MA30047

Mathematical Biology 1

Ben Adams[1]

Centre for Mathematical Biology
Department of Mathematical Sciences
University of Bath

2019/20, Semester 1

[1]These notes are an evolution of earlier versions by Nicholas Britton and Frank Hilker. The copyright rests with the authors. Please do not distribute them beyond the University of Bath.

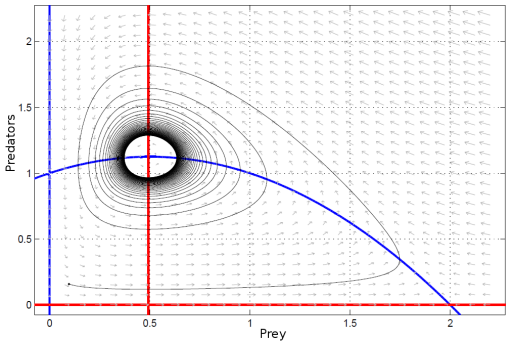
Introduction and Overview

Philosophy and science

> "This is the foundation of all. We are not to imagine or suppose, but to *discover*, what nature does or may be made to do."
> – Francis Bacon (1561–1626)

> "Mathematics is biology's next microscope, only better.
> Biology is mathematics' next physics, only better."
> – Joel Cohen (1944 –)

Mathematical biology is Bacon's process of discovery using Cohen's microscope. It involves constructing, analysing and interpreting mathematical models of biological process. It uses mathematics to provide insights into/ask questions about/test hypotheses regarding biological systems. It typically involves complex and nonlinear mathematical problems



Course description

Aims

- Introduce problems in population biology that can be tackled with applied maths
- Cover both mathematical modelling and mathematical analysis
- Emphasise the interplay between the mathematics and the underlying biology

Intended Learning Outcomes

- Handle mathematical modelling issues for problems in population biology
- Analyse models written as ordinary differential equations or difference equations
- Give a qualitative and quantitative account of their solution
- Interpret the results in terms of the original biological problem

Skills

- Mathematical modelling in biology
- Problem solving
- Written communication and group work

Content

- Single species population dynamics:
  - continuous-time and discrete-time models
  - harvesting
  - spruce budworm model: hysteresis effect
- Interacting populations:
  - predation
  - competition
  - host–parasitoid systems
- Infectious diseases:
  - epidemics
  - endemic circulation
  - vector-borne diseases

Recommended textbooks

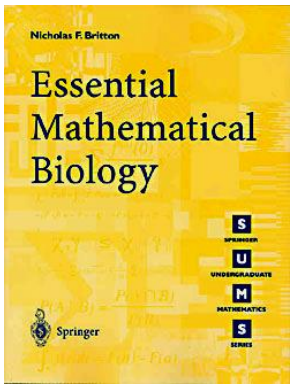
The recommended textbook is

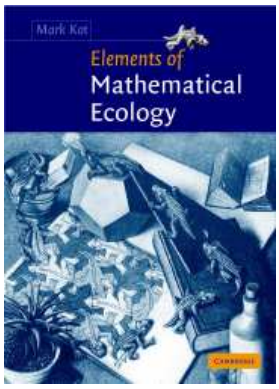
N.F. Britton, Essential Mathematical Biology, Springer, 2003.

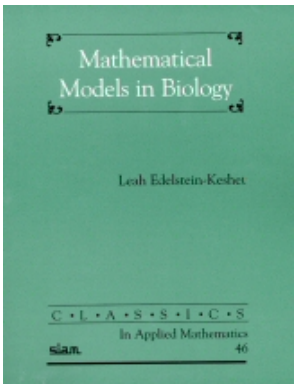
There are many other excellent mathematical biology textbooks

M. Kot, Elements of Mathematical Ecology, Cambridge University Press, 2001.

L. Edelstein-Keshet, Mathematical Models in Biology, Random House, 1988, reprinted by SIAM, 2005.







Chapter 1

Single-species population dynamics in continuous time

1.1 Introduction

How do birth and death processes determine the rate of growth or decay of a population? The goal is to reveal underlying principles of population dynamics and to deduce management strategies e.g. for harvesting. We shall consider models in continuous time. We shall assume that there is no spatial variation in the population (or at least that it is not important), and ignore any other form of structure, such as age or life stages. The advantage of unstructured models is their simplicity. As we add more biological detail the models become more realistic but also more challenging. A good strategy to begin with is KEEP IT SIMPLE.

To model population dynamics we use book-keeping for the population i.e. (for $t_1 > t_0$)

$$\begin{aligned}\text{Population at } t_1 = \text{Population at } t_0 \\ + \text{Births in } (t_0, t_1) - \text{Deaths in } (t_0, t_1) \\ + \text{Immigration in } (t_0, t_1) - \text{Emigration in } (t_0, t_1).\end{aligned}$$

Equivalently, using the Fundamental Theorem of Calculus as in MA20221,

$$\begin{aligned}\text{Rate of change of population} \\ = \text{Birth rate} - \text{Death rate} \\ + \text{Immigration rate} - \text{Emigration rate}.\end{aligned}$$

We shall assume that the population is closed, i.e. there is no immigration or emigration. Let $N(t)$ be the population size (measured in numbers or density) at time t. Then we can translate the 'word model' above into the ordinary differential equation (ODE):

$$\dot{N} = \frac{dN}{dt} = \beta(N) - \delta(N) = f(N)$$

where $\beta(N)$ is the birth rate (and $\beta(N)/N$ the per capita birth rate) of the population at size N, $\delta(N)$ is the death rate (and $\delta(N)/N$ the per capita death rate) of the population at size N, $f(N) = \beta(N) - \delta(N)$ is the (net) growth rate of the population at size N.

For isolated populations, there can be no growth when population size is zero. That is, we require $\frac{dN}{dt} = f(N) = 0$ when $N = 0$. This is also called the *axiom of parenthood*: each organism has parents (Hutchinson 1978). In order to highlight this, we write the ODE as

$$\dot{N} = \frac{dN}{dt} = NF(N)$$

where $F(N)$ is the (net) *per capita* growth rate.

If the per capita growth rate is indeed a function of population size, the population dynamics are *density-dependent*. However, we begin with an example of density-independent growth.

1.2 Malthusian (density-independent) growth

Let us assume that the per capita birth and death rates do not depend on population size N i.e. $\beta(N)/N = b$ and $\delta(N)/N = d$ where b and d are positive constants. Then the growth model becomes

$$\frac{dN}{dt} = bN - dN = rN$$

with $r = b - d$. This is known as the *Malthusian* equation in continuous time (Malthus 1798).

1.2.1 Solution of the Malthusian equation

The Malthusian equation is a linear ODE that can be solved explicitly. With initial condition $N(0) = N_0$, the solution is

$$N(t) = N_0\, e^{r\,t}.$$

We can distinguish three cases: $r > 0$ - population grows exponentially (could apply to over a limited time period), $r = 0$ - population remains constant at N_0, (unlikely biologically), $r < 0$ - population decays exponentially (could apply to decay of a drug circulating in the blood).

1.2.2 Basic reproduction number

A key concept in demography (and in epidemiology) is the basic reproduction number (or basic reproductive ratio) R_0 defined as *the expected number of offspring produced by an individual over its lifetime* i.e. R_0 is the average number of offspring produced by an individual per unit time $\times$ average lifespan of an individual, $R_0 = b \cdot \frac{1}{d}$. Note that r may be written in terms of R_0 as $r = b\frac{R_0 - 1}{R_0}$. If $R_0 > 1$, the population grows. If $R_0 < 1$, the population declines to zero. In more complex cases, we shall define the *reproduction number* or *effective reproduction number* R to be the expected number of offspring produced by an individual in a given situation, and R_0 to be the reproduction number in some ideal setting e.g. in the absence of density-dependent effects.

1.3 Logistic growth

The growth of most natural populations is limited. An S-shaped growth curve is typical. For small populations growth is (approximately) Malthusian. For large populations growth

is restricted by an environmental carrying capacity K. Below K the population growth rate is positive. Above K the population growth rate is negative. The logistic growth model gives very good fits to data from numerous populations including bacteria, yeast, rats, sheep and Americans. It assumes that the net per capita growth rate $F(N)$ decreases linearly with population size. Biologically, this could be due to intraspecific competition or overcrowding effects. Note that we do not specify whether the density dependence occurs in the births or the deaths or both.

$$\dot{N} = \frac{dN}{dt} = NF(N) = rN\left(1 - \frac{N}{K}\right)$$

where $r > 0$ is the intrinsic or maximum per capita growth rate, and $K > 0$ is the carrying capacity. This model is also known as the Pearl–Verhulst equation (Verhulst 1838, Pearl 1920).

1.3.1 Solution of the logistic equation

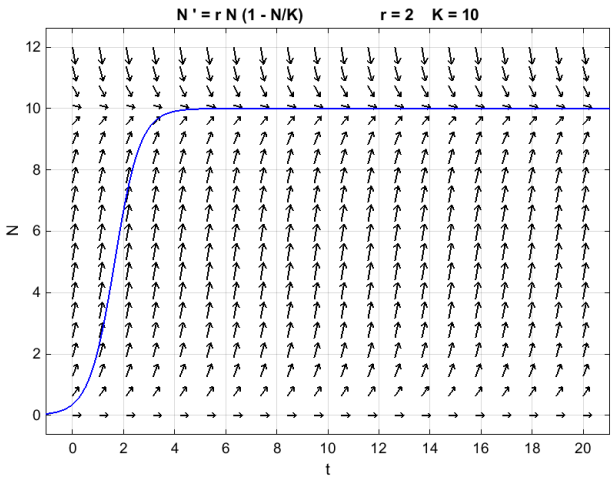
The logistic equation is one of the simplest models exhibiting density-dependent growth and may be solved explicitly (e.g. by treating it as a separable equation or a Bernoulli equation) with solution

$$N(t) = \frac{KN_0e^{rt}}{K - N_0 + N_0e^{rt}}$$

for an initial condition $N(0) = N_0$. In contrast to Malthusian growth, where (for $r > 0$) the population grows unboundedly, a logistically growing population converges to K as $t \to \infty$ (but does not reach it in finite time). The existence of a carrying capacity is a core feature of density-dependent growth.

1.3.2 Graphical analysis with vector fields

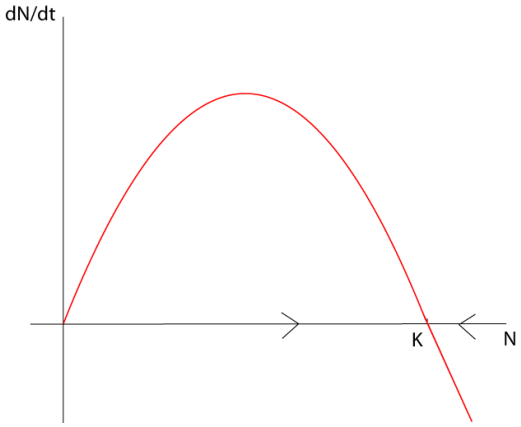
A qualitative understanding of the solution behaviour can be gained by plotting the *vector field*. For this, we indicate the slope $\dot{N}$ of the solution for many points in the (t, N)-plane by drawing arrows reflecting the slope. Note that the slope is given by the right-hand side of the ODE, $\dot{N} = f(N)$. We can now pick an initial condition $N(t_0) = N_0$, corresponding to a point (t_0, N_0) in the (t, N) plane, and draw the solution. The solution curve is tangential to the short arrows.



By the Picard–Lindelöf Theorem the initial value problem consisting of the ODE $\dot{N} = f(N)$ and the initial condition $N(t_0) = N_0$ is well posed (under reasonable assumptions such as continuously differentiable f) and therefore has a unique solution. It follows that solutions of the ODE with different initial conditions do not intersect in the (t, N) plane. This theorem also holds for systems of ODEs in state space.

1.3.3 Phase-line analysis

A graphical tool to analyse the long-term behaviour of a single ODE is *phase-line analysis*, the one-dimensional version of phase-plane or phase-space analysis. We plot the direction field (the sign of the population growth rate $\dot{N} = f(N)$) as a function of the population size N. For the logistic model (1.3) the growth rate is positive if $N < K$ and negative if $N > K$. Hence, the population will tend to K.



For the Malthusian growth model with $r > 0$ we see that population growth is always positive meaning the population size will grow to infinity.

1.3.4 Equilibria and their stability

An equilibrium or steady state of the ODE $\dot{N} = f(N)$ is defined to be a point $N = N^*$ satisfying $f(N^*) = 0$. At an equilibrium $N = N^*$ the rate of population change is zero. The solution of $dN/dt = f(N)$ with initial condition $N(0) = N^*$ is $N(t) = N^*$ for all t. To be biologically realistic, an equilibrium must be non-negative.

If a population is at an equilibrium point, it remains there for all time. However, what happens if a population is not exactly at the equilibrium point, but in its vicinity? Local stability of an equilibrium point means that solutions starting sufficiently close to the point stay close (for all time). Local asymptotic stability of an equilibrium point means that all sufficiently small initial deviations eventually return to the equilibrium, or tend towards it as $t \to \infty$. For global asymptotic stability the same is true whatever the size of the initial deviation. An equilibrium point that is not stable is unstable. Stability is determined by the slope $f'(N^*) = \frac{df}{dN}(N^*)$. To see this, define a (small) perturbation n from the steady state N^* by $n = N - N^*$. Then, by Taylor's Theorem, n satisfies

$$\frac{dn}{dt} = \frac{dN}{dt} = f(N) = f(N^* + n) = f(N^*) + f'(N^*)n + \text{higher order terms.}$$

So approximately

$$\frac{dn}{dt} = f'(N^*)n = \lambda n,$$

say, with solution $n(t) = n_0 \exp(\lambda t)$. So, *if* we can neglect nonlinear terms, N^* is locally asymptotically stable for $\lambda = f'(N^*) < 0$ (since then $n(t) \to 0$ so $N(t) \to N^*$ as $t \to \infty$), and unstable for $\lambda = f'(N^*) > 0$. This neglect of nonlinear terms in determining stability is called linear stability analysis. It is valid as long as the steady state is hyperbolic i.e. as long as $\lambda \neq 0$.

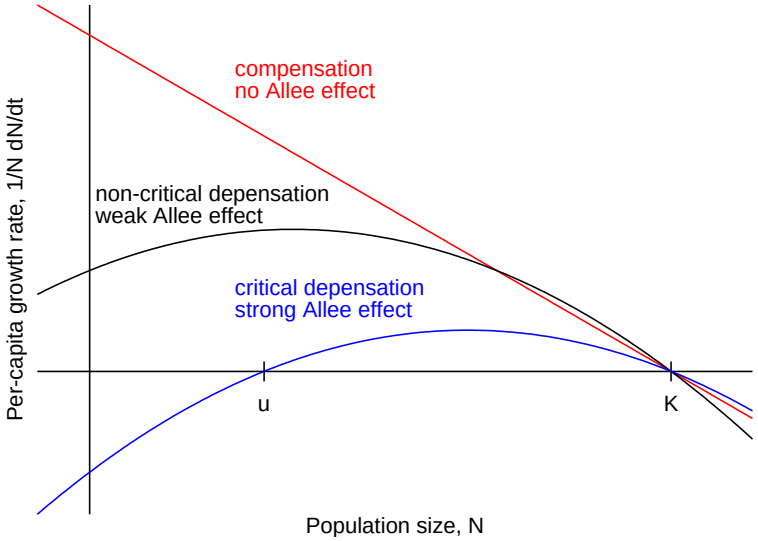
For the logistic growth model $dN/dt = f(N) = rN(1 - N/K)$, the equilibria are $N = 0$ and $N = K$, biologically interpreted as extinction and carrying capacity. Furthermore

$$f'(N) = \frac{df}{dN}(N) = r\left(1 - \frac{2N}{K}\right).$$

Then $f'(0) = r > 0$ so the extinction state is unstable, while $f'(K) = -r < 0$ so the carrying capacity is locally asymptotically stable.

1.4 Allee effects

The Allee effect describes the phenomenon of reduced per capita population growth at low population sizes or densities.



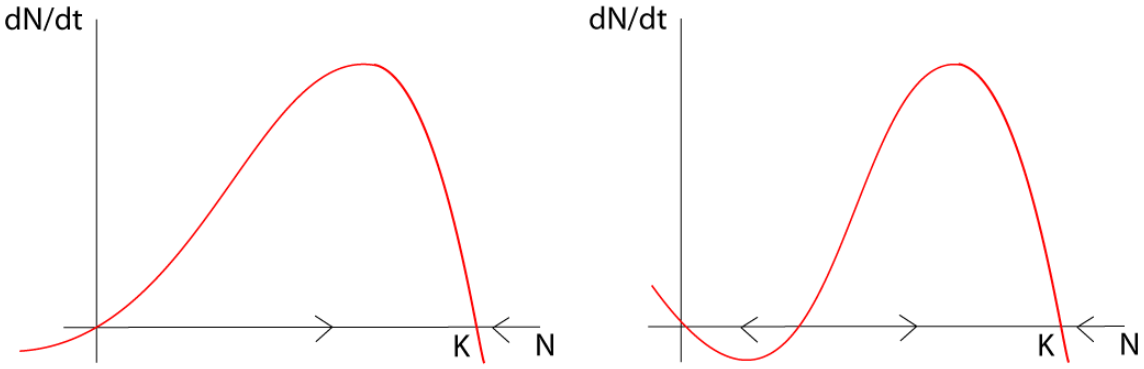
Logistic growth exhibits negative density dependence (in the net per capita growth rate, that is) for all population sizes,

$$\frac{dF}{dN}(N) < 0 \quad \text{for all} \quad N.$$

By contrast, Allee effects express themselves by positive density dependence at small population sizes. That is, there are fitness increases for larger populations (before the negative density dependence from resource competition takes over). Biological mechanisms giving rise to Allee effects include the frequency of mating encounters and cooperative protection against predation. One of the most general models of an Allee effect is

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)\left(\frac{N}{K} - \frac{u}{K}\right).$$

Parameters $r, K > 0$ are defined as before, while u is called the *Allee threshold* and can take values $-K < u < K$.



If $-K < u \leq 0$, the Allee effect is *weak*. The per capita growth rate is reduced at small densities, but remains positive. The main difference between this and logistic growth is that population growth is slowed down at small densities. If $0 < u < K$, the Allee effect is *strong*. The per capita growth rate becomes negative at small densities, i.e. the population declines. The Allee threshold u is an unstable equilibrium which acts as a minimum viable population size. If $0 < N_0 < u$ then $N(t) \to 0$ for $t \to \infty$ (extinction). If $u < N_0$ then $N(t) \to K$ for $t \to \infty$ (persistence). In general we can distinguish three growth forms.

Compensation. Per capita growth rate $F(N)$ is a decreasing function of N for all N, i.e. $\frac{dF}{dN}(N) < 0$ for all N. There is no Allee effect.

Depensation. When N is small, the per capita growth rate $F(N)$ is positive and an increasing function of N i.e. $F(N) > 0$ and $\frac{dF}{dN}(N) > 0$ for small N. There is a weak Allee effect.

Critical depensation. When N is small, the per capita growth rate is negative and an increasing function of N i.e. $F(N) < 0$ and $\frac{dF}{dN}(N) > 0$ for small N. There is a strong Allee effect.

Chapter 2

Harvesting

2.1 Introduction

What are ecologically acceptable strategies for harvesting renewable resources? On the one hand, the goal is to maximise the yield. On the other hand, the population must be maintained. We are therefore looking for the *maximum sustainable yield*. Broadly, one can distinguish two different harvest rates: a constant harvest rate independent of population size (*constant yield strategy*) and a harvest rate proportional to population size (*constant effort strategy*).

