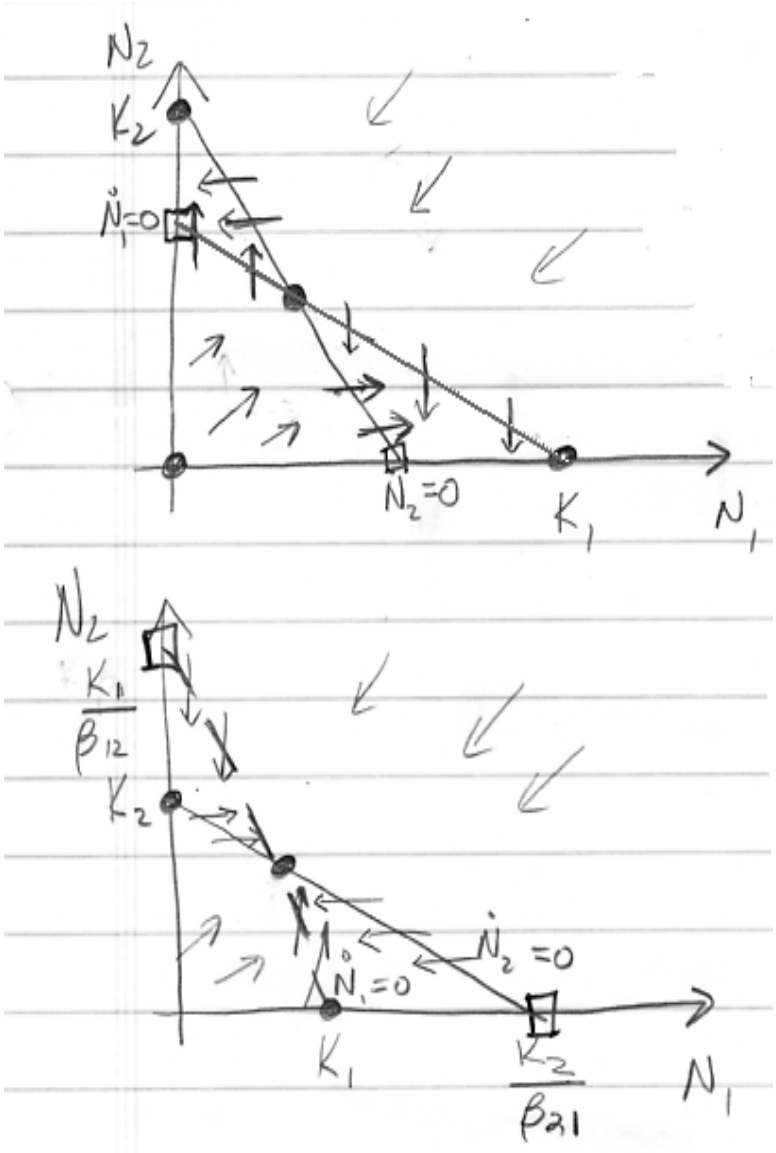


Figure 4.2: Nullclines in the Lotka-Volterra competition model, for the cases where the nullclines cross. Arrows indicate the direction of flow implied by the nullclines. In the upper panel there is contingent exclusion – one or the other goes extinct, depending on initial conditions – and in the bottom there is coexistence.

So coexistence of both species occurs only when

$$\begin{aligned} &\frac{K_2}{\beta_{21}} > K_1 \iff \frac{1}{K_1} > \frac{\beta_{21}}{K_2} \\ &\text{and} \\ &\frac{K_1}{\beta_{12}} > K_2 \iff \frac{1}{K_2} > \frac{\beta_{12}}{K_1} \end{aligned} \tag{4.4}$$

We can say this in words: coexistence occurs when within-species effects of density – the impact of one individual on the per-capita growth rate of its own species – is larger than the between-species effect.

Exercise 4.1 [from Hastings 1997] One nonobvious prediction of the Lotka-Volterra competition model is that nonselective mortality – "rarefaction" – can permit coexistence when there would otherwise be competitive exclusion. That is, consider the model

$$\begin{aligned} \dot{N}_1 &= r_1 N_1(1 - N_1/K_1 - \beta_{12} N_2/K_1) - mN_1 \\ \dot{N}_2 &= r_2 N_2(1 - N_2/K_2 - \beta_{21} N_1/K_2) - mN_2 \end{aligned} \tag{4.5}$$

(a) Find the equations for the nullclines in this model.
(b) Show, by plotting nullclines, that there are parameter values such that exclusion occurs for $m = 0$, but that coexistence occurs when m is increased to a suitable positive value (e.g. by experimental removals) with all other parameters remaining the same.

4.2 Mechanistic competition models

Although historically important and influential, the Lotka-Volterra models are now viewed as *phenomenological* because the interspecific effects are imposed without regard to the underlying mechanism – we simply posit that for some unspecified reason each species impedes the other's growth by some amount. Some argue that Lotka-Volterra models therefore should be totally discarded; others still use them as the basis for modeling multispecies interactions; but virtually all agree that it is preferable to use models based explicitly on the mechanisms of interaction, whenever possible. In this section we discuss two examples: competition for resources, and competition for space.

4.2.1 Competition for resources

Suppose that two species are both limited by some one resource that is in short supply, all other resources being relatively more abundant. We can model this mechanistically by adding resource abundance (R) as a state variable in the model. For example,

$$\begin{aligned} \dot{n}_1 &= \chi_1 f_1(R) n_1 - d_1 n_1 \\ \dot{n}_2 &= \chi_2 f_2(R) n_2 - d_2 n_2 \\ \dot{R} &= R_0 - f_2(R) n_1 - f_2(R) n_2 - \delta R \end{aligned} \tag{4.6}$$

Here d_1, d_2 are the per-capita death rates of the species, f_1, f_2 are their per-capita rates of nutrient uptake, and χ_1, χ_2 are the conversion rates between nutrient uptake and offspring production. The limiting is supplied at rate R_0 and degrades (or is lost) at rate δ. This model is *well-mixed* – we don't take account of spatial variability in resource availability or the density of the two species.

More species (e.g. grazers) or multiple nutrients can be included. Aquatic ecologists place great faith in such models because they have fared remarkably well in experimental tests using laboratory-scale

microcosms with plankton species (e.g., Tilman 1982, Fussmann et al. 2000). A classic result for the simple two-species, one-nutrient model is that you always have competitive exclusion:

> Let R_1^*, R_2^* be the nutrient levels at which each species can just barely survive ($\chi_i f_i(R_i^*) = d_i$). Then whichever species has the lower R^* drives the other to extinction.

Thus, coexistence of competitors must involve more than one limiting resource. Tilman, Matson, and Langer (1981) measured the R^*'s for silica of two species of freshwater diatoms (*Asterionella formosa, Synedra Ulna*) by growing each separately in a laboratory system corresponding to the model above with only one species at a time present, and waiting for it to reach equilibrium (at which point $R = R^*$ for the species). They found that *Synedra* had the lower R^*. Subsequent experiments confirmed that with both species in the system, regardless of the initial population densities, *Synedra* drove *Asterionella* to extinction. Tilman (1982) reports similarly good results for predictions of coexistence versus exclusion in systems with more than one limiting nutrient.

Exercise 4.2 (a) To the resource-competition model above, add a *toxin* – a chemical that each species releases (at a species-specific per-capita rate), and which has a detrimental effect on both of the species. Let L denote the toxin concentration (i.e. we assume that both species are releasing the same toxic substance) and assume that each species' death rate is an increasing function of L (though the impact of L is not necessarily the same on each species). Write out and explain a system of differential equations corresponding to these assumptions. (b) Do you think it might be possible for two species to coexist in the model with a toxin? Why, or why not?

4.2.2 Competition for space

Sessile or territorial organisms require a unit of space or substrate in order to complete their life cycle and reproduce. Population growth is then limited by the availability of such "sites". Suppose that there are N sites available (e.g. room for N adult barnacles on a stretch of rocky shoreline), and two species in a strict competitive hierarchy. [This example and the exercises are mostly taken from Hastings (1997)]

Species 1 is the top competitor, and as far as it is concerned there is no difference between an empty site and one occupied by species 2. Its dynamics are

$$\dot{n}_1 = M_1 n_1 (N - n_1) - e n_1$$

representing a balance between colonization of new sites (empty or occupied by species 2) and extinction of occupied sites. The colonization model implicit in this equation is that each established individual of species 1 has probability M_1 per unit time of having an offspring land in a given site, which it then occupies unless some other individual of species 1 is already in the site.

For species 2, a site occupied by species 1 is not available for colonization, so

$$\dot{n}_2 = M_2 n_2 (N - n_1 - n_2) - e n_2 - M_1 n_1 n_2$$

Here the last term represents the rate at which species 1 takes over sites occupied by species 2.

Exercise 4.3 Show how the model above can be rescaled into the form

$$\begin{aligned} \dot{x}_1 &= m_1 x_1(1 - x_1) - e x_1 \\ \dot{x}_2 &= m_2 x_2(1 - x_1 - x_2) - e x_2 - m_1 x_1 x_2 \end{aligned} \tag{4.7}$$

and give expressions for the x_i and m_i in terms of the original model variables and parameters. [Note: we could scale out one more parameter, but for interpreting the results of the following exercises it's better not to].

Exercise 4.4 Find the nonzero equilibrium for the dominant species by setting $\dot{x}_1 = 0$. What conditions on the parameters are necessary for the equilbrium to be positive? Give an intuitive interpretation.

Exercise 4.5 Assuming parameter values such that the dominant competitor can survive, find the nonzero equilibrium for species 2 by substituting the equilibrium value for x_1 into $\dot{x}_2$ and solving for the value of x_2 that makes $\dot{x}_2 = 0$. What must be true about the relative values of m_1 and m_2 for both species to persist? Does this make intuitive sense?

Exercise 4.6 Suppose that parameters are such that both species can coexist, but the extinction rate e is then slowly increased. Which species goes extinct first (i.e., which has its positive equilibrium become negative at the lower value of e)?

4.2.3 Competition for light

Jef Huisman and colleagues (see citations at the end of this chapter) have developed and tested a mechanistic theory of competition for light among light-limited plankton. Although similar in many ways to resource competition theory, these models also take account of the spatial variability of light, in particular its depth-dependence in the water column and how that is affected by the abundance of the competing species. **more needed here**.

4.3 Local stability analysis: continuous time

Nullclines don't always tell us the whole story. More often we need to use local stability analysis. For multi-species models this requires a bit of matrix and vector algebra, which is reviewed in the Appendix at the end of this chapter.

Local stability analysis for models with more than one state variable is based on a multivariate Taylor series. For two variables, the leading terms in the series are

$$f(x_1 + e_1, x_2 + e_2) = f(x_1, x_2) + e_1 \frac{\partial f}{\partial x_1} + e_2 \frac{\partial f}{\partial x_2} + \cdots \tag{4.8}$$

where the derivatives are evaluated at (x_1, x_2), and $\cdots$ indicates higher order terms in the deviations

e_1, e_2. Similarly with m state variables

$$f(x_1 + e_1, x_2 + e_2, \cdots, x_m + e_m) = f(x_1, x_2) + e_1 \frac{\partial f}{\partial x_1} + e_2 \frac{\partial f}{\partial x_2} + \cdots + e_m \frac{\partial f}{\partial x_m} + \cdots \tag{4.9}$$

We now apply this to a general multispecies model

$$\dot{n}_i = f_i(n_1, n_2, \cdots n_m) = f_i(n), \qquad i = 1, 2, \cdots, m \tag{4.10}$$

An equilibrium occurs at $\hat{n}$ where all $f_i(\hat{n}) = 0$.

Let $x_i = n_i - \hat{n}_i \qquad \Rightarrow \dot{x}_i = \dot{n}_i = f_i(n) = f_i(\hat{n} + x)$ We now use Taylor series to approximate $f_i(\hat{n} + x)$ using the fact that $f_i(\hat{n}) = 0$. The resulting equation for the local dynamics is

$$\dot{x}_i = \sum_j \frac{\partial f_i}{\partial n_j} x_j \tag{4.11}$$

where the derivatives are evaluated at $\hat{n}$. The sum in the last equation has the form of a matrix-vector multiplication. We therefore define the *Jacobian matrix* $\mathbf{J}$ by

$$\mathbf{J}[i, j] = \frac{\partial f_i}{\partial n_j} \text{ evaluated at } \hat{n}. \tag{4.12}$$

The local dynamics are therefore the linear system

$$\dot{x} = \mathbf{J}x. \tag{4.13}$$

How does this system behave?

Consider a linear system $\dot{x} = \mathbf{A}x$, with equilibrium $\hat{x} = 0$. In the 1-variable case ($\dot{x} = ax$) we get exponential solutions, $e^{at}x(0)$.

In the multivariable case, we get something similar from the eigenvectors and eigenvalues of the matrix $\mathbf{A}$. Recall the definition:

$$\lambda \text{ is an eigenvalue of } \mathbf{A}, \text{ and } w \neq 0 \text{ a corresponding eigenvector, if } \mathbf{A}w = \lambda w.$$

Note that eigenvectors are only defined up to constant: if w is an eigenvector for $\mathbf{A}$ so is cw for any constant $c \neq 0$. The requirement that $w \neq 0$ is important. Without it any number c would be an "eigenvalue" corresponding to $w = 0$, because $\mathbf{A}0 = c0 = 0$.

Eigenvectors are important because they give exponentially growing or decaying solutions, $x(t) = e^{\lambda t}w$. This can be demonstrated by direct substitution:

$$\begin{aligned} \dot{x} &= \lambda e^{\lambda t} w \\ \mathbf{A}x &= e^{\lambda t}\mathbf{A}w = e^{\lambda t}\lambda w = \lambda e^{\lambda t} w. \end{aligned} \tag{4.14}$$

The *typical* situation for a matrix $\mathbf{A}$ of size $k \times k$ is that it will have have k distinct eigenvalues and corresponding eigenvectors. Then the general solution to $\dot{x} = \mathbf{A}x$ is a sum of exponential solutions:

$$x(t) = c_1 e^{\lambda_1 t} w_1 + c_2 e^{\lambda_2 t} w_2 + \cdots + c_k e^{\lambda_k t} w_k. \tag{4.15}$$

The fact that (4.15) is a solution can be verified directly:

$$\begin{aligned}\dot{x}(t) &= c_1\lambda_1 e^{\lambda_1 t}w_1 + \cdots + \lambda_k c_k e^{\lambda_k t}w_k \\ &= c_1 e^{\lambda_1 t}\lambda_1 w_1 + \cdots + c_k e^{\lambda_k t}\lambda_k w_k \\ &= c_1 e^{\lambda_1 t}\mathbf{A}w_1 + \cdots + c_k e^{\lambda_k t}\mathbf{A}w_k \\ &= \mathbf{A}\left(c_1 e^{\lambda_1 t}w_1 + \cdots + c_k e^{\lambda_k t}w_k\right) \\ &= \mathbf{A}x(t)\end{aligned} \tag{4.16}$$

But for the stability of a fixed point (which is at $x = 0$ for linear systems) only one term in the sum really matters.

- **Example:** $x(t) = c_1 e^{2t}w_1 + c_2 e^{-3t}w_2$. The e^{-3t} term doesn't matter in the long run, since it goes to 0.
- **Example:** $x(t) = c_1 e^{2t}w_1 + c_2 e^{t}w_2$. The e^{t} terms doesn't matter in the long run, since the other term grows much faster ($e^t/e^{2t} \to 0$).

What if there are complex eigenvalues, $\lambda = \alpha + i\omega$? The general solution (4.15) still applies, but we need to work out its implications.

$$e^{\lambda t} = e^{\alpha t}e^{i\omega t} = e^{\alpha t}(\cos(\omega t) + i\sin(\omega t)) \tag{4.17}$$

The $(\cos + i\sin)$ term always has magnitude 1 (it rotates around the unit circle in the complex plane with period $2\pi/\omega$). Therefore the magnitude of $e^{\lambda t}$ is $e^{\alpha t}$, and the change in magnitude over time is accompanied by sinusoidal oscillations.

The conclusion is that *local stability of a fixed point for a system of differential equations depends on the eigenvalue of the Jacobian with largest real part:*

- Stable if all eigenvalues of the Jacobian have negative real part.
- Unstable if any eigenvalues of the Jacobian have positive real part.
- If the largest real part of any eigenvalue is exactly 0, the equilibrium could either be stable or unstable – local linearization is inconclusive.

To reach this conclusion we have been assuming that the eigenvalues of the Jacobian are distinct, but the same holds for a Jacobian with non-distinct eigenvalues (this can be proved using Jordan Canonical form, if you've had an abstract linear algebra class).

4.3.1 The 2 × 2 case

There is a useful formula for the eigenvalues of a 2×2 matrix $\mathbf{A}$. If T is the **trace** (sum of diagonal elements) and Δ is the **determinant** ($\mathbf{A}[1,1]\mathbf{A}[2,2] - \mathbf{A}[1,2]\mathbf{A}[2,1]$) then the eigenvalues are

$$\lambda_{1,2} = \frac{1}{2}\left(T \pm \sqrt{T^2 - 4\Delta}\right) \tag{4.18}$$

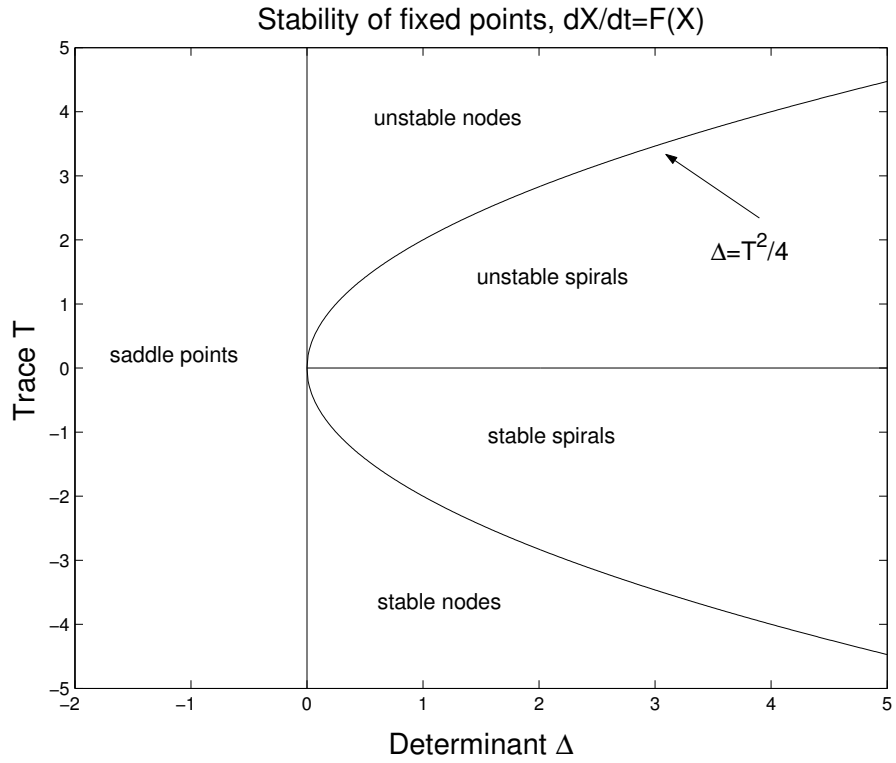


Figure 4.3: Stability analysis of 2-variable system $\dot{x} = F(x)$ in terms of the determinant D and trace T of the Jacobian matrix $\mathbf{J}$ at an equilibrium. At nodes and saddle points there are two real eigenvalues and therefore two eigenvectors; the equilibrium is a node if both eigenvalues have the same sign (negative or positive) and a saddle if they have opposite signs. At spirals the eigenvalues are complex conjugates, and solutions rotate.

This makes it possible to do a complete local stability analysis of two-dimensional differential equation systems in terms of the T and Δ (Figure 4.3). In consequence we obtain a simple criterion for local stability in the 2×2 case. In order for all eigenvalues of the Jacobian to have negative real part, we must have

$$\Delta > 0, T < 0. \tag{4.19}$$

The qualitative behavior of trajectories is also determined by the local analysis (see Figure 4.4). That Figure shows solution trajectories for the linear system. However, the Hartman-Grobman Theorem guarantees us that locally – within some small distance of the fixed point – solutions of the nonlinear system are qualitatively the same as those of the linear system, so long as none of the eigenvalues of the Jacobian has exactly zero real part.[1]

[1] A fixed points with this property is called *hyperbolic*. The precise meaning of "qualitatively the same" in the Hartman-Grobman Theorem is that the solution curves of the nonlinear system are the image of those for the linear system under some invertible continuous mapping.

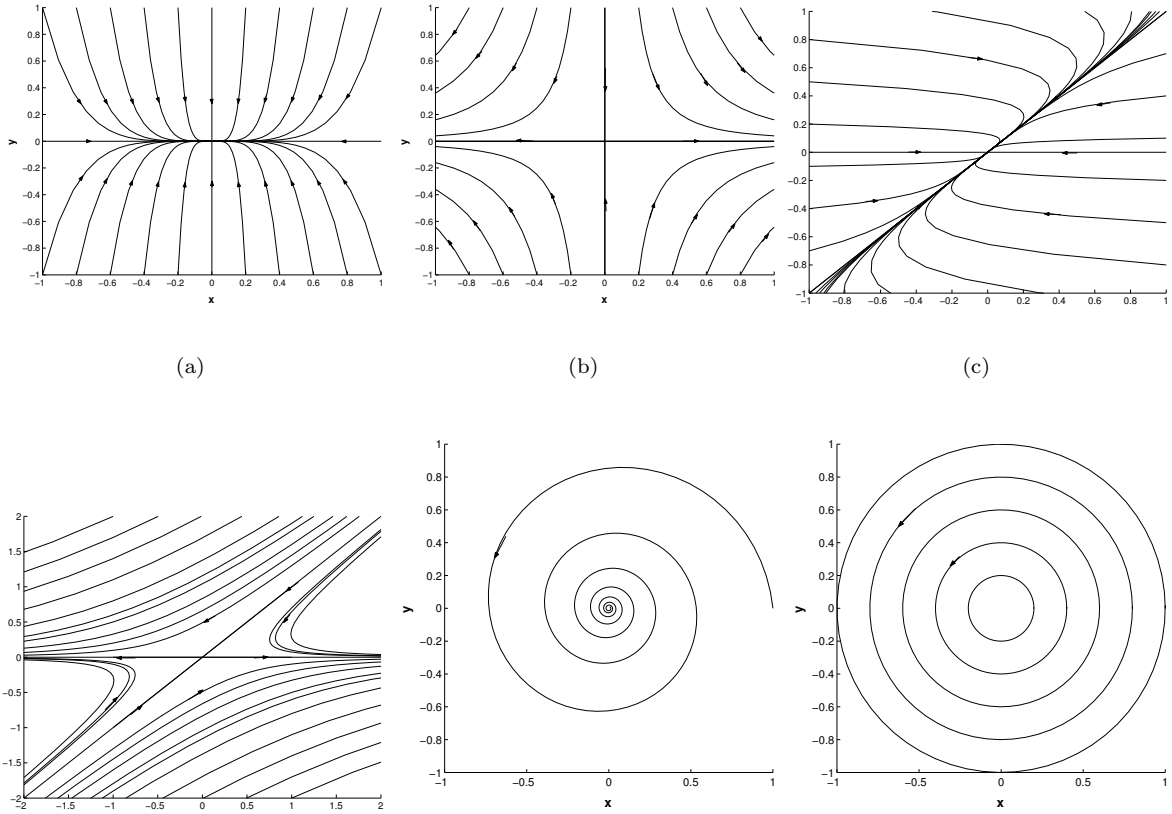


Figure 4.4: Phase portraits of two dimensional linear vector fields with a stable node (a and c) and a saddle (b and d). In (a) and (b) the eigenvectors are the coordinate axes, and in (c) and (d) the eigenvectors are not perpendicular. Panels (e) shows a stable spiral, and (f) a neutrally stable spiral where the real part of the eigenvalues is 0. An unstable node looks like a stable node, and an unstable spiral like a stable spiral, except that the arrows indicating the direction of flow point in the opposite direction. Figures provided by John Guckenheimer and incorporated without permission.