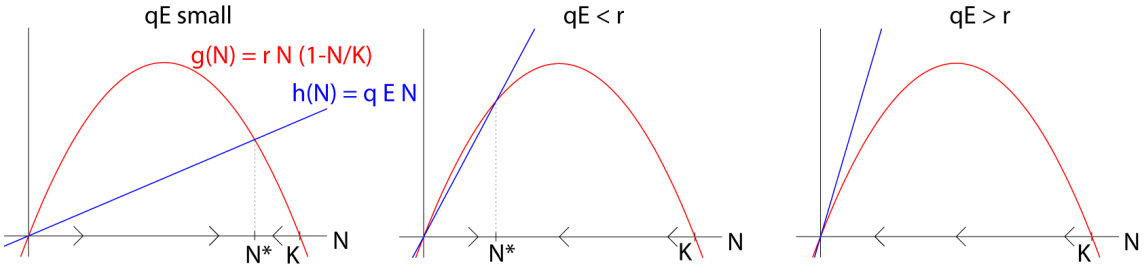
2.2 Constant effort harvesting

This harvesting strategy typically arises in fisheries, where it is assumed that the catch (the number of fish caught) is proportional to E, the effort expended in fishing. The effort may be measured in terms of the number of fishing vessels in use at a given time, for instance.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) - qEN = f(N) = g(N) - h(N).$$

The underlying growth rate $g(N)$ is logistic. The harvest rate $h(N)$ is the product of: $E > 0$, harvesting effort; $q > 0$, proportionality constant measuring catchability; N, population size. The location and the stability of equilibria can be understood by sketching $g(N)$ and $h(N)$ and using phase-line analysis. At equilibria these functions are equal, and the logistic growth rate equals the harvest rate.

For small harvesting efforts, the nontrivial equilibrium is asymptotically stable. On the right-hand side of N^*, $h(N) > g(N)$, the harvest rate is greater than the growth rate, which is why the population size decreases. Conversely, on the left-hand side of N^*, $h(N) < g(N)$, the harvest rate is smaller than the growth rate, which is why the population size increases. Increasing the harvest effort E shifts the nontrivial equilibrium point to the left, but maintains its stability as long as $qE < r$. For harvest efforts E above a critical level, $qE > r$, the harvest exceeds population growth at all population sizes N. The only equilibrium is the trivial solution, which is stable. There is no positive equilibrium. Over-harvesting leads to population extinction.



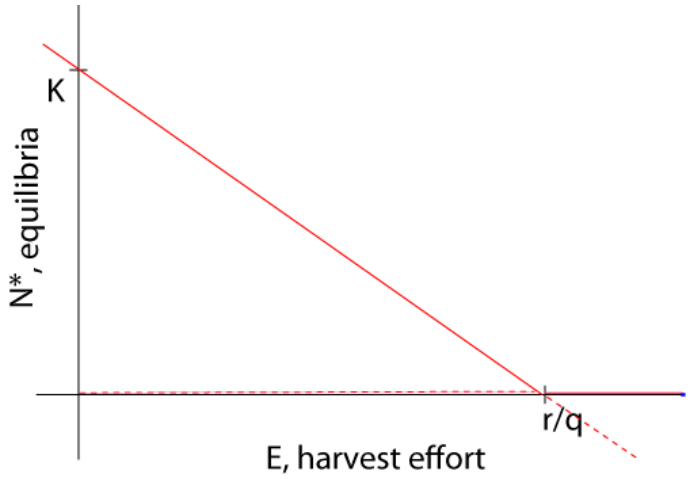
We can find N^* analytically. Rearranging the differential equation for N to

$$\frac{dN}{dt} = rN\left(1 - \frac{qE}{r} - \frac{N}{K}\right)$$

makes it easy to see that the net population growth rate retains its quadratic shape. There can be two equilibria, a trivial one at 0 and a nontrivial one at $N^* = K\left(1 - \frac{qE}{r}\right)$ realistic as long as $qE < r$ confirming that harvesting reduces population size.

2.2.1 Bifurcation diagram

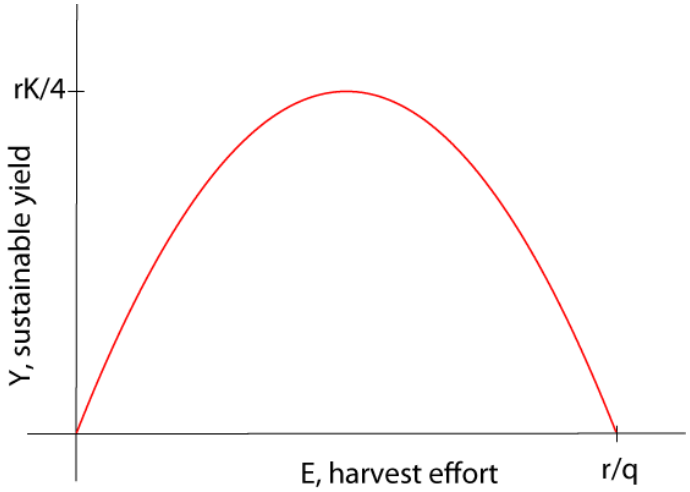
The equilibria can be plotted as a function of the harvesting effort.



At $E = r/q$ there is a transcritical bifurcation. When the nontrivial equilibrium point becomes negative (i.e. biologically unfeasible), it exchanges stability with the trivial solution.

2.2.2 Yield–effort curve and the maximum sustainable yield

How should one choose a harvesting effort to maximise yield? It must of course guarantee population persistence. The equilibrium harvest rate is the sustainable yield $Y = qEN^* = qEK\left(1 - \frac{qE}{r}\right)$.



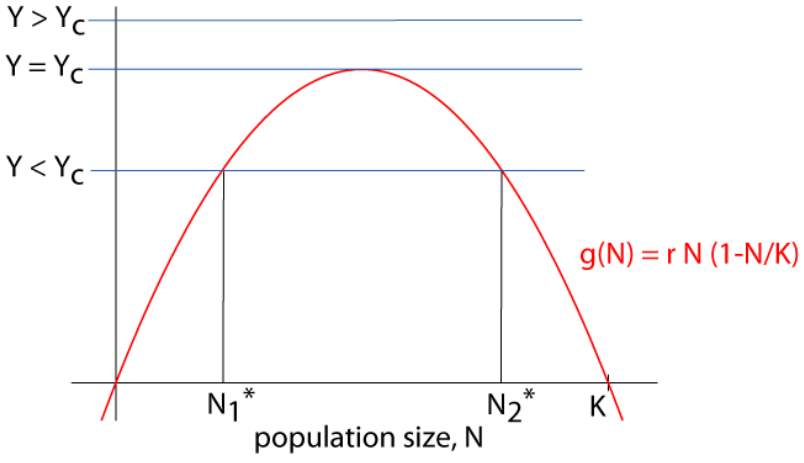
Initially, the sustainable yield increases with harvesting effort, but then the population becomes increasingly overexploited and eventually depleted. For a fixed catchability q, the *maximum sustainable yield* (Y_{opt}) occurs for $\frac{dY}{dE} = qK\left(1 - \frac{2qE}{r}\right) = 0$. The optimal harvesting effort is $E_{\text{opt}} = \frac{r}{2q}$, giving rise to $Y_{\text{opt}} = \frac{rK}{4}$.

2.3 Constant-yield harvesting

This type of harvesting arises when a quota is specified, e.g. by fisheries agreements or hunting permits.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) - Y = g(N) - h(N) = f(N).$$

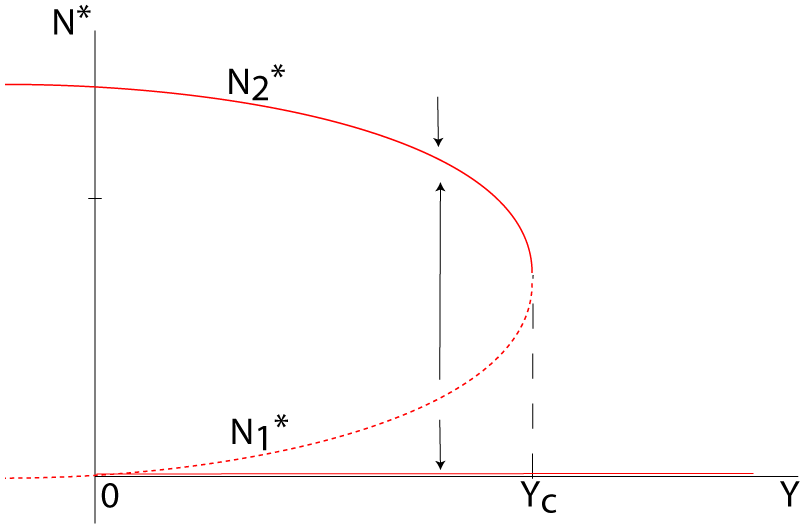
The harvest rate $h(N)$ is now a constant, $Y > 0$, which corresponds to the constant yield. Strictly speaking the equation is only valid for $N > 0$. For $N = 0$, $\dot{N} = 0$. We shall analyse solutions using phase-line analysis.



If $Y > Y_c = \max g(N) = rK/4$ then $f(N) < 0$ for all N, so $\dot{N} < 0$ for all $N > 0$, so the population will go extinct through over-harvesting. If $Y = Y_c = rK/4$, then $f(N) = 0$ at $N = N^* = \frac{1}{2}K$ but $f(N) < 0$ elsewhere. The only steady state N^* is unstable, sometimes confusingly called semistable. If $Y < Y_c = rK/4$, then the steady states are $N_0^* = 0$,

stable, and two positive roots $N_1^* < N_2^*$, N_1^* unstable and N_2^* stable. The unstable equilibrium acts as a *breakpoint* between the stable equilibrium and the stable trivial solution. A system with two alternative stable steady states is called bistable.

The asymptotic behaviour of the system may be shown as a bifurcation diagram. The change in behaviour at Y_c is called a saddle–node bifurcation. If N were at N_2^* and Y were then to increase, there would be an abrupt transition to extinction at $Y = Y_c$, where the two nontrivial equilibria collide and annihilate each other.



The two steady states N_1^* and N_2^* (for $Y < Y_c$) may be found analytically by solving $f(N^*) = 0$, or (multiplying by K/r and rearranging) So $N^{*2} - KN^* + KY/r = 0$, which gives $N_{1,2}^* = \frac{1}{2}\left(K \mp \sqrt{K^2 - 4KY/r}\right)$.

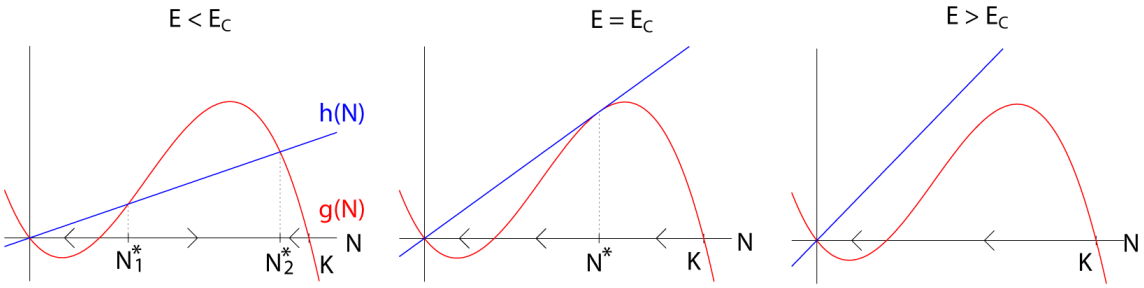
2.4 Constant effort harvesting and depensation

The previous models assumed logistic population growth. We now consider the case of a strong Allee effect in the population (critical depensation). We have

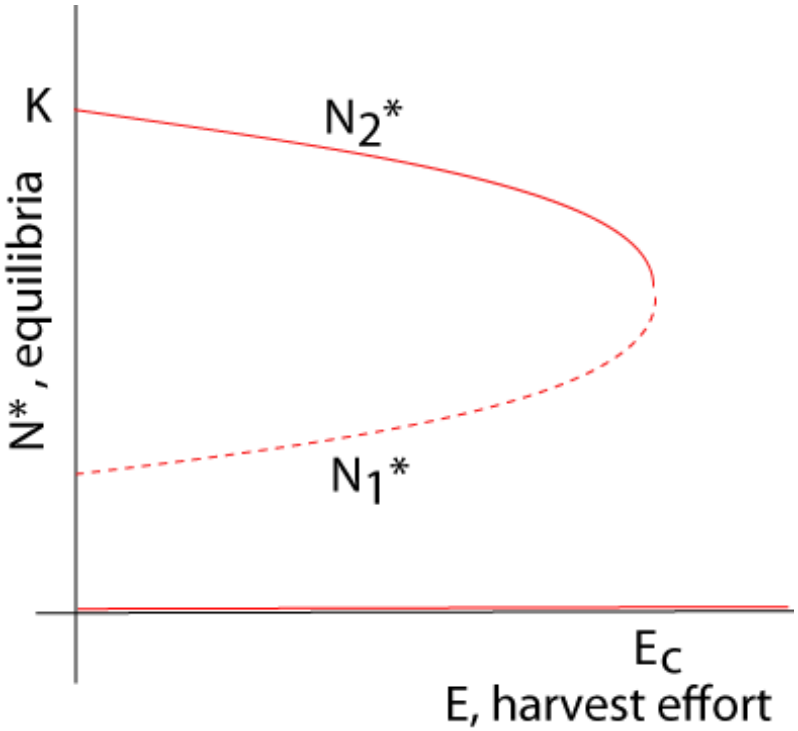
$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)\left(\frac{N}{K} - \frac{u}{K}\right) - qEN = f(N) = g(N) - h(N)$$

with $0 < u < K$.

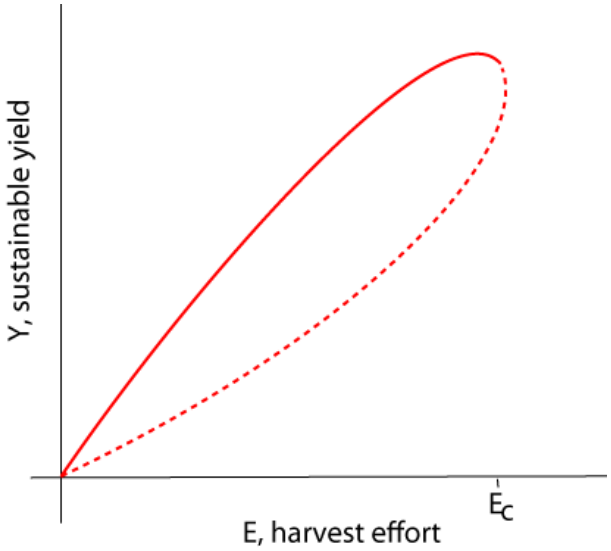
We use phase-line analysis. For a low harvesting effort ($E < E_c$) there are steady states at $0 = N_0^* < N_1^* < N_2^*$, with N_0^* stable, N_1^* unstable, and N_2^* stable. The system is bistable. For a high harvesting rate ($E > E_c$) there is only one steady state, the trivial one $N_0^* = 0$, and it is stable. The population goes to extinction. The transition from sustainable harvesting to population extinction takes place when there is a unique intersection in the first quadrant (when $E = E_c$). There is then a single nontrivial steady state which is semistable (and therefore unstable).



The asymptotic behaviour of the system may be shown as a bifurcation diagram, which is similar to the constant–yield bifurcation diagram with a saddle–node bifurcation at $E = E_c$. There is an abrupt transition to extinction if E increases past $E = E_c$.



The sustainable yield is $Y = qEN_2^*$ with N_2^* being the stable nontrivial equilibrium. We can find the nontrivial equilibrium by solving $f(N^*) = 0$, which reduces to a quadratic equation, or sketch Y as a function of E without doing this.



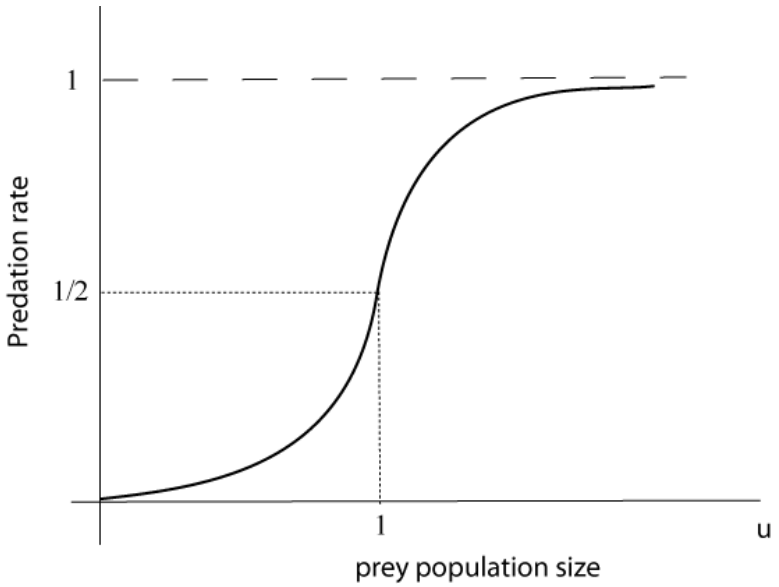
Close to the saddle–node bifurcation point, there is a catastrophic collapse from the maximum sustainable yield. Small changes can have large effects.

2.5 Spruce budworm model

Harvesting can lead to multiple steady states. Predation can have a similar effect. We now model an insect pest, the spruce budworm, that lives in the spruce and fir forests of Canada and the northeastern USA. Every 30–70 years, there is an outbreak of spruce budworms and their numbers kill up to 80% of the mature trees in a forest. A model for the spruce budworm population dynamics, in dimensionless form, is (Ludwig et al. 1978)

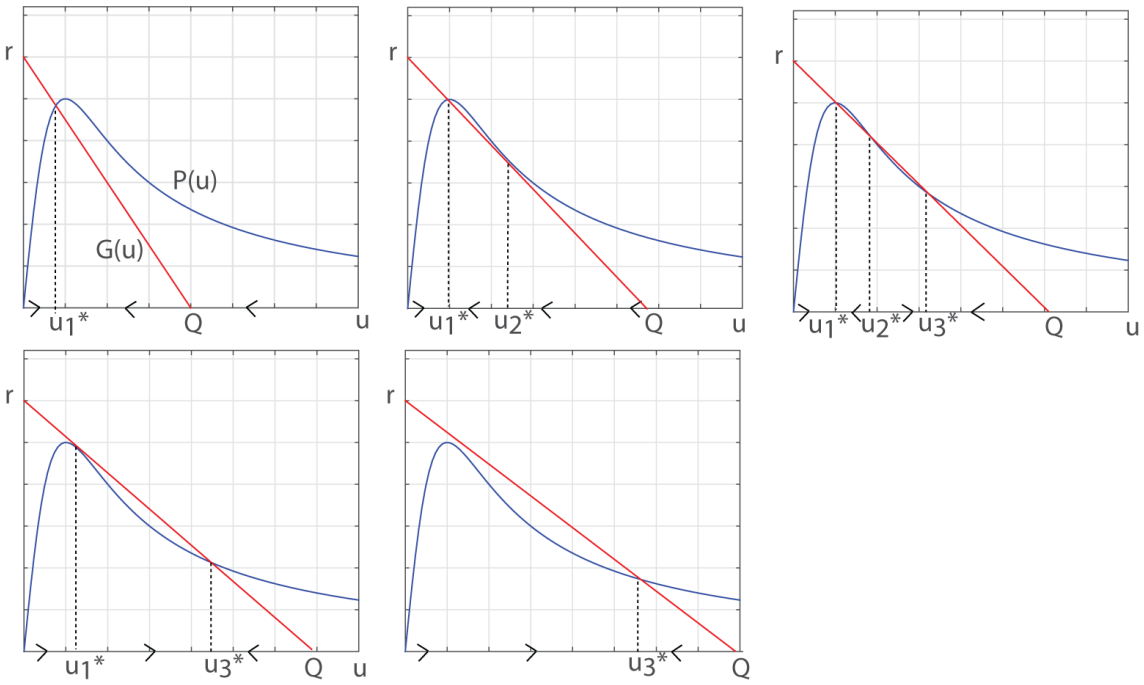
$$\frac{du}{dt} = f(u) = ru\left(1 - \frac{u}{Q}\right) - \frac{u^2}{1+u^2} = g(u) - p(u)$$

where u is spruce budworm population density and r and Q are positive parameters. The growth rate function $g(u)$ is logistic, with intrinsic per capita growth rate r and carrying capacity Q. The function $p(u)$ represents the rate of predation by birds (with the bird population density assumed constant). The predation rate is a sigmoid function of prey density. At low u, predation is low; the birds tend to seek different food. As u increases, birds switch their 'search image' and predation increases. The predation rate saturates to a maximum non-dimensionalised to 1 as $u \to \infty$.

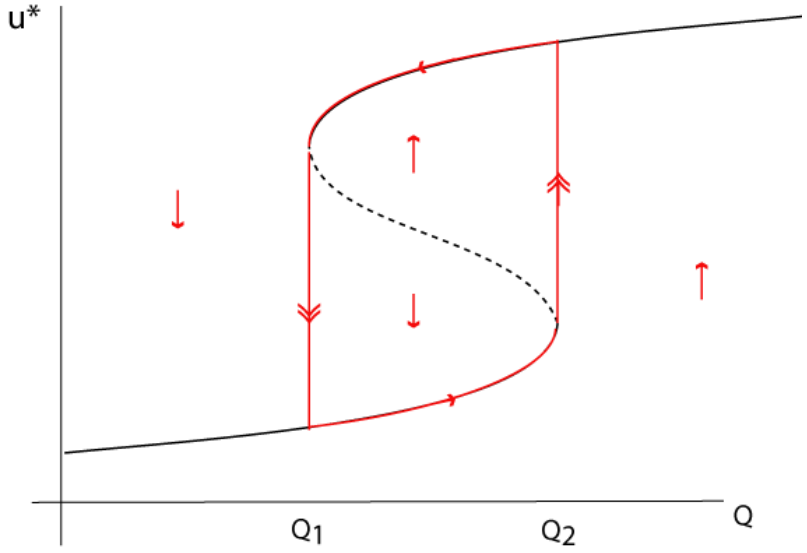


2.5.1 Analysis and the hysteresis effect

We write the equation in the form $\dot{u} = uF(u) = u(G(u) - P(u))$, so that $G(u) = r(1 - \frac{u}{Q})$ and $P(u) = \frac{u}{1+u^2}$ are the per capita growth and predation rates. We shall analyse this equation in terms of the parameters r and Q. One equilibrium is easily found to be $u_0^* = 0$. Other equilibria are found by solving $F(u) = G(u) - P(u) = 0$. This is a cubic polynomial and finding the roots is arduous. It is more convenient to find the equilibria graphically as the points of intersections of the two curves $G(u)$ and $P(u)$. If we fix a value of r and vary Q, there can be one, two or three equilibria.



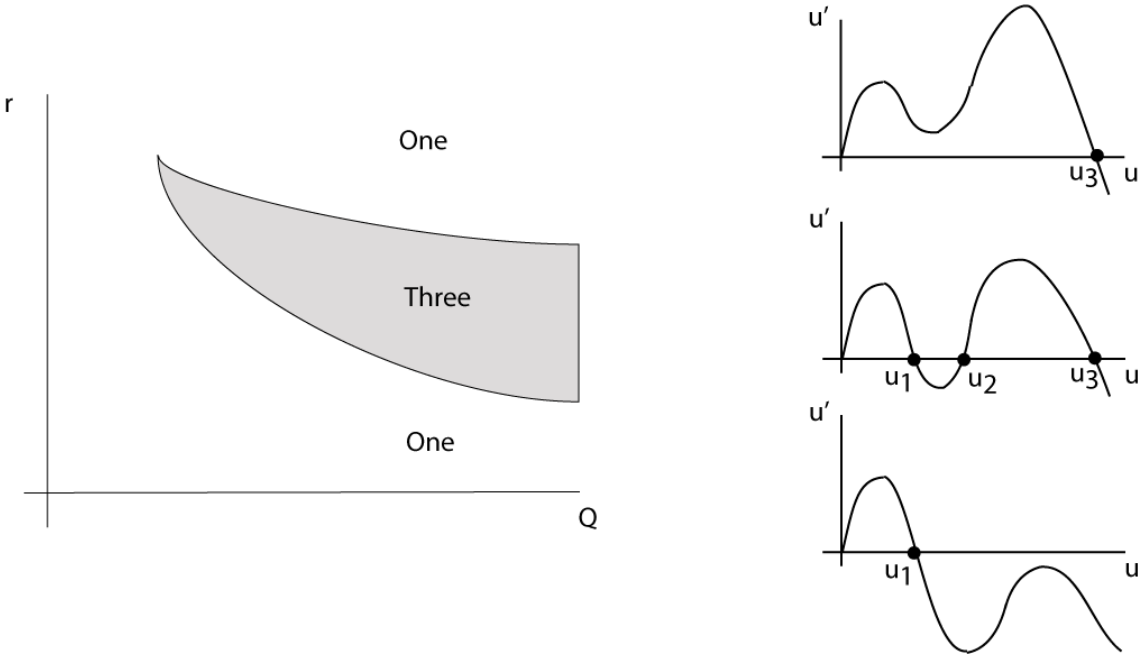
When there are three equilibria, say $u_1^* < u_2^* < u_3^*$, two of them (u_1^* and u_3^*) are stable. The smallest one, u_1^*, is the refuge equilibrium. The largest one, u_3^*, is the outbreak equilibrium. There is an unstable threshold equilibrium (u_2^*) in between.



We can now sketch the bifurcation diagram for bifurcation parameter Q with r fixed. The saddle–node bifurcation at Q_1 occurs when u_2^* and u_3^* collide and annihilate each other as Q decreases, while the saddle–node bifurcation at Q_2 occurs when u_1^* and u_2^* collide and annihilate each other as Q increases. These bifurcations occur when $G(u)$ is tangent to $P(u)$.

Let us now fix Q and allow r to vary. If Q is fixed sufficiently large then the situation is similar, with one steady state for r small, one for r large, and three for r in an intermediate region. However if Q is fixed at a small value then there can only be one steady state, whatever the value of r. The two-parameter bifurcation diagram is therefore as shown. In

some parts of parameter space there are three steady states, and in others there is only one. The two curves separating these regions may be shown to meet at a cusp.



This analysis suggests some possible methods of pest control. Application of insecticides (reduce Q) is a temporary and expensive control, but if large enough it could force the dynamics to have only one equilibrium (the refuge). Biological control methods (introducing sterile males or new predators) to decrease reproduction (reduce r) could be effective if they move r below the value where u_3^* exists.

Chapter 3

Population dynamics of interacting populations

Species interactions may be classified as below (May 1974). Complex food-webs may be thought of as compositions of these modules.

- Competition $(-,-)$. Each species has a negative effect on the growth of the other.
- Mutualism or symbiosis $(+,+)$. Each species has a positive effect on the growth of the other.
- Predation $(+,-)$. One species, the predator, has a negative effect on the growth of the other, the prey; the prey has a positive effect on the predator.
- Commensalism $(+,0)$: One species has a positive effect on the growth of the other, there is no interaction in the other direction.
- Amensalism $(-,0)$: One species has a negative effect on the growth of the other, there is no interaction in the other direction.
- Neutralism $(0,0)$: Two species interact without affecting each other.

3.1 Interspecific competition

We have already (implicitly) considered intraspecific competition, with density-dependent growth models such as the logistic differential equation. Now we shall look at interspecific competition, which involves two (or more) species. Interspecific competition occurs when the growth rate of a population of one species is reduced in the presence of a population of another species. This may be due to (i) shared use of limiting resources (*exploitative competition*) or (ii) active inhibition (*interference competition*). Models for interference competition just include equations for the populations of the species, while models for exploitative competition start by including an equation for the resources themselves, but may reduce to equations for the populations only.

The following model of two competing species N_1 and N_2 is due to Lotka (1932) and Volterra (1926).

$$\frac{dN_1}{dT} = r_1 N_1 (1 - \alpha_{11} N_1 - \alpha_{12} N_2)$$

$$\frac{dN_2}{dT} = r_2 N_2 (1 - \alpha_{21} N_1 - \alpha_{22} N_2)$$

where all parameters are positive. Each species has its own intrinsic growth rate (r_1 and r_2) and carrying capacity ($K_1 = 1/\alpha_{11}$ and $K_2 = 1/\alpha_{22}$). In the absence of its competitor, each species will grow to this carrying capacity. The parameters α_{11} and α_{22} may be thought of as intraspecific competition coefficients. We assume that the effect of interspecific competition is similar to that of intraspecific competition, parameterised by the pair of competition coefficients α_{12} and α_{21}. They describe the competitive effect of species 2 on species 1 and of species 1 on species 2 respectively. We can non-dimensionalise by writing $u = \alpha_{11} N_1 = N_1/K_1, v = \alpha_{22} N_2 = N_2/K_2, t = r_1 T$, $s = r_2/r_1, a = \alpha_{12}/\alpha_{22}, b = \alpha_{21}/\alpha_{11}$. Then the model becomes

$$\begin{aligned} \frac{du}{dt} &= u(1 - u - av) = f(u, v) \\ \frac{dv}{dt} &= s\, v (1 - bu - v) = g(u, v). \end{aligned}$$

These are two coupled nonlinear ODEs, and we can only dream of finding their solution.

3.1.1 Steady states and their stability

A *steady state* of the competition system above is a point $S = (u^*, v^*)$ in state space (i.e. (u, v) space) satisfying $f(u^*, v^*) = u^*(1-u^*-av^*) = 0$ and $g(u^*, v^*) = sv^*(1-bu^*-v^*) = 0$. Note that these equations must be solved *simultaneously*. So there are four steady states $S = (u^*, v^*)$, the first three of which always exist in the nonnegative quadrant, and are therefore biologically realistic. $S_0 = (0, 0)$ is the trivial steady state, both species are absent. $S_1 = (1, 0)$ is a semitrivial steady state, species 2 is absent and species 1 is at its carrying capacity. $S_2 = (0, 1)$ is a semitrivial steady state, species 1 is absent and species 2 is at its carrying capacity. $S^* = (u^*, v^*)$ is a nontrivial steady state, the two species coexist such that $u^* = \frac{1-a}{1-ab}, v^* = \frac{1-b}{1-ab}$. The nontrivial steady state is biologically realistic (feasible) if both $a < 1$ and $b < 1$ or both $a > 1$ and $b > 1$.

The *linear stability* of a steady state (u^*, v^*) of the system is determined by the eigenvalues of the Jacobian matrix $J^* = J(u^*, v^*)$ of the system at the steady state, given by

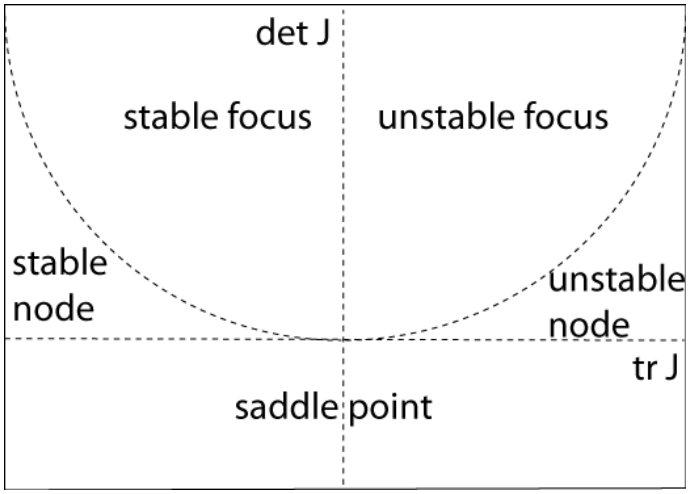
$$J^* = \begin{pmatrix} f_u^* & f_v^* \\ g_u^* & g_v^* \end{pmatrix}$$

where for example $f_u^* = \frac{\partial f}{\partial u}(u^*, v^*)$. The steady state is linearly asymptotically stable if both eigenvalues of J^* have negative real part, unstable if either has positive real part.

Nonlinear stability (or instability) follows from linear stability (or instability) as long as neither eigenvalue has zero real part. The *character* of the steady state is as follows.

Eigenvalues	$++$	$--$	$+-$	$a \pm ib$ $a > 0$,	$a \pm ib$ $a < 0$	$\pm ib$
Linear system	unstable node	stable node	unstable saddle	unstable focus	stable focus	centre
Nonlinear system	unstable node	stable node	unstable saddle	unstable focus	stable focus	centre/ unstable focus/ stable focus

If J^* is diagonal or triangular then the eigenvalues may be found and stability of the steady state deduced immediately. If not, it may be easier to use the Routh–Hurwitz criteria to determine stability. These criteria follow from the fact that the eigenvalues are the roots of the *characteristic equation* $P(\lambda) = 0$, where $P(\lambda) = \lambda^2 - \operatorname{tr} J^* \lambda + \det J^*$. They state that the eigenvalues of J^* both have negative real part if and only if both $\operatorname{tr} J^* < 0$ and $\det J^* > 0$. The *discriminant* $\operatorname{disc} J^*$ of the matrix J^* is defined by $\operatorname{disc} J^* = (\operatorname{tr} J^*)^2 - 4 \det J^*$. It follows from the formula for the roots of the quadratic equation $P(\lambda) = 0$ that the eigenvalues of J^* are real if $\operatorname{disc} J^* \geq 0$ and complex if $\operatorname{disc} J^* < 0$. The character of the steady states is summarised below.



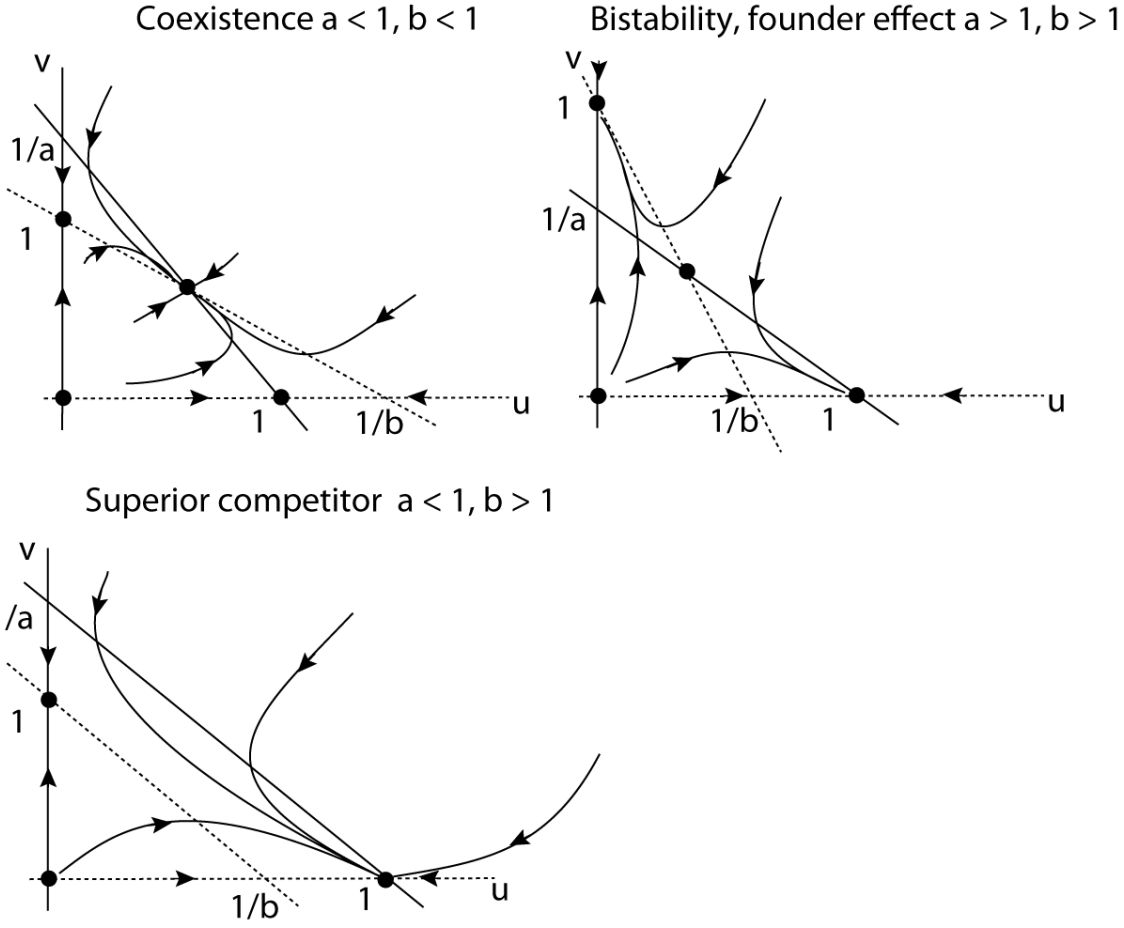
3.1.2 Graphical phase plane analysis

It is instructive to study the behaviour of two-dimensional systems in the *phase plane* i.e. the (u, v) space. Phase plane analysis was introduced in MA20221 and in MA20202. Recall the crucial idea that the vector field $(f(u, v), g(u, v))$ determines the velocity $(\dot{u}, \dot{v})$ given the position (u, v). We apply the following method:

- Sketch the *nullclines* $f(u, v) = 0$ and $g(u, v) = 0$ in the phase plane. Mark the intersection points; these are the steady states.
- Find where in the phase plane $f > 0$, $f < 0$, $g > 0$ and $g < 0$. Sketch the direction of the vector field, the direction field, on the nullclines and in the regions between them.
- If you wish to sketch some solution trajectories, make use of the direction field to do so. The solution of a two-dimensional system with given initial conditions is unique,

so that the trajectories cannot cross. If necessary for important details, analyse the steady states by linearisation, and determine their character (nodes, saddle points, centres, or foci).

For the Lotka-Volterra competition model, the u nullclines are $u = 0$ and $1 - u - av = 0$. The v nullclines are $v = 0$ and $1 - bu - v = 0$. The nontrivial u nullcline is a straight line through $(1, 0)$ and $(0, 1/a)$, while the nontrivial v nullcline is a straight line through $(1/b, 0)$ and $(0, 1)$. There are four cases of their relative location, depending on whether $a > 1$ and whether $b > 1$. Taking into account the direction of flow, we arrive at the schematic phase portraits below. At this stage, the problem is solved for all practical purposes. We can draw sensible phase planes in only one distinct way for each case. Determining the linear stability of the steady states is not required.



(a) **Coexistence.** If $a < 1$ and $b < 1$, interspecific competition for both species is relatively weak (compared to intraspecific competition). The single-species equilibria on the boundaries are unstable saddle points. Note that initial conditions starting on one of the boundaries will remain there, but any small perturbation (such as immigration) will grow into the interior. All trajectories starting in the interior are drawn towards the stable coexistence equilibrium S^*. At this equilibrium, the population size of each species is reduced in comparison to the respective carrying capacities. We shall show later that the coexistence state is in fact globally asymptotically stable.

(b) **Bistability.** If $a > 1$ and $b > 1$, the interspecific effects are large for both species. The two single-species equilibria S_1 and S_2 are each stable nodes, so the system

is *bistable*. S_1 and S_2 correspond to the exclusion of the other species. Which of the two species will be outcompeted depends on the initial condition. Each of the single-species steady states has a *domain (or basin) of attraction*. Note that the interior equilibrium S^* is a saddle point, located in between the two stable nodes. The stable manifold of this saddle point is the *separatrix* i.e. the boundary between the two distinct domains of attraction. This scenario is also called a *founder effect*. The species that manages to establish itself first (the founder) has an advantage and will exclude the other.

(c) **Species 1 excludes species 2.** If $a < 1$ and $b > 1$, species 1 has a relatively large effect on species 2, while species 2 has only a relatively small effect on species 1. We expect that the superior competitor u will exclude the inferior competitor v, so that species 2 will go extinct and species 1 will tend to its carrying capacity.

(d) **Species 2 excludes species 1.** If $a > 1$ and $b < 1$, we have the reverse situation of (c). We expect species 1 to go extinct and species 2 to approach carrying capacity.

3.1.3 Principle of competitive exclusion

In many natural systems, interference competition is usually asymmetric, with one species dominating the other. This dominance is often associated with larger body size.