Example: The linear system $\dot{x} = \mathbf{A}x, \quad \mathbf{A} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$. To determine the behavior of solutions we can

- Write out system: $\dot{x}_1 =, \quad \dot{x}_2 =.$
- Find the eigenvalues using (7.21).
- Find the eigenvectors directly from $\mathbf{A}w_i = \lambda_i w_i$
- Plot eigenvectors and infer "phase portrait" (shape of solution curves)

Example: The Lotka Volterra system

$$\begin{aligned} \dot{n}_1 &= n_1(3 - 2n_1 - n_2) \\ \dot{n}_2 &= n_2(5 - n_1 - 4n_2) \end{aligned} \tag{4.20}$$

has interior fixed point $n_1 = n_2 = 1$. To determine stability we

- Write out f_1, f_2.
- Derive the general form of the Jacobian matrix.
- Evaluate the Jacobian at the fixed point $(1, 1)$.
- Compute the determinant and trace.
- Determine local stability from Figure (4.3).

Exercise 4.7 Use local stability analysis to verify the conclusions shown in Figure 3.3 (a) and (d) of Bulmer Chapter 3, for the stability of the interior equilibrium for the Lotka-Volterra competition equations. Note that Bulmer gives you a formula for the fixed point (equation 2, p. 31). So your tasks are to
(a) Compute the Jacobian matrix for the Lotka-Volterra model, in Bulmer's notation
(b) Evaluate the Jacobian at the interior fixed point with 3.3(a) parameters and show that the equilibrium is stable.
(c) Ditto with 3.3(d) parameters, and show that it is unstable.

Exercise 4.8 The following is a modified Lotka-Volterra model for two species interacting mutualistically

$$\begin{aligned} \dot{x} &= rx(1 - \frac{x}{K + ay}) \\ \dot{y} &= ry(1 - \frac{y}{K + bx}) \end{aligned} \tag{4.21}$$

(a) Show that if $ab < 1$ this model has an equilibrium $(\hat{x}, \hat{y})$ with $\hat{x} > 0$ and $\hat{y} > 0$. **(b)** Is the positive equilibrium from part (a) stable or unstable?
(c) Your conclusions in (a) and (b) did not depend on the values of r or K (if you got the right answers). Why?

Exercise 4.9 In this exercise, local stability analysis will be used to conjecture the global behavior of a "resource competition" model with one species present:

$$\begin{aligned} \dot{R} &= \delta(R_0 - R) - N\frac{VR}{K + R} \\ \dot{N} &= \chi N\frac{VR}{K + R} - \delta N \end{aligned} \tag{4.22}$$

with R_0, V, K, χ, δ positive constants. This represents a flow-through system (e.g. a pond or laboratory mesocosm with equal inflow and outflow) with δ being the flow rate (e.g., if 10% of the pond water is lost

to outflow each day and replaced by inflowing stream water, then $\delta = 0.1/d$). $\frac{VR}{K+R}$ is the per-individual rate at which organisms (N) take up the limiting nutrient (R), and χ is the conversion rate between nutrient uptake and birth rate.

(a) Show how the model can be re-scaled to dimensionless form

$$\begin{aligned} \dot{r} &= r_0 - r - x\frac{r}{1+r} \\ \dot{x} &= \varepsilon x\frac{r}{1+r} - x \end{aligned} \tag{4.23}$$

and give the expressions for its state variables (r, x) and parameters (r_0, ε) in terms of those of the original model.

(b) Show that there is always an equilibrium with $\hat{x} = 0$; what does this correspond to biologically [hint: this is easy].

(c) Use local stability analysis to find the conditions on parameters under which this first equilibrium is locally stable versus unstable.

(d) Show that for some (but not all) values of the parameters, there is a second equilibrium with $\hat{x} > 0$; in particular find this second equilibrium and the conditions under which it exists with $\hat{x} > 0$.

(e) Use local stability analysis to find the conditions on parameters under which this second equilibrium is locally stable versus unstable.

(f) Identify the type of the second equilibrium on the basis of Figure (4.3), as a function of parameter values (i.e. what possible types can it be, and for which range of parameters does each occur?).

(g) Put these pieces together to make a reasonable guess about what happens in general in this system, for different possible values of the parameters. That is: imagine an experiment in which the system is started at some values of $R(0), N(0)$ with $N(0) > 0$. What are the possible long-term outcomes, for which different values of the parameters?

4.4 Predator-prey interactions: functional response and the paradox of enrichment

The classic model for predator-prey dynamics is again the Lotka-Volterra model,

$$\begin{aligned} \dot{n}_1 &= r_1 n_1 - \alpha_1 n_1 n_2 \\ \dot{n}_2 &= \alpha_2 n_1 n_2 - r_2 n_2 \end{aligned} \tag{4.24}$$

This was proposed as a minimal model for fish stocks in the Adriatic. A fisheries biologist, Umberto D'Ancona, compared catch statistics before, during and after the first World War. Fishing in the Adriatic dropped nearly to nil during the war. D'Ancona observed that predatory fish (which feed on other fish) became more common during the War, relative to the periods just before and after the war. Prey fish species showed the reverse pattern. D'Ancona knew who to ask for an explanation: his father-in-law

Vita Volterra, who was (and remains – or at least his remains remain) a famous mathematician. Volterra developed the model (4.24) and showed that it could explain D'Ancona's observations (Volterra 1926).

Model (4.24) also makes some untenable predictions. In particular, the positive equilibrium is a neutrally stable center surrounded by a family of neutrally stable periodic orbits, qualitatively like the neutrally stable center of a linear system in Figure 4.4. Fortunately these go away when the model is made more realistic. The classic general form for an unstructured predator-prey model is

$$\begin{aligned} \dot{x} &= f(x) - yg(x) \\ \dot{y} &= eyg(x) - \mu y \end{aligned} \tag{4.25}$$

The change in the numbers of prey x is a balance between intrinsic births and deaths at net rate $f(x)$, and predation with $g(x)$ being the per-predator rate of prey capture. Predator y change is a balance between births (with breeding proportional to eating) and natural mortality. It would typically be assumed that f is logistic-like, so that prey in the absence of predators would reach a steady-state limited by their resource supply. Until very recently the assumption in (4.25) that g only depends on x was nearly universal – each predator in the model responds independently to the density of prey and is unaffected by other predators. The function $g(x)$ would be taken as monotonically increasing, and saturating. Typical forms are shown in Figure 4.5.

The classical form (4.25) obscures a crucial feature of any consumer-resource model – the prey and predator birth rates are *both* proportional to the *prey* abundance, when breeding is proportional to eating. To see why this holds, we need a different parsing of the terms in (4.25). The predator net birth rate $f(x)$ can be broken down as

$$f(x) = (\#\text{ prey}) \times (\text{prey per capita birth-death rate}) = x \times \frac{f(x)}{x}. \tag{4.26}$$

The capture rate $yg(x)$ can be broken down as

$$(\#\text{ prey}) \times (\text{probability of capture per unit time}) = x \times \frac{yg(x)}{x} \tag{4.27}$$

and the predator birth rate is the same multiplied by the conversion efficiency e. So $g(x)/x$ is the probability per unit time that any one given prey individual will be captured and consumed by any one given predator.

Model (4.25) is bare-bones at best, but it's a place to start. Making it more realistic might involve

- a time delay between eating and producing new offspring (Harrison 1995, Jost and Ellner 2000).
- mechanistic prey dynamics rather than a logistic-type equation.
- predator self-limitation.
- predator interference, replacing $g(x)$ with $g(x, y)$ that is a decreasing function of y.

The last of these is potentially the most important omission from the basic model. Recent studies suggest that in many cases – perhaps even most – the rate of consumption by each predator is a decreasing

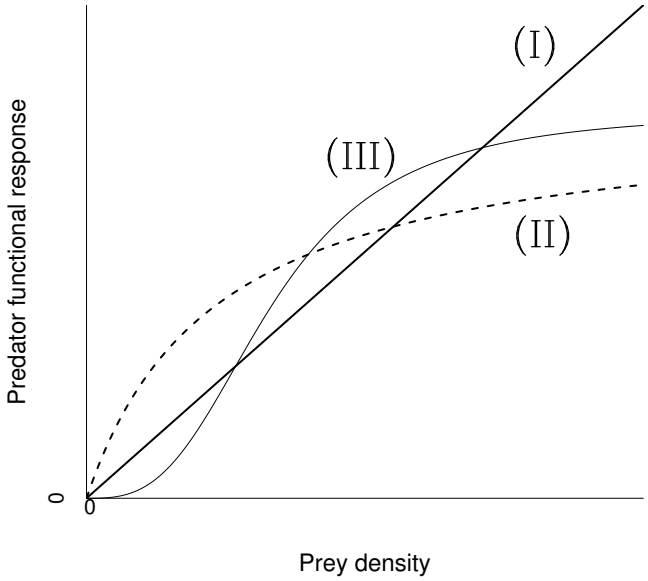


Figure 4.5: Predator functional response functions $g(x)$. The linear, saturating, and sigmoid forms are referred to as type I, II, and III respectively.

function of predator density, all else being equal. This has been found even in systems such as microbial microcosms where it is not at all clear how interference might occur (Skalski-Gilliam, Ginzburg-Abrams, Jost-Ellner).

Example: The *Rosenzweig-MacArthur model* uses logistic f and a type-II g. Note that type-II is the consumption rate equation we derived long ago (the prey model of optimal foraging theory) for a consumer hunting for a single kind of resource.

For analysis we consider the generic model (4.25) without predator interference. The typical qualitative shapes of the rate functions are shown in Figure 4.6. As usual we start by looking for steady states, via the nullclines. We have $\dot{y} = 0$ when $g(x) = \mu/e$ so the predator nullcline is the line

$$x = x^* \equiv g^{-1}(\mu/e). \tag{4.28}$$

The prey nullcline is the curve

$$y = f(x)/g(x) \tag{4.29}$$

which is well-defined except possibly at $x = 0$, depending on how f and g behave at the origin.

Do the nullclines intersect in the first quadrant – are there any equilibria? The predator nullcline is a vertical line at $x = x^*$. The prey nullcline is a curve that is positive for $x < K$ and negative for $x > K$. The nullclines therefore intersect in the first quadrant if and only if $x^* < K$. This condition can be interpreted biologically. K is the density that the prey will reach on their own, with predators

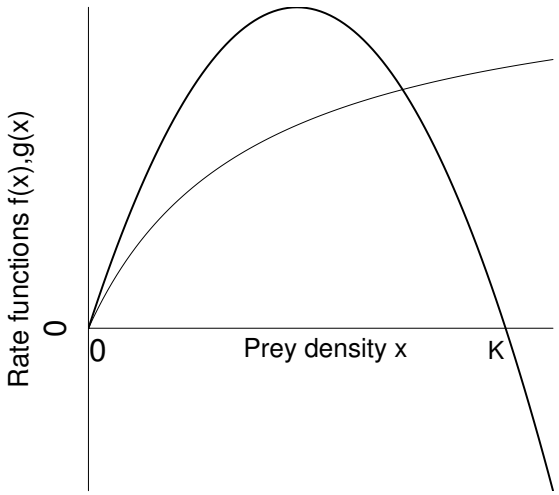


Figure 4.6: Rate functions for the basic predator-prey model. The bold curve is prey self-limitation $f(x)$, the light curve is the predator functional response $g(x)$. Though f is not necessarily logistic, we use K to denote the prey density at which f goes from positive to negative, representing a stable equilibrium for the prey in the absence of predators.

absent. x^* is the prey density at which predator births and deaths balance exactly – any lower and the predator population declines. So the relative values of x^* and K determine whether or not a few rare predators could increase, if they were introduced into a population of prey at steady state. If so, there's a steady state (x^*, y^*) with both prey and predators present at positive density. If not, then the only steady states are $(0, 0)$ and $(K, 0)$. And since the prey can never get above K (having predators around only makes things worse), the predators must decline and go to extinction. So in order for the prey and predators to coexist we must assume that $x^* > K$, and the prey and predator nullclines intersect at the steady state

$$x^* = g^{-1}(\mu/e), \quad y^* = f(x^*)/g(x^*) = ef(x^*)/\mu. \tag{4.30}$$

To determine the stability of the steady state we compute the Jacobian matrix of (4.25),

$$J = \begin{bmatrix} f'(x) - yg'(x) & -g(x) \\ eg'(x)y & eg(x) - \mu \end{bmatrix}. \tag{4.31}$$

At the steady state (x^*, y^*) this becomes

$$J^* = \begin{bmatrix} f'(x^*) - y^*g'(x^*) & -\mu/e \\ eg'(x^*)y^* & 0 \end{bmatrix}. \tag{4.32}$$

We therefore have determinant and trace

$$\begin{aligned} \Delta &= \mu g'(x^*) y^* > 0 \\ T &= f'(x^*) - y^* g'(x^*) \end{aligned} \tag{4.33}$$

The trace is positive (implying instability of the fixed point) if

$$f'(x^*) > y^* g'(x^*);$$

what does that mean? Since $y^* = y^* = f(x^*)/g(x^*)$, at (x^*, y^*) we have

$$\begin{aligned} T = f' - \frac{f}{g} g' = \frac{f'g - g'f}{g} &= g\left(\frac{f'g - g'f}{g^2}\right) \\ &= g \times \frac{d}{dx}\left(\frac{f}{g}\right) \end{aligned} \tag{4.34}$$

and $y = f(x)/g(x)$ is the prey nullcline. So *the trace has the same sign as the slope of the prey nullcline*, and the steady state undergoes a Hopf bifurcation whenever (as parameters change) the slope of the prey nullcline goes from negative to positive.

The behavior of f/g therefore determines whether the steady state is stable or unstable. We know f/g goes from positive when x is small to negative when $x > K$ – the question is: how does it get there? If f/g goes down monotonically, the steady state is always stable. If it isn't monotonic we have the potential for cycles.

Cycling versus stability therefore depends crucially on the form of the predator functional response. With a type-I functional response (Figure 4.5) $g(x) = ax$, the prey nullcline is proportional to $f(x)/x$, which is the per-capita growth rate of the prey. In order to have cycles, this must have a maximum at some intermediate value of x, rather than decreasing monotonically at higher prey density. That is, there has to be an Allee effect in the prey, otherwise the fixed point is always stable.

With a type-II functional response the prey nullcline can have a hump even without an Allee effect in the prey. For this reason it is often said that "a type-II functional response is destabilizing". For example, consider the Rosenzweig-MacArthur model

$$\begin{aligned} f(x) &= rx(1 - x/K) \\ g(x) &= ax/(1 + bx) \end{aligned}$$

for which the prey nullcline is

$$\frac{f(x)}{g(x)} = \frac{rx(1 - x/K)(1 + bx)}{ax} \propto (1 + bx)\left(1 - \frac{x}{K}\right). \tag{4.35}$$

The prey nullcline will have a hump if its derivative at $x = 0$ is positive. This will be proportional to the derivative at 0 of the last expression in (4.35), which is $b - \frac{1}{K}$. The prey nullcline therefore has a hump (and cycles are possible) if $bK > 1$. If bK there is no hump and the fixed point is stable. A bit of calculus shows that the hump in the prey nullcline is at

$$x_h = \frac{bK - 1}{2b}. \tag{4.36}$$

So an increase in K moves the hump in the prey nullcline further and further to the right. However, it has no effect on the predator nullcline, which remains at $x^* = g^{-1}(\mu/e)$ So if K keeps increasing, eventually $x_h > x^*$ and the fixed point goes from unstable to stable via a Hopf bifurcation. This behavior is known as the *paradox of enrichment*. Enrichment of the system's resource base – reflected in a higher prey carrying capacity – is beneficial up to a point, but eventually it destabilizes the system.

Another interpretation of this result is that it predicts a *suppression-stability* tradeoff (Murdoch et al. 2003): cycles occur if $K \gg x^*$. That is, if the level to which the predator can suppress the prey (at steady state) is far below the intrinsic carrying capacity of the prey, then cycles are predicted. This prediction is especially interesting because there are many striking counterexamples: natural populations that are held stably by predators at numbers far lower than they can reach when predators are absent (Murdoch et al. 2003). It also has the interesting interpretation that *biological control of pest species should be impossible:* any control agent that can reduce pest numbers appreciably should be unable to do so permanently, instead there should be periodic outbreaks of the pest. Biological control has had mixed success, and perhaps this is one reason for it – but biological control has also had some successes, which are therefore hard to explain. We'll come back to this later in this chapter.

The form of the Jacobian (4.31) explains why consumer-resource interactions and other exploiter-victim interactions are generally prone to cycle. The nature of the interaction means that the resource abundance has a strong positive impact on the consumer, while the consumer abundance has a strong negative impact on the resource species. We therefore get Jacobians of the form

$$J = \begin{bmatrix} r & + \\ - & c \end{bmatrix} \tag{4.37}$$

where r and c represent the self-regulatory effects of resource and consumer species density on their own species' growth. If the within-species effects are weak compared to the trophic interaction, we have

$$\begin{aligned} \Delta &= (\text{small})(\text{small}) - (+)(-) > 0 \\ T &= \text{ small} + \text{small } \approx 0 \end{aligned} \tag{4.38}$$

A fixed point for the system is therefore sitting near the boundary between stable and unstable spirals – that is, between damped and undamped oscillations around the fixed point. Even if the fixed point is stable, random perturbations (weather, accidents of demography in finite populations, and so on) will continually kick it away from the fixed point, so the observed dynamics would often be somewhat cyclic.

Exercise 4.10 Here is a more general consideration of how a type-II functional response is destabilizing relative to a type-I. As we noted above, the prey nullcline must have a hump if the curve $y = f(x)/g(x)$ has positive slope at $x = 0$. Show that if $g'(0) > 0$ this slope has the same sign as

$$g'(0)f''(0) - f'(0)g''(0).$$

(Note: this will require two applications of L'Hopital's rule to resolve 0/0 limits.) So all else being equal, a negative second derivative in the predator functional response g will tend to increase the slope at 0 and favor a hump in the prey nullcline.

Exercise 4.11 What happens with a type-III (accelerating) functional response by the predator – Is it stabilizing or destabilizing relative to a type-I (linear) functional response? A typical type-III response equation would be $g(x) = ax^2/(1+bx^2)$, where x is the prey density. This eventually saturates for large x, and before that starts looking like a type-II (concave down rather than up). So to see what can be "new and different" in type-III, for this exercise consider instead $g(x) = ax^2$.

4.5 Discrete time models

4.5.1 Nicholson-Bailey model

The prime example of an ecologically interesting discrete-time model for interacting populations is the Nicholson-Bailey model for host-parasitoid dynamics. Parasitoids are insect species (some terrestrial, some extraterrestrial as in the *Aliens* movies) whose larvae develop by feeding on the bodies of other arthropods, usually killing them. Larvae emerge from the host and develop into free-living adults. The adults then lay their eggs in a subsequent generation of hosts. Most parasitoid larvae require a specific life-stage of the host, so parasitoid and host generations are linked to one another. Consequently host-parasitoid models often use a discrete time step corresponding to the common generation length of host and parasitoid.

Every important idea in theoretical ecology has been applied to the study of host-parasitoid dynamics, for two main reasons:

- Practical importance. Biological control of insect pests – as a part of integrated pest management aimed at reducing chemical pesticide use – is largely through introduction of parasitoids. The goal is to produce low, stable numbers of the pest.
- Theoretical interest. There are a lot of things about real host-parasitoid systems that simple models cannot explain. So it becomes an arena for adding all sorts of complications - spatial dynamics, metapopulations, stage structure, foraging theory - to see what can happen.

The classic model was derived by Nicholson and Bailey (1935). There were hoping to publish a book on the subject, but their proposal was rejected due to the expectation that ecologists would not be interested in a book full of equations. Nicholson and Bailey's only real problem was being 30 years ahead of their time, and ecology might be a very different field today had their book been published. The assumptions of the Nicholson-Bailey model are as follows:

- Hosts are distributed at random, at density H_t per unit area in generation t
- Parasitoids search at random and independently, each having an "area of discovery" a, and lay an egg in each host found.
- Each parasitized host gives rise to 1 new parasitoid in generation $t+1$
- Each unparasitized host gives rise to $R > 1$ new hosts in generation $t+1$

Each parasitoid attacks the hosts found in a units of area, so the expected number of hosts attacked is by each parasitoid is aH. The expected total number of attacks is aPH. The total number of attacks can also be written as the sum over hosts of the number of attacks on each host. All hosts have the same expected number of attacks, so the expected number of attacks on any given host must be aP. Under the assumptions listed above, the number of eggs per host has a Poisson distribution (see the Appendix at the end of this chapter). Consequently, the expected fraction of hosts that are not parasitized is the probability that a Poisson random variable with mean aP takes the value zero, which $e^(- aP)$. The resulting population dynamics are

$$\begin{aligned} H_{t+1} &= RH_t e^{-aP_t} \\ P_{t+1} &= H_t \left(1 - e^{-aP_t}\right) \end{aligned} \tag{4.39}$$

4.5.2 Local stability analysis

To see how the Nicholson-Bailey model behaves, we (once again) start by finding fixed points and analyzing their stability. Consider a general multispecies difference equation model

$$n_i(t+1) = F_i(n_1(t), n_2(t), \cdots, n_k(t)), i = 1, 2, \cdots, k \tag{4.40}$$

which we can write compactly as

$$n(t+1) = F(n(t)). \tag{4.41}$$

A fixed point of the model is then a solution of

$$\hat{n} = F(\hat{n}). \tag{4.42}$$

To study the dynamics near a fixed point we (once again) use linearization. Writing $p(t) = n(t) - \hat{n}$ and proceeding as we did above for a differential equation model, the local dynamics are approximated by

$$p(t+1) = \mathbf{J}p(t) \tag{4.43}$$

where $\mathbf{J}$ is again the Jacobian matrix for F evaluated at the fixed point, i.e. the $(i,j)^{th}$ element of $\mathbf{J}$ is $\frac{\partial F_i}{\partial n_j}$ evaluated at $\hat{n}$. In the *typical* case where the eigenvalues are all distinct, the general solution to (4.43) will (once again) be a sum of exponential terms, but these are not quite the same as for a differential equation model:

$$p(t) = c_1\lambda_1^t w_1 + c_2\lambda_2^t w_2 + \cdots + c_k\lambda_k^t w_k. \tag{4.44}$$

The conclusion from (4.44) is that the fixed point is locally stable if all eigenvalues of the Jacobian are less than 1 in absolute value, and is locally unstable if any eigenvalues are greater than 1 in absolute value – recall that for for a complex eigenvalue $\lambda = a + ib$ the absolute value is

$$|\lambda| = \sqrt{a^2 + b^2}. \tag{4.45}$$

As for differential equation models, the local stability criterion also applies when there are repeated eigenvalues.

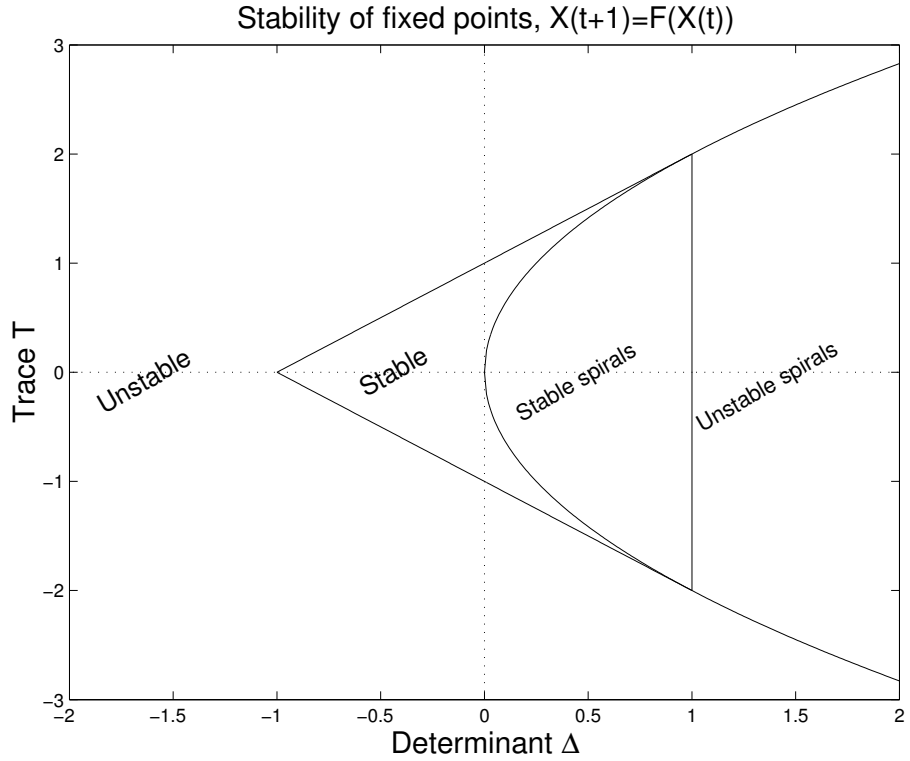


Figure 4.7: Stability analysis of 2-variable system $X_{t+1} = F(X_t)$ in terms of the determinant Δ and trace T of the Jacobian matrix $\mathbf{J}$ at an equilibrium. At spirals the eigenvalues are complex conjugates, and solutions rotate. Otherwise both eigenvalues are real; if $\Delta > 0$ they have the same sign (equal to that of the Trace), and if $\Delta < 0$ they have opposite signs.