Principle of competitive exclusion: if two species occupy the same *ecological niche*, then they cannot coexist.

Two species are said to occupy the same ecological niche if they exhibit the same ecological traits e.g. they live in the same habitat, interact in the same way with other species, and compete for the same resources. Generally, we expect the interspecific competition coefficients to be asymmetric. This leads to one of the scenarios (c) or (d) with the exclusion of the weaker competitor. However, ecosystems frequently consist of many competitors. Theoretical and experimental research has therefore tried to identify mechanisms that can allow species to share the same niche and to coexist. Some of these mechanisms include spatial distribution e.g. territoriality, seasonality and other variations in time and space, effects of additional species such as predator or pathogens.

3.1.4 Systems in Kolmogorov form

We may generalise the principle of competitive exclusion by looking at more general competition systems. A system of two differential equations for two interacting species is said to be in Kolmogorov form if it may be written as

$$\frac{du}{dt} = uF(u,v), \quad \frac{dv}{dt} = vG(u,v)$$

where F and G are per capita growth rates of u and v. Steady states of such systems satisfy $u^*F(u^*,v^*) = 0, \quad v^*G(u^*,v^*) = 0$. There are four possibilities: $u^* = v^* = 0$ – trivial steady state, $u^* = 0$ and $G(u^*,v^*) = G(0,v^*) = 0$ – semitrivial steady state, $v^* = 0$ and $F(u^*,v^*) = F(u^*,0) = 0$ – semitrivial steady state, $F(u^*,v^*) = G(u^*,v^*) = 0$ – nontrivial steady state, coexistence if realistic. A linear stability analysis involves finding the Jacobian matrix

$$J(u,v) = \begin{pmatrix} F(u,v) + uF_u(u,v) & uF_v(u,v) \\ vG_u(u,v) & G(u,v) + vG_v(u,v) \end{pmatrix}.$$

J is diagonal at the trivial steady state, so we can read off the eigenvalues. It is triangular at the semitrivial steady states, so we can read off the eigenvalues. Now consider a realistic nontrivial steady state (u^*, v^*) with $u^* > 0$, $v^* > 0$. Here $F(u^*, v^*) = G(u^*, v^*) = 0$, so

$$J^* = J(u^*, v^*) = \begin{pmatrix} u^* F_u(u^*, v^*) & u^* F_v(u^*, v^*) \\ v^* G_u(u^*, v^*) & v^* G_v(u^*, v^*) \end{pmatrix}.$$

We shall use the Routh-Hurwitz criteria. Assuming there is no Allee effect, population dynamics are compensatory and the per capita growth rates F and G decrease monotonically with population density. Hence $F_u^* < 0$ and $G_v^* < 0$ so that $\mathrm{tr} J^* = u^* F_u^* + v^* G_v^* < 0$. The equilibrium is therefore stable as long as $\det J^* > 0$ or $F_u^* G_v^* > F_v^* G_u^*$. For stability, the product of the intraspecific competition coefficients (left-hand side) must be greater than the product of the interspecific competition coefficients (right-hand side). We do not expect this to occur if the two species occupy the same ecological niche.

3.2 Global stability and Lyapunov functions

Systems of nonlinear ordinary differential equations may have several steady states. Close to a steady state, the behaviour is approximated by the linearised equations and this is used in linear stability analysis to determine local asymptotic stability (MA20221 or MA20202). An equilibrium is *locally asymptotically stable* if a system *near* the equilibrium approaches it, i.e. the equilibrium is locally attracting. An equilibrium is *globally asymptotically stable* if a system approaches the equilibrium regardless of its initial condition. An equilibrium is *unstable* if it is not stable, i.e. there is at least one point near the equilibrium from where the system moves away from it.

Lyapunov functions may be used to prove stability or asymptotic stability of an equilibrium in a given region, i.e. not only in the neighbourhood of the equilibrium. Hence, they may be used to prove global asymptotic stability. The technique (also known as the *direct method of Lyapunov*) may be applied to systems of n differential equations.

Positive definite functions
Let Ω be an open subset of $\mathbb{R}^2$. A function $\Phi : \Omega \to \mathbb{R}$ is *positive definite* in Ω about a point $(x^*, y^*) \in \Omega$ if $\Phi(x^*, y^*) = 0$ and $\Phi(x, y) > 0$ for all $(x, y) \in \Omega \setminus \{(x^*, y^*)\}$. A function Φ is *negative definite* if $-\Phi$ is positive definite.

Example
The function

$$\Phi(x, y) = s\left(x - x^* - x^* \ln \frac{x}{x^*}\right) + y - y^* - y^* \ln \frac{y}{y^*}$$

where s, x^* and y^* are all positive, is positive definite in $\Omega = \mathbb{R}^2_+ = \{(x, y) | x, y > 0\}$ about (x^*, y^*).

To see this note that $\Phi(x^*, y^*) = 0$. Then show that $\Phi(x, y)$ is strictly positive for all $(x, y) > 0$ with $(x, y) \neq (x^*, y^*)$. The first partial derivatives of Φ are

$$\Phi_x(x, y) = s(1 - x^*/x) \quad \text{and} \quad \Phi_y(x, y) = 1 - y^*/y.$$

Then, for $(x, y) \in \mathbb{R}^2_+$,

$$\Phi(x, y) = \Phi(x^*, y^*) + \int_{x^*}^{x} \Phi_x(\xi, y) d\xi + \int_{y^*}^{y} \Phi_y(x, \eta) d\eta \geq \Phi(x^*, y^*)$$

with equality if and only if $(x, y) = (x^*, y^*)$. Hence Φ is positive definite in $\mathbb{R}^2_+$.

The derivative of a function along trajectories of a system

Now let us assume that Φ is a function of (x, y) as before, but now x and y satisfy the two-dimensional system

$$\frac{dx}{dt} = f(x, y), \qquad \frac{dy}{dt} = g(x, y) \tag{$\star$}$$

which has a steady state (x^*, y^*). Hence x and y are functions of t, and so Φ is also a function of t. How does Φ change with t as x and y move along solution trajectories of the system? By the chain rule, the derivative of the function Φ along these solution trajectories is

$$\frac{d\Phi}{dt}(x(t), y(t)) = \frac{\partial \Phi}{\partial x}(x(t), y(t))\, \frac{dx}{dt}(t) + \frac{\partial \Phi}{\partial y}(x(t), y(t))\, \frac{dy}{dt}(t)$$

or

$$\frac{d\Phi}{dt}(x, y) = \frac{\partial \Phi}{\partial x}(x, y)\, f(x, y) + \frac{\partial \Phi}{\partial y}(x, y)\, g(x, y).$$

Lyapnuov function

A continuously differentiable function Φ that is positive definite in Ω about (x^*, y^*) is a *Lyapunov function* for system $(\star)$ if

$$\frac{d\Phi(x, y)}{dt} \leq 0 \text{ for all } (x, y) \in \Omega \setminus \{(x^*, y^*)\}.$$

The function Φ is a *strict Lyapunov function* if

$$\frac{d\Phi(x, y)}{dt} < 0 \text{ for all } (x, y) \in \Omega \setminus \{(x^*, y^*)\}.$$

Lyapunov's Stability Theorem

Let (x^*, y^*) be an equilibrium of system $(\star)$ and let Φ be a positive definite continuously differentiable function in a neighbourhood Ω about (x^*, y^*). Then

(i) If $\dot{\Phi} \leq 0$ for $(x, y) \in \Omega \setminus \{(x^*, y^*)\}$, then (x^*, y^*) is stable.

(ii) If $\dot{\Phi} < 0$ for $(x, y) \in \Omega \setminus \{(x^*, y^*)\}$, then (x^*, y^*) is asymptotically stable.

(iii) If $\dot{\Phi} > 0$ for $(x, y) \in \Omega \setminus \{(x^*, y^*)\}$, then (x^*, y^*) is unstable.

Obviously, in case (i) the function Φ is a Lyapunov function for $(\star)$ and in case (ii) Φ is a strict Lyapunov function for the system. If the region Ω is the set of all possible initial conditions (e.g. $\Omega = \mathbb{R}^2_+ = \{(x, y) | (x, y) > 0\}$), then we say that the equilibrium (x^*, y^*) is *globally asymptotically stable* in Ω.

Usually, the problem is to find a suitable Lyapunov function. This problem has been solved for Lotka–Volterra systems composed of n ordinary differential equations of the form

$$\frac{dx_i}{dt} = r_i x_i \left(1 + \sum_{j=1}^{n} a_{ij} x_j\right)$$

for $i = 1, 2, \ldots, n$. It has been shown that in many such systems with a unique nontrivial equilibrium $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)$, there exists a Lyapunov function in $\mathbb{R}_+^n$ which has the form

$$\Phi(x_1, \ldots, x_n) = \sum_{i=1}^{n} c_i \left[x_i - x_i^* - x_i^* \ln\left(\frac{x_i}{x_i^*}\right) \right],$$

where the constants c_i are positive and are chosen dependent on the parameters of the particular system.

Example
Consider the non-dimensional Lotka–Volterra competition model

$$\frac{du}{dt} = u(1 - u - av), \quad \frac{dv}{dt} = sv(1 - bu - v).$$

The nontrivial equilibrium (u^*, v^*) with $u^* = \frac{1-a}{1-ab}, v^* = \frac{1-b}{1-ab}$ is globally asymptotically stable in $\mathbb{R}_+^2$ for $a < 1$ and $b < 1$ (the coexistence case). The function Φ defined by

$$\Phi(u, v) = s\left(u - u^* - u^* \ln\frac{u}{u^*}\right) + v - v^* - v^* \ln\frac{v}{v^*}$$

is a strict Lyapunov function about (u^*, v^*) for the system.

Obviously, Φ is continuously differentiable in $\mathbb{R}_+^2$. From the example above Φ is positive definite in $\mathbb{R}_+^2$ about (u^*, v^*). Now, recall that $1 - u^* - av^* = 1 - bu^* - v^* = 0$ and therefore

$$\begin{aligned} 1 - u - av &= (1 - u - av) - (1 - u^* - av^*) = -(u - u^*) - a(v - v^*) \\ 1 - bu - v &= (1 - bu - v) - (1 - bu^* - v^*) = -b(u - u^*) - (v - v^*). \end{aligned}$$

Now consider the derivative of Φ along trajectories of the system. For $(u, v) = (u^*, v^*)$, $\dot{\Phi} = 0$. Otherwise,

$$\begin{aligned} \frac{d\Phi}{dt} &= \frac{\partial \Phi}{\partial u}\frac{du}{dt} + \frac{\partial \Phi}{\partial v}\frac{dv}{dt} \\ &= s\frac{u - u^*}{u} u \underbrace{(1 - u - av)}_{-(u-u^*)-a(v-v^*)} + s\frac{v - v^*}{v} v \underbrace{(1 - bu - v)}_{-b(u-u^*)-(v-v^*)} \\ &= s\Big[-(u - u^*)^2 - \underbrace{(a + b)}_{<2}(u - u^*)(v - v^*) - (v - v^*)^2 \Big] \\ &< s\Big[-|u - u^*|^2 + 2|u - u^*||v - v^*| - |v - v^*|^2 \Big] \\ &= -s\Big[|u - u^*| - |v - v^*| \Big]^2. \end{aligned}$$

Note that $\frac{d\Phi}{dt}(u^*, v^*) = 0$ and $\frac{d\Phi}{dt}(u, v) < 0$ for all other $(u, v) > 0$ with $(u, v) \neq (u^*, v^*)$. Hence Φ is a strict Lyapunov function about (u^*, v^*) and the coexistence state is globally asymptotically stable in $\mathbb{R}_+^2$.

3.3 Predator–prey models

The spruce budworm model considered the situation of a dynamic prey (budworm) population and a static predator (bird) population. The predators were assumed to be *generalists*, because their switching behaviour was the basis of the predation model. We now consider the case of *specialist* predators. Their only food source is the prey we consider. Hence, depending on the availability of prey and how much of the prey they consume, the predators change in time as well. We therefore get a system of two ODEs, one for the prey and one for the predators.

Predator–prey models can be written in a general form, called *Gause-type models*,

$$\begin{aligned} \frac{dN}{dt} &= NF(N) - \varphi(N)P \\ \frac{dP}{dt} &= (\psi(N) - m)P \end{aligned}$$

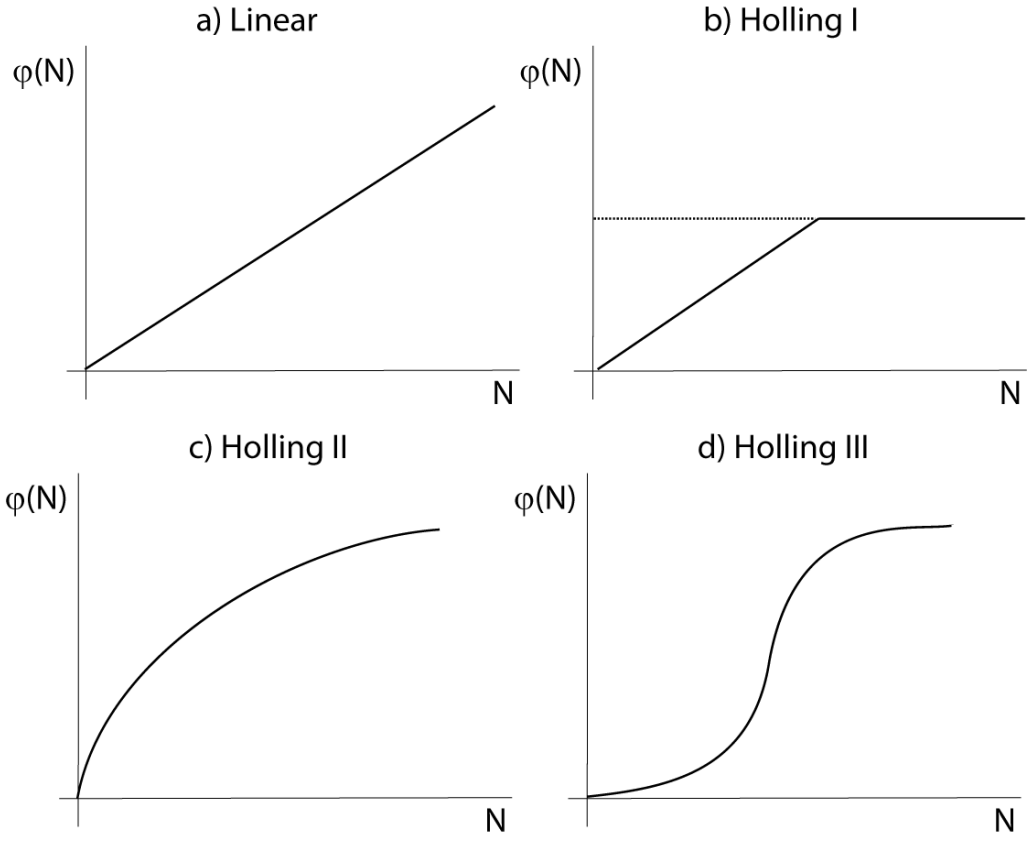
where $F(N)$ is the per capita growth rate of the prey in the absence of predators, m is the per capita death rate of the predators, $\varphi(N)$ is the functional response, $\psi(N)$ is the numerical response. The functional response is the relationship between the per capita predation rate (number of prey consumed per unit time be a single predator, also called the consumption rate) and the prey population size. The numerical response is the per capita gain in predator population size per unit time as a consequence of prey consumption. We typically assume that it is proportional to the numerical response i.e. $\psi(N) = e\,\varphi(N)$, where $0 < e < 1$ is the dimensionless efficiency with which predators convert prey biomass into their own biomass. Four archetypal functional responses were identified by Holling (1959).

(a) Linear (type 0). $\varphi(N) = aN$ where a is the effective search rate. The linear functional response is based on the law of mass action. It is used in the Lotka–Volterra prey–predator model.

(b) Holling type I. $\varphi(N) = \min\{aN, \varphi_{\max}\}$ where a is the effective search rate. There is a linear increase in consumption rate until satiation. This is considered rare in nature. But filter feeders provide an example: intake increases with density of food in water, up to the point at which the filtering apparatus is working at top speed.

(c) Holling type II (hyperbolic), also known as Holling's disc equation. $\varphi(N) = \dfrac{aN}{1 + ahN} = \dfrac{cN}{d + N}$ where a is the maximum effective search rate, h is the handling time, c is the maximum killing rate and d is the half-saturation constant. Saturation takes place gradually. This is common in nature. It can be derived mechanistically by accounting for prey handling times - the time it takes to catch, subdue, kill and eat a prey item.

(d) Holling type III (sigmoid). $\varphi(N) = \dfrac{cN^2}{d^2 + N^2}$ where c is the maximum killing rate and d the half-saturation constant. This form is common in nature. It applies to generalist predators. The search rate is an increasing function of prey density due to, for example, the presence of alternative food sources and prey switching, predator learning and the development of search images, the existence of prey refuge/cover.



3.3.1 The Rosenzweig–MacArthur model and the paradox of enrichment

The Rosenzweig–MacArthur system (1963) is a simple model that produces limit-cycle oscillations, defined as isolated (stable or unstable) periodic solutions, and so can explain sustained (and asymptotically stable) predator–prey oscillations which are frequently observed in nature. The system is:

$$\begin{aligned} \frac{dN}{dT} &= rN\left(1-\frac{N}{K}\right)-\frac{aNP}{1+ahN} \\ \frac{dP}{dT} &= \frac{eaNP}{1+ahN}-mP \end{aligned}$$

with all parameters positive.

The equations can be non-dimensionalised by defining $x = ahN$, $y = aP/r$, $t = rT$. Then

$$\begin{aligned} \frac{dx}{dt} &= x\left(1-\frac{x}{\kappa}\right)-\frac{xy}{1+x} \\ \frac{dy}{dt} &= \beta\left(\frac{x}{1+x}-\alpha\right)y. \end{aligned}$$

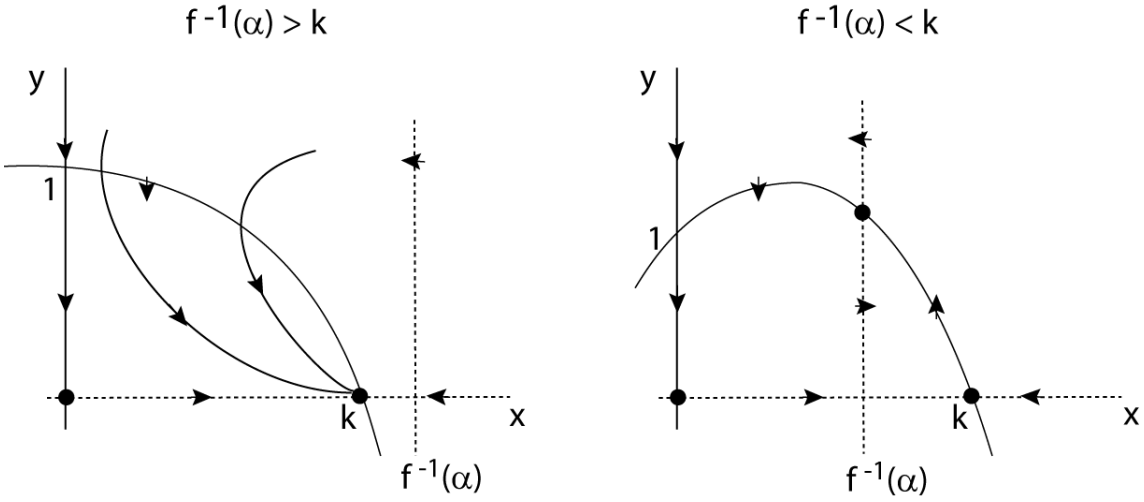
with parameters $\kappa = ahK, \beta = \frac{e}{rh}, \alpha = \frac{mh}{e}$. Note that we have reduced the number of parameters from six to three (which are all positive).

3.3.2 Nullclines and equilibria

The system may be written as

$$\begin{aligned}\frac{dx}{dt} &= \frac{x}{1+x}\left\{(1+x)\left(1-\frac{x}{\kappa}\right)-y\right\}\\ &= f(x)(g(x)-y)\\ \frac{dy}{dt} &= \beta\left(\frac{x}{1+x}-\alpha\right)y\\ &= \beta(f(x)-\alpha)y\end{aligned}$$

where $f(x) = x/(1+x)$, $g(x) = (1+x)(1-x/\kappa)$. The prey nullclines are $x = 0$ and $y = g(x) = (1+x)\left(1-\frac{x}{\kappa}\right)$. The predator nullclines are $y = 0$ and $f(x) = \alpha$ or $x = f^{-1}(\alpha) = \frac{\alpha}{1-\alpha}$. Possible phase planes are sketched below, depending on whether $f^{-1}(\alpha) < \kappa$ or $f^{-1}(\alpha) > \kappa$. We shall assume from now on that $\alpha < 1$, so that the nontrivial predator nullcline passes through the positive quadrant.



There are three steady states: the trivial steady state $S_0 = (0,0)$, the semitrivial steady state $S_1 = (\kappa, 0)$ and the nontrivial steady state $S^* = (x^*, y^*) = (x^*, g(x^*))$ with $x^* = f^{-1}(\alpha)$ which is biologically realistic for $x^* = f^{-1}(\alpha) < \kappa$, or equivalently $f(x^*) = \alpha < f(\kappa)$. It is clear that the predators cannot survive if $\alpha > f(\kappa) = \kappa/(1+\kappa)$, because then they could not be sustained even if their prey were at carrying capacity.

3.3.3 Stability analysis

The Jacobian J at a general point (x, y) is given by

$$\begin{aligned}J(x,y) &= \begin{pmatrix}1-2x/\kappa - yf'(x) & -f(x)\\ \beta y f'(x) & \beta(f(x)-\alpha)\end{pmatrix}\\ &= \begin{pmatrix}f(x)g'(x)+f'(x)(g(x)-y) & -f(x)\\ \beta y f'(x) & \beta(f(x)-\alpha)\end{pmatrix}.\end{aligned}$$

At S_0, the Jacobian $J_0 = J(0,0)$ is given by

$$J_0 = \begin{pmatrix}1 & 0\\ 0 & -\alpha\beta\end{pmatrix}$$

with eigenvalues 1 and $-\alpha\beta$. So S_0 is always a saddle point.

At S_1, the Jacobian $J_1 = J(\kappa, 0)$ is given by

$$J_1 = \begin{pmatrix} -1 & -f(\kappa) \\ 0 & \beta(f(\kappa) - \alpha) \end{pmatrix}$$

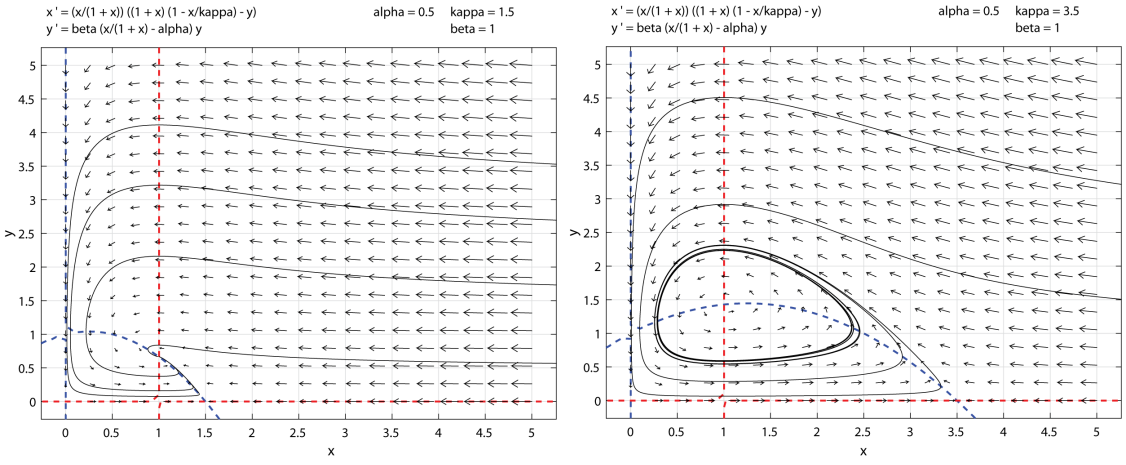
with eigenvalues $\lambda_1 = -1 < 0$ and $\lambda_2 = \beta(f(\kappa) - \alpha)$. So S_1 is a saddle point if the nontrivial steady state S^* is biologically realistic and a stable node otherwise.

At S^*, the Jacobian $J^* = J(x^*, g(x^*))$ is given by

$$J^* = \begin{pmatrix} f(x^*)g'(x^*) & -f(x^*) \\ \beta y^* f'(x^*) & 0 \end{pmatrix} = \begin{pmatrix} \alpha g'(x^*) & -\alpha \\ \beta y^* f'(x^*) & 0 \end{pmatrix}.$$

By the Routh–Hurwitz criteria, the equilibrium is locally asymptotically stable if and only if both $\mathrm{tr}J = \alpha g'(x^*) < 0$ and $\det J = \alpha\beta y^* f'(x^*) > 0$. Since α and β are positive parameters and $f'(x) > 0$ for all x then the second condition holds if and only if $y^* > 0$, or S^* is biologically realistic. Thus, the nontrivial equilibrium is asymptotically stable if $g'(x^*) < 0$ and unstable if $g'(x^*) > 0$. If $g'(x^*) = 0$, the eigenvalues are purely imaginary, and we therefore expect them to be complex for nearby values of x^*. The Hopf Bifurcation Theorem (discussed later) guarantees that a limit cycle occurs close to this borderline between oscillatory stability and oscillatory instability.

Note that $g'(x^*)$ is the slope of the parabolic prey nullcline at its intersection with the predator nullcline. That is, the nontrivial equilibrium is stable if the vertical predator nullcline is to the right of the maximum of the prey nullcline. A Hopf bifurcation occurs as the predator nullcline passes through the peak of the parabola.



The instability condition $g'(x^*) > 0$ is equivalent to $\kappa > \frac{1+\alpha}{1-\alpha}$. So, in summary:

	$\kappa < \frac{\alpha}{1-\alpha}$	$\frac{\alpha}{1-\alpha} < \kappa < \frac{1+\alpha}{1-\alpha}$	$\kappa > \frac{1+\alpha}{1-\alpha}$
$S_0 = (0,0)$	saddle point	saddle point	saddle point
$S_1 = (\kappa, 0)$	stable node	saddle point	saddle point
$S^* = (x^*, g(x^*))$	unrealistic (saddle point)	stable	unstable

What is causing the instability and thus the occurrence of limit cycles? With increasing prey density the additional prey consumption becomes smaller as the predations rate saturate. This limits the growth of the predators and allows the prey to grow until their negative density dependence kicks in. In some sense, saturation thus introduces a kind of delay effect, which typically tends to be destabilising.

3.3.4 Paradox of enrichment

This term was coined by Rosenzweig (1971). In the model, enrichment means increaseing the carrying capacity K (i.e. non-dimensional κ. This moves the peak of parabolic prey nullcline to the right. A large enough increase in K destabilises the coexistence state via a Hopf bifurcation. Then oscillations occur, and the limit cycle may come close to the axes and thus to extinction. *So enrichment is not always good for ecosystems.*

3.4 Limit cycles and periodic orbits

3.4.1 Hopf Bifurcation Theorem

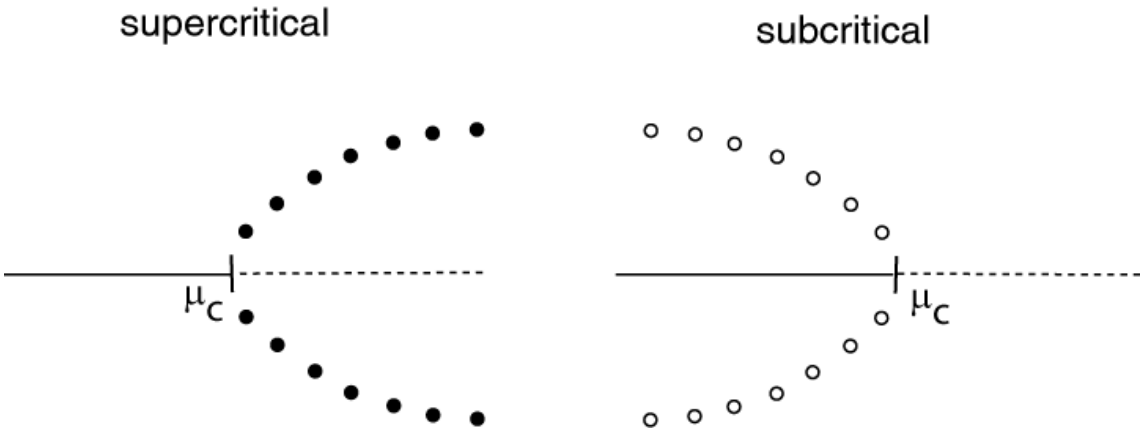
The simple Lotka–Volterra predator–prey model possesses a continuum of periodic orbits that are centre oscillations. They are not limit cycles because they are not isolated. They are structurally unstable because they change in response to perturbations of the system. The Rosenzweig–MacArthur model, by contrast, exhibits limit-cycle oscillations, which are *robust* periodic orbits and emerge in Hopf bifurcations. The Hopf Bifurcation Theorem states sufficient conditions for the existence of such limit cycles.

Consider the system

$$\frac{dx}{dt} = f(x, y, \mu), \qquad \frac{dy}{dt} = g(x, y, \mu).$$

Suppose that i) (x^*, y^*) is an equilibrium of the system depending on the parameter μ, ii) the equilibrium is stable in a certain range of μ but becomes unstable as μ passes through a bifurcation value μ_c, iii) linearisation about the equilibrium gives rise to a complex conjugate pair of eigenvalues $\lambda(\mu) = \mathrm{Re}\lambda(\mu) \pm \mathrm{i}\ \mathrm{Im}\lambda(\mu)$ in a neighbourhood of μ_c such that $\mathrm{Re}\lambda(\mu) < 0$ if $\mu < \mu_c$; $\mathrm{Re}\lambda(\mu) = 0$ and $\mathrm{Im}\lambda(\mu) \neq 0$ if $\mu = \mu_c$, $\mathrm{Re}\lambda(\mu) > 0$ if $\mu > \mu_c$. Then there exists a periodic solution of the system, unique up to phase shifts, in a one-sided neighbourhood of μ_c.

There are two possibilities. In the supercritical case, a stable periodic limit-cycle solution exists for $\mu > \mu_c$, where the equilibrium solution is unstable. In the subcritical case, an unstable periodic limit-cycle solution exists for $\mu < \mu_c$, where the equilibrium solution is stable. The limit-cycle solutions have small amplitude near $\mu = \mu_c$, and approximate period $2\pi/\omega_c$, where $\omega_c = \mathrm{Im}(\lambda(\mu_c))$. The periodic solutions will only be observed in the supercritical case, although often subcritical Hopf bifurcations become stable later through another bifurcation. So for a Hopf bifurcation we have that $\mathrm{Re}\lambda < 0$ and $\mathrm{tr}J < 0$ when $\mu < \mu_c$, $\mathrm{Re}\lambda = 0$ and $\mathrm{tr}J = 0$ when $\mu = \mu_c$, $\mathrm{Re}\lambda > 0$ and $\mathrm{tr}J > 0$ when $\mu > \mu_c$ while $\mathrm{Im}\lambda \neq 0$ and $\det J > 0$. Hence, as a parameter is varied, the dynamics of the system change from stable focus $\rightarrow$ centre $\rightarrow$ unstable focus. The eigenvalues change from having negative real parts to zero real parts to positive real parts, and there exist periodic oscillations.



Hopf bifurcations can occur only in second or higher order systems of ordinary differential equations (because they require complex eigenvalues). A Hopf bifurcation may also be called a Poincaré–Andronov–Hopf bifurcation.

3.4.2 Poincaré–Bendixson Theorem

The Poincaré–Bendixson Theorem states conditions for the existence of periodic solutions of a two-dimensional system. Consider the system

$$\frac{dx}{dt} = f(x, y), \qquad \frac{dy}{dt} = g(x, y).$$

Let f and g be continuously differentiable. Let the solution trajectory $(x(t), y(t))$ be bounded as $t \to \infty$. Then $(x(t), y(t))$ is at an equilibrium, or it tends to an equilibrium as $t \to \infty$, or it tends to a periodic orbit.

3.4.3 Dulac's Criterion

Dulac's criterion gives conditions sufficient to rule out the possibility of periodic solutions of systems of two ODEs. Consider the ODE system above. Let Ω be a simply connected region of the plane. Let the functions f and g be continuously differentiable in Ω. Let the function B be continuously differentiable in Ω such that

$$\frac{\partial(Bf)}{\partial x}(x, y) + \frac{\partial(Bg)}{\partial y}(x, y)$$

is not identically zero and does not change sign in Ω. Then there are no periodic orbits of the system in Ω.

Note that Dulac's Criterion is a sufficient but not necessary condition for the nonexistence of periodic solutions. If it is not satisfied there may or may not be periodic solutions. Usually the difficulty is finding the function B, which is called a Dulac function. The special case $B(x, y) = 1$ for all $(x, y) \in \Omega$ is called Bendixson's Criterion.

Example
The Lotka–Volterra prey–predator model (MA20221 and MA20202) in non-dimensional form is

$$\frac{dx}{dt} = x(1 - y), \qquad \frac{dy}{dt} = a(x - 1)y$$

with a positive. If prey population growth in the absence of predators is logistic the model becomes:

$$\frac{dx}{dt} = f(x,y) = x(1-\epsilon x - y), \qquad \frac{dy}{dt} = g(x,y) = a(x-1)y$$

where $\epsilon > 0$ and $1/\epsilon$ is the (non-dimensional) prey carrying capacity. This modified model has no periodic solution. To see this, let $B(x,y) = \frac{1}{xy}$. This is often a good choice for models of interacting populations. Let Ω be the positive quadrant, and note that B is continuously differentiable there. Thus,

$$\begin{aligned} \frac{\partial(Bf)}{\partial x}(x,y) + \frac{\partial(Bg)}{\partial y}(x,y) &= \frac{\partial}{\partial x}\,\frac{(1-\epsilon x - y)}{y} + \frac{\partial}{\partial y}\,\frac{a(x-1)}{x} \\ &= -\frac{\epsilon}{y} < 0 \quad \text{in } \Omega. \end{aligned}$$

Dulac's criterion implies that no periodic solution exists in the positive quadrant. Recall that the unmodified system (with $\epsilon = 0$) has infinitely many periodic solutions, while we have just shown that the system with $\epsilon > 0$ has none, however small ϵ may be. This qualitative change of solution behaviour as a result of arbitrarily small changes in the model is called *structural instability*.

Chapter 4

Discrete-time models

Discrete-time models may be appropriate if a population is censused at intervals, or reproduction occurs during a well defined breeding season, rather than continuously. The general form for a single species discrete-time model is the difference equation

$$N_{t+1} = f(N_t) = N_t F(N_t)$$

where N_t is the population size (or density) at time t.

Time is measured in discrete steps $t \in \mathbb{Z}$. The interval between t and $t+1$ is called the *projection interval* and may, for example, be a year or a generation or the time between censuses. We will refer to this interval as a year. The *net production* of a population of size N in a year is given by $f(N) \geq 0$, with $f(0) = 0$ by the axiom of parenthood, and has the same dimension as the population size. The *net per capita production* of a population of size N in a year is given by $F(N) \geq 0$, and is dimensionless. This model assumes that the population size at time $t+1$ is completely determined by the population size at time t. Note that neither the production nor the per capita production are rates. Rates occur exclusively in continuous-time models where they describe the change of population size per unit time. The production gives the population size in the next time step.

4.1 Linear difference equations

If the per capita production does not depend on population size (so that $F(N)$ is a constant) the population growth is *density-independent*. This gives a linear difference equation

$$N_{t+1} = \lambda N_t$$

where λ is the net per capita production, also called the *net growth ratio*. It can be thought of as being made up of birth and death processes with $\lambda = 1 + b - d$ where the *per capita reproduction* $b \geq 0$ is the average number of offspring for each individual in the population in a year that survive to be counted at the end of that year, and the *per capita mortality* $0 \leq d \leq 1$ is the expected fraction of the population alive at the beginning of the year that is dead by the end of the year, or the probability that a given individual dies during the year. Note that $\lambda \geq 0$. The parameters b and d, and hence λ, are dimensionless.

$$N_{t+1} = N_t + \Delta N_t = \underbrace{N_t}_{\substack{\text{present} \\ \text{population}}} + \underbrace{b\,N_t}_{\substack{\text{total number} \\ \text{of births}}} - \underbrace{d\,N_t}_{\substack{\text{total number} \\ \text{of deaths}}}$$

The solution to the system with initial condition N_0 is $N_t = \lambda^t N_0$. If $\lambda > 1$ then the population grows geometrically (Malthusian growth) while if $0 < \lambda < 1$ the population decays geometrically. Note that for $\lambda > 0$ we may write $N_t = N_0 \exp(t \ln \lambda)$, so the growth or decay may also be described as exponential.

Non-overlapping generations

Consider a univoltine insect, with well defined annual non-overlapping generations, so that $d = 1$. Then the average number R_0 of offspring produced by each insect over its lifetime is equal to the average number produced over a year, so $R_0 = b$, and $\lambda = 1 + b - d = R_0$. We usually assume that R_0 refers to the expected number of offspring produced under ideal conditions e.g. in the absence of competition or other density-dependent effects which may reduce reproduction or survival.

4.2 Nonlinear difference equations

Density dependence occurs if the number of offspring per adult (per capita production $F(N)$) depends on the population size, often due to intraspecific competition. This gives rise to a *nonlinear* difference equation

$$N_{t+1} = F(N_t)N_t = f(N_t).$$

Steady states N^* of such equations satisfy $N^* = f(N^*)$. Stability of a steady state N^* is often determined by *linear stability analysis*. Defining $n_t = N_t - N^*$, we have

$$n_{t+1} = N_{t+1} - N^* = f(N_t) - f(N^*) = f(N^* + n_t) - f(N^*).$$

Using a Taylor expansion for $f(N^* + n_t)$ this equation is approximated by the *linearisation*

$$n_{t+1} = f'(N^*)n_t = \lambda n_t.$$

Hence, *if* we can neglect higher order terms, we may summarise the behaviour of n_t as follows.

$\lambda < -1$	$-1 < \lambda < 0$	$0 < \lambda < 1$	$1 < \lambda$
oscillatory growth	oscillatory decay	monotonic decay	monotonic growth

Since n_t is a perturbation from the steady state N^*, monotonic or oscillatory stability or instability of N^* follows.

$f'(N^*) < -1$	$-1 < f'(N^*) < 0$	$0 < f'(N^*) < 1$	$1 < f'(N^*)$
oscillatory unstable	oscillatory stable	monotonic stable	monotonic unstable

It can be shown that we can indeed neglect higher order terms as long as we are not on the borderline between different behaviours, where the behaviour that changes may be affected by these terms. For example, if $f'(N^*) = 0$ the steady state N^* is stable for the nonlinear equation, but may be oscillatory or monotonic depending on the nonlinearity. Corresponding results in terms of $F(N)$ are not so complete, but may be found by differentiating $f(N) = NF(N)$ to obtain $f'(N) = F(N) + NF'(N)$. Since $f'(0) = F(0)$, the trivial steady state is monotonically stable for $0 < F(0) < 1$ and monotonically unstable for $F(0) > 1$, while the case $F(0) < 0$ (negative per capita production) is unrealistic. A nontrivial steady state (where $F(N^*) = 1$) is monotonically unstable if $F'(N^*) > 0$ (since

then $f'(N^*) = 1 + N^*F'(N^*) > 1$), but it can be monotonically stable, oscillatory stable, or oscillatory unstable if $F'(N^*) < 0$, and stability is better analysed in terms of $f'(N^*)$.