For 2×2 systems we still have the eigenvalue formula

$$\lambda_{1,2} = \frac{1}{2}\left(T \pm \sqrt{T^2 - 4\Delta}\right) \tag{4.46}$$

where T is the trace and Δ the determinant of the Jacobian matrix at a fixed point. Using this, the condition for local stability of a fixed point (both eigenvalues < 1 in absolute value) is

$$|T| < 1 + \Delta < 2 \tag{4.47}$$

The curve $\Delta = T^2/4$ separates real versus complex eigenvalues, so as in the continuous time-case we get spirals when $4\Delta > T^2$. These results are summarized in Figure 4.7.

4.5.3 Analysis of Nicholson-Bailey

First we need to find steady states $(\hat{H}, \hat{P})$. The host equation gives use

$$\hat{H} = R\hat{H}e^{-a\hat{P}} \Rightarrow Re^{-a\hat{P}} = 1 \Rightarrow \hat{P} = \log(R)/a. \tag{4.48}$$

Then since $e^{-a\hat{P}} = 1/R$ the parasitoid equation gives

$$\hat{H} = \frac{R}{R-1}\hat{P} = \frac{R}{R-1}\log(R)/a. \tag{4.49}$$

We therefore have a positive steady state so long as $R > 1$.

Next we need the Jacobian. For the NB the component maps are

$$F_1(H,P) = RHe^{-aP} \qquad F_2(H,P) = H\left(1 - e^{-aP}\right).$$

Taking the necessary derivatives and simplifying, we find that the Jacobian at the fixed point is

$$J = \begin{bmatrix} 1 & -a\hat{H} \\ \frac{R-1}{R} & \frac{a\hat{H}}{R}. \end{bmatrix} \tag{4.50}$$

The determinant of the Jacobian is

$$\Delta = \frac{a\hat{H}}{R} + \frac{a\hat{H}(R-1)}{R} = a\hat{H} = \frac{R\log R}{R-1}. \tag{4.51}$$

A numerical plot shows that $\frac{R\log R}{R-1}$ is an increasing function of R for $R > 1$, and converges towards 1 as R decreases to 1. The fact that this quantity is always > 1 is equivalent to

$$q(r) \equiv R\log R - R + 1$$

always being positive for $R > 1$ – which is true because $q(1) = 0$ and $q'(R) = 1 + \log R - 1 > 0$ for all $R > 1$. Consequently, a fixed point of the Nicholson-Bailey model is always unstable because $1 + \Delta > 2$ whenever the fixed point exists. Numerical solutions show that the actual dynamics are highly unstable (Figure 4.8). Oscillations diverge away from the fixed point, and lead to extreme outbreaks and crashes that could not be sustained in the real world.

Exercise 4.12 Show that the eigenvalues of the Jacobian for the Nicholson-Bailey model's steady state are always complex, by showing that $T^2 - 4\Delta < 0$ for all $R > 1$.

Exercise 4.13 (a) Show how the Nicholson-Bailey model can be rescaled to eliminate the parameter a. (b) Use numerical simulation to study the behavior of the Nicholson-Bailey model for different values of R.

4.5.4 Stabilizing host-parasitoid models

The stability of actual host-parasitoid systems – contrasted with the extreme instability of the Nicholson-Bailey model – is therefore something of a mystery. As modelers we can modify the model until it stabilizes. Theoreticians have come up with a number of different modifications that can achieve stability, that fall mainly in two categories:

- Additional forms of density dependence, such as parasitoid interference (with the per-parasitoid search area a replaced by some some decreasing function of parasitoid density) or intrinsic self-limitation in the host.

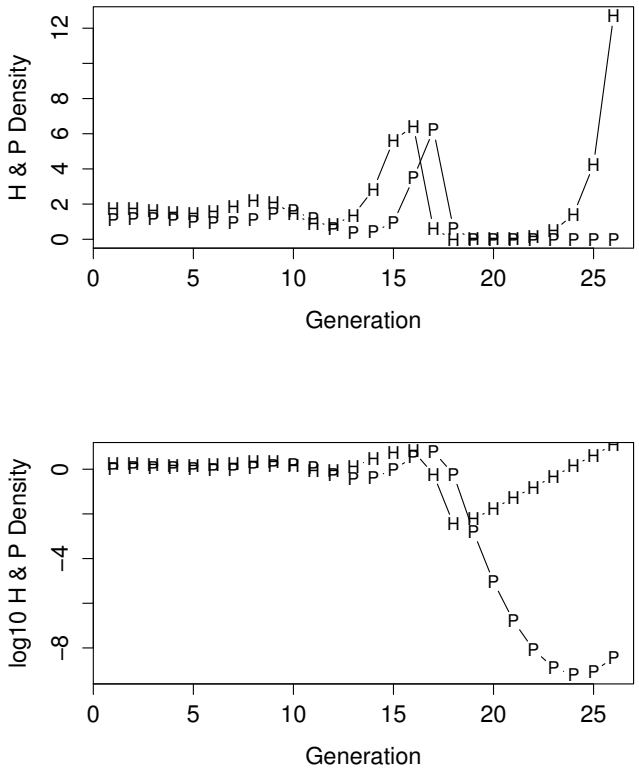


Figure 4.8: Dynamics of the Nicholson-Bailey model with $R = 3$.

- Spatial population dynamics, ranging from simple (a host refuge) to baroque (spatial chaos in systems of coupled host-parasitoid models).

The existence of theoretical solutions is not satisfying without evidence to justify the changes, and that has proved surprisingly difficult. Spatial dynamics has been the main focus of work, which began in the 1970s. May (1978) proposed a phenomenological spatial model, based on assuming that hosts are distributed evenly into a set of discrete patches, and are attacked only by the parasitoids in that patch. If parasitoids are distributed among patches independently, the Nicholson-Bailey model results. But if parasitoids aggregate (so patches tend to either have a lot of parasitoids, or just a few), more stable dynamics can result. One influential example assumed a gamma distribution for the parasitoid density in each patch, which led to a negative binomial distribution for the number of attacks per host. May (1978) showed that in that case one could write a pair of simple equations for the dynamics of the total host and parasite densities in the system of patches:

$$\begin{aligned} H_{t+1} &= RH_t(1 + aP_t/k)^{-k} \\ P_{t+1} &= H_t(1 - (1 + aP_t/k)^{-k}) \end{aligned} \tag{4.52}$$

where k is equal to the inverse of the squared coefficient of the number of parasitoids per patch, $k = 1/CV_p^2$. Note that as $k \to \infty$ this converges to the Nicholson-Bailey model (using the result from calculus that $(1 + x/n)^n \to e^x$).

Stability analysis of this model (May 1978) yielded a very simple result: the interior fixed point is stable if $k < 1$. The condition for stability is thus

$$CV_p^2 > 1. \tag{4.53}$$

This equation then proceeded to take on a life of its own. It was initially interpreted as saying that stability would ensue if parasitoids aggregate on patches with the most hosts. Eventually it was realized that **any** kind of aggregation would be stabilizing – for example, if parasitoids aggregate on patches with the fewest hosts. The key to stability in this model is that some hosts are at relatively high risk of attack, others at relatively low risk.

By the early 1990s, theoreticians had shown that the $CV_p^2 > 1$ rule applied, at least approximately, to a wide range of models. Empirical ecologists took advantage of the opportunity this created – to publish papers just by counting bugs. The punchline, unfortunately, was "sorry, try again". Reviewing 34 data sets with adequate data, Taylor (1993) found only 9 with $CV_p^2 > 1$, so as a general explanation for persistence of host-parasitoid systems it fails.

In the models considered so far space is *implicit* - the model specifies how parasitoids are distributed across patches, but not how they got there. Theoreticians also considered explicitly spatial models. For example, consider a regular rectangular lattice of patches, with Nicholson-Bailey dynamics within the patch and symmetric dispersal among neighboring patches (i=patch index, t=time):

$$\begin{aligned} H^*_{i,t} &= RH_{i,t}e^{-aP_{i,t}} \\ P^*_{i,t} &= H_t(1 - e^{aP_{i,t}}) \\ H_{i,t+1} &= (1 - \mu_H)H^*_{i,t} + \mu_H \bar{H}_{i,t} \\ P_{i,t+1} &= (1 - \mu_P)P^*_{i,t} + \mu_P \bar{P}_{i,t} \end{aligned} \tag{4.54}$$

where $\bar{H}_i, \bar{P}_i$ are the averages of H^*, P^* over the 8 patches surrounding patch i(Hassell et al. 1991, Comins et al. 1992). This model was found to exhibit coexistence either as a "crystal lattice" (an unchanging pattern of spatially varying densities, which occurs when μH is low and μ_P is high), spiral waves, or spatial chaos (Figure 4.9). In the latter two cases coexistence occurs through hosts dispersing into relatively unoccupied patches, and being consumed by a following wave of parasitoids which then dies out – but not before some hosts have moved on. Both of these mechanisms depend on the within-patch dynamics being unstable. *Local* instability generates *large-scale persistence* – host and parasite are both abundant at all times, but in different places.

Unfortunately, explicitly spatial modeling did not provide a refuge from the conflict with empirical studies. For cases where coexistence occurs in the explicitly spatial models, the resulting spatial distribution of parasitoids satisfies the condition derived in the phenomenological models: $CV_p^2 > 1$. So the empirical evidence against the phenomenological models applied equally well to explicitly spatial models – not to

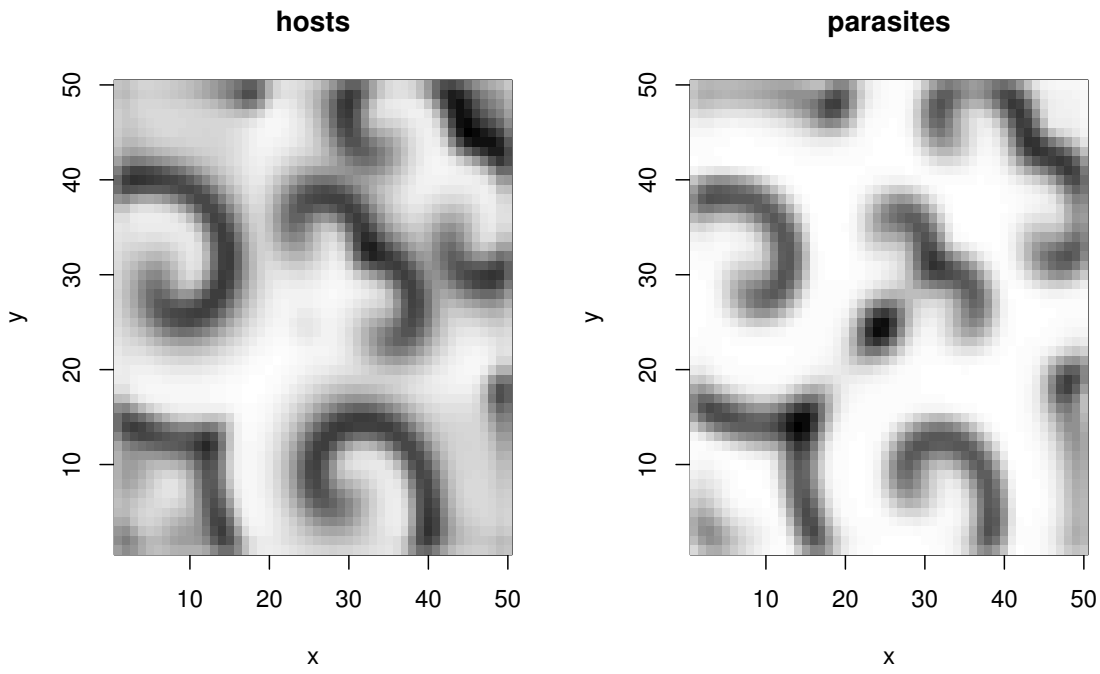


Figure 4.9: Spiral waves produced by the spatial Nicholson-Bailey model (4.54). Darkness is proportional to the square root of host and parasitoid density. This simulation had $a = 1, R = 2, \mu_H = \mu_P = 0.8$ on a 50×50 square lattice with absorbing boundaries – individuals who migrate off the lattice are lost from the population

.

mention the fact that spiral waves or spatial chaos have not yet been observed in any real host-parasitoid system (Rohani et al. 1997). Undaunted, theoreticians went on to consider 3-species systems (e.g., two parasitoids on one host) in the same kind of lattice system (Comins and Hassell 1996, with the admission that "theory is far ahead of experiment or observation").

Exercise 4.14 : (based on exercise 9 in Bulmer Ch. 3). Consider the Nicholson-Bailey model proposed by Hassell and Varley, with interference between parasitoids such that the search area per parasitoid is inversely proportional to $\sqrt{P_t}$:

$$\begin{aligned} H_{t+1} &= RH_t e^{-a\sqrt{P_t}} \\ P_{t+1} &= H_t(1 - e^{-a\sqrt{P_t}}) \end{aligned} \tag{4.55}$$

(a) Show how the model can be rescaled to eliminate a (i.e. so that the equations look just like the ones above, but with $a = 1$).
(b) Show that for $R > 1$ this model has an equilibrium $(\bar{H}, \bar{P})$ with $\bar{H} > 0, \bar{P} > 0$.
(c) Show that this internal equilibrium is stable for $R < 4.92$, by numerically evaluating and plotting the absolute value of the dominant eigenvalue of the Jacobian at $(\bar{H}, \bar{P})$ as a function of R. The ideal

solution for this problem would consist of three parts: (1) a derivation of the Jacobian matrix at $(\bar{H}, \bar{P})$, (2) a script file that evaluate the dominant eigenvalue as a function of R, and (3) a plot of results from the script file.
(d) Propose a more general version of this model, and derive a prediction about it using methods like those above. For example, your more general model will have at least one more parameter. What effect does that parameter have on existence and stability of the internal equilibrium?

Exercise 4.15 We consider here a simple discrete-time host-pathogen model for spread of an infectious disease with permanent immunity. The time-step is equal to the duration of the infectious phase. The state variables are S_t, the number of susceptible individuals at time t, and C_t representing the number of individuals getting the disease (new cases) between times $t-1$ and t.

$$\begin{aligned} C_{t+1} &= S_t(1 - e^{-\beta C_t}) \\ S_{t+1} &= S_t + B - C_{t+1} = S_t e^{-\beta C_t} + B \end{aligned} \tag{4.56}$$

Here B is the number of births between t and $t+1$, all added to the susceptible class and assumed to be constant over time. So the second line of the model is just "conservation of mass" for the Susceptible class. The first line is like Nicholson-Bailey: it comes from assuming that each susceptible escapes infection with probability $e^{-\beta C_t}$ – the more infectives there are, the lower the chance of escape. The model ignores mortality in the susceptible class, on the assumption that everyone gets the disease while young and mortality occurs later in life.
(a) Analyze this model with regard to existence and local stability of steady states, and use simulations to check your conclusions.

(b) Re-interpret the results of (a) for host-parasitoid dynamics.

(c) Generalize your analysis to the case where there is host mortality at rate $d > 0$, and the susceptible dynamics become $S_{t+1} = (1-d)S_t + B - C_t$.

4.6 Understanding stability: red scale

Murdoch (1994) reviewed efforts over 15 years to understand stability of the interaction between California red scale (host) and its introduced parasitoid *Aphytis melitus*. Red scale is a citrus pest that has been successfully controlled by the parasitoid. It has been held for decades at about a 99% reduction relative to its pre-control density, and is one of the least variable insect populations on record (Murdoch 1994). There is no evidence of local extinctions and recolonizations (as occur in "spatial chaos" kinds of models): each local population seems to be stable on its own.

Over (by now) 2 decades of study, Murdoch and collaborators "tested and failed to find evidence for 8 hypotheses that might account for the system's stability" (Murdoch et al. 1994).

1. Density dependent parasitoid aggregation.

2. Density independent parasitoid aggregation.

3. Density dependence in the parasitism rate (e.g. parasitoid interference as in the Hassell-Varley model above)

4. Density-dependent sex ratio in the parasitoid.

5. Refuge for the host: if the interior of the tree is relatively safe for hosts, some of whom could "leak" out and stabilize the exterior population in the foliage. There was a significant refuge (over 90% of reproductive female red scale were in the refuge), but in experiments where the refuge population was eliminated, the variability of the exterior population did not go up.

6. Spatial population dynamics of various sorts (no: isolating individual trees from one another did not increase the variabilty of populations).

7. Density dependence in the host (an odd hypothesis, since the host is kept at such low levels).

8. Limitation of the host by predators (also odd, since predators would be expected not to focus on such a rare species).

So what else might it be? Murdoch (1994) considered some other explanations based on stage structure in the two populations.

1. Adult hosts are invulnerable to attack.

2. Host feeding: in small host individuals the parasitoid lays male eggs (or eats the host instead), in large hosts it lays female eggs. For this model, estimated parameters are very close to the stability border.

3. Cannibalism by the parasitoid, as a result of eating small hosts that have already been parasitized.

These ideas were subsequently tested, and the outcome is summarized by Murdoch et al. (2006). A detailed model for the red scale system identified two factors as crucial for stability:

1. The presence of invulnerable stages in the host life cycle.

2. The maturation time of the parasitoid is shorter than that of the host.

These mechanisms depend on the specifics of this system, and would not provide a general explanation for the persistence of host-parasitoid interactions. However, Murdoch et al. (2006) tabulate evidence that the cases where biological control by a single natural enemy has maintained stable reductions in coccid insects besides red scale all have the two crucial features: invulnerable stages and rapid maturation of the enemy.

The lack of a simple general explanation for host-parasitoid stable coexistence has not decreased interest in host-parasitoid dynamics (see Hochberg and Ives 2000), but it means that theory is now dominated by more complicated models.

Another important change in direction resulted from a paper by Murdoch and Stewart-Oaten (1989) that questioned the emphasis on parasitoid aggregation that started with May (1978). MSO noted that the spatial extensions of the Nicholson-Bailey framework all make the unrealistic assumption that parasitoids only move once per generation. As an opposite extreme, MSO considered a continuous-time version of the Nicholson-Bailey model where parasitoids could re-aggregate continuously across a set of patches differing in host density. They discovered that

- Parasitoid aggregation *per se* has no effect whatsoever on stability. What matters is how parasitoids respond to spatial variability in host density, and its consequences for how the degree of parasitoid aggregation varies with mean host density.
- Aggregation of parasitoids in response to host density can be either stabilizing or destabilizing, depending on the nature of the response.

This says that any static measure of spatial variability such as CV_p^2 is irrelevant. What matters for stability is the variability is generated by the behavioral decisions of hosts and parasitoids. On the other hand, the MSO model's assumption of "ideal free parasitoids" who redistribute themselves instantaneously (based on perfect knowledge of the host distribution) is no more realistic than the alternative of once-per-generation redistribution. As a result, current models for spatial host-parasitoid dynamics also tend to involve many details of host and parasitoid life-cycles, dispersal and foraging behavior.

Exercise 4.16 This and the following exercises develop the model of Murdoch and Stewart-Oaten (1989) for a host-parasitoid system with continuous attack rather than discrete generations, and continuous redistribution of parasitoids based on host density. For a single patch, in continuous time the Nicholson-Bailey assumptions lead to the model

$$\frac{dH}{dt} = aH - bHP \qquad dPdt = cHP - dP. \tag{4.57}$$

with $a, b, c, d > 0$ (there is actually one additional assumption: parasitized hosts instantly die and produce new parasitoids). This is a Lotka-Volterra predator-prey model, whose fixed point (when it exists) is always neutrally stable. Now consider a multi-patch system, and let h and p denote the average host and parasitoid densities across patches. Show that

$$\frac{dh}{dt} = ah - b(hp + Cov(H, P)) \qquad \frac{dp}{dt} = c(hp + Cov(H, P)) - dp \tag{4.58}$$

where $Cov(H, P) = E(HP) - E(H)E(P)$ and E denotes the average across patches. So if parasitoid aggregation is independent of host density (implying that $Cov(H, P) = 0$), it has no effect on stability: the spatial system is identical to the one-patch system, so the value of CV_p^2 has no impact on stability.

Exercise 4.17 To go further we need an expression for $Cov(H, P)$ in terms of h and p. Murdoch and Stewart-Oaten assumed that parasitoids respond instantly and omnisciently to host distribution such that $P_i = p(1 + G(H_i, h))$, where the subscript i indicates densities in patch i. They also assume that the spatial distribution of hosts depends only on the mean density h (e.g., a Poisson distribution). This gives

$$E(HP) = E(Hp(1 + G(H, h)) = hp + pE(HG(H, h)) = hp + pQ(h) \tag{4.59}$$

where $Q(h) = E(HG(H,h))$ – this implies $Cov(H,P) = pQ(h)$. The spatial model is then

$$\frac{dh}{dt} = ah - bp(h + Q(h)) \qquad dpdt = cp(h + Q(h)) - dp. \tag{4.60}$$

Show that a positive steady state $(\hat{h}, \hat{p})$ is locally stable if $Q'(\hat{h}) > Q(\hat{h})/\hat{h}$ and locally unstable if the reverse inequality holds. Interpret this condition in terms of the shape of $Q(h)$ and of the predator functional response $g(h) = h + Q(h)$ (note that $Q'(h) > Q(h)/h$ is equivalent to $g'(h) > g(h)/h$). It is safe to assume that $g(0) = 0$ and that g is monotonically increasing.

Exercise 4.18 Suppose that there is density-dependent parasitoid aggregation, with a linear response to relative host density: $G(H,h) = m(H-h)/h$ where m is a constant. Suppose further that $Var(H) = Ah^x$ for $A > 0$ and some $x > 0$ – this form for the relationship between the mean and variance of density across habitat patches is called *Taylor's power law*. Show that the steady state may be either stable or unstable, depending on the value of x. The conclusion is that if parasitoids redistribute continuously in response to host density, density-dependent parasitoid aggregation can either be stabilizing or destabilizing.

4.7 Understanding cycles: larch budmoth

To end on a more upbeat note, we echo the theme of Zimmer (1999): the unreasonable *quantitative* success of relatively simple nonlinear models that 25 years ago would have been defended as "metaphors" for real ecological systems and tools for exploring ideas. As a case in point we consider the population cycles of larch budmoth, based on work of Turchin and collaborators (Turchin et al. 2002, 2003; Turchin 2003).

The larch budmoth is a classic population cycle. Budmoth density varies by nearly 5 orders of magnitude, with remarkably constant period and amplitude (the data are graphed in the previous chapter). For the last few decades the predominant explanation, and consequently the focus of ongoing research, has been the interaction between budmoth and its food supply, larch needles. The cycle is interpreted as a consumer-resource interaction with budmoth the consumer and larch needles the resouce. Outbreaks of budmoth leave the trees shorn of high-quality needles, and lacking the resources to produce high quality foliage the next year. The needles are then shorter and have a higher raw fiber content. Feeding on such needles results in lower larval survival and female fecundity. It takes several years for the quality of foliage to recover.

Models based on the budmoth–tree interaction can generate cycles that resemble those observed, but analysis of time-series data on budworm and needles revealed a serious discrepancy between models and data. In model solutions there is a strong reciprocal influence between budmoth density and needle quality, each affecting the other's changes. In the data one only sees a one-way effect – budmoth have a large impact on the needles, but not vice-versa. In addition, a recent outbreak (early 1990s) never reached levels at which defoliation occurred. Foliage quality remained high but the outbreak collapsed nonetheless. This indicates that foliage quality is not causal, but "along for the ride" as budmoth fluctuates for some other reason(s).

The other potential mechanism known to be operating in this system was attack by parasitoids. The parasitism attack rate can be quite high (up to 80-90%), but this mechanism had been rejected based on the timing of peak attacks. Maximal parasitism rates occur 2 to 3 years after the peaks of budmoth density, and it was argued that if parasitoids cannot even limit budmoth increase, they were are unlikely to play an important role in the decrease phase of the cycle. However this was strictly a verbal argument and proved to be flawed. In a model having both direct density dependence (e.g. a finite food supply for budnoth that sets a limit to growth) and parasitoids, the host-parasitoid interaction can generate cycles that match the observed lag between the peaks in budmoth density and parasitoid attack (Turchin et al. 2003). In fact a simple model based on the Nicholson-Bailey framework was able to account for 83% of the variance in budworm numbers, compared to only 44% explained by the best-fitting model for the budworm-foliage interaction. A more careful statistical analysis (Turchin et al. 2003) showed that the budworm-foliage interaction does play a role in the cycle: although the goodness of fit to the data is only marginally improved by including that interaction, the improvement is statistically significant.