Example
The Beverton–Holt (1957) model is

$$N_{t+1} = \frac{R_0 N_t}{1 + aN_t}$$

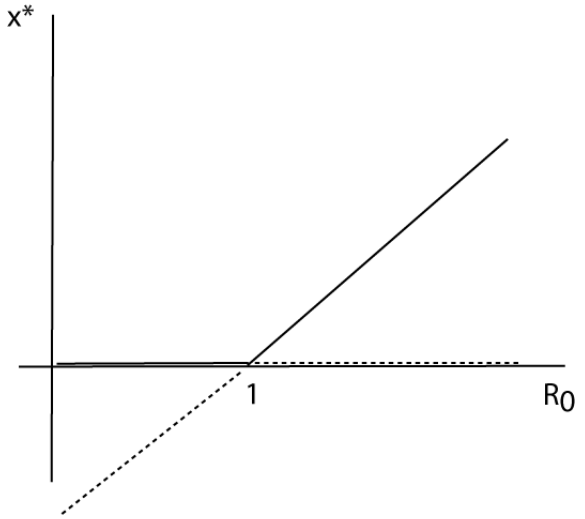
where R_0 and a are positive parameters. The parameter R_0 is dimensionless, while a has the dimension [population size]$^{-1}$. We can simplify the model by rescaling. Defining $x_t = aN_t$ we obtain

$$\frac{x_{t+1}}{a} = \frac{R_0(x_t/a)}{1 + a(x_t/a)}, \quad \text{or} \quad x_{t+1} = f(x_t) = \frac{R_0 x_t}{1 + x_t}.$$

Note that $f(x) > 0$ for $x > 0$, $f'(x) = \dfrac{R_0}{(1+x)^2} > 0$ for $x \geq 0$ and $f(x) \to R_0$ as $x \to \infty$. Steady states x^* of this map are given by $x^* = f(x^*) = \dfrac{R_0 x^*}{1 + x^*}$ so the steady states are $x^* = 0$, the trivial steady state, and $x^* = R_0 - 1$ the nontrivial steady state, which is biologically meaningful for $R_0 > 1$. We have $f'(0) = F(0) = R_0$ and $f'(R_0 - 1) = 1/R_0$, so linear stability analysis yields the following result

	$0 < R_0 < 1$	$1 < R_0 < \infty$
$x^* = 0$	monontonic stable	monotonic unstable
$x^* = R_0 - 1$	unrealistic	monotonic stable

There is a transcritical bifurcation at $R_0 = 1$, analogous to those we have seen in continuous-time equations. In continuous-time systems we had $\lambda = f'(x^*) = 0$ at such bifurcations, while in discrete-time systems we shall always have $\lambda = f'(x^*) = 1$. The Beverton-Holt model has a unique nontrivial steady state which is always approached monotonically. There is no possibility of oscillatory behaviour since here $f'(x) > 0$ for all x and oscillatory behaviour only occurs when $f'(x^*) < 0$.



Other examples

Beverton and Holt (1957)	$N_{t+1} = \frac{R_0\, N_t}{1+a\, N_t}$
Hassell (1975)	$N_{t+1} = \frac{R_0\, N_t}{(1+a\, N_t)^\beta}$
Hassell variant	$N_{t+1} = \frac{R_0\, N_t}{1+a\, N_t^b}$
Allee effect, Myers et al (1995)	$N_{t+1} = \frac{R_0\, N_t^d}{1+a\, N_t^d}$
Quadratic (or logistic) map	$N_{t+1} = r\, N_t \left(1 - \frac{N_t}{K}\right)$
Ricker (1954)	$N_{t+1} = N_t\, e^{r\left(1-\frac{N_t}{K}\right)}$

4.3 Cobwebbing: a geometric technique

Consider the first-order equation

$$x_{t+1} = f(x_t).$$

Plot on the same diagram in the $(x, y)-$plane the curve $y = f(x)$ and the diagonal $y = x$. Let $(x, y) = (x^*, y^*)$ be a point of intersection of the curve and the diagonal. Then $y^* = x^*$ (since the point is on the diagonal) and $y^* = f(x^*)$ (since the point is on the curve). Hence $x^* = f(x^*)$, and x^* is a steady state of the system. We shall use the diagram to predict the behaviour of solutions of the system as t increases. The idea is to plot

$$(x_0, x_0) \to (x_0, x_1) \to (x_1, x_1) \to (x_1, x_2) \to \ldots.$$

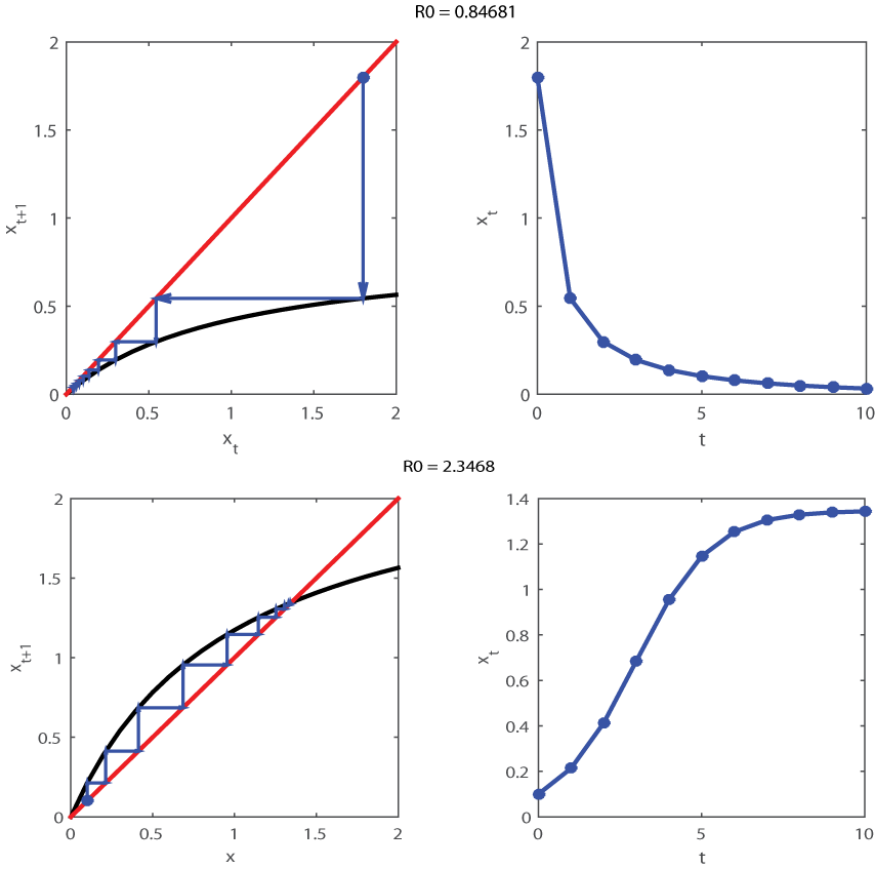
- STEP 0: choose an initial condition x_0 and plot (x_0, x_0).
- STEP 1: $(x_0, x_0) \to (x_0, x_1)$; since $x_1 = f(x_0)$, move vertically (up or down) to the curve $y = f(x)$.
- STEP 2: $(x_0, x_1) \to (x_1, x_1)$; move horizontally (left or right) to the diagonal $y = x$.
- REPEAT steps 1 and 2 until the asymptotic (large-time) behaviour becomes clear.

Note: a common alternative notation is to label the x axis as x_t, and the y axis as x_{t+1}, and to plot the curve $x_{t+1} = f(x_t)$ and the diagonal $x_{t+1} = x_t$.

Example
Cobweb plots for the Beverton–Holt model,

$$x_{t+1} = f(x_t) = \frac{R_0 x_t}{1 + x_t},$$

with $R_0 > 1$ and $R_0 < 1$. Note that the stability properties of the steady states $x^* = 0$ and $x^* = R_0 - 1$ are as expected.



4.4 Density dependence in discrete-time models

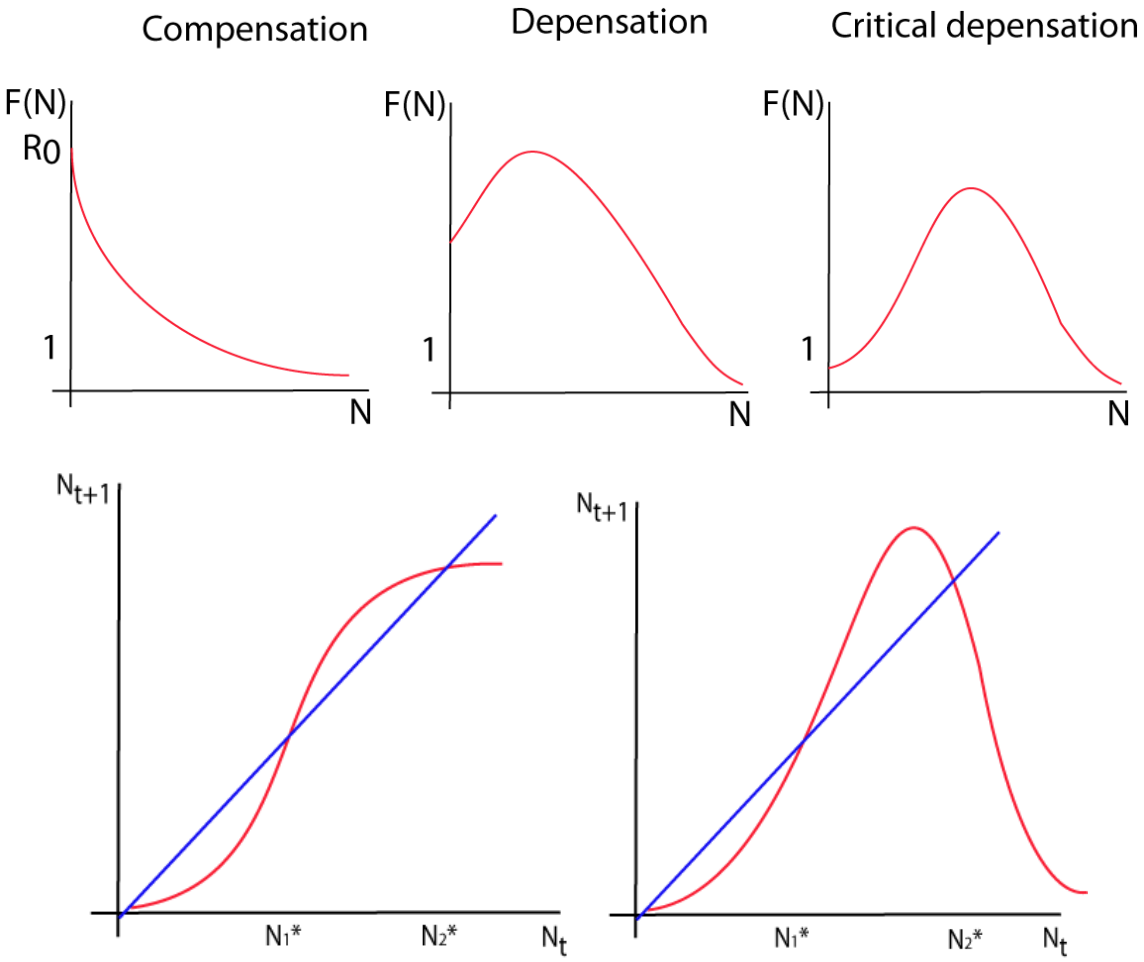
Just as for continuous-time models, we can distinguish general growth forms according to the impact of density-dependence.

4.4.1 Depensation, $F'(N) > 0$ for small N

Depensation, or an Allee effect, means that a decrease in N leads to a decrease in per capita production $F(N)$, or $F'(N) > 0$. We do not expect this to hold for large populations, because of intraspecific competition, so we require $F'(N) > 0$ for small N.

Critical depensation, or a strong Allee effect, is an extreme case of depensation where there exists a critical population size below which $F(N) < 1$ (or equivalently $f(N) < N$), so that production does not replace the current population: $F(N) < 1$ and $F'(N) > 0$ for small N. If the population does not go extinct, then there exists a point N^* with $F(N^*) = 1$ and $F'(N^*) > 0$, which must therefore be a monotonically unstable steady state. This is the critical population size.

For a strong Allee effect (critical depensation) there is a threshold N_1^* below which the population in the next generation is smaller than in the current generation. Any population that drops below N_1^* will continue to fall in size until it becomes extinct. On the other hand, any population starting above N_1^* will grow. Typically there will be another non-trivial steady state N_2^* that may be monotonic stable, oscillatory stable or oscillatory unstable.



4.4.2 Compensation, $F'(N) < 0$ for all N

Compenstation means that per capita production $F(N)$ is a decreasing function of N, $F'(N) < 0$ for all $N \geq 0$. We would always expect this to hold for large N, where intraspecific competition is dominant, but in compensatory dynamics it holds for all N. There is no Allee effect. We have already seen the Beverton–Holt example where $F(N) = R_0/(1+aN)$ is a decreasing function of N.

Compensatory dynamics can be further divided into two categories according to the behaviour of $f(N)$ as $N \to \infty$.

Undercompensation where $f(N) \to \infty$ as $N \to \infty$.

Overcompensation where $f(N) \to 0$ as $N \to \infty$.

Exact (or perfect) compensation is the dividing line between these, where $f(N) \to c > 0$ as $N \to \infty$.

With under- or exact compensation we typically have $f'(N) > 0$ for all N, so any steady state N^* is monotonically stable or unstable (with $\lambda = f'(N^*) > 0$), while with overcompensatory dynamics there is the possibility of oscillatory behaviour near a steady state (with $\lambda = f'(N^*) < 0$).

How do these different forms of compensation arise? Ecologists distinguish between two idealised and extreme forms of (intraspecific) competition. Let us assume that the population grows geometrically, $N_{t+1} = R_0 N_t$, in the absence of intraspecific competition but individuals compete for a finite number of resources (e.g. safe refuges or exclusive use of a feeding territory) such that the survival function $S(N)$ that gives the fraction of offspring that survive the competition and $N_{t+1} = R_0 N_t S(N_t)$.

Contest competition. If an individual obtains one of the limited resources it survives and reproduces; otherwise it dies. Let R be the number of available resources. Then

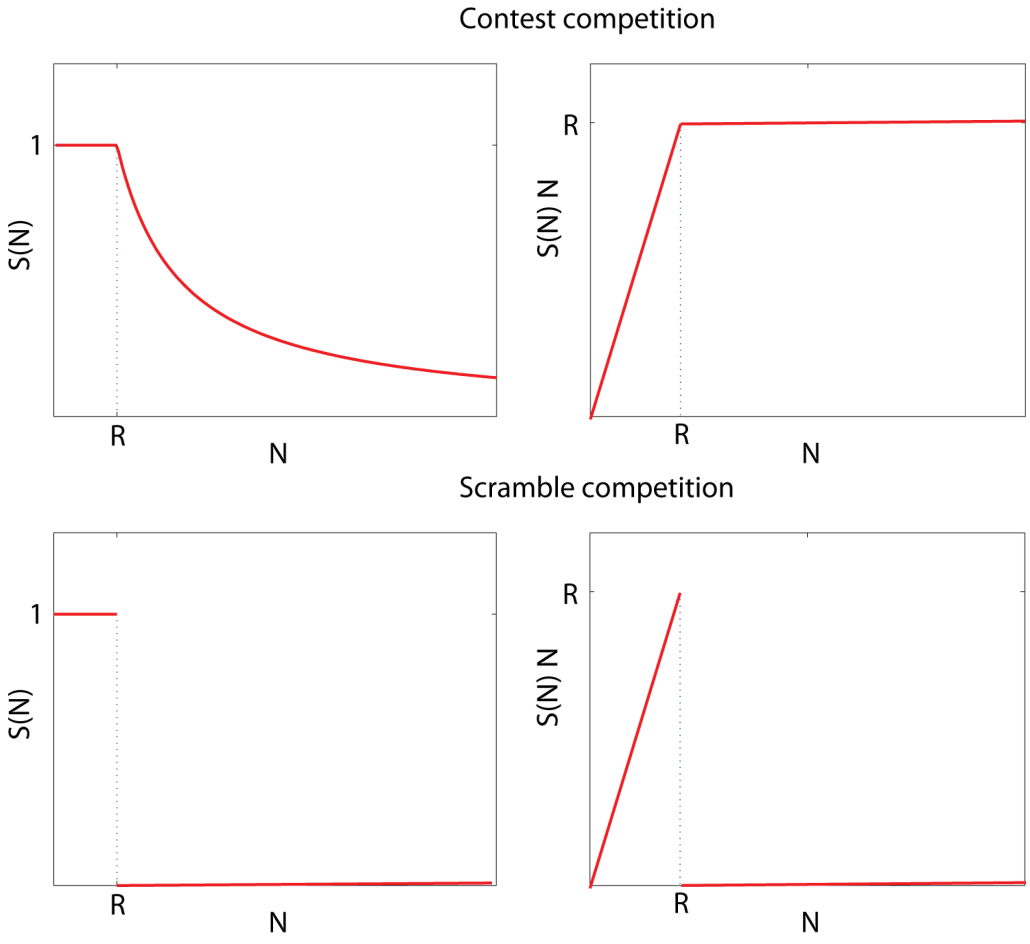
$$S(N) = \begin{cases} 1 & \text{for } N \leq R, \\ R/N & \text{for } N > R. \end{cases}$$

Some individuals win, while others lose. With increasing density recruitment levels off, but does not actually decrease.

Scramble competition. Each individual is obtains an equal share of the available resources. If this share is enough to survive and reproduce, all individuals will do so equally; otherwise, all individuals die.

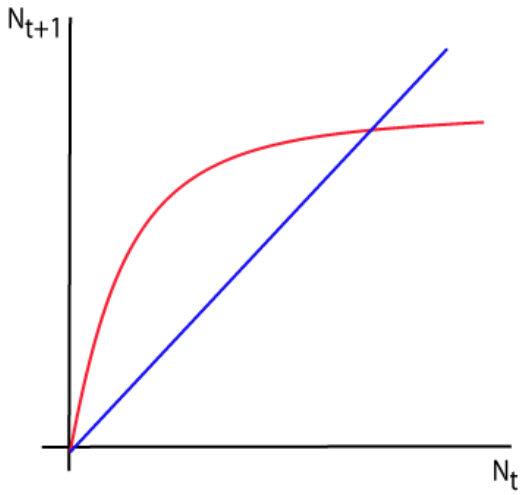
$$S(N) = \begin{cases} 1 & \text{for } N \leq R, \\ 0 & \text{for } N > R. \end{cases}$$

Populations typically crash if they become too large.



Under- or exact compensation

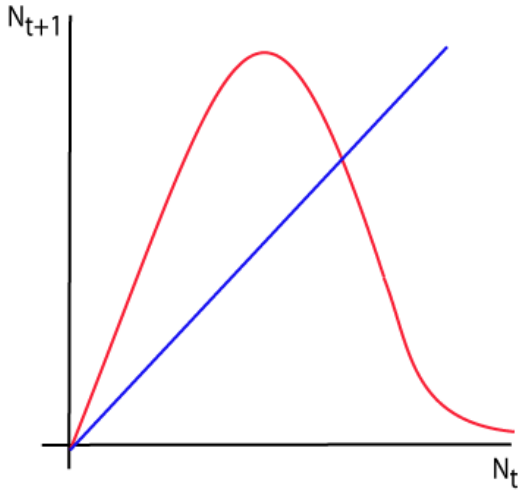
If $N_{t+1} = f(N_t)$ with $f(N)$ monotonic increasing then the dynamics exhibit under- or exact compensation. For exact compensation ($f(N) \to c > 0$, a constant, as $N \to \infty$) if, in addition, $f'(0) > 1$ then it is easy to see geometrically that there exists a nontrivial steady state N^* with $f'(N^*) < 1$, so that N^* is monotonically stable.



The Beverton–Holt model exhibits exact compensation, while the Hassell model and its variant exhibit under-compensation for β or b less than 1.