The importance of parasitoids is supported by subsequent work on the spatial dynamics of budworm outbreaks by Bj**o**rnstad et al 2002. Budworm outbreaks across Europe form a spatial "wave" moving at about 210km per year from southwest to northeast. A spatial extension of the Turchin et al (2003) host-parasitoid model, along the lines of (4.54), was able to reproduce the spatial wave under certain assumptions about moth dispersal and the spatial variation in habitat quality. However a model based on the budmoth–tree interaction was only able to produce waves travelling in the wrong direction (Bj**o**rnstad et al 2002).

The books edited by Berryman (2002) and written by Turchin (2003) review a number of other cases where quantitative modeling has been able to identify a single mechanism that is best able to account for the available data on population variability. In many cases this success depends on the availability of more information than total population counts. For example the larch budmoth data series is accompanied by less extensive data on needle length (a surrogate measure of needle quality) and the fraction of budmoth attacked by parasitoids.

So far most of the well-studied cycles seem to be driven by trophic interactions – consumer-resource, host-parasitoid, host-pathogen, or tritrophic. However in one case the most likely explanation seems to be *maternal effects*. This means that the conditions experienced by members of one generation affect the quality (and therefore the reproductive success) of their offspring. The additional state variable "individual quality" behaves much like the density of a resource species. When quality is high there is rapid population growth, but high population density brings quality down, as if it were being "consumed". As in genuinely trophic interactions, cycles are a likely outcome of population regulation by maternal effects. Analyzing and modeling data on the pine looper moth in UK forests, Kendall et al. (in press) concluded that the host-parasitoid interaction – which has been the most-studied hypothesis for the cycles in that system – is just along for the ride, while the cycles are driven by strong effects of the mother's environment on the viability of her offspring.

Because relatively few cases have been resolved, it is probably premature to generalize about the factors driving population variability. However, the ability to reach strong conclusions about well-studied

systems is encouraging.

Exercise 4.19 Something about the Ginzburg-Taneyhill model for maternal effects.

4.8 References

Bj**o**rnstad, O.N., M. Peltonen, A.M. Liebhold, W. Baltensweiler. 2002. Waves of larch budmoth outbreaks in the European Alps. Science 298: 1020-1023.

Comins, H.N., M.P. Hassell, and R.M. May. 1992. The spatial dynamics of host-parasitoid systems. J. Anim. Ecology 61: 735-748.

Comins, H.N. and M.P. Hassell. 1996. Persistence of multispecies host-parasitoid interactions in spatially distributed models with local dispersal. J. Theor. Biol. 183: 19-28.

Ellner, S. and P. Turchin. 1995. Chaos in a noisy world: new methods and evidence from time series analysis. American Naturalist 145: 343-375.

Hassell, M.P., J.H. Lawton, and R.M. May. 1976. Patterns of dynamical behavior in single-species populations. Journal of Animal Ecology 45: 471-486.

Hassell, M.P., H.N. Comins, and R.M. May. 1991. Spatial structure and chaos in insect population dynamics. Nature 353: 255-258.

Hastings, A. 1997. Population Biology: Concepts and Methods. Springer, NY.

Hubbell, S, S. B. Hsu and P. Waltman. 1977. A mathematical theory for single-nutrient competition in continuous cultures of micro-organisms. SIAM Journal on Applied Mathematics 32: 366-383.

Huisman, J. 1999. Population dynamics of light-limited phytoplankton: Microcosm experiments. Ecology 80:202-210.

Huisman, J., and F. J. Weissing. 1994. Light-Limited Growth and Competition for Light in Well-Mixed Aquatic Environments - an Elementary Model. Ecology 75:507-520.

Huisman, J., R. R. Jonker, C. Zonneveld, and F. J. Weissing. 1999. Competition for light between phytoplankton species: Experimental tests of mechanistic theory. Ecology 80:211-222.

Huisman, J., J. Sharples, J. M. Stroom, P. M. Visser, W. E. A. Kardinaal, J. M. H. Verspagen, and B. Sommeijer. 2004. Changes in turbulent mixing shift competition for light between phytoplankton species. Ecology 85:2960-2970.

Kendall, B.E., J. Prendergast, and O.N. Bj**o**rnstad. 1998. The macroecology of population cycles: taxonomic and biogeographic patterns in population cycles. Ecology Letters 1: 160-164.

Kendall, B.E., S.P. Ellner, E. McCauley, S.N. Wood, C.J. Briggs, W.W. Murdoch, and P. Turchin. Population cycles in the pine looper moth *Bupalus piniarius*: Dynamical tests of mechanistic hypotheses.

Ecological Monographs *in press*.

Ludwig, D., D.D. Jones, and C.S. Holling. 1978. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. Journal of Animal Ecology 47: 315-332.

Murdoch, W.W. 1994. Population regulation in theory and practice. Ecology 75: 271-287.

Murdoch, W.W. and A. Stewart-Oaten. 1989. Aggregation by parasitoids and predators: effects on equilibrium and stability. Amer. Natur. 134: 288-310.

Murdoch, W.W., B.E. Kendall, R.M. Nisbet, C.J. Briggs, E. McCauley, and R. Bolser. 2002. Single-species models for many-species food webs. Nature 417: 541-543.

Murdoch, W.W., S.L. Swarbrick, and C.J. Briggs. 2006. Biological control: lessons from a study of California red scale. Population Ecology 48: 297-305.

Pulliam, H. R. 1988. Sources, sinks and population regulation. American Naturalist 132: 652-661.

Rohani, P., T.J. Lewis, D. Gruenbaum, and G.D. Ruxton. 1997. Spatial self-organization in ecology: pretty patterns or robust reality? Trends in Ecology and Evolution 12: 70-74.

Strogatz, S. 1994. Nonlinear Dynamics and Chaos. Perseus Books, Reading Mass.

Taylor, A.D. 1993. Heterogeneity in host-parasitoid interactions: 'aggregation of risk' and the '$CV^2 > 1$' rule. Trends in Ecology and Evolution 8: 400-405.

Tilman, D. 1982. Resource Competition and Community Structure. Princeton University Press, Princeton NJ.

Turchin, P., S. N. Woods, S. P. Ellner, B. E. Kendall, W. W. Murdoch, A. Fischlin, J. Casas, E. McCauley, and C. J. Briggs. 2003. Dynamical effects of plant quality and parasitism on population cycles of larch budmoth. Ecology 84:1207-1214.

Turchin, P. 2003. Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton University Press, Princeton NJ.

Veilleux, B. G. 1976. The analysis of a predatory interaction between *Didinium* and *Paramecium*. Master's thesis, University of Alberta.

Veilleux, B. G. 1979. An analysis of the predatory interaction between *Paramecium* and *Didinium* Journal of Animal Ecology 4: 787-803.

Volterra, V. 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118: 558-560.

Zimmer, C. 1999. Life after chaos. Science 284: 83-86.

4.9 Appendix: Poisson distribution of eggs per host

The Poisson distribution for the numbers of parasitoid eggs per host can be derived by a limiting process. Imagine the habitat divided into n cells of size $1/n$, having chosen units so that the total habitat size is 1. In a time interval of length T, each parasitoid searches through cT units of area, consisting of ncT randomly chosen cells. There are a total of P parasitoids doing this, hence a total of $ncTP$ events of a parasitoid searching a cell and attacking any occupants. The number of attacks on any one host is equal to the number of searches through the cell that it's in. Under the assumption that parasitoids search randomly and independent of one another, the number of searches through any one cell is a Binomial random variable with $N = ncTP$ "trials" (the total number of searches by all parasitoids), and "success probability" $p = 1/n$ (the probability that the focal cell is chosen for search). The mean number of attacks on a given cell is $Np = cTP$. As $N \to \infty$, a Binomial(N, p) random variable with $Np = \lambda$ converges to a Poisson(λ) distribution. So as $n \to \infty$ the number of attacks on a given host converges to a Poisson random variable with mean cTP. Since each parasitoid searches an area of size c per unit time, cT is the parasitoid "area of discovery" that is denoted a in (4.39).

Chapter 5

Structured populations: discrete time

Population structure refers to consistent demographic differences among individuals as a function of some other attribute such as age, size, physiological state, or the history of conditions experienced as they develop. A structured population model tracks the dynamics not only of total population, but also the distribution of the variables that differentiate individuals. Change in the population are then predicted from the collective behavior of the individuals comprising the population, which therefore depends on the current composition of the population (e.g., the fraction of individuals who are old enough and large enough to breed).

Over the last 30 years structured population models have become a central modeling formalism in theoretical ecology, and one of the most widely used. Ecological theory was initially grounded mostly in unstructured models like those we have been studying so far. These models ignore within-species variability, so model predictions result from interactions between species, or between species and their environment. The current approach is to start at the level of individual organisms. Higher-level properties are derived as consequences of what happens to individual organisms. A natural result is a focus on the causes and consequences of differences between individuals.

This chapter covers some classical material on structured population models, rather tersely, so that we can get on to newer stuff. For more details, you can see the chapter on matrix models in Ellner and Guckenheimer (2006) which has considerable overlap with this chapter. For complete details and applications, there is no better source than the authoritative, entertaining and highly readable monograph by Caswell (2001).

5.1 The life table and Leslie matrix models

We consider first a population in which only individual age matters, and without any effects of population density. This is the simplest structured-population analog to $n(t+1) = \lambda n(t)$. The traditional model is female-based. If the species reproduces sexually, we tacitly assume that reproduction is never limited by a shortage of mates. We assume (to begin with) an annual birth-pulse – all births for the year happen "at once". We census the population immediately after the birth-pulse (this is called a "post-breeding" census), so that the age of newborn individuals is 0.

Definitions for the life table model

- $n_a(t)$ = number of females age a at time t
- m_a = per-capita fecundity (daughters) of age-a females
- p_a = survivorship of age-a females to age $a+1$
- $l_a = p_0 p_1 p_2 \cdots p_{a-1}$ = survival from birth to age a
- $f_a = p_a m_{a+1}$ = average fecundity at age $a+1$ of females now age a = number of newborns "next year" per age-a female "now".

m_a is daughters "now" – this year's birth pulse – for a female whose age now is a. It's about the present, not the future. f_a is daughters next year for a female whose age now is a. So for forecasting the population, the relevant quantity is f_a.

Population dynamics

$$\begin{aligned} n_0(t+1) &= f_0 n_0(t) + f_1 n_1(t) + \cdots = \sum_{a \geqslant 0} f_a n_a(t) \\ n_a(t+1) &= p_{a-1} n_{a-1}(t) \end{aligned} \tag{5.1}$$

There is an another convention in which the population is censused just before the birth pulse, so all individuals are at least 1 year old. This is called a "pre-breeding census". Pre-breeding census also changes the interpretation of f_a in the equations above - instead of the definition above, we have

$$\text{pre-breeding } f_a = m_a p_0$$

. Births occur "now" (immediately after the current census) – m_a daughters per age-a female now), and then p_0 is the fraction of those kids who survive until the next census (when they are 1 year old and first appear in a population count). So when the model is expressed in terms of m's and p's rather than f's and p's, there are two different versions depending on whether a pre-breeding or post-breeding census is assumed. This is very confusing, especially to experts. One often finds the two versions intermingled in a single paper or book, and even cases where the author state one convention, e.g. that newborns are age 0, but uses the other one. Beware and read carefully!

Leslie matrix Suppose that there is a finite age A beyond which no individuals can live, i.e. $l_A > 0, l_{A+1} = 0$. The population dynamics (5.1) then consists of finitely many equations and can be written in matrix form:

$$\begin{bmatrix} n_0(t+1) \\ n_1(t+1) \\ \vdots \\ \\ n_A(t+1) \end{bmatrix} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_A \\ p_0 & 0 & 0 & \cdots & 0 \\ 0 & p_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & & p_{A-1} & 0 \end{bmatrix} \begin{bmatrix} n_0(t) \\ n_1(t) \\ \vdots \\ \\ n_A(t) \end{bmatrix} \tag{5.2}$$

or simply

$$\mathbf{n}(t+1) = \mathbf{L}n(t) \tag{5.3}$$

$\mathbf{L}$ is called the *Leslie matrix* after P.H. Leslie. Leslie popularized its use in animal population ecology in the mid 20^{th} century, initially as an approximation to models with continuous age structure from human demography.

We have seen (5.3) before, in the context of local stability analysis, so we know what to expect: exponential solutions constructed from the eigenvalues λ_j and corresponding eigenvectors w_j of the Leslie matrix,

$$n(t) = c_1\lambda_1^t w_1 + c_2\lambda_2^t w_2 + \cdots + c_A\lambda_A^t w_A$$

and for large t the solution is dominated by the term from the dominant eigenvalue (the one with the largest absolute value):

$$n(t) \sim c_1\lambda_1^t w_1.$$

The existence of a dominant eigenvalue is guaranteed so long as the matrix $\mathbf{L}$ is power-positive: some power L^m has all entries > 0, where the exponentiation here is in the sense of matrix multiplication, not entry-by-entry. For example, the matrix $\mathbf{L} = \begin{bmatrix} 0.5 & 3 \\ 0.8 & 0 \end{bmatrix}$ is power positive with $m = 2$

```
> L=matrix(c(0.5,0.8,3,0),2,2); L%*%L;
     [,1] [,2]
[1,] 2.65  1.5
[2,] 0.40  2.4
```

If a non-negative matrix $\mathbf{L}$ is power-positive, the Perron-Frobenius Theorem implies that $\mathbf{L}$ has a unique dominant eigenvalue λ_1 which is positive, real, and strictly larger in magnitude than any other eigenvalue of $\mathbf{L}$. Moreover, there is a stictly positive eigenvector w_1 corresponding to λ_1. As $t \to \infty$ the population vector $\mathbf{n}(t)$ converges a stable age distribution proportional to w_1 in the sense that

$$\frac{\mathbf{n}(t)}{\lambda_1^t} \to cw_1 \tag{5.4}$$

for some constant c that depends only on $\mathbf{n}(0)$, unless $\mathbf{n}(0) = 0$. Since $\log(\lambda_1^t) = t\log(\lambda_1)$,

Note that "stable" is used here in a distinctive sense. The population itself is growing, but the *relative numbers* in different age classes are becoming stable. In order for power-positivity to hold, we have to restrict the model to reproductive age-classes, i.e. $f_A > 0$ must be true (proof: Otherwise a population started with only post-reproductive individuals would die out, implying that for any m and any sufficiently large k we would have $\mathbf{L}^{km}n(0) = 0$. But if all entries of $\mathbf{L}^m$ are positive so are all entries of $\mathbf{L}^{km}$, hence $\mathbf{L}$ cannot be power positive). If $f_A > 0$, and any two consecutive $f's$ are positive, then $\mathbf{L}$ will be power-positive. Note that it is safe to treat post-reproductives as if they were dead (for purposes ot calculating λ) because they make no contribution to future population growth.

To illustrate these properties, consider the simpest possible age-structured case, a 2×2 matrix with $A = 1$. For example, consider

$$\mathbf{L} = \begin{bmatrix} .2 & 1.5 \\ .8 & 0 \end{bmatrix} \tag{5.5}$$

Using **R** we find that the eigenvalues are $\lambda_1 = 1.2, \lambda_2 = -1$, and the eigenvector associated with λ_1 is $w_1 = (0.832, 0.555)$. To get the proportions of the population in the different ages, we normalize w_1 so that its entries sum to 1:

```
w1=eigen(L)$vectors[,1]; w1=w1/sum(w1); w1
[1] 0.6 0.4
```

In the long run, we should see 60% of the population being age 0, and 40% being age 1, regardless of what the population state is initially, and the total population size should grow by 20% per year. Figure 5.1 shows that these occur within about the first 20 generations, even from extreme initial states (all age 0, or all age 1).

Some other properties of the stable age distribution and long-term growth rate λ_1 are developed in the following exercises.

Exercise 5.1 Explain in words why

$$n_0(t+1) = \sum_{a=0}^{A} f_a l_a n_0(t-a) \tag{5.6}$$

Exercise 5.2 Once the population is growing at rate λ_1 (i.e. once it converges to the stable age distribution) then in particular $n_0(t) = C\lambda_1^t$ for some constant $C > 0$. Use this observation and (5.6) to deduce the Euler-Lotka equation

$$\sum_{a=0}^{A} \lambda^{-(a+1)} l_a f_a = 1. \tag{5.7}$$

Exercise 5.3 Show that the Euler-Lotka equation can be written in the form (which is the most often-used version)

$$\sum_{x=1}^{A} \lambda^{-x} l_x m_x = 1. \tag{5.8}$$

Figure 5.1: Convergence to stable age distribution for the Leslie matrix model (5.5), with two different initial populations: 1 newborn and no adults (solid lines), 5 adults and no newborns(dashed lines). (a) Growth of total population size, on log scale. (b) Convergence of the age distribution; values plotted are the number of newborns (age 0) divided by the total population (age 0 + age 1) (c) Annual population growth rate, $N(t+1)/N(t)$, where $N(t) = n_0(t) + n_1(t)$. A high fraction of newborns in a given year implies a low growth rate that year, and vice versa.

Exercise 5.4 Assuming again that the population has converged to its stable age distribution, explain why the population's age distribution is proportional to

$$(1, l_1/\lambda_1, l_2/\lambda_1^2, \cdots, l_A/\lambda_1^A). \tag{5.9}$$

[HINT: to be 19 years old now, you must have been born (WHEN?) and then (WHAT?)]

The left-hand side of the Euler-Lotka equation is a decreasing function of λ_1 for $0 < \lambda_1 < \infty$. Conse-

quently we have that

$$\lambda > 1 \text{ if and only if } R_0 > 1 \text{ where } R_0 = \sum_{a=0}^{A} l_a f_a.$$

R_0 is the expected number of offspring to a newborn female over her lifetime; so we find (not surprisingly) that the population grows in the long run if and only if each newborn female can expect to have more than 1 daughter, on average.

The survival and fecundity parameters can also be used to derive a number of summaries of the life history, such as the mean age at reproduction, generation time, life expectancy, and so on (see e.g. Chapter 4 of Bulmer). For our purposes here these are not important.

One common extension of the Leslie matrix is to let age A represent "old" individuals – age A or older – and assume that survival and fecundity are the same for all "old" individuals. The only change in the Leslie matrix is that the bottom-right entry is then p_A rather than 0, where p_A is the probability that an "old" individual survives an additional year. Formally this allows individuals to live forever, but since there is no actual distinction by age in the "old" category this is not a problem for the model. And, since by definition few individuals reach extreme old age, an approximate treatment of the very oldest segment of the population will not be a major cause of error in population forecasts.

5.2 Stage structured matrix models

In place of $\mathbf{L}$ we can consider populations classified by any discrete variable and governed by a population projection matrix $\mathbf{M}$,

$$\mathbf{n}(t+1) = \mathbf{M}\mathbf{n}(t) \tag{5.10}$$

Perron-Frobenius still applies so long as $\mathbf{M}$ is power-positive:

$$n(t) \sim c\lambda_1^t w_1$$

That is, we a long-term growth rate given by the dominant eigenvalue λ_1 of the matrix $\mathbf{M}$, and convergence to a stable stage distribution (proportions of individuals in the different stages) given by the corresponding eigenvector w_1.

The terms "stage" or "stage-class" are often used for the different groups of individuals recognized in the model. Sometimes they really are distinct life stages, but more often they are categories imposed by the modeler on a continuously varying trait such as size (or age). Caswell (2000) presents size-based models for sea turtles, desert tortoise, killer whales, geese, striped bass, bryozoan corals, and spotted owls.

The matrix $\mathbf{M}$ then describes all the ways that an individual in one stage class "now" can contribute to another stage class "next year" (or whatever the time step is in the model). To see how this works in detail it is (for once) convenient to use the algebraic definition of matrix multiplication. The vector

form of the model (5.10) decodes into the entry-by-entry formulas

$$n_i(t+1) = \sum_j M_{ij} n_j(t) \tag{5.11}$$

Thus M_{ij} quantifies how many stage-i individuals are produced "next time" by each stage-j individual now. This provides a recipe for "reading" a projection matrix. Each stage-1 individual "now" produces

$$\begin{array}{c} M_{11} \text{ type 1 individuals} \\ M_{21} \text{ type 2 individuals} \\ M_{31} \text{ type 2 individuals} \\ \vdots \\ M_{k1} \text{ type } k \text{ individuals} \end{array}$$

at the next time step. The M values listed above comprise the first column of the projection matrix $\mathbf{M}$. Similarly, what you get from a stage-2 individual will be given by the entries in the second column.

Matrix models are extremely popular because they are simple, easy to construct and give a lot of information. Elasticity analysis of these matrix models has become a major tool in conservation biology (elasticities describe how the value of λ_1 changes in response to changes in different matrix entries; entries with large elasiticites are identified as good "targets" management intervention, because relatively small changes have relatively large effects). Because all this is covered extensively in the Conservation Biology course (NTRES 405) here we only give a few examples.

5.2.1 Teasel, Werner and Caswell (1977)

Matrix for Field A teasel population (corrected version, Caswell 2001 p. 60).

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 322.38 \\ .966 & 0 & 0 & 0 & 0 & 0 \\ .013 & 0.01 & .125 & 0 & 0 & 3.448 \\ .007 & 0 & .125 & .238 & 0 & 30.170 \\ .008 & 0 & 0 & .245 & .167 & 0.862 \\ 0 & 0 & 0 & .023 & .750 & 0 \end{pmatrix}$$

1=Dormant seeds, year 1; 2=Dormant seeds, year 2
3,4,5 =Small,Medium,Large rosettes
6=Flowering plants

Exercise 5.5 . Using the recipe given above for "reading" a transition matrix, decode and state in words what the first column of the matrix for teasel says about dormant seeds, and what the last column says about flowering plants.

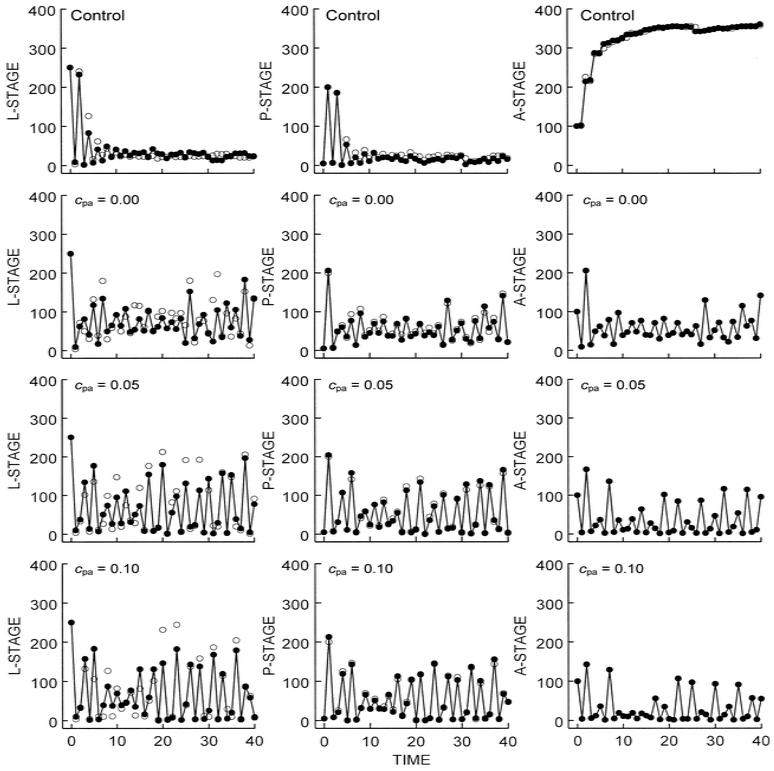


Figure 5.2: Time-series data (open circles) and one-step predictions from the LPA model (closed circles) for representative cultures from four experimental treatments: control, and cpa (adultpupa cannibalism coefficient) = 0.00, 0.05, and 0.10. The unit of time is two weeks. This figure is from Dennis et al. (2001).

5.2.2 Flour beetles *Tribolium*

Costantino et al. (1995), Dennis et al. (2001) describe experiments on laboratory populations of flour beetles. The time step is 2 weeks, roughly equal to the durations of the egg+larval and pupal stages.

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & F_3(n) \\ P_1 & 0 & 0 \\ 0 & P_2(n) & P_3 \end{pmatrix}$$

1=Larvae, 2=Pupae, 3=Adults

$$F_3(n) = b\exp(-c_{ea}n_3 - c_{el}n_1) \qquad P_2(n) = \exp(-c_{pa}n_3)$$

The density-dependence is due to cannibalism of immobile stages (eggs, pupae) by mobile stages (larvae, adults) with rates based on random encounter (as in Nicholson-Bailey, survival is the zero term of a Poisson distribution for the number of mobile individuals encountering the spot where the immobile one is sitting). In more familiar terms, it's as if butterflies were to eat any cocoons they encounter (including

their own offspring), and caterpillars ate any butterfly eggs they happened to find; it's no wonder that flour beetles are generally regarded as a "pest", except by scientists. Based on the stage names, this is called the LPA model. Figures 5.2 and 5.3 show the kinds of dynamics it can produce, and how well the LPA model (either deterministic as above, or with random perturbations to represent finite population effects) succeeds in predicting the outcome of experiments (these being genuine predictions: the model was not fitted to these data, rather its parameters were estimated from independent data and then one parameter (c_{pa}) was varied in the model and in experimental cultures).

5.2.3 Vegetation dynamics in England

Usher (1972): The "individuals" are plots of land, and the stages are 1=Bog, 2=Heath, 3=Woodland, 4=Grazed.

$$\mathbf{M} = \begin{pmatrix} .65 & .30 & 0 & 0 \\ .29 & .33 & .28 & .4 \\ .06 & .30 & .69 & .2 \\ 0 & .07 & .03 & .4 \end{pmatrix}$$

Each column sums to 1, indicating that the total amount of area neither grows nor shrinks.

Exercise 5.6 . What is the dominant eigenvalue λ_1 for the vegetation dynamics matrix, and why?

5.2.4 Sea turtles: Crouse,Crowder,Heppell et al.

Stage	**Description**	**Duration(yr)** T_i	**Survival** σ_i	**Fecundity** F_i
1	Eggs	1	0.6747	0
2	Small juveniles	~ 7	0.75	0
3	Large juveniles	~ 8	0.6758	0
4	Subadults	~ 6	0.7425	0
5	Adults	> 32	0.8091	76.5

Because stages are really age-classes in this model, individuals can grow or stay the same "size", but not shrink. This leads to a simple form for the transition matrix.

Transition Matrix

$$A = \begin{bmatrix} P_1 & 0 & 0 & 0 & F_5 \\ G_1 & P_2 & 0 & 0 & 0 \\ 0 & G_2 & P_3 & 0 & 0 \\ 0 & 0 & G_3 & P_4 & 0 \\ 0 & 0 & 0 & G_4 & P_5 \end{bmatrix}$$

$P_i = \sigma_i(1-\gamma_i)$ survive and remain in stage
$G_i = \sigma_i\gamma_i$ survive and grow to next stage

Growth probabilities γ_i were computed by assuming fixed stage durations, and stable age distribution

within each stage, so γ_i is the fraction of individuals in the last year of the stage:

$$\gamma_i = (z^{T_i} - z^{T_i-1})/(z^{T_i} - 1), \quad \text{where} \quad z = \sigma_i/\lambda_1$$

A general size-based model is more complicated to parameterize, because there is the possibility of individuals shrinking as well as growing.

The model (in various versions) was used to compare the effectiveness of headstarting (save turtles on the beach) versus TEDs (turtle excluder devices on shrimping boats, that can catch and kill juveniles at sea; see the papers by Heppell, Crouse, Crowder et al. given below). This story is an often-repeated "poster child" for matrix models in conservation biology, because a model-based change in conservation strategy has apparently halted and even reversed a longstanding decline of Atlantic sea turtle populations.

5.3 Limitations and alternatives

The main intrinsic limitations of the matrix model come from the fact that individual state is discrete – a finite list of options. All individuals within a given stage therefore are assumed to be identical, in terms of their possible fates and the odds for each possible fate. Sometimes this is reasonable, and sometimes it is less reasonable. For example in the sea turtle models, it is not really the case that all small juveniles have the same chance of maturing into the large juvenile class – one who hatched last year has no chance, one who hatched 6 years years ago has a very good chance. Similar issues are present if individuals are classified based on some physiological variable – such as size – that really varies continuously: an individual that just grew into a size category will probably stay there for a while, but an individual who is close to the upper size limit for the category has a good chance of growing out of it soon.

Dealing with these problems requires models having a continuous individual-level state variable. If we modify the basic matrix model in this way alone, the result is an integral model (Easterling et al. 2000). The state of the population is described by a distribution function $n(x,t)$ that can be thought of (loosely) as the number of state-x individuals at time t. More precisely, $n(x,t)$ is the state density function such that

$$\int_{x_1}^{x_2} n(x,t)dx$$

is the number of individuals between sizes x_1 and x_2 at time t. The matrix iteration (5.11) is then replaced by

$$n(x,t+1) = \int_{\Omega} K(y,x)n(y,t)dy$$

where the *kernel function* $K(y,x)$ gives (loosely) the number of state-y individuals "next year", per state-x individual "now", and Ω is the set of possible sizes. The papers by Rees and Rose (2002) and Childs et al. (2003, 2004) describe some applications of integral models to plant populations structured by size and age, with size as a continuous variable. The change to a continuous structuring variable turns out

to have very little impact on the behavior of the model. For example, if the kernel function is continuous and power-positive then there is a dominant eigenvalue and corresponding dominant eigenvector that give the long-term growth rate and structure of the population, and formulas similar to those in the matrix model are available to compute the sensitivity of the dominant eigenvalue to changes in the kernel (Ellner and Rees 2005).

Integral models are a relatively recent innovation, so it remains to be seen whether they will catch on. Like matrix models, they make it easy to represent situations where growth is not deterministic – i.e. size "next year" is not perfectly predicted by size "now". If growth is modeled as deterministic, it is more convenient to use models with continuous time as well as continuous state. Those are the subject of the following chapter.

5.4 References

Special Feature: Elasticity Analysis in Population Biology. Ecology 81: 605-708 (2000)

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd edition. Sinauer Associates, Sunderland MA.

Childs, D. Z., M. Rees, K. E. Rose, P. J. Grubb, and S. P. Ellner. 2003. Evolution of complex flowering strategies: An age and size-structured integral projection model approach. Proceedings of the Royal Society 270:1829-1838.

Childs, D. Z., M. Rees, K. E. Rose, P. J. Grubb, and S. P. Ellner. 2004. Evolution of size dependent flowering in a variable environment: Construction and analysis of a stochastic integral projection model. Proceedings of the Royal Society 271:425-434.

Crouse, D.T., L.B. Crowder, and H. Caswell. 1987. A stage-based population model for loggerhead sea turtles and implications for conservation. Ecology 68: 1412-1423.

Crowder, L.B., D.T. Crouse, S.S. Heppell, and T.H. Martin. 1994. Predicting the impact of turtle excluder devices on loggerhead sea turtle populations. Ecological Applications 4: 437 - 445.

Costantino, R.F., J.M. Cushing, B. Dennis, and R. A. Desharnais. 1995. Experimentally induced transitions in the dynamic behavior of insect populations. Nature 375: 227-230

Brian Dennis, Robert A. Desharnais, J. M. Cushing, Shandelle M. Henson, and R. F. Costantino. 2001. Estimating chaos and complex dynamics in an insect population. Ecological Monographs 71: 277303.

Easterling, M. R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new structured population model. Ecology 81:694-708.

Ellner, S.P. Ellner and J. Guckenheimer. 2006. Dynamic Models in Biology. Princeton University Press, Princeton NJ.

Ellner, S.P. and M. Rees. 2005. Integral projection models for species with complex demography. American Naturalist 167: 410-428.

Ellner, S.P. and M. Rees. 2007. Stochastic stable population growth in integral projection models. Journal of Mathematical Biology 54:227256

Heppell, S.S., L.B. Crowder, and D.T. Crouse. 1996. Model to evaluate headstarting as a management tool for long-lived turtles. Ecological Applications 6: 556-565.

Rees, M., and K. E. Rose. 2002. Evolution of flowering strategies in Oenothera glazioviana: an integral projection model approach. Proceedings of the Royal Society Series B 269: 1509-1515.

Usher, M.B. 1972. Developments in the Leslie Matrix Model. pp. 29-60 in J.N.R. Jeffers (ed) Models in Ecology. Blackwell, London.

Werner, P.A. and H. Caswell. 1977. Population growth rates and age versus stage-distribution models for teasel (*Dipsacus sylvestris* Huds). Ecology 58: 1103-1111.

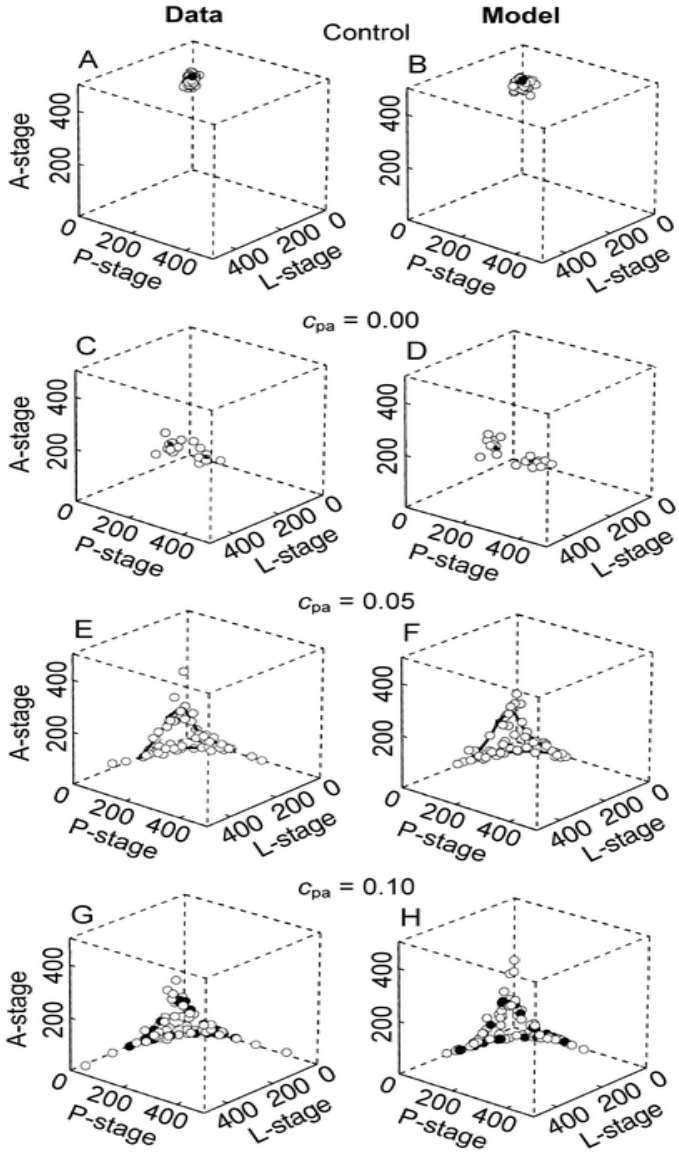


Figure 5.3: Plots in (L, P, A) space of the data (A, C, E, G) and stochastic model (B, D, F, H) for weeks 4080 (open circles) together with the deterministic model attractor (solid circles or lines) for four experimental treatments: control, and cpa = 0.00, 0.05, and 0.10. Model simulations used the same initial conditions, number of replicates, and experiment duration as the real experiments data. Data from the start of each experiment (weeks 0–38 for control, cpa=0.05 and 0.10, weeks 0–66 for cpa=0.00) were discarded as representing transient behavior, in order to focus attention on comparing the long-run behavior of the experiments and model. This figure is from Dennis et al. (2001).

Chapter 6

Structured populations: continuous state and time

The goal of this chapter is to develop some important aspects of current approaches to modeling structured populations in continuous time. We begin with some classical models that every theoretical ecologist should know about, and then move to more contemporary work.

As in the last chapter, the goal of these models is to understand the consequences of variability among individuals in their "state": age, size, or any other attribute or combination of attributes that has implications for birth and death rates or for the attributes of offspring. Once we move to continuous time, it becomes natural to also imagine that the state variables characterizing individuals also change continuously, rather than having a discrete set of possible states as in a matrix model.

The combination of continuous state and continuous time puts us in the realm of partial differential equations, which are both numerically and analytically harder to deal with than the models we've dealt with so far. Ways to reduce model complexity, while retaining as much biological realism as possible, are therefore very useful. The approach that we will be emphasizing originated in attempts to understand the experimental results shown in Figure 6.1. Both of these are laboratory insect populations, limited by the rate of food supply. These were grown under constant conditions, so the oscillations are internally generated.

A striking feature of the blowfly data is the emergence of nearly discrete generations, with each new cohort of flies generated by a separate burst of egg production. In this experiment the limiting factor was the food supply to adults – food for larvae was provided in excess. Note that eggs were only produced when adult density is low. When there are too many adults relative to the food supply, they can eat enough to survive but aren't able to reproduce. The period of the population cycles is roughly 2-3 times the maturation time (which is roughly constant, because immatures have as much food as they want).

In the *Plodia* data, the cycles are more irregular, and their dominant period is close to the generation

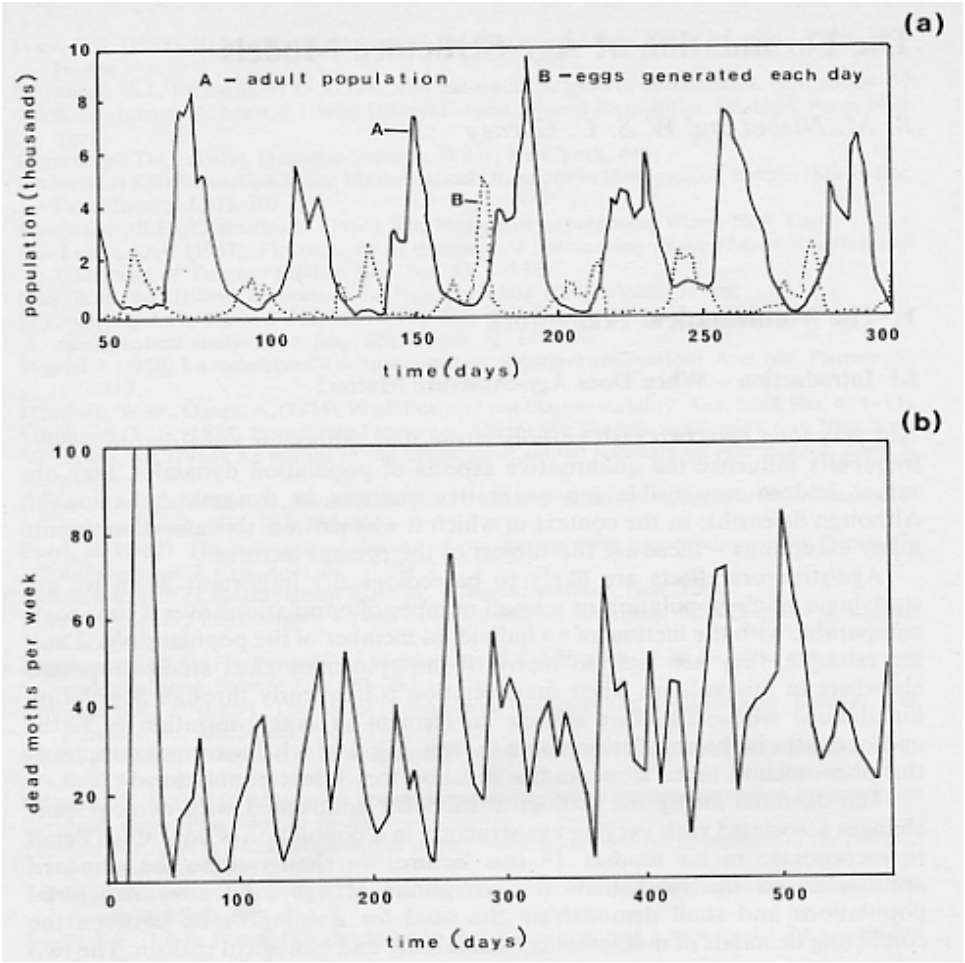


Figure 6.1: Experimental data on insect population fluctuations under constant conditions resulting from age-structure effects. (a) Nicholson's (1954,1957) blowflies. The solid curve is the total adult population, and the dotted curve is the rate of egg production. These data first appeared as Figure 3 in Nicholson (1954) and are from an experiment where the population was "governed by the daily supply of 0.5g of ground liver for the adults" (Nicholson 1954 p. 21). (b) Lawton's experiments on *Plodia interpunctella* (Indian meal moth) with the population regulated by limited food supply for larvae. The data are the number of dead adult moths, which is a proxy for the adult population. This figure is taken from Nisbet and Gurney (1986).

time. In this experiment population was limited by the food used by larvae, while food for adults was supplied as needed.

So there are two things in need of explanation:

1. How can discrete generations can emerge spontaneously in a continuously breeding organism growing under constant conditions?

2. How do different mechanisms of population regulation result in very different relations between the period of population cycles and the generation length of the organism?

Both of these questions have to be addressed using models where time and individual state are both continuous. In a discrete time model, by specifying a discrete time interval we are imposing the *a priori* restriction that cycle periods must be multiples of the model's time step. So to fully understand these patterns we have to start with a model where time is continuous, so temporal patterns cannot be artifacts of an arbitrary choice of time step. Continuous state is needed so that quantities like "maturation time" can be well defined in the model. In a model with discrete stages (e.g. egg-larva-pupa-adults), individuals within a stage at a given time necessarily all have the same probability of moving up to the next stage.

6.1 Age structure in continuous time

The simplest starting point is the continuous-time analog of the Leslie matrix, in which vital rates depend on individual age (a). As in the Leslie model we ignore (to begin with) effects of population density and environmental factors, consider a single species, and count only females. The state of the population is then characterized by the *age distribution* $n(a,t)$ that we can loosely think of as being the number of age-a individuals at time t. Technically n is the probability density of the age distribution, and

$$\int_a^b n(s,t)ds = \text{ number of individuals of age } a \text{ to } b. \tag{6.1}$$

However if $n(s,t) \approx n(a,t)$ over the whole interval $a \leq s \leq a+1$, then $n(a,t)$ will be approximately the number of individuals between age a and age $a+1$.

The dynamics of $n(a,t)$ are generated by the age-specific per-capita birth rate $b(a)$ and death rate $\mu(a)$ – these are the "primitives" of the model. The basic balance law is that in order to be age $a > 0$ at time t, an individual must have been age $a - dt$ at time $t - dt$, and not died between time $t - dt$ and time t. That is:

$$n(a,t) = n(a-dt, t-dt)\left(1 - \mu(a-dt)dt\right). \tag{6.2}$$

With a bit of algebra this rearranges to

$$\frac{n(a,t) - n(a-dt, t-dt)}{dt} = -\mu(a-dt)n(a-dt, t-dt). \tag{6.3}$$

The left-hand side of (6.3) is a difference quotient, so letting $dt \to 0$ we get

$$\frac{\partial n}{\partial t} + \frac{\partial n}{\partial a} = -\mu(a)n(a,t) \tag{6.4}$$

which is the dynamic equation for $n(a,t)$. To complete the model we only need to supply the *boundary conditions*

$$n(0,t) = \int_0^{infty} b(a)n(a,t)da. \tag{6.5}$$

To study model solutions and their properties we need to calculate the survival curve $l(a)$, the fraction of individuals that survive from birth (at age 0) to age a. Starting with N_0 age-0 individuals at time 0, let $N(t)$ be the number alive at time t. The instantaneous per-capita mortality rate for this cohort is $\mu(t)$, because all individuals are age t at time t. So

$$\frac{dN}{dt} = -\mu(t)N(t)$$

and therefore

$$\frac{d}{dt}\log N(t) = -\mu(t).$$

Integrating both sides of the last equation, we get

$$\log N(t) = \int_0^t -\mu(s)ds + C$$

where C is a constant depending on initial conditions. Setting $t = 0$ we see that $C = \log N(0)$, so

$$N(t) = N_0 e^{-\int_0^t \mu(s)ds}$$

and therefore

$$l(a) = e^{-\int_0^a \mu(s)ds}. \tag{6.6}$$

Sometimes $l(a, b)$ is used to denote the survival from age a to age b, and then $l(a) = l(0, a)$. A derivation like that above gives

$$l(a, b) = e^{-\int_a^b \mu(s)ds}. \tag{6.7}$$

For any $b > a$ we have $l(0, b) = l(0, a)l(a, b)$, so $l(a, b) = l(b)/l(a)$.

Exercise 6.1 Explain in words why $l(0, b) = l(0, a)l(a, b)$ whenever $0 < a < b$.

Exercise 6.2 If X is a non-negative random variable, and $F(x)$ is its cumulative distribution function $F(x) = Prob(X \leq x)$, then

$$E[X] = \int_0^\infty (1 - F(x))dx.$$

Using this fact, show that the life expectance (i.e. the mean age at death) for an individual is $\int_0^\infty l(a)da$.

6.1.1 Solving the age-structure model

The solution to the age-structure model $n(a, t)$ can be expressed in terms of the survival function $l(a, b)$, and the initial population $n(a, 0)$. The form of the solution depends on whether or not $a > t$. If so, any age-a individuals must have been alive at time 0. If not, age-a individuals were born at $t - a > 0$.

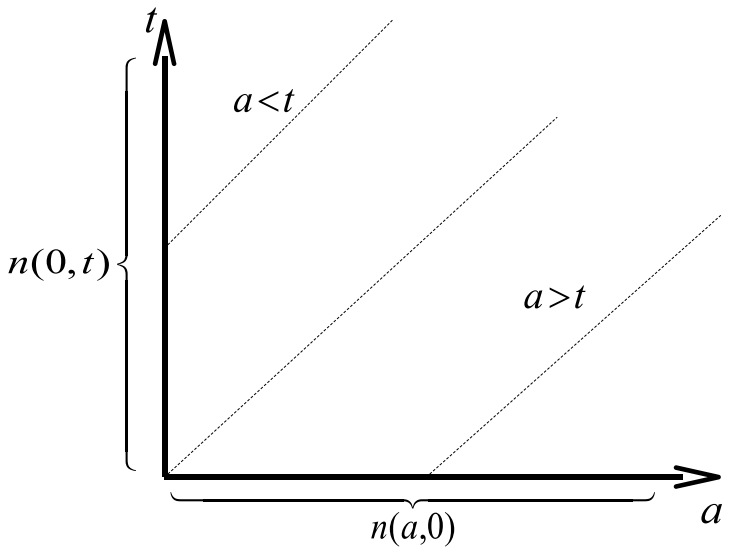


Figure 6.2: The so-called Lexis Diagram representing the solution (6.8) to the age-structure model.

In either case, $n(a,t)$ is the number of those "founders" who survive up to time t. So the form of the solution is

$$\begin{aligned} a < t: &\quad n(a,t) = n(0,t-a)l(0,a) \\ a > t: &\quad n(a,t) = n(a-t,0)l(a-t,a) \end{aligned} \tag{6.8}$$

See Figure 6.2; following back lines of slope 1 (representing time and age changing at the same rate), points with $a < t$ trace back to the t axis, representing newborns (age 0) at time $t - a > 0$. Points with $a > t$ trace back to the a axis, representing individuals who were alive (and age $a - t$) at time 0.

This is not really a complete solution to the model, because the number of births $n(0, t-a)$ is not something we know – it has to be calculated from (6.5). Nonetheless it helps us determine the model's longterm behavior. In the long run we expect to see exponential growth or decline at some rate r,

$$n(a,t) \sim Cn^*(a)e^{rt} \tag{6.9}$$

where $n^*(a)$ is the *stable age distribution* and C is a constant depending on initial conditions. Like any eigenvector (pause a moment until that makes sense to you, or ask your instructor) n^* is only defined up to a multiplication by a constant. It is convenient to scale the stable age distribution so that $n^*(0) = 1$, so $n^*(a)$ is then the number of age-a individuals relative to the number of age-0 individuals.

For t large, we can substitute the second line of (6.8) into (6.5), getting

$$n(0,t) = \int_0^\infty b(a)n(0,t-a)l(a)da. \tag{6.10}$$

Then substituting (6.9) on both sides of this equation:

$$\begin{aligned} Ce^{rt} &= \int_0^\infty b(a)Ce^{r(t-a)}l(a)da \\ \implies 1 &= \int_0^\infty e^{-ra}l(a)b(a)da. \end{aligned} \tag{6.11}$$

The last line above is the famous Euler-Lotka equation. Because the right-hand side is a monotonic decreasing function of r, with limiting values $+\infty$ as $r \to -\infty$ and 0 as $r \to \infty$, there is always a unique solution.