Over-compensation

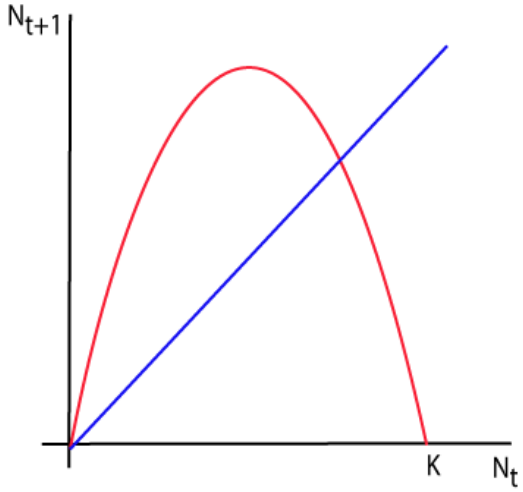
If $N_{t+1} = f(N_t)$ with $f(N)$ over-compenstaory ($f(N) \to 0$ as $N \to \infty$), unimodal (i.e. $f(N)$ is monotonic increasing for $N < N^\dagger$ and monotonic decreasing for $N > N^\dagger$) and $f'(0) > 1$ it is easy to see that there exists a nontrivial steady state N^*, and $f'(N^*)$ may be negative if it is located on the decreasing branch of the map. The Hassell model and its variant are over-compensator for β or b greater than 1.



The quadratic or logistic map $N_{t+1} = rN_t(1 - N_t/K)$ must be modified to

$$N_{t+1} = \max\left\{rN_t\left(1 - \frac{N_t}{K}\right), 0\right\}$$

to give a guarantee that the population size does not become negative. It *looks* similar to the logistic growth equation in continuous time, but it exhibits very different behaviour as it is over-compensatory.



4.5 Asymptotic dynamics of over-compensatory maps

If a steady state is stable, solutions starting near to it stay close to it in the future. The term *asymptotic* behaviour refers to the long-term dynamics as time goes to infinity. In contrast, *transient* behaviour refers to pre-asymptotic dynamics.

Consider the scaled Hassell model given by

$$x_{t+1} = f(x_t) = \frac{R_0 x_t}{(1+x_t)^\beta}$$

where $R_0 > 0$ and $\beta > 1$. Since $f(x) \to 0$ as $x \to \infty$ this is an over-compensatory map, so steady states x^* with $f'(x^*) < 0$ are a possibility. In fact

$$x^* = f(x^*) = \frac{R_0 x^*}{(1+x^*)^\beta}$$

gives $x^* = 0$, the trivial steady state, and $(1+x^*)^\beta = R_0$, equivalently $x^* = R_0^{1/\beta} - 1$, the nontrivial steady state, which is biologically meaningful for $R_0^{1/\beta} > 1$ or $R_0 > 1$. Now

$$f'(x) = R_0 \frac{1 + x - \beta x}{(1+x)^{\beta+1}},$$

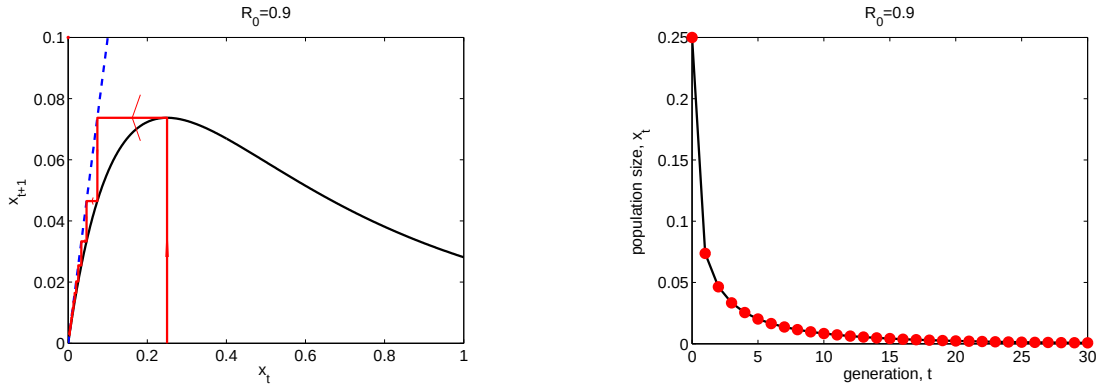
so $f'(0) = R_0$ and

$$f'(x^*) = 1 - \beta \frac{R_0^{1/\beta} - 1}{R_0^{1/\beta}}.$$

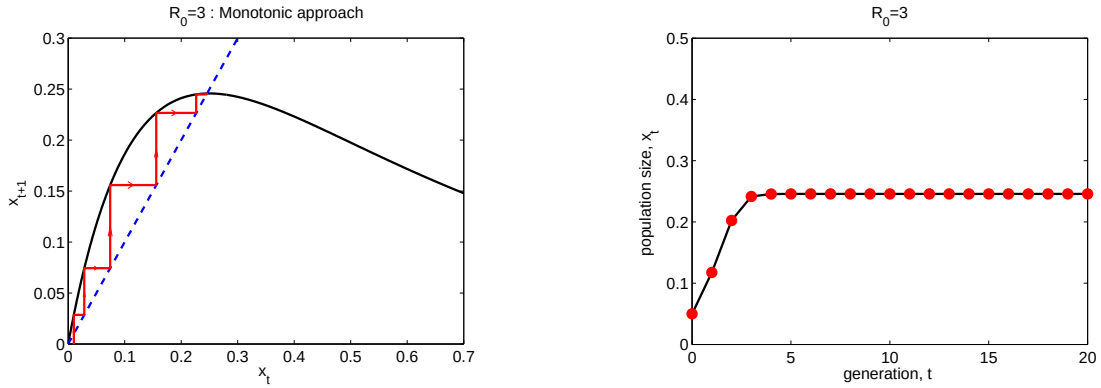
using the fact that $(1+x^*)^\beta = R_0$. Linear stability analysis yields the following result, where R_1 and R_2 are determined by the conditions $f'(x^*) = 0$ and $f'(x^*) = -1$, respectively.

x^*	$0 < R_0 < 1$	$1 < R_0 < R_1$	$R_1 < R_0 < R_2$	$R_0 > R_2$
0	monotonic stable	monotonic unstable	monotonic unstable	monotonic unstable
$R_0^{1/\beta} - 1$	unrealistic, monotonic unstable	monotonic stable	oscillatory stable	oscillatory unstable

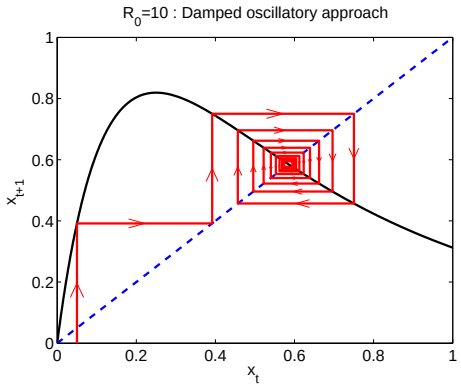
If R_0 is sufficiently large in the Hassell model, both the trivial and the nontrivial steady states are unstable. We shall investigate its asymptotic behaviour numerically. We shall keep $\beta = 5$ fixed and vary $R_0 > 0$ as a control (bifurcation) parameter. For $R_0 < 1$, there is no realistic nontrivial steady state. The trivial solution is asymptotically stable.

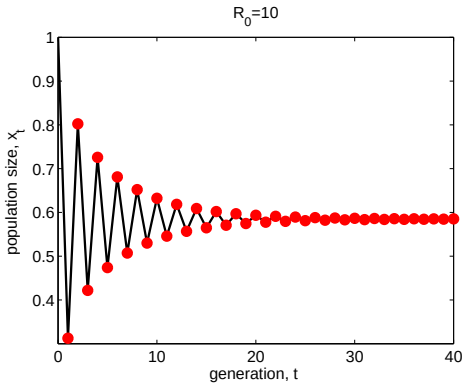


At $R_0 = 1$, we have a *transcritical bifurcation*. The nontrivial steady state crosses the trivial solution to become positive and exchanges stability with it. This is indicated by $\lambda = f'(x^*) = f'(0) = 1$. For $1 < R_0 < R_1 \approx 3.05$, the nontrivial steady state is monotonically asymptotically stable, while the trivial solution is unstable.

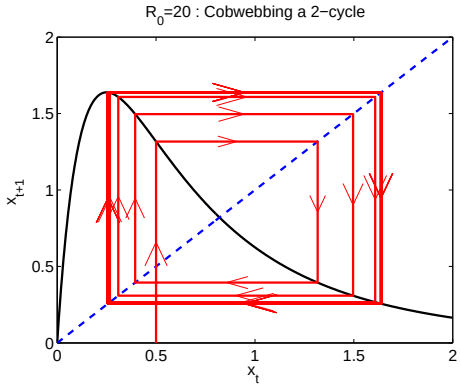


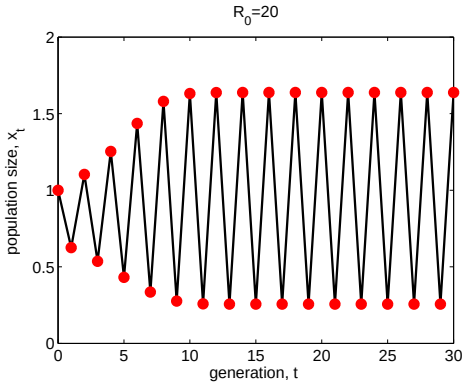
For $R_1 < R_0 < R_2 \approx 12.86$, the approach to the nontrivial steady state is oscillatory. Note the different scales of the axes throughout this series of figures, and how the branch of the map intersecting the identity line becomes steeper.



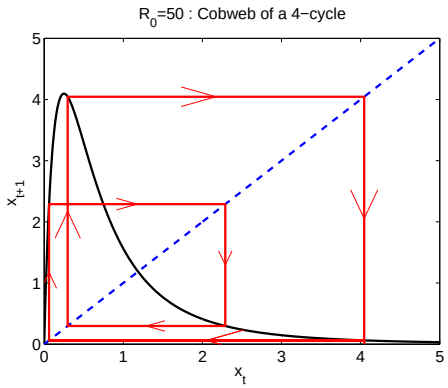


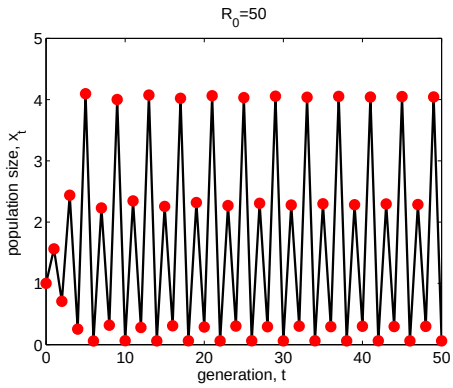
At $R_0 = R_2$, there is a *period-doubling bifurcation*. This is indicated by $\lambda = f'(x^*) = -1$. For $R_0 > R_2$, the nontrivial steady state becomes unstable, while the trivial steady state remains unstable. We can observe a variety of different behaviour. The first one is a 2-cycle.



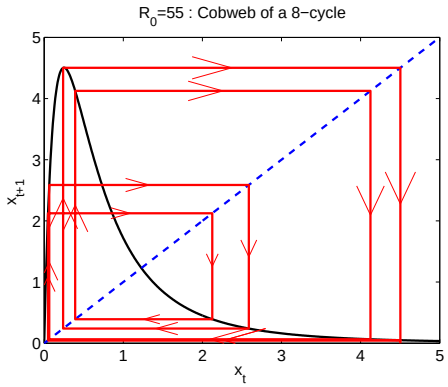


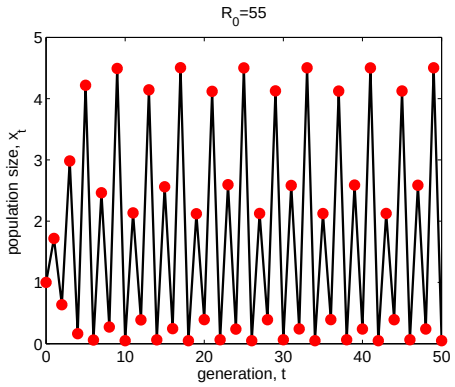
The next one is a 4-cycle...



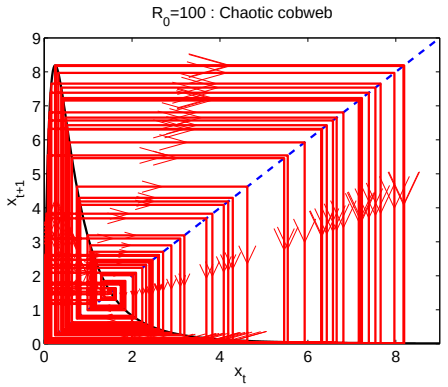


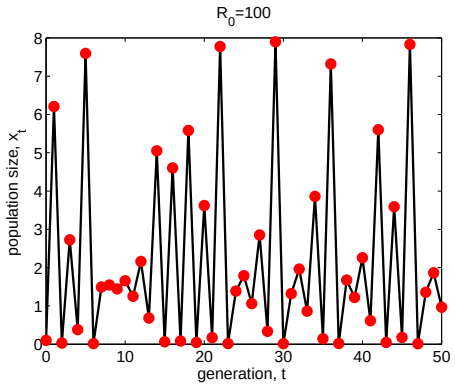
...followed by an 8-cycle...



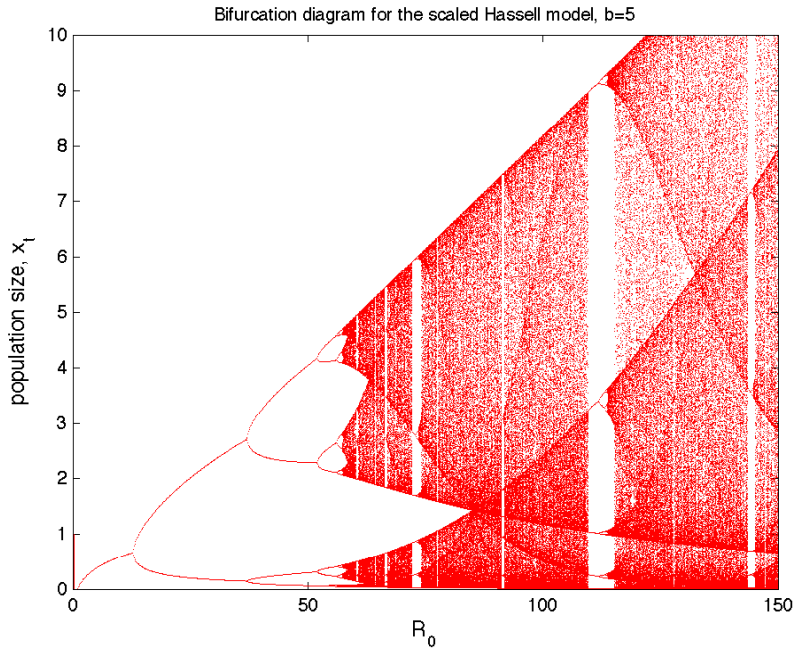


...and so on... A *period-doubling* cascade takes place. In fact, there are infinitely many cycles of period 2^p. Beyond a critical value R_3, we can find aperiodic and irregular dynamics, which we may call *chaotic*.





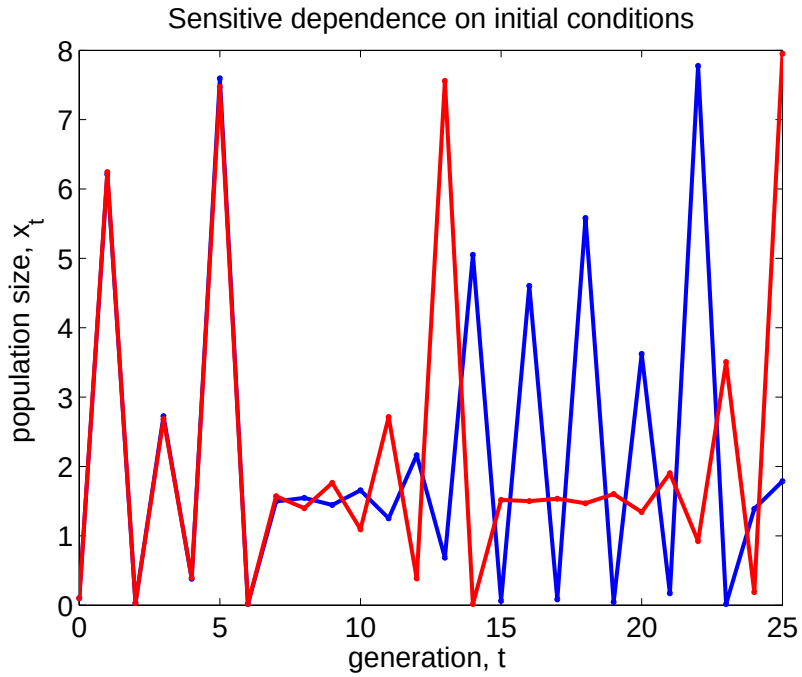
The following figure is a *bifurcation diagram*.



For each value of R_0 we plot the attractor starting from a fixed initial condition. We do not indicate unstable steady states. A p-cycle appears as p different points for a fixed value of R_0. The asymptotic behaviour is indicated by: single point - stable steady state, two points - stable 2-cycle, four points - stable 4-cycle. Period-doubling bifurcations appear as forks, and that the chaotic region is interwoven with periodic cycles called 'periodic windows'.

Deterministic chaos

Chaos can be defined as 'deterministic, aperiodic, bounded dynamics with sensitive dependence on initial conditions". Deterministic: there are no random terms in the governing equations. Aperiodic: iterations do not repeat themselves. Bounded: state variables remain in finite ranges; they do not approach $\pm\infty$ but small differences in the initial conditions may be amplified exponentially. Sensitive dependence on initial conditions: small perturbations can become large, in fact as big as the signal itself. There is short-term predictability, but long-term predictions are worthless due to inevitable measurement and round-off errors. This is sometimes called the *butterfly effect*. Dynamics can be classified as chaotic using the (dominant) *Lyapunov exponent*. A positive Lyapunov exponent indicates exponential divergence of nearby trajectories and is a hallmark of chaos.



4.6 Discrete-time models for two interacting populations

Consider two interacting species with populations sizes X_t and Y_t in year t. A general discrete-time model for the system is

$$X_{t+1} = f(X_t, Y_t), \quad Y_{t+1} = g(X_t, Y_t).$$

Steady states (X^*, Y^*) of such systems satisfy

$$X^* = f(X^*, Y^*), \quad Y^* = g(X^*, Y^*)$$

which must be satisfied simultaneously. Stability of a steady state is determined by defining $x_t = X_t - X^*$, $y_t = Y_t - Y^*$, and linearising about the steady state. We obtain

$$\begin{aligned} x_{t+1} &= -X^* + X_{t+1} = -X^* + f(X_t, Y_t) = -X^* + f(X^* + x_t, Y^* + y_t) \\ &= -X^* + f(X^*, Y^*) + \frac{\partial f}{\partial X}(X^*, Y^*)x_t + \frac{\partial f}{\partial Y}(X^*, Y^*)y_t + \text{ higher order terms,} \end{aligned}$$

with a similar equation for y_{t+1}. Using $f(X^*, Y^*) = X^*$, $g(X^*, Y^*) = Y^*$ the linearisation is given by

$$x_{t+1} = f_X^* x_t + f_Y^* y_t, \quad y_{t+1} = g_X^* x_t + g_Y^* y_t$$

with the usual notation for partial derivatives. We may write this as

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} f_X^* & f_Y^* \\ g_X^* & g_Y^* \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix},$$

or $\boldsymbol{z}_{t+1} = J^* \boldsymbol{z}_t$ where $\boldsymbol{z}_t = (x_t, y_t)^T$ and $J^* = J(X^*, Y^*)$ is the Jacobian matrix at the steady state. Stability is determined by the eigenvalues of the Jacobian, just as for continuous-time systems, but now solutions are of the form λ^t rather than $\exp(\lambda t)$, so the dividing line between stability and instability is $|\lambda| = 1$ rather than $\mathrm{Re}\lambda = 0$. There are conditions similar to the Routh–Hurwitz conditions that are necessary and sufficient for stability, the Jury conditions.