Exercise 6.3 Define

$$R_0 = \int_0^\infty l(a)b(a)da. \tag{6.12}$$

R_0 can be interpreted as the average lifetime number of (female) offspring to a (female) newborn individual. Show that the asymptotic population growth rate r, which is the solution to the Euler-Lotka equation, is positive if $R_0 > 1$, and negative if $R_0 < 1$.

We can also derive the form of the stable age distribution using (6.9). For t large

$$\begin{aligned} n(a,t) &= n(0,t-a)l(a) \\ Ce^{rt}n^*(a) &= Ce^{r(t-a)}n^*(0)l(a) = Ce^{r(t-a)}l(a) \\ n^*(a) &= e^{-ra}l(a) \end{aligned} \tag{6.13}$$

There are several different ways that people define the *generation time*, all denoted T. These are different answers to the question: what is the average time between successive generations in the population, once it reaches stable age distribution? The simplest is to define T by the equation

$$e^{rT} = R_0$$

giving $T = \log(R_0)/r$. The motivation for this definition is that in each generation, a mother replaces herself by (on average) R_0 female offspring. When each mother has been replaced by her R_0 offspring, the population has grown (or shrunk) by a factor R_0. T as defined above is the amount of time it takes (at stable age distribution) for the population to grow or shrink by that amount.

Another way of defining T is to take all the females born at some moment, and look at their average age at reproduction (i.e. each time one of them has a child, write down the mother's age; when all the females being tracked have died, average all the recorded ages at childbirth). This gives

$$T = \frac{1}{R_0}\int_0^\infty al(a)b(a)da$$

(remember, the last two equations are two *different* ways of defining T, and will generally give different values of T).

Exercise 6.4 Another way of defining T is as the mean age of the mother, for all of the children born "now" (at some large time t, when the population has reached stable age distribution). What is the formula for T under this definition?

6.2 Size structure in continuous time

It is not always sufficient or desirable to use age as the sole variable distinguishing individuals. An approach similar to the age-structured model can be developed for individuals classified by any continuous state variable m that changes as an individual develops over time. We will think of m as "size", but it could really be any trait that changes deterministically over the course of an individual's lifetime.

The state of the population is given by $n(m,t)$, the size-distribution of individuals at time t. The formal meaning of n is that $\int_a^b n(m,t)dm$ is the total number of individuals with size between a and b at time t. More useful are the approximate consequences of this definition:

The number of individuals in the size range $[m, m+h]$ *is* $n(m,t)h$

The number of individuals in the size range $[m, m-h]$ *is* $n(m,t)h$

These are approximations that hold as $h \to 0$, with error proportional to h^2, but you generally won't go wrong by thinking of them as being exactly true for h small.

In matrix models for size-structured populations we could easily allow changes in size to be random: some individuals of a given size-class now grow, while others shrink. In continuous time this is more difficult, so we will begin by only considering *deterministic* size dynamics. That is, we assume that all individuals of size m have size-dependent growth rate $g(m)$. The size of an individual therefore obeys the differential equation

$$dm/dt = g(m), \qquad m(0) = m_0 \tag{6.14}$$

where m_0 is the individual's size at birth. Size at birth can be allowed to vary in this model (i.e. not everyone will have the same m_0), but the growth function $g(m)$ is assumed to be the same for all individuals. Similarly, we assume that all size-m individuals have the same per capita mortality rate $\mu(m)$.

As in the age-structured case, we can derive a partial differential equation for the changes in $n(m,t)$ by tracking how many the individuals enter and leave a particular small category – in this case a size range $[m_1, m_2]$ – in a short interval of time of duration $\tau \ll 1$. *For now we assume* that the size range of new

offspring does not overlap $[m_1, m_2]$. Then we can apply the basic balance law, and assert that

$$\begin{aligned}\text{Number in } [m_1, m_2] \text{ at } t+\tau = {} & \text{Number in } [m_1, m_2] \text{ at } t \\ & + \text{Number of smaller individuals that reach size } m_1 \text{ by } t+\tau \\ & - \text{Number that grow to sizes larger than } m_2 \text{ by } t+\tau \\ & - \text{Number that die by } t+\tau.\end{aligned} \tag{6.15}$$

- The number in $[m_1, m_2]$ at time t can be expressed as $n(m,t)(m_2 - m_1)$ where m is a "typical" size in $[m_1, m_2]$. We don't have to say exactly what m is, because we will soon let $m_2 \to m_1$, so that $m \to m_1$ also.

- At the left endpoint m_1, the growth rate is $g(m_1)$ per unit time, so individuals will reach size m_1 if their starting size is at least

$$m_1 - g(m_1)\tau$$

  (Note: we will eventually let $\tau \to 0$, and all such individuals will have sizes within $O(\tau)$ of m_1, so if we pretend that they are all size m_1 we make an error of size $O(\tau)$ in their growth rate and therefore an error of size $O(\tau^2)$ in the starting size required to reach size m_1 – and when we let $\tau \to 0$ this error will be negligible).

- Similarly, at the right endpoint individuals grow out of the size interval if their starting size it at least $m_2 - g(m_2)\tau$.

- The number that die is $\mu(m)\tau$ times the number in the interval, which is $\mu(m)\tau n(m,t)(m_2 - m_1)$.

Substituting these expressions into (6.15) we have

$$\begin{aligned}n(m, t+\tau)(m_2 - m_1) = {} & n(m,t)(m_2 - m_1) \\ & + n(m_1, t)g(m_1)\tau \\ & - n(m_2, t)g(m_2)\tau \\ & - \mu(m)n(m,t)(m_2 - m_1)\tau.\end{aligned} \tag{6.16}$$

We now divide through by $(m_2 - m_1)$ in (6.16) and then let $m_2 \to m_1$; the result is

$$n(m_1, t+\tau) = n(m_1, t) - \frac{\partial(gn)}{\partial m}(m_1, t)\tau - \mu(m)n(m,t)\tau. \tag{6.17}$$

Bringing $n(m_1, t)$ over to the left-hand side, dividing through by τ and letting $\tau \to 0$ we get the final form of the model:

$$\frac{\partial n}{\partial t} + \frac{\partial(gn)}{\partial m} = -\mu n \tag{6.18}$$

This is called the *McKendrick-vonFoerster equation* after the people who first applied it to structured population dynamics. But it is actually identical to – and derived in exactly the same way as – the

equation for advective particle transport in fluid mechanics (with m being particle location and g the location-dependent flow velocity). Taking advantage of this analogy, it can be shown that, exactly as in fluid transport, changes in size with a random component can be modeled by adding a diffusion term (involving second derivatives of n with respect to m) to the right-hand side of the equation. In the simplest case, where the variance in growth rate is independent of size, the equation becomes

$$\frac{\partial n}{\partial t} + \frac{\partial (gn)}{\partial m} = -\mu n + D\frac{\partial^2 n}{\partial m^2} \tag{6.19}$$

with $D > 0$ the *diffusion coefficient*. At the level of the basic mathematical description, it doesn't matter whether m refers to the size of an individual or its location in space, so long as $g(m)$ gives the rate of change for all individuals currently "at m". This is one of the reasons why a degree in applied mathematics (or theoretical physics) is good preparation for a career in anything.

To put this into more general perspective: the "primitives" of these structured population models are

- A set of state variables (m, L and W, or whatever) describing the state of the individual (so-called i-state variables)
- A model for how the i-state variables of each individual change over time

From these we derive a "macroscopic" model that tracks the total population in terms of the frequency distribution of i-state variables. At the macroscopic level we are no longer tracking individuals one-by-one, but instead looking at the frequency distribution of the i-state variable across the population as a whole. But the model is derived by making assumptions at the individual level, not at the population level.

Exercise 6.5 Suppose that individuals are classified by two variables, length L and weight W, with growth rate functions $g_L(L, W)$ and $g_W(L, W)$. By generalizing the derivation above (using a box in the (L, W) plane rather than an interval on the m line) show that the population distribution $n(L, W, t)$ satisfies the partial differential equation

$$\frac{\partial n}{\partial t} + \frac{\partial (g_L n)}{\partial L} + \frac{\partial (g_W n)}{\partial W} = -\mu n. \tag{6.20}$$

6.3 Dynamics of stage classes

The general PDEs for age or size-structure dynamics are not simple as they stand. And to understand population dynamics, we have to make things even more complicated by adding density-dependence, interspecific interactions, or other forms of nonlinearity. If we do this and leave the models at the current level of generality, it is hard to derive anything at all about their behavior, and it is not even simple to solve them numerically.

To get around this problem, Gurney, Nisbet and Lawton (1983) proposed simplifying the models by assuming stage-specific vital rate functions. This is similar to, but less extreme than, the assumption in

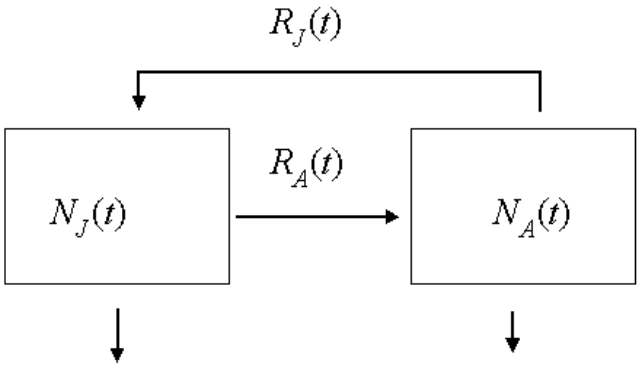


Figure 6.3: Age-structured model in continuous time with a Juvenile stage of fixed duration τ, and an adult stage with constant mortality. R_J and R_A denote the rates of recruitment into the Juvenile and Adult stages, respectively.

a matrix model that all individuals within a stage are identical. The distinction is this: we *will* assume that all juveniles have the same growth rate, possibly as a function of external variables such as food supply. We *will not* assume (as a matrix model would) that all juveniles are the same size, so (for example) smaller juveniles might have a lower probability of maturing into the adult class.

The simplest case is when stages correspond to age classes. As an example to illustrate what kind of model is produced by assuming stage-specific vital rate functions, consider the following simple model that was proposed for Nicholson's blowly experiments with adult food limitation:

- Ages 0 to τ are Juveniles, with constant mortality rate $\mu(a,t) = \mu_J$ and birth rate $b(a,t) = 0$.
- Ages τ and above are Adults, with constant mortality rate $\mu(a,t) = \mu_A$ and birth rate $b(a,t) = qe^{-cN_A(t)}$ where $N_A(t)$ is the total number of adults at time t.

Under these assumptions, differential equations can be derived for the total numbers of Juveniles and Adults, $N_A(t)$ and $N_J(t)$. In the original article (and earlier renditions of these lectures) those equations were derived formally from the definitions

$$N_J(t) = \int_0^\tau n(a,t)da, \quad N_A(t) = \int_\tau^\infty n(a,t)da.$$

You differentiate with respect to t, pull the derivative inside the integrals, use the differential equation

$$\frac{\partial n}{\partial t} + \frac{\partial n}{\partial a} = -\mu(a)n(a,t)$$

to replace the t-derivative with an a-derivative, integrate by parts, $\cdots$. It's not pretty.

Fortunately none of that is necessary. The stage structure of the population, and the rates of transfer between stages, are shown in Figure 6.3. Formally, we have the system of differential equations for the total numbers in each age class

$$\begin{aligned} dN_J/dt &= R_J(t) - R_A(t) - \mu_J N_J(t) \\ dN_A/dt &= R_A(t) - \mu_A N_A(t) \end{aligned} \tag{6.21}$$

The remaining task is to specify R_J and R_A. By assumption we have

$$R_J(t) = qN_A(t)e^{-cN_A(t)}.$$

Moreover, since the juvenile stage lasts exactly τ time units and the mortality rate is constant,

$$\begin{aligned} R_A(t) &= R_J(t-\tau) \times \text{survival through the juvenile stage} \\ &= R_J(t-\tau)e^{-\tau\mu_J} \end{aligned} \tag{6.22}$$

To simplify notation define $S_J = e^{-\tau\mu_J}$; we then have

$$R_A(t) = S_J q N_A(t-\tau)e^{-cN_A(t-\tau)}. \tag{6.23}$$

Since this depends only on N_A we can write a "standalone" equation for the adult population:

$$dN_A(t)/dt = S_J q N_A(t-\tau)e^{-cN_A(t-\tau)} - \mu_A N_A(t). \tag{6.24}$$

To reach this simple model we had to assume that all juveniles have the same rate functions. However, they are still allowed to differ in state: some are nearly mature and will soon become adults, others are recently born and will not become mature for some time. Although the final model only involves the total numbers in each class, its structure reflects the fact that newborns all wait τ time-units before maturing into Adults. The presence of the time delay τ in (6.24) is the price we pay for allowing individuals within stages to differ in state.

6.3.1 Modeling Nicholson's blowflies with adult food-limitation

Gurney et al. (1983) were able to use Nicholson's data to estimate the parameters for model (6.24):

- using the duration of each stage and the stage-specific mortality to estimate the egg-to-adult delay time $\tau \doteq 15.6$ and survival $S_J \doteq 0.91$.
- estimating the egg-production rate by combining data on egg production versus food supply with the assumption that food is divided evenly among adults, to get $b(N_A) \doteq 8.5e^{-N_A/600}$ for the experiments being modeled.
- using the rate of decline in adult population when no recruitment is occurring to estimate the adult mortality rate $\mu_A \doteq 0.27/d$.

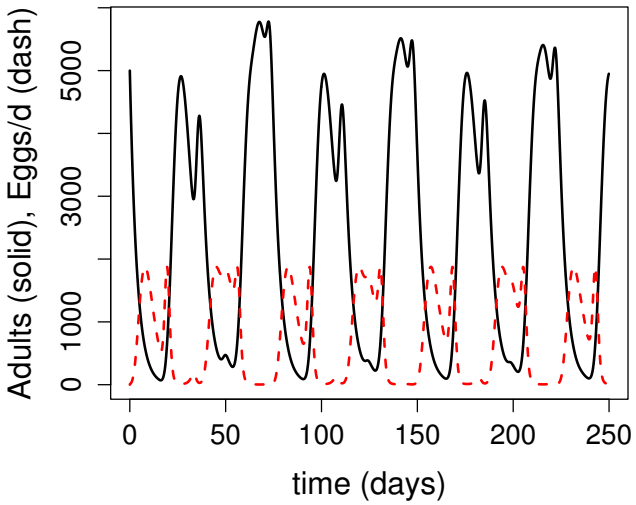


Figure 6.4: Simulation of the blowfly model with adult food limitation, by expressing it as an age-structured model with time and age increments of 0.1 days. The solid line shows the total adult population, and the dashed line is the rate of egg production (eggs/d)

With these estimates, the model produces sustained cycles with a period of about 37 days (compared to an average observed period of about 38 days), and adult population varying between a minimum of 150 and a maximum of 5400 (compared to observed mins and maxes of 270 ± 120 and 7500 ± 500) – pretty good for a model with *zero* free parameters adjusted to fit the adult population data. Moreover, model solutions exhibit the "double peak" that usually occurred in the data.

To simulate the model – without having to face the difficulties of solving delay-differential equations – we can express it as an age-structured model. Because the maturation time is 15.6 days, it is convenient to use time and age increments of 0.1 days. Then individuals of ages 0-156 are juveniles, with zero fecundity and survival probability $S_J^{1/157}$ per time increment. For adults we need only a single age-class: their fecundity per time step is $b(N_A)/10$ and survival probability (remaining in the adult class) $e^{-0.027}$ per time step. These rates specify a density-dependent Lefkovitch matrix of size 158×158. That's larger than what you usually see in the literature, but with a bit of attention to vectorizing a 500-day simulation (Figure 6.4 takes about 1 second in R.

It is instructive to contrast this model with one having discrete stage structure – a continuous-time analog of the stage-based matrix models in the previous chapter. The defining feature of such models is that individuals within a stage are assumed to be identical in *all* respects, including their odds of moving to the next-larger stage between one population census and the next. This is true for adults in (6.24), but juveniles differ in state. The closest possible analog to (6.24) with discrete stage structure would have juveniles leaving the juvenile stage at rate $1/\tau$ – so that the mean duration of the juvenile

stage is τ – and a fraction S_J of those who exit the stage mature into adults while the remainder die.

$$\begin{aligned} dN_J/dt &= qN_A exp(-cN_A) - N_J/\tau \\ dN_A/dt &= S_J N_J/\tau - \mu_A N_A \end{aligned} \tag{6.25}$$

This model can be rescaled into the form

$$\dot{J} = QAe^{-A} - J \qquad \dot{A} = J - \delta A \tag{6.26}$$

and then analyzed by the methods of chapter 4. The conclusion is that there is a positive steady state so long as $Q > \delta$, it is always stable.

So regardless of parameter values, the model with discrete stage structure cannot possibly explain the observed population cycles. The difference in behavior between (6.24) and (6.25) illustrates the principle that a fixed time delay is usually destabilizing – here, relative to a model with an exponential distribution of the maturation time.

Exercise 6.6 Verify the statements made above about (6.26) by finding the steady state and computing the trace and determinant of the Jacobian matrix. Find the conditions on parameters under which the steady state is a stable spiral. For parameters in that range, use simulations of the model to show that cycles damp down very quickly onto the steady state, so that the sustained cycles observed in Nicholson's experiments cannot be explained by the transient behavior of model (6.26).

6.3.2 Modeling *Plodia*

Plodia requires a different model because larvae were food limited rather than adults. It was assumed that egg and pupa stages were short enough to ignore, so the model structure is the same as figure (6.3) except that the J compartment is now labeled A. Because adults are not food-limited, the model assumes constant per-capita mortality and fecundity for adults. Thus the rate of loss from the Adult compartment is $\mu_A A(t)$ and the recruitment rate of new larvae is $R_L(t) = qA(t)$.

The food limitation on larvae was modeled by assuming density-dependent larval mortality,

$$\mu_L(t) = \alpha L(t).$$

This also complicates the calculation of the recruitment into the adult class. We can write

$$R_A(t) = R_L(t-\tau) \times \text{ (larval survival from } t-\tau \text{ to } t).$$

The first term on the right-hand side is $qA(t-\tau)$. The second is

$$S_L(t) \equiv e^{-\int_{t-\tau}^{t} \mu_L(s)ds} = e^{-\alpha \int_{t-\tau}^{t} L(s)ds}. \tag{6.27}$$

So we seem to have a *delay nonlinear integrodifferential equation model* on our hands. Fortunately we don't, thanks to a new trick: writing a differential equation for $S(t)$.

$$\frac{dS_L}{dt} = S_L(t)\left[-\alpha(L(t) - L(t-\tau))\right] = \alpha S_L(t)\left[L(t-\tau) - L(t)\right]. \tag{6.28}$$

Combining this with the balance equations for L and A (below) gives a complete model:

$$\begin{aligned} \frac{dA}{dt} &= qA(t-\tau)S_L(t) - \mu_A A(t) \\ \frac{dL}{dt} &= qA(t) - \alpha L(t)^2 - qA(t-\tau)S_L(t) \end{aligned} \tag{6.29}$$

The initial conditions for this model are somewhat tricky. To start the model running at $t = 0$ we need $A(t)$ and $L(t)$ from times $-\tau$ to 0. Gurney et al. suggest modeling inoculation of the population with new larvae: $A(t) = L(t) = 0$ for $-\tau \leq t \leq 0$ followed by

$$\frac{dL}{dt} = qA(t) + I(t) - \alpha L(t)^2, 0 < t < \tau$$

where $I(t)$ is positive and constant for short time period $0 < t < \epsilon$, and zero thereafter.

Gurney et al. also considered two variants of the model. The first modification is to assume that adults live for a fixed lifespan of τ_A days, and then die. The loss term $\mu_A A(t)$ is then replaced by $R_A(t - \tau_A)$ – deaths now are those individuals who became adults τ_A days ago. The second variant is to assume "cohort competition" among larvae: rather than competing with all larvae for food, larvae compete only with other larvae of the same age. This second variant is equivalent to the blowfly model, because a reduction in the survival of larvae of a given age has exactly the same effect as a reduction in the adult fecundity at their moment of birth.

The "target" that the *Plodia* model aims to hit is the cycle period being nearly the same as the generation length, roughly 42 days (Figure 6.5). The results are pretty good (Figure 6.6); even though the period is not predicted very closely, the qualitative result is right: the larval competition in the *Plodia* model leads to a much shorter cycle period (relative to the lifespan of the organism) than the adult competition in the blowfly model, matching the experimental findings.

6.4 Characteristic cycle periods

The blowfly and *Plodia* models suggest that different modes of population regulation lead to different characteristic values for the ratio between cycle period and maturation time. This appears to be true, as an "empirical fact" in the sense that it holds in a range of different models. The essential difference, first identified by Gurney et al. (1985) is whether the self-regulatory mechanism acts directly on numbers in the regulated stage (e.g.,larval density causes higher larval mortality, hence the numbers of larvae are affected) or acts on the numbers in a later stage (e.g. adult density affects adult fecundity and hence the numbers of individuals in the egg/larval stages, or larval density affects how fecund those individuals will be when they mature).

Figure 6.7 summarizes their findings for 4 models (2 each of immediate/delayed). The basis for these results is an audacious approximation, validated by simulation. The approximation is to use linear stability analysis to compute the period of oscillation at the point where cycles arise by a Hopf bifurcation from a stable equilibirum. The cycle period of the linear system is determined by the point on the

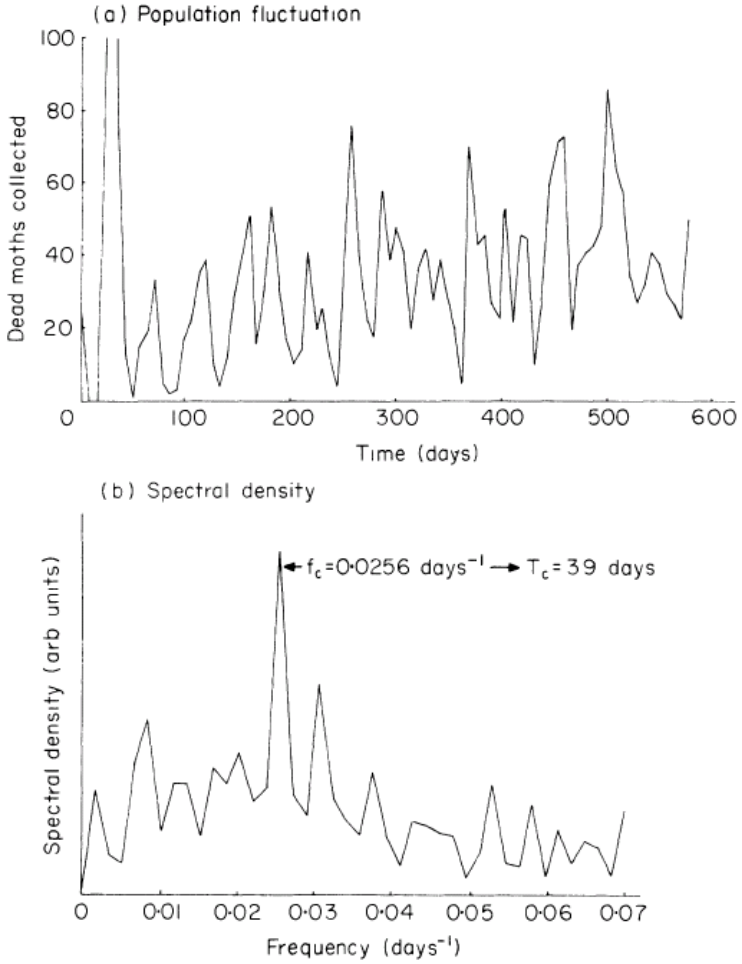


FIG. 5. (a) Weekly counts of dead adult moths in long-term cultures of *Plodia interpunctella*.

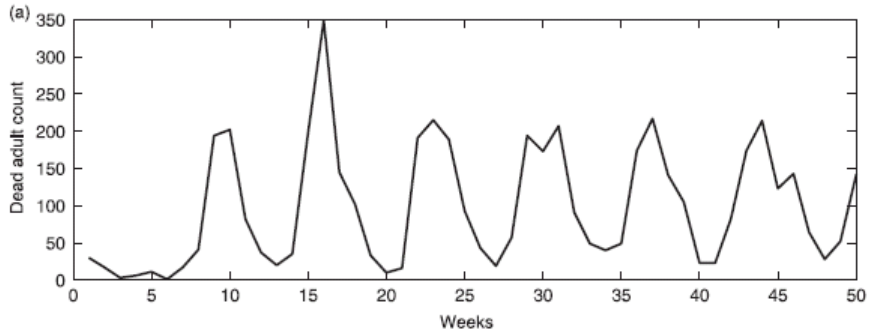


Figure 6.5: Data on laboratory populations of *Plodia interpunctella* with larval food limitation. Top panels are from Gurney et al. (1983), bottom from Wearing et al. (2004).

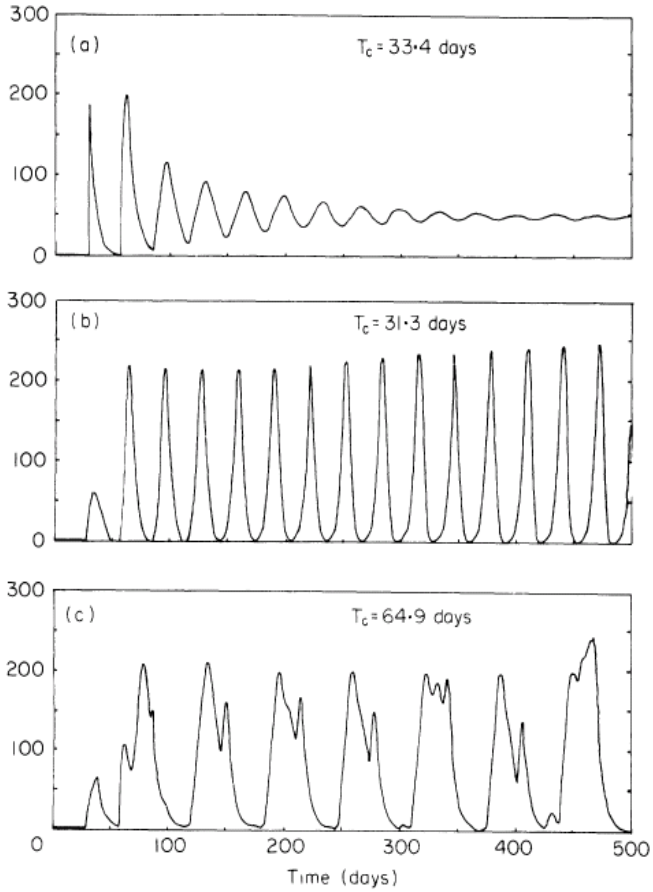


FIG. 4. Numerical solutions of the larval competition model with parameters appropriate to JHL's cultures of *Plodia interpunctella*: (a) uniform competition model with constant adult death rate (Variant (i)); (b) uniform competition model with fixed adult lifetime (Variant (ii)); (c) cohort competition model.

Figure 6.6: Simulations of the three variants of the *Plodia* model, from Gurney et al. 1983.

imaginary axis where the real part of the eigenvalues (a complex conjugate pair) crosses from negative to positive. Just into the unstable region, the cycles amplitude is small (proportional to the square-root of the real part of the eigenvalue, genericall), so the cycle period is approximately that of the linear system. There is no reason why cycle periods must remain more or less the same once the system is in the range of large-scale oscillations, but in fact this is often what happens in population models that undergo a Hopf bifurcation giving rise to limit cycles. As a mathematical exercise it is easy to construct examples where cycle period changes quickly when cycle amplitude increases, but "empirically" in structured population models the cycle period at a Hopf bifurcation point is a good approximation to the period well beyond the bifurcation point.

The difference between immediate and delayed feedback can be understood intuitively, as follows (see Figure 6.8). In each case, we assume that there is a "limiting stage" in the life cycle whose density has a detrimental affect on survival. With direct feedback the limiting stage affects its own abundance; with

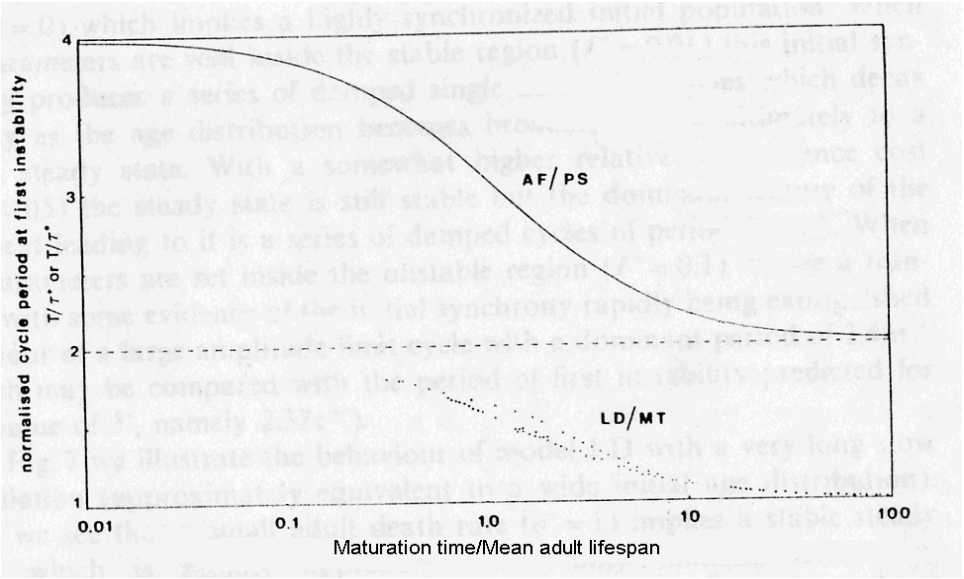


Figure 6.7: Scaled cycle period (relative to the maturation time) in continuous delay-differential models for age-structured populations with competition for limiting food. The x-axis is the ratio between the maturation time and the mean adult lifespan. In the "LD/MT" models the effects of food limitation act directly on the food-limited stage of the life cycle (e.g., food-deprived larvae have higher mortality, or a longer maturation time and hence lower survival to the next stage); these result in cycle periods "1 and a bit" times the maturation time. In the "AF/PS" models the effects of food limitation on the number of individuals is delayed to a later stage in the life cycle (food-deprived larvae grow up to be adults with lower fecundity, or pupae with lower survival); these result in cycle periods "2 and a bit" times the maturation time, up to 4 times the maturation time if mean adult lifespan is much longer than the maturation time.

delayed feedback it affects the abundance of another stage later in the life cycle. This distinction is more subtle than it appears. Consider adults competing amongst themselves for food. If this results in higher adult mortality when food is scarce, that would be direct feedback. If it results in lower adult fecundity when food is scarce, that is delayed feedback: it directly affects the number of eggs, so only after some time is there an effect on the number of adults.

1. With direct feedback, suppose there is a big burst of recruitment into the limiting stage at time t, creating a "dominant cohort". There is then high mortality, killing off everyone else in the stage (and most of the dominant cohort, but not all of them, because they outnumber everyone else). This continues until the dominant cohort matures out of the limiting stage. So a dominant cohort "clears out" everyone else within an age band whose width is twice the duration of the limiting stage. If that's everyone (because the limiting stage is a large fraction of the lifespan), the result is one cycle per generation: the the next big burst of recruitment into the limiting stage will be the offspring of the dominant cohort. If the limiting stage is shorter, one can get shorter periods (e.g., two bursts of recruitment per generation time).

2. With delayed feedback, let τ be the "maturation time" – the time elapsed between leaving the affected stage and entering the limiting stage. If there are many individuals in the limiting stage "now", that clears out the affected stage "now", so there will be few in the limiting stage after the "maturation time" τ. That will allow high survival in the affected stage at time τ, and so there will again be many in the limiting stage after another generation time has passed. So the period of cycles is about 2τ, twice the "maturation time". If we assume that adults are the limiting class and younger juveniles are affected, then τ really is the maturation time, and it will be approximately the generation time.

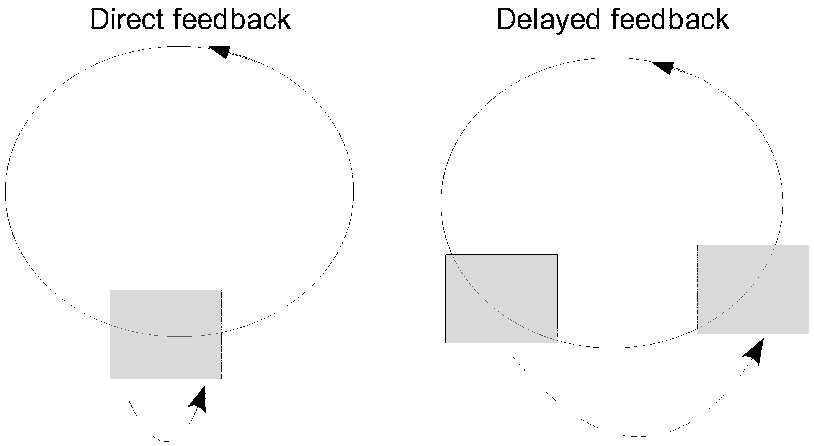


Figure 6.8: Diagrams of life cycles with direct versus delayed feedback on population growth.

6.4.1 Characteristic period for the blowfly model

Here we illustrate how the characteristic cycle period for a given class of structured models can be derived analytically, following Gurney and Nisbet (1998, section 8.6.1). Consider a general version of the blowfly model written in delay differential equation form,

$$\frac{dA}{dt} = R(A(t-\tau)) - D(A(t)) \tag{6.30}$$

where R is recruitment rate and D is death rate. Near and equilibrium A^* we can set $x = A(t) - A^*$ and linearize the model, to get

$$\frac{dx}{dt} = \alpha x(t-\tau) - \gamma x(t) \tag{6.31}$$

where

$$\alpha = \frac{dR}{dA}(A^*), \qquad \gamma = \frac{dD}{dA}(A^*) > 0.$$

Solutions of the linearized models will be sums of exponential terms $e^{\lambda t}$, for values of λ such that the term satisfies (6.31). Substituting $x(t) = e^{\lambda t}$ into (6.31) we get the characteristic equation for λ:

$$\lambda + \gamma = \alpha e^{-\lambda \tau}. \tag{6.32}$$

If $\alpha > \gamma$, then (6.32) has a real, positive root (proof: at $\lambda = 0$ the left-hand side is smaller than the right-hand side. As $\lambda \to \infty$ the right-hand side shrinks to 0 while the left hand side grows to ∞). That kind of monotonic instability is irrelevant to cycles, so for a study of cycles we can assume $\alpha < \gamma$, implying that all real roots are negative.

Cycles arise when a complex conjugate pair of eigenvalues cross the imaginary axis, i.e. $\lambda = \pm i\omega$ (here $i = \sqrt{-1}$). The period of cycles is determined by the value of ω, which we now set out to find. Setting $\lambda = i\omega$ and equating the real and imaginary parts of (6.32), we get

$$\cos(\omega\tau) = \frac{\gamma}{\alpha}, \qquad \sin(\omega\tau) = -\frac{\omega}{\alpha}. \tag{6.33}$$

Because the cosine function is always between -1 and 1, the first condition says that $|\alpha| > |\gamma|$ must hold. And since $\alpha < \gamma$, we conclude that α must be negative and less than $-\gamma$. α is the slope of the recruitment function at the fixed point, so the condition $\alpha < -\gamma$ means that there must be strong *overcompenstation*: adult competition is so severe that higher adult density results in a drop in the total rate of offspring production. Using this in we conclude that

$$\cos(\omega\tau) < 0, \quad \sin(\omega\tau) > 0 \qquad \Longrightarrow \qquad \frac{\pi}{2} \leq \omega\tau \leq \pi.$$

The cycle period is $\dfrac{2\pi}{\omega}$, so the period is between 2τ and 4τ, i.e. 2 to 4 times the maturation period. The maturation delay in those experiments was $\tau \approx 15$ days, so a cycle period of 30 to 60 days is predicted, in line with the observed period of about 38 days.

A strength of this analysis is that it does not require specification of the functional forms for R and D – everything comes from the slopes of these functions at the equilibrium. Its main limitation is that results have to be extrapolated

- from local linearization into the nonlinear domain
- from the bifurcation point (where $Re(\lambda) = 0$) to parameters where $Re(\lambda) > 0$, giving rise to stable oscillations.

6.5 Other applications

This chapter has introduced continuous-time structured population models by way of their applications to the study of population cycles, but this is only one out of many, many current applications of structured

population models. The books edited by Metz and Diekmann (1986) and Tuljapurkar and Caswell (1997) contain a number of examples and reviews of such applications, along with additional aspects of the theory. Here we mention a few examples to indicate the range of possibilities, but for a serious look at structure population models in action you'll need to do some more reading on your own.

Levin and Paine (1974) proposed a model, motivated by rocky intertidal communities, in which a landscape is structured by external disturbances that create patches of open space. Patch creation is assumed to be followed by a deterministic local within-patch succession process. The local community is then viewed as a mosaic of patches in different stages of succession. In their model patches are classified by age a and size ξ, so the state variable is $n(\xi, a, t)$ and the model structure is

$$\frac{\partial n}{\partial t} + \frac{\partial n}{\partial a} + \frac{\partial (gn)}{\partial z} = -\mu(\xi, a) n(\xi, a, t)$$

(SAL may still have the T-shirt with this equation on it presented by his graduate students for his 35th birthday). Paine and Levin (1981) parameterized the model based on Paine's studies of mussel beds on Tattoosh Island, Washington from 1970-1979. In that system patches are created when mussels are dislodged by external disturbances, creating new patches of various sizes. The mortality term μ represents patch disappearance. Small and medium-size patches disappear through incursions by surrounding mussels, while larger patches fill in through recruitment and establishment of larval mussels. With empirically based estimates of patch birth, growth, and disappearance rate functions, the model was able to make accurate predictions of total patch area. Predictions of the age×size distribution were good for larger patches, but small patches were influenced strongly by stochastic effects that are not included in the model.

Murdoch et al. (1996) used stage-structured models for host-parasitoid dynamics to explain the competetive displacement of one parasitoid by another, both having been introduced in order to control red scale in California. The original control agent was unsuccessful. But another member of the same genus, once introduced, displaced the original control agent and provided successful control of red scale. Murdoch et al. (1996) show that this could be explained by a simple difference between the two introduced parasitoids: whichever parasitoid is able to produce female offspring on smaller host individuals is predicted to be the winner in competition and the more successful control agent.

Keeling and Grenfell (1997) examined the failure of conventional epidemic models to explain the "critical community size" (CCS) for measles in Great Britain. The CCS is a phenomenon first noted in the 1950's by Bartlett. Comparing case reports from cities of different sizes, one finds that in cities above a threshold size (about 250,000 to 400,000) the disease is always present: in every weekly reporting period, there are new cases. In cities below the threshold, there are fadeouts (defined by Keeling and Grenfell as 3 consecutive weeks with no case reports, which is long enough for anyone who caught the disease before the fadeout to have recovered from it). Similar patterns have been observed in the US and elsewhere.

Standard epidemic models use a discrete classification by disease state: individuals are either Susceptible, Latent, Infective, or Recovered. Consequently, the standard models do not have the property that individuals who have been sick for some time will recover sooner than recently-infected individuals. Despite numerous elaborations, including age structured disease transmission and spatial and social

structure, these models uniformly overpredict the occurrence of fadeouts, and predict a CCS of about 1 million. Keeling and Grenfell (1997) extended an existing model by incorporating time-within-state, so that individuals are structured both disease state and also by how long they have been in that state, e.g. latent individuals (infected but not yet symptomatic) are represented by $n_L(a,t)$, the where a is the time since the individual was infected by the disease. Compared to the standard models, their model has a lower between-individual variance for the time spent in each disease state. As a result, its dynamics are less irregular than that of the standard models, and it predicts fewer fadeouts. Simulations of the model show that it can account accurately not only for the observed CCS in Great Britain, but also for how the frequency of fadeouts goes up as city size decreases below the CCS.

Size structure is also known to be important for predator-prey interactions between fish species, because of gape limitation: a fish can't eat another fish that is too big for its jaws. Small differences in timing – in predator abundance, how quickly juveniles grow, or whether climatic conditions give them a "head start" in outgrowing potential predators – can therefore be magnified into large differences in juvenile survival (e.g., Persson et al. 1996). As a result, size-structured models are important for predicting the survival of juvenile cohorts in prey species through their period of high vulnerability to predation.

Exercise 6.7 Consider the basic size-structured McKendrick-vonFoerster model

$$\frac{\partial n}{\partial t} + \frac{\partial (gn)}{\partial m} = -\mu n \tag{6.34}$$

that holds when individuals are all born at some minimum size m_0. Suppose instead that individuals give birth only through *fisson* – splitting into two individuals of size αm where m is the size of the "parent" and $0 < \alpha \leq \frac{1}{2}$. Let $\phi(m)$ be the rate (splits/individual/time) at which individuals undergo fission. The range of possible sizes for this model is $(0, \infty)$. Derive the modified equation that applies for this model. Note that when an individual undergoes fission, one individual of size m vanishes, and two individuals of size αm appear. You can pattern your derivation on the one for the basic model, but remember that fission adds a new way to exit, and to enter, any given size range.

Exercise 6.8 Consider the "*Plodia*" model where larval competition causes there to be an increased death rate of larvae. Suppose instead that larval death rate (per unit time) is constant, but that the transition from juvenile to adult occurs at a certain size rather than a certain age, so the duration of the juvenile stage can vary over time. This gives a stage-structured rather than an age-structured model. Specifically, suppose that

- Individuals of size 0 to 1 are *juveniles* with zero reproduction, and constant per-capita mortality rate $\mu_J > 0$.

- Individuals of size $\geq$1 are *adults*, with constant mortality rate $\mu_A > 0$ and constant per-capita fecundity $q > 0$.

- Juveniles grow at rate $g(t) = e^{-cN_J(t)}$; thus the duration of the juvenile stage for individuals

maturing to adulthood at time t, which we will call τ(t), is implicitly defined by

$$\int_{t-\tau(t)}^{t} g(s)ds = 1. \tag{6.35}$$

In this model the effect of crowding is to slow the growth of juveniles, resulting in fewer of them surviving to adulthood.

(a) By differentiating both sides of (6.35) with respect to t, show that $\frac{d\tau}{dt} = 1 - \frac{g(t)}{g(t-\tau)}$.

(b) Show that $\frac{dN_A}{dt} = qN_A(t-\tau)e^{-\mu_J \tau(t)}\frac{g(t)}{g(t-\tau)} - \mu_A N_A(t)$.

(c) Derive the corresponding equation for $\frac{dN_J}{dt}$.

It is not a good idea to try deriving these results from the McKendrick-vonFoerster equation. Think in terms of Figure 6.3 and the rates at which individuals enter and exit each stage. Remember that the juveniles who become adults between times t and $t+dt$ are those whose size at time t is between $1-g(t)dt$ and 1. How much time elapsed between the birth times of the smallest and largest such individuals?

6.6 References

Gurney, W.S.C., R.M. Nisbet, and J.H. Lawton. 1983. The systematic formulation of tractable single-species population models incorporating age structure. Journal of Animal Ecology 52: 479-495.

Gurney, W.S.C. and R.M. Nisbet. 1985. Fluctuation periodicity, generation separation, and the expression of larval competition. Theoretical Population Biology 28: 150 - 180.

Keeling, M.J. and B.T. Grenfell. 1997. Disease persistence and community size: modeling the persistence of measles. Science 275: 65-67.

Levin, S.A. and R. T. Paine. 1974. Disturbance, Patch Formation, and Community Structure. PNAS 71: 2744-2747.

Metz, J.A.J. and O. Diekmann, eds. 1968. The Dynamics of Physiologically Structured Populations. (Lecture Notes in Biomathematics vol. 68). Springer-Verlag.

Murdoch, W. W., C. J. Briggs, and R. M. Nisbet. 1996. Competitive displacement and biological control in parasitoids: a model. American Naturalist 148:807-826

Nicholson, A.J. (1954). An outline of the dynamics of animal populations. Australian Journal of Ecology 2: 9-65.

Nicholson, A.J. (1957). The self-adjustment of populations to change. Cold Spring Harbor Symposia 22: 153-173.

Nisbet, R.M. and W.S.C. Gurney. 1986. The formulation of age-structure models. pp. 95-115 in: T.G. Hallam and S.A. Levin (eds.) Mathematical Ecology: an Introduction (Biomathematics vol. 17) Springer-Verlag, NY.

Paine,R.T. and S. A. Levin. 1981. Intertidal Landscapes: Disturbance and the Dynamics of Pattern Ecological Monographs 51: 145-178.

Persson, L., J. Andersson, E. Wahlstrom and P. Eklov. 1996. Size-Specific Interactions in Lake Systems: Predator Gape Limitation and Prey Growth Rate and Mortality. Ecology 77: 900-911.

Tuljapurkar, S. and H. Caswell, eds. 1997. Structured-population Models in Marine, Freshwater, and Terrestrial Systems. Chapman and Hall, New York.

Wearing, H., S.M. Sait, T.C. Cameron and P. Rohani. 2004. Stage-structured competition and the cyclic dynamics of host-parasitoid systems. Journal of Animal Ecology 73: 706-722.

Chapter 7

Some mathematical background

7.1 Some differential calculus

This section won't serve as a substitute for a calculus class, but it should serve to remind you of some things that you need to remember from your calculus class.

Differential calculus is a mathematical language for working with quantities that change continuously. We are concerned in these lectures with two main examples:

1. The change over time (t) in a variable $n(t)$ describing some aspect of an ecological system
2. changes in some system property $M(p)$ as the value of some parameter p (predator death rate, harvesting effort, etc.) is altered.

Our first question, for some quantity $f(x)$, is how a small change in the value of x affects the value of f. The answer is given by the *derivative* of f,

$$df/dx = \text{limiting value as } \varepsilon \to 0 \text{ of } \frac{f(x+\varepsilon) - f(x)}{\varepsilon}. \tag{7.1}$$

The sign of the derivative tells us whether f is increasing at x (if $df/dx > 0$) or decreasing (if $df/dx < 0$). The numerical value of the derivative tells us how fast this is happening (the "velocity" of f).

Other notations for the derivative include $f'(x)$, $Df(x)$, $D_x f$ and $\dot{f}$ (the last indicating a derivative with respect to time). All of these are commonly used.

We also need the *second derivative*, which is the derivative of the derivative. The second derivative is the rate of change in the rate of change, i.e., the acceleration. Notations for the second derivative includes $f''(x)$, $D^2 f(x)$, $D_{xx} f$ and $\ddot{f}$ for second derivatives with respect to time.

Some important properties for us are:

Function $f(x)$	Derivative $f'(x)$
x^n	nx^{n-1}
e^x	e^x
$\log(x)$	$1/x$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$

Table 7.1: Derivatives of some functions. Note that log means *natural* logs with base e.

1. At a value of x where $f(x)$ has a minimum or maximum, we have $f'(x) = 0$. If $f'' > 0$ it's a minimum, and if $f'' < 0$ it's a maximum.
2. If $f(x)$ is constant on an interval $a < x < b$ then $f'(x) = 0$ on that interval, and vice-versa.
3. If $f'(x)$ is positive for all x in an interval $[a, b]$ then f is increasing on that interval — if $a \leq x < y \leq b$ then $f(y) > f(x)$. Similarly, if $f'(x)$ is negative for all x in $[a, b]$ then f is decreasing on that interval.

Property (1) is needed in analyzing optimality models (as in: which value of p_j gives a forager the highest rate of energy gain per unit time?). We use property (2) in analyzing differential equation models.

In order to use these properties of derivatives, we sometimes need to be able to find the derivative of a specific function. Table 7.1 lists a few important simple examples. More complicated cases can often be reduced to these using a few simple rules:

1. **Scalar Rule:** If c is a constant, $\frac{d}{dx}(cf(x)) = cdf/dx$
2. **Sum Rule:** $\frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx} + \frac{dg}{dx}$ (derivative of sum = sum of derivatives)
3. **Product Rrule:** $\frac{d}{dx}(f(x)g(x)) = f\frac{dg}{dx} + g\frac{df}{dx}$
4. **Quotient Rule:** $\frac{d}{dx}\frac{u}{v} = \frac{v(du/dx) - u(dv/dx)}{v^2}$
5. **Chain Rule:** $\frac{d}{dx}(f(g(x)) = \frac{df}{dg}(g(x)) \times \frac{dg}{dx}(x)$

The Chain Rule is often written in the easy-to-memorize form

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}. \tag{7.2}$$

To illustrate the basic rules: we know that $\frac{d}{dx}(x^2) = 2x$ and $\frac{d}{dx}(\sin(x)) = \cos(x)$. From these we get

1. $\frac{d}{dx}(3x^2) = 3 \times 2x = 6x$ (Scalar Rule)
2. $\frac{d}{dx}(x^2 + \sin(x)) = 2x + \cos(x)$ (Sum Rule)

3. $\dfrac{d}{dx}(x^2\sin(x)) = x^2\cos(x) + 2x\sin(x)$ (Product Rule)

4. $\dfrac{d}{dx}\dfrac{\sin(x)}{x^2} = \dfrac{x^2\cos(x) - \sin(x)\times 2x}{(x^2)^2} = \dfrac{x\cos(x) - 2\sin(x)}{x^3}$ (Quotient Rule)

We need the Chain Rule for functions like $f(x) = \sin(x^2)$. Here $g(x) = x^2$, and $f(g) = \sin(g)$. Using the Chain Rule in the form of (7.2), we have

$$\frac{df}{dg} = \cos(g) = \cos(x^2), \qquad \frac{dg}{dx} = 2x, \tag{7.3}$$

and therefore

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx} = \cos(x^2)\times 2x = 2x\cos(x^2). \tag{7.4}$$

When using (7.2) it's important to remember that $\frac{df}{dg}$ is a function of x and has to be written that way using the definition of g. In our example above, this was done in the step $\cos(g) = \cos(x^2)$.

7.1.1 Partial derivatives

"Partial derivative" means that we have a function of several variables, and we need to know its rate of change in response to each variable that affects it. How you find partial derivatives is simple: one variable at a time, treating all other variables as constant. A partial derivative is indicated by ∂, for example $\partial f/\partial x$ denotes the partial derivative of f with respect to x

For example, consider $F(x, y) = x^2y$. To take its partial derivative with respect to x, we regard y as a constant. The first of our Rules above then applies:

$$\partial F/\partial x = y\times\partial(x^2)/\partial x = y\times 2x = 2xy \tag{7.5}$$

Note that for functions of a single variable (such as x^2) the partial derivative is the same as the ordinary derivative. Now what is $\partial F/\partial y$?

We can also take partial second derivatives, which can be "pure" (e.g., $\partial^2 F/\partial x^2$) or "mixed" ($\partial^2 F/\partial x\partial y$). In the latter case, the order doesn't matter: you can first find $\partial F/\partial x$ and then take the derivative of the result with respect to y, or do it in the reverse order (first y, then x).

Of course this new concept brings some new notation. $\partial f/\partial x$ can also be written as $D_x f$ or f_x. For higher partials we have (with the obvious meanings) $D_{xx}f, D_{xy}f$ and so on, or f_{xx}, f_{xy} and so on.

Exercise 7.1 Find the derivatives of the following functions (note that unless otherwise specified log

means the base-e "natural" logarithm, and anything else will be specified explicitly, e.g. $\log_{10}$):

(1) $f(t) = t^3$

(2) $g(t) = 2 + t^3 - 5t$

(3) $H(x) = x^2 \log(x)$

(4) $P(x) = 3x^3 - x^2 \log(x)$

(5) $x(t) = \dfrac{1+7t}{1+t^2}$

(6) $f(x) = 2x(1 - x/6)$

Exercise 7.2 Find the partial derivatives of each function with respect to each of its variables.
(7) $f(x, y) = 1 + 2x^2 + 3x \sin(y)$
(8) $g(x, y) = 1 + ax + by + cx^2 + dxy + ey^2$ where a, b, c, d, e are constants
(9) $P(n, V, T) = nRT/V$, where R is a constant.

Exercise solutions

1. $f'(t) = 3t^2$

2. $g'(t) = 3t^2 - 5$

3. $H'(x) = x^2 \left(\dfrac{1}{x}\right) + 2x \log(x) = x + 2x \log(x)$

4. $P'(x) = 9x^2 - H'(x) = 9x^2 - x - 2x \log(x)$

5. $x'(t) = \dfrac{(1+t^2)(7) - (1+7t)(2t)}{(1+t^2)^2} = \dfrac{7 - 2t - 7t^2}{1 + 4t^2 + t^4}$

6. Using the product rule $f'(x) = 2(1 - x/6) + 2x(-1/6) = 2 - x/3 - x/3 = 2 - 2x/3$

7. $\dfrac{\partial f}{\partial x} = 4x + 3\sin(y) \quad \dfrac{\partial f}{\partial y} = 3x \cos(y)$

8. $\dfrac{\partial g}{\partial x} = a + 2cx + dy \quad \dfrac{\partial g}{\partial y} = b + dx + 2ey$

9. $\dfrac{\partial P}{\partial n} = RT/V$ since RT/V is treated as a constant

   $\dfrac{\partial P}{\partial T} = nR/V$ since nR/V is treated as a constant

   $\dfrac{\partial P}{\partial V} = nRT \times \dfrac{\partial}{\partial V}\dfrac{1}{V} = -\dfrac{nRT}{V^2}$

7.2 Complex numbers

A complex number z consists of a *real part*, which is an ordinary real number, and an *imaginary part* which is a multiple of $i = \sqrt{-1}$; for example $3 + 5i$ has real part 3, and imaginary part $5i$.

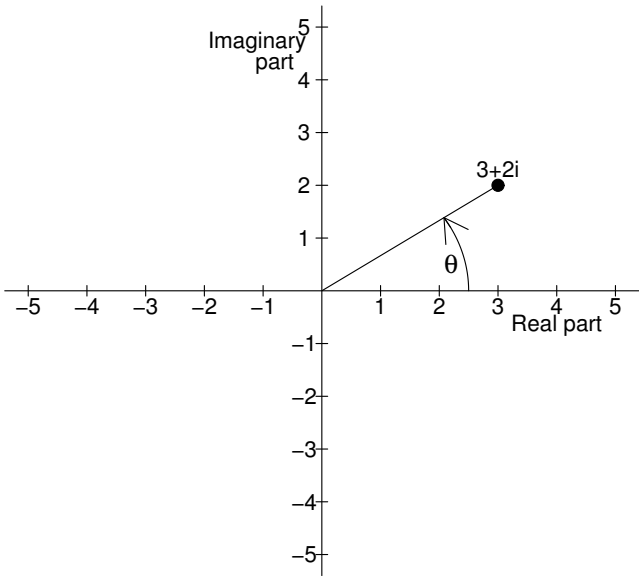


Figure 7.1: Representation of complex numbers as points in the plane.

Complex numbers can be visualized as points in the *complex plane* (Figure 7.1), with the real part on the horizontal axis, and imaginary part on the vertical. This depiction of complex numbers also clears up the question "what is the square root of -1, anyway?" Complex numbers are a way of taking the arithmetic operations on real numbers – addition, subtraction, multiplication and division – and defining them for points in the plane. $\sqrt{-1}$ is the vector such that you get -1 (i.e. $-1+0i$ in the complex plane) when you multiply that vector by itself.

The arithmetic of complex numbers is defined by treating them like ordinary numbers and applying the rule $i^2 = -1$.

$$\begin{aligned}
(a + bi) + (c + di) &= (a + c) + (b + d)i \\
(a + bi) - (c + di) &= (a - c) + (b - d)i \\
(a + bi) \times (c + di) &= (a \times c) + (a \times di) + (bi \times c) + (bi \times di) \\
&= ac + (ad + bc)i + bd(-1) \\
&= (ac - bd) + (ad + bc)i
\end{aligned} \tag{7.6}$$

For division it is best to use an alternate representation for complex numbers using polar coordinates and the exponential function,

$$z = re^{i\theta} \tag{7.7}$$

where r is the distance from the origion to z, and θ is the counter-clockwise angle between the real axis and the line running from the origin to z (See Figure 7.1).

To derive (7.7), recall the Taylor series expansion

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots .$$

Applying this to an imaginary number θi, and using the Taylor series for the sine and cosine functions, we get

$$\begin{aligned} e^{\theta i} &= 1 + \theta i - \frac{\theta^2}{2!} - i\frac{\theta^3}{3!} + \frac{\theta^4}{4!} + \cdots \\ &= (1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots) + i(\theta - \frac{\theta^3}{3!} + \cdots) \\ &= \cos(\theta) + i\sin(\theta), \end{aligned} \tag{7.8}$$

Therefore

$$re^{i\theta} = r(\cos(\theta) + i\sin(\theta)) = a + bi$$

with

$$\begin{aligned} a = r\cos(\theta), \quad b = r\sin(\theta), \\ \text{and therefore} \quad r = \sqrt{a^2 + b^2}, \quad \theta = \tan^{-1}(b/a). \end{aligned} \tag{7.9}$$

The quantity $r = \sqrt{a^2 + b^2}$ is often called the *absolute value*, the *modulus* or the *magnitude* of the complex number $z = a + bi$, and θ is sometimes called the *argument*.

Finally, we can use the polar-coordinates representation to define division for complex numbers. If $z_1 = r_1 e^{i\theta_1}, z_2 = r_2 e^{i\theta_2}$, then

$$\frac{z_1}{z2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2}\frac{e^{i\theta_1}}{e^{i\theta_2}} = \frac{r_1}{r_2}e^{i(\theta_1 - \theta_2)}. \tag{7.10}$$

The polar representation is also convenient for multiplication:

$$z_1 z_2 = r_1 e^{i\theta_1} \times r_2 e^{i\theta_2} = r_1 r_2 e^{i\theta_1} e^{i\theta_2} \quad = r_1 r_2 e^{i(\theta_1 + \theta_2)}. \tag{7.11}$$

Stability analysis for differential equations leads to complex numbers of the form $e^{\lambda t}$ where $\lambda = a + bi$ is an eigenvalue of the Jacobian matrix. The behavior of these as $t \to \infty$ is a combination of exponential growth or decay, governed by the real part of λ, and sinusoidal oscillations whose rate depends on the imaginary part:

$$e^{\lambda t} = e^{(a+bi)t} = e^{at+bti} = e^{at}(\cos(bt) + i\sin(bt)), \tag{7.12}$$

which implies that

$$|e^{\lambda t}| = e^{at}.$$

So if $a > 0$ there is exponential growth of $e^{\lambda t}$ as $t \to \infty$, and if $a < 0$ there is exponential decay to 0.

This is one of the many reasons why complex numbers are useful. Solutions to linear dynamic models involve combinations of exponential growth/decay and sinusoidal oscillations (think of a damped pendulum, for example), and complex numbers allow us to express these together in a simple form and work with them using arithmetic operations that behave just like ordinary arithmetic.

Stability analysis for difference equations leads to complex numbers of the form λ^t where λ is an eigenvalue of the Jacobian matrix. In this case it is better to use the polar representation $\lambda = re^{i\theta}$:

$$\lambda^t = (re^{i\theta})^t = r^t e^{i\theta t} = r^t(\cos(\theta t) + i\sin(\theta t)). \tag{7.13}$$

which implies that

$$|\lambda^t| = r^t = |\lambda|^t.$$

So if $|\lambda| > 1$ there is exponential growth of λ^t, and if $|\lambda| < 1$ there is exponential decay to 0.

7.3 Matrix algebra

A matrix is a rectangular array of numbers; a vector is a list of numbers.

Addition of matrices or vectors is element-by-element (which requires that the items being added are the same size and shape); ditto subtraction, and multiplication of a vector or matrix by a number. Examples:

$$\begin{gathered}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 6 & 10 \\ 4 & 8 & 12 \end{bmatrix} = \begin{bmatrix} 1+2 & 2+6 & 3+10 \\ 4+4 & 5+8 & 6+12 \end{bmatrix} \\ 2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \qquad 2[1,1,2,2] = [2,2,4,4]\end{gathered} \tag{7.14}$$

Two vectors v, w of the same size, or two matrices of the same size, can be multiplied element-by-element to yield another vector or matrix. This is called the *Hadamard product* and is usually indicated by an open circle, $\circ$. Example:

$$(1,2,3) \circ (2,3,4) = (1 \times 2, 2 \times 3, 3 \times 4) = (2,6,12) \tag{7.15}$$

The **dot product** or **inner product** of two vectors v, w of the same size is the sum of the entries in their product. Notation varies: $\langle v, w \rangle$, (v, w) and $v \bullet w$ are often used.

$$\langle (1,2,3), (2,3,4) \rangle = (1 \times 2 + 2 \times 3 + 3 \times 4) = (2 + 6 + 12) = 20. \tag{7.16}$$

Matrix-vector and matrix-matrix multiplication are **not** done element-by-element.

Matrix-vector multiplication: If $\mathbf{A}$ is a matrix and x a vector, then

$y = \mathbf{A}x$ is a vector whose i^{th} entry is $y[i] = \langle \mathbf{A}[i,], x \rangle$.

That is: to compute $\mathbf{A}x$, take the inner product of x with each row of $\mathbf{A}$, and combine those values into a vector. Consequently

- $\mathbf{A}x$ is only defined if the length of x equals the number of columns of $\mathbf{A}$
- The length of $\mathbf{A}x$ will equal the number of rows of $\mathbf{A}$
- In brief: an $m \times n$ matrix takes n-vectors to m-vectors.

$$\text{Example}: \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} \langle (1,2,3),(1,3,5) \rangle \\ \langle (4,5,6),(1,3,5) \rangle \end{bmatrix} = \begin{bmatrix} 1+6+15 \\ 4+15+30 \end{bmatrix} = \begin{bmatrix} 22 \\ 49 \end{bmatrix} \tag{7.17}$$

Here a 2×3 matrix (2 rows, 3 columns) has taken a 3-vector to a 2-vector.

We can see from the above that there is another, equivalent way to define matrix-vector multiplication: $\mathbf{A}x$ is the vector

$$(x[1] \times \text{1st column of } \mathbf{A}) + (x[2] \times \text{2nd column of } \mathbf{A}) + \cdots (x[n] \times \text{last column of } \mathbf{A}).$$

Matrix-vector multiplication is a **linear operation**. That is, if x, y are vectors and a, b are numbers, then

$$\mathbf{A}(ax + by) = a(\mathbf{A}x) + b(\mathbf{A}y) \tag{7.18}$$

and the same interchange of operations holds for any number of summed terms. This is a *very important property*. In fact, any linear operation on vectors can be expressed as multiplication by some matrix.

Matrix multiplication: the product $\mathbf{C} = \mathbf{AB}$ is defined if the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$. If $\mathbf{A} = (a_{ij})$ has size $m \times n$ and $\mathbf{B} = (b_{ij})$ has size $n \times r$, then $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ has size $m \times r$ and

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} \tag{7.19}$$

Note that if $\mathbf{B}$ has only one column, this reduces to the definition of matrix-vector multiplication (**??**). Thus, another definition of matrix multiplication is the following:

$$k^{th} \text{ column of } \mathbf{AB} = \mathbf{A} \times (k^{th} \text{ column of } \mathbf{B}) \tag{7.20}$$

For these lectures, the best attitude is that matrix multiplication is done by the computer. It is important to understand the conceptual definition (7.20) and the formula (7.19). But when you are working with actual numbers and anything bigger than a 2×2 matrix, it is easier and less error-prone to enter the matrices and do the multiplication using a matrix programming language such as R or Matlab.

Matrix operations share many properties with the familiar arithmetic of real numbers. For example,

- Matrix addition is associative $[\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}]$ and commutative $[\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}]$.
- Matrix multiplication is associative $[\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}]$ and distributive over addition $[\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}, (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}]$.

However, *matrix multiplication is not commutative* – typically $\mathbf{AB} \neq \mathbf{BA}$. Indeed, unless $\mathbf{A}$ and $\mathbf{B}$ are square matrices of the same size, either one of the products $\mathbf{AB}$ and $\mathbf{BA}$ will be undefined, or the two products will be matrices of different sizes. But even in the case of square matrices commutativity typically does not hold. Here is a simple example:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

but

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

However, scalar multiplication is commutative in the sense that $\mathbf{A}(c\mathbf{B}) = c(\mathbf{AB})$.

7.3.1 Eigenvectors and eigenvalues

We will be dealing exclusively with *square* matrices, meaning that the number of rows equals the number of columns. A square matrix with n rows and n columns can multiply n-vectors, and the result is another n-vectors.

A number (possibly complex) λ is an *eigenvalue* of a square matrix $\mathbf{A}$ if there is a non-zero vector $\mathbf{w}$ such that $\mathbf{Aw} = \lambda\mathbf{w}$, and $\mathbf{w}$ is called the corresponding eigenvector. Eigenvectors are only defined up to scaling factors: if $\mathbf{w}$ is an eigenvector for λ then so is $c\mathbf{w}$ for any number $c \neq 0$. An $n \times n$ matrix $\mathbf{A}$ must have at least 1 eigenvalue-eigenvector pair, and it can have up to n. The typical situation is to have n distinct eigenvalues each with a corresponding eigenvector. This is typical in the sense that if matrix entries are chosen at random according to some smooth probability distribution, the probability that the resulting matrix has n distinct eigenvalues is 1.

There is a useful formula for the eigenvalues of a 2×2 matrix $\mathbf{A}$. If $T = a_{11} + a_{22}$ is the *trace* (sum of diagonal elements) and $\Delta = a_{11}a_{22} - a_{12}a_{21}$ is the *determinant* then the eigenvalues are

$$\lambda_{1,2} = \frac{1}{2}\left(T \pm \sqrt{T^2 - 4\Delta}\right) \tag{7.21}$$

7.4 Random variables

The bet-hedging models in Chapter 2 are based on viewing environmental conditions and the birth and death rates that result from them as *random variables*: quantities whose value cannot be predicted in advance with certainty. Random variables are characterized by the set of possible outcomes, and the associated *probability* of different possible outcomes.

Random variables come mainly in two flavors, discrete and continuous. A **discrete random variable** X has a list of possible values $\{x_1, x_2, x_3, \cdots\}$ and is characterized by its density function $p_k = p(x_k) = P[X = x_k]$.

A **continuous random variable** X can take any value in some interval (L, U) which may be finite or infinite in length, and is characterized by its density function $p(x)$ which has the property that

$$P[a \leqslant X \leqslant b] = \int_a^b p(x)dx \tag{7.22}$$

Some important continuous random variables are the **Uniform**, **Normal** (also called **Gaussian**) and **Exponential**.

- Uniform on $[a, b]$: $p(x) = 1/(b-a), a \leq x \leq b$.
- Exponential with rate λ: $p(x) = \lambda \exp(-\lambda x), x \geq 0$.
- Gaussian(μ, σ^2): $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-x^2/(2\sigma^2))$

Two flavors of random variable imply a third: their mixture (just like with ice cream). So it is possible to have random variables that can take values in a discrete list, or in an interval. We don't have any need for these, so from here out we'll assume that any random variable is either discrete or continuous.

Two random variables X, Y are **independent** if

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]. \tag{7.23}$$

This says that X and Y do not interact: the relative frequency of different outcomes for Y is unaffected by the value of X.

The **mean** or **expected value** of X, denoted by $E[X]$, μ_X, or $\bar{X}$, represents the long-term average over many repetitions of the same "random experiment" (average rainfall, average temperature, etc. are examples of expected values). The mean is computed as

$$\begin{aligned} E[X] &= \sum_j p_j x_j \qquad \text{(discrete)} \\ E[X] &= \int_L^U xp(x)dx \qquad \text{(continuous)} \end{aligned} \tag{7.24}$$

The **variance** of X, denoted Var(X) or σ_X^2, is

$$\sigma_X^2 = E[(X - \mu_X)^2]. \tag{7.25}$$

The square root of the variance, σ_X, is called the **standard deviation**. These measure how far X tends to stray from its mean value. The standard deviation is on the same scale as the variable itself (i.e., it has the same units as the variable); the variance has the units of the variable squared.

Rules about Means and Variances

1. If c is a constant, $E[c] = c$ and $Var(c) = 0$.

2. $E[cX] = cE[X]$ for any constant c.
3. $E[X_1 + X_2 + \cdots X_n] = E[X_1] + E[X_2] + \cdots E[X_n]$ **always**.
4. $E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n]$ **if the X_i are mutually independent.**
5. $Var(X) = E[X^2] - E[X]^2$ (so $E[X] = 0 \Longrightarrow Var(X) = E[X^2]$)
6. $Var(cX) = c^2 Var(X)$ for any constant c
7. $Var[X_1 + X_2 + \cdots X_n] = Var[X_1] + Var[X_2] + \cdots Var[X_n]$ **if the $X_i$are mutually independent.**
8. $(X - \mu_X)/\sigma_X$ has mean=0, and variance=1.

The Law of the Unconscious Statistician: the random variable $f(X)$ has mean given by

$$\begin{aligned} E[f(X)] &= \sum_j p_j f(x_j) \qquad \text{(discrete)} \\ E[f(X)] &= \int_L^U f(x)p(x)dx \qquad \text{(continuous)} \end{aligned} \tag{7.26}$$

This is applied in Chapter 2 to compute expected logarithm of population growth rate.

Small Variance Approximations: if $0 < \sigma_X \ll 1$, then

$$E[f(X)] \approx f(\mu_X) + \frac{1}{2} f''(\mu_X)\sigma_X^2 \tag{7.27}$$

and

$$Var(f(X)) \approx f'(\mu_X)\sigma_X^2 \tag{7.28}$$

Derivation: from Taylor Series expansion for f about μ_X:

$$f(x + \varepsilon) = f(x) + f'(x)\varepsilon + \frac{f''(x)}{2!}\varepsilon^2 + \cdots . \tag{7.29}$$

For a random variable X, let

$$z = (X - \mu_X)/\sigma_X .$$

Then we have $X = \mu_X + \sigma_X z$, and our rules for means and variances imply that z has mean 0, and variance 1. Then using the first 3 terms in the Taylor series with $x = \mu_X, \varepsilon = \sigma_X z$ we have

$$f(X) \doteq f(\mu_X) + f'(\mu_X)\sigma_X z + \frac{f''(\mu_X)}{2}\sigma_X^2 z^2. \tag{7.30}$$

Then to get (7.27) we take the mean of both sides of (7.30). Deriving (7.28) is even easier: keep only the first two terms in (7.29) and compute the variance of both sides.

Jensen's Inequality If $f''(x) \leq 0$ for all possible values of X, then $E[f(X)] \leq f[E(X)]$ (with strict inequality if $f'' < 0$ and $Var(X) > 0$)

The Small Variance Approximation says that Jensen's Inequality holds for small variance, but in fact it is always true. The intuituve meaning of Jensen's inequality is that risk-aversion is always beneficial when the utility function has diminishing returns.

Exercise 7.3 (a) Derive the small-variance approximations for $E[e^X]$ and $E[\sin(X)]$, where X is a Gaussian random variable with mean $E[X] = 1$ and variance σ^2. (b) Use simulations to see how well this works. That is, for a range of σ values going from "small" to "large", estimate $E[e^X]$ and $E[sin(X)]$ by the average over 50,000 randomly generated values of X (using **rnorm**), and explore how the error of the small variance approximation drops as σ goes down. A graphical summary of your results (with verbal interpretation and conclusions) will be especially effective.

7.4.1 Limit theorems for random variables

Let $\{X_1, X_2, \cdots\}$ be **independent, identically distributed** random variables with common mean μ and variance σ^2. Then our rules above give,

$$\begin{aligned} E\left[\frac{X_1 + X_2 + \cdots X_n}{n}\right] &= \mu \\ Var\left[\frac{X_1 + X_2 + \cdots X_n}{n}\right] &= \frac{\sigma^2}{n}. \end{aligned} \tag{7.31}$$

As more and more independent X's are averaged, the average of their average stays constant, but the variance of their average goes down. As a result we have (though it takes some work to prove these results mathematically):

Strong Law of Large Numbers: $\left[\frac{X_1 + X_2 + \cdots X_n}{n}\right] \to \mu$ as $n \to \infty$.

Weak Law of Large Numbers: For any $\varepsilon > 0$, $P\left[\left|\frac{X_1+X_2+\cdots X_n}{n} - \mu\right| > \varepsilon\right] \to 0$ as $n \to \infty$.

We can also do a different scaling of the sums in the numerator of (7.31) so that the variance remains constant:

$$\begin{aligned} E\left[\frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}}\right] &= 0, \\ Var\left[\frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}}\right] &= 1. \end{aligned} \tag{7.32}$$

What happens in this scaling is that the shape of the variability (for the scaled sums on the left hand side of (7.32)) converges to a constant.

Central Limit Theorem. As $n \to \infty$, the distribution of $\left[\frac{X_1+X_2+\cdots X_n-n\mu}{\sigma\sqrt{n}}\right]$ converges to $N(0,1)$, the Normal (Gaussian) distribution with mean=0, variance=1.

These limit theorems can be summarized informally by writing

$$X_1 + X_2 + \cdots X_n \sim n\mu + \sqrt{n}\sigma Z \tag{7.33}$$

where Z is an Gaussian random variable with mean 0 and variance 1. The meaning of $\sim$ is (roughly)

that the *relative* difference between the two sides goes to zero, however the *absolute* difference generally does not go to zero.