Jury conditions
Both eigenvalues of a 2×2 matrix J^* satisfy $|\lambda| < 1$, so that (X^*, Y^*) is asymptotically stable, if and only if *all three* of the following conditions hold:

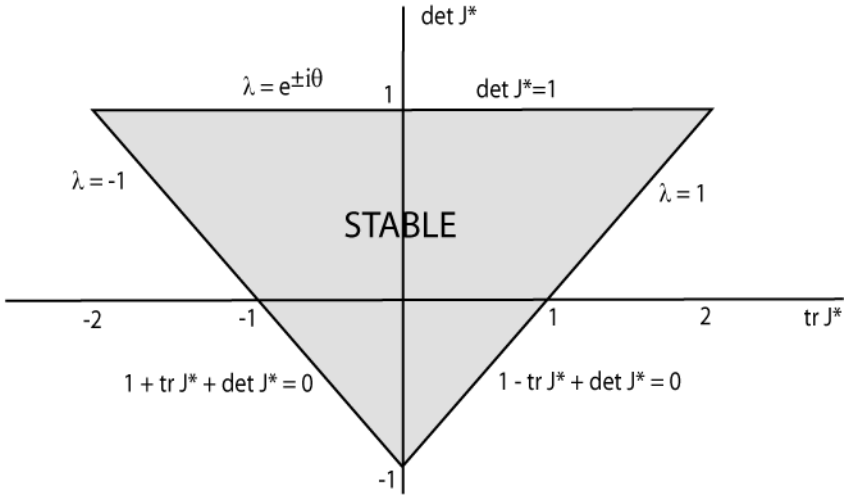
$$-\mathrm{tr}\ J^* < 1 + \det J^*, \quad \mathrm{tr}\ J^* < 1 + \det J^*, \quad \text{and} \quad \det J^* < 1.$$

These are often written $|\mathrm{tr}\ J^*| < 1 + \det J^* < 2$, but it is important to remember that they involve checking *three* inequalities.

The Jury conditions may be thought of in terms of the eigenvalues of the Jacobian. The eigenvalues satisfy the *characteristic equation* $P(\lambda) = 0$, where

$$P(\lambda) = \lambda^2 - \mathrm{tr}\ J^* \lambda + \det J^* = 0.$$

If $P(\lambda) = 0$ has a root $\lambda = 1$, then $1 - \mathrm{tr}\ J^* + \det J^* = 0$. If $P(\lambda) = 0$ has a root $\lambda = -1$, then $1 + \mathrm{tr}\ J^* + \det J^* = 0$. Finally, if $P(\lambda) = 0$ has a pair of complex conjugate roots $\lambda_\pm$ with $|\lambda| = 1$, which we can write $\lambda_\pm = \exp(\pm i\theta)$, then $\det J^* = 1$ (and $-2 \leq \mathrm{tr}\ J^* \leq 2$). Crossing one of the boundaries determined by the Jury conditions has clear consequences in terms of the behaviour of the eigenvalues.



4.6.1 Host–parasitoid systems

Parasitoids are parasites that live freely as adults but lay eggs in the larvae or pupae of the host. Unparasitised hosts develop normally. Parasitised hosts die, but the eggs laid by the parasitoid may survive to be the next generation of parasitoids. A general host-parasitoid model is

$$\begin{aligned} H_{t+1} &= R_0 H_t f(H_t, P_t) \\ P_{t+1} &= c H_t \left(1 - f(H_t, P_t)\right) \end{aligned}$$

where R_0 is the host basic reproduction number i.e. the per capita production in absence of parasitism, c is the average number of eggs laid by an adult parasitoid in a single host that survive to breed in the next generation, and $f(H, P)$ is the fraction of hosts *not* parasitised. This model assumes that in the absence of parasitism, the host grows geometrically, the parasitoid cannot survive without the host,and parasitism takes place before reproduction.

4.6.2 Nicholson–Bailey model

The Nicholson–Bailey model (1935) is the classical host-parasitoid sytstem. In addition to the general framework above it assumes that the number of encounters between hosts and parasitoids follows the *law of mass action*, and encounters occur at random and independently (so that parasitoids do not distinguish between hosts that have been parasitised and those that have not). Therefore the total number of encounters is aH_tP_t, where a is the searching efficiency. The *average* number n of encounters per host is $n = \frac{aH_tP_t}{H_t} = aP_t$. But the actual number of encounters a host experiences is desribed by a Poisson process. So we can use the Poisson distribution for the probability $p(i)$ that a host experiences i encounters $p(i) = \frac{n^i \mathrm{e}^{-n}}{i!}$. Since only the first encounter between a host and a parasitoid is important, the fraction of hosts *not* being parasitised is

$$f(H_t, P_t) = p(0) = \mathrm{e}^{-n} = \mathrm{e}^{-aP_t}.$$

Moreover, a parasitised host will bear exactly c parasitoid progeny, independently of the number of encounters. Hence

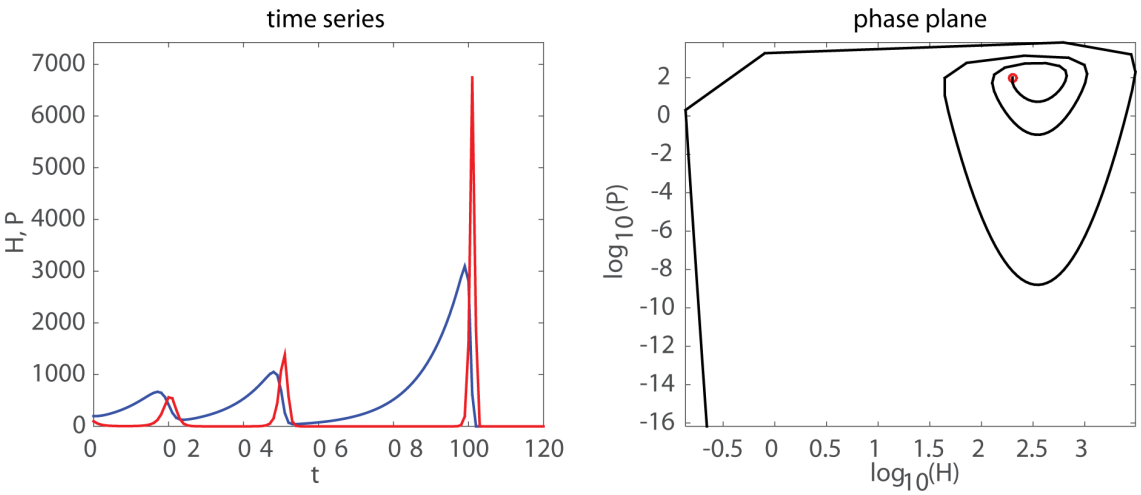
$$H_{t+1} = R_0 H_t \mathrm{e}^{-aP_t}, \quad P_{t+1} = cH_t(1 - \mathrm{e}^{-aP_t}).$$

Analysis

Steady states (H^*, P^*) satisfy $H^* = R_0 H^* \exp(-aP^*)$ and $P^* = cH^*(1 - \exp(-aP^*))$. Solutions are $(0, 0)$, the trivial steady state, and (H^*, P^*), the coexistence steady state, where $P^* = \frac{1}{a} \ln R_0$, $H^* = \frac{P^* R_0}{c(R_0 - 1)}$, which is biologically meaningful for $R_0 > 1$. The first two Jury conditions are satisfied, so the condition for stability is given by $\det J^* = \frac{R_0 \ln R_0}{R_0 - 1} < 1$ for $R_0 > 1$. Equivalently, we require $g(R_0) := R_0 - 1 - R_0 \ln R_0 > 0$ for $R_0 > 1$, but this does not hold because $g(1) = 0$ and

$$g'(R_0) = 1 - \left(\ln R_0 + R_0 \frac{1}{R_0} \right) = - \ln R_0 < 0$$

for $R_0 > 1$. Hence, the coexistence state in the Nicholson–Bailey model is unstable. Simulations show oscillations of ever increasing amplitude.



The ever increasing oscillations of the Nicholson–Bailey model are clearly unrealistic. How can we improve the model? Constant per capita production of the host could be replaced with density dependence. The random search for hosts by parasitoids may be replaced by a model in which parasitoids aggregate where there are many hosts, possibly attracted by chemical cues. The homogeneous environment may be replaced by a patchy environment with refuges in which hosts are safe from attacks by parasitoids. Additional interacting species may be incorporated. Probably the easiest way to stabilise the Nicholson–Bailey model is to introduce self-limitation (density-dependence) of the host population e.g. in the form of the Ricker map. The resulting model is:

$$H_{t+1} = H_t \mathrm{e}^{r(1 - H_t/K) - aP_t}, \quad P_{t+1} = cH_t(1 - \mathrm{e}^{-aP_t}).$$

The are similar to those of the discrete-time predator–prey model (Neubert and Kot, 1992) described below.

4.6.3 Predator-prey model

Consider the following discrete-time predator–prey model (Neubert and Kot, 1992). Let N_t be the prey population size in year t and P_t the predator population size. Assume the prey population grows logistically with net per capita reproduction r and carrying capacity K. Assume all predators die after one generation. Assume mass action predation with effective search rate c. Then

$$N_{t+1} = N_t + rN_t(1 - N_t/K) - cN_tP_t,$$
$$P_{t+1} = bN_tP_t.$$

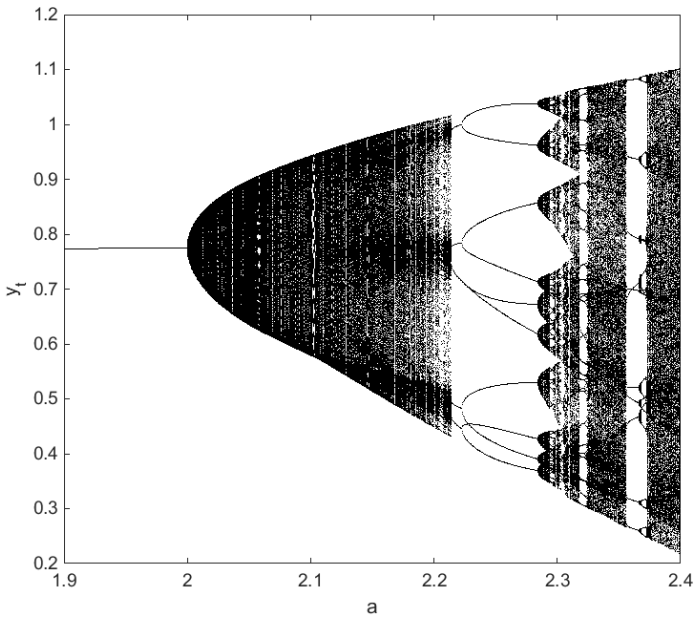
Rescaling with $x_t = N_t/K$, $y_t = cP_t/(bK)$, $a = bK$ gives

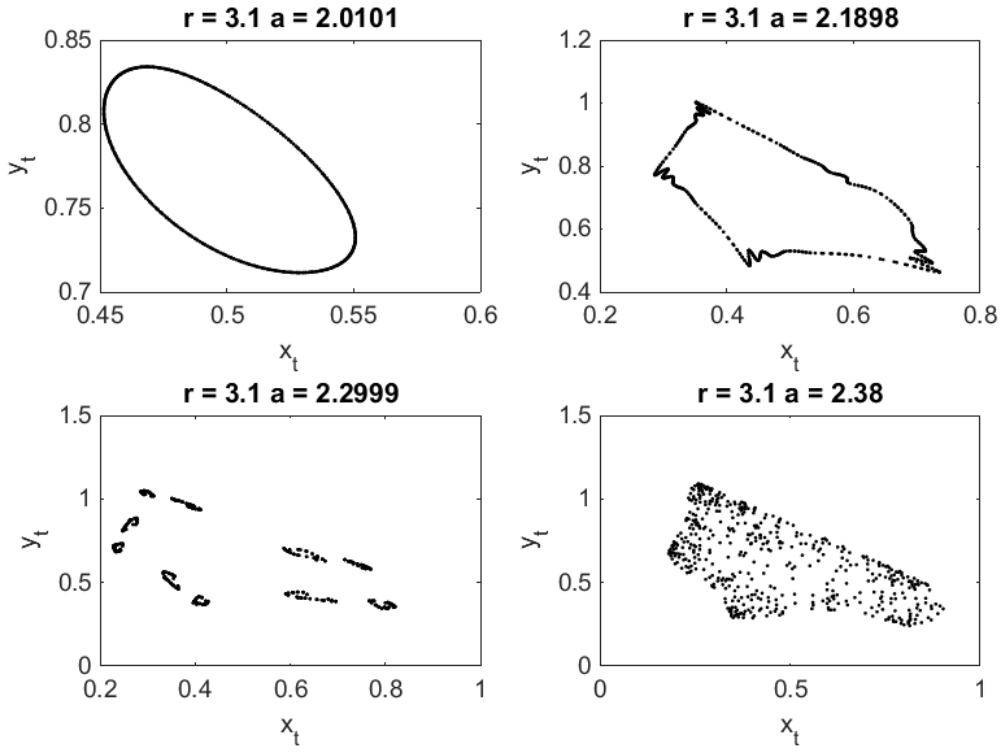
$$x_{t+1} = (1 + r)x_t - rx_t^2 - ax_ty_t, \quad y_{t+1} = ax_ty_t.$$

4.6.4 Neimark–Sacker bifurcation

The discrete-time predator–prey model by Neubert and Kot serves to illustrate a *Neimark–Sacker bifurcation*, often referred to as a Hopf bifurcation for difference equations. At a Neimark–Sacker bifurcation point, there are two complex conjugate eigenvalues λ_1 and λ_2 with $|\lambda_1| = |\lambda_2| = 1$. In terms of the Jacobian matrix, the bifurcation conditions are $-\mathrm{tr}\ J^* < 1 + \det J^*$, $\mathrm{tr}\ J^* < 1 + \det J^*$, $\det J^* = 1$. The steady state changes stability via a pair of complex eigenvalues with unit modulus. The Neimark–Sacker bifurcation gives birth to a *closed invariant curve* ('limit cycle') from a steady state. Note the difference between a closed invariant curve and a 2-cycle. The bifurcation can be *supercritical* or *subcritical*, resulting in a stable or unstable closed invariant curve, respectively.

The diagrams below shows the supercritical Neimark–Sacker bifurcation in the Neubert–Kot model (followed by other bifurcations). As a increases, a closed invariant curve appears, and then collapses. First it becomes kinked, then it undergoes a secondary Neimark–Sacker bifurcation, and eventually it becomes a chaotic attractor.




r = 3.1 a = 2.0101
r = 3.1 a = 2.1898
r = 3.1 a = 2.2999
r = 3.1 a = 2.38
y_t
x_t

Chapter 5

Infectious diseases

Parasites and pathogens can be broadly divided into microparasites (viruses, bacteria, protozoa) and macroparasites (helminths, ticks). Microparasites produce a sustained immunological response in the host, so models effectively divide the host into distinct epidemiological classes (like susceptible, infected, immune). Macroparasites do not produce a sustained immunonological response; the degree of infestation is important, so mathematical models track the distribution of individual macroparasites within the host population. We shall consider diseases caused by microparasites. Transmission of microparasites can occur directly between definitive hosts (referred to as *direct life cycle*; *contagious disease*) or involve one or more intermediate host species (referred to as *indirect life cycle*; *vector-borne disease*). An *epidemic* is an outbreak of a disease in a population at a particular time. An *endemic* disease is chronically present in a population. The infection *prevalence* is the *fraction* of the population infected. The infection *incidence* is the *rate* at which infections occur.

5.1 The simple SI epidemic

The simplest epidemic model assumes that the population is homogeneous and well-mixed, the disease is contagious (spread by direct contact between a susceptible and an infectious individual), upon infection a susceptible individual immediately becomes infectious and remains so indefinitely (there is no recovery), infection takes place over a short time period compared to the life span of the host population (we ignore births and deaths and assume a constant (and closed) population). Let N be the host population size, $S(T)$ be the number of susceptible individuals at time T, $I(T)$ be the number of infectious (and infected) individuals at time T, $f(S,I)$ be the incidence function, or number of new infected individuals arising per unit time. The *transfer diagram* for the model is

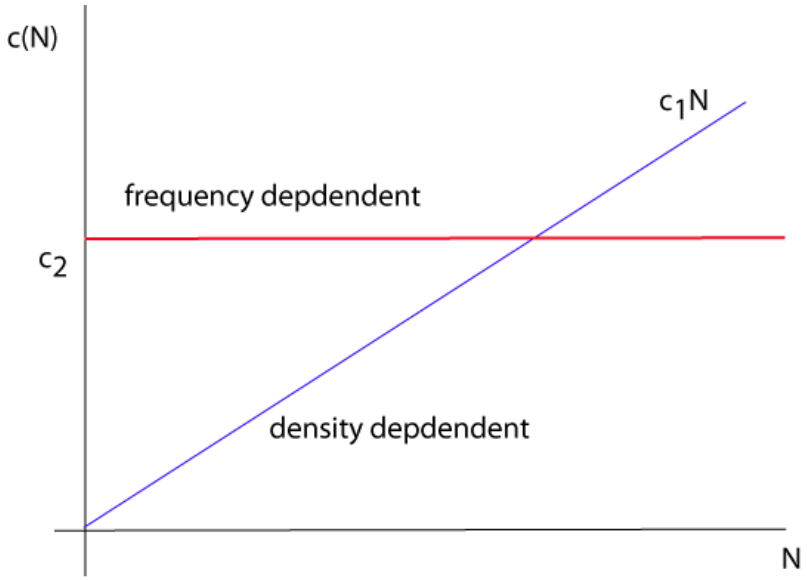
$$S \xrightarrow{f(S,I)} I$$

The corresponding differential equations are

$$\frac{dS}{dT} = -f(S,I), \qquad \frac{dI}{dT} = f(S,I).$$

We can write the incidence function as $f(S,I) = \Lambda(I)S$ where $\Lambda(I)$ is the *force of infection*, which is the product of the contact rate of susceptible individuals $c(N)$, the proportion of these contacts that are infected I/N and the probability p that a contact with an infected individual actually leads to transmission: $\Lambda(I) = pc(N)\frac{I}{N}$.

Density- and frequency-dependent transmission



Density-dependent transmission
Sometimes called *mass-action transmission*. If the contact rate between individuals increases linearly with population size, $c(N) = c_1 N$ the incidence function $f(S, I) = \beta_1 SI$ with $\beta_1 = pc_1$. Density-dependent transmission is considered to be appropriate for airborne diseases.

Frequency-dependent transmission
Sometimes called *proportionate mixing* or *standard incidence*. If the contact rate between individuals is constant $c(N) = c_2$ the incidence function is $f(S, I) = \beta_2 \frac{SI}{N}$ with $\beta_2 = pc_2$. Frequency-dependent transmission is considered to be appropriate for sexually transmitted diseases.

Note that the transmission parameters β_1 and β_2 have different dimensions. β_1 is individuals^{-1} $\times$ time^{-1}, β_2 is time^{-1}, so $c(N)$ is time^{-1} for both. In models with constant population size, the two incidences are equivalent: $\beta_1 = \beta_2/N$ with N being another constant.

SI model with density-dependent transmission

Let us now assume density-dependent incidence βSI. As the total population is constant, we can substitute $S = N - I$ and deal with the single equation

$$\frac{dI}{dT} = \beta(N - I)I = \beta NI\left(1 - \frac{I}{N}\right).$$

This corresponds to the logistic equation with intrinsic growth rate βN and carrying capacity N. (With frequency-dependent incidence we arrive at the logistic equation with carrying capacity N but with intrinsic growth rate β.) Hence, an SI epidemic will always spread and eventually infect all individuals: $I(T) \to N, S(T) \to 0$ as $T \to \infty$. This behaviour is not what is normally observed in data. We have to revise the model.

5.2 The SIS epidemic

We now introduce recovery from the disease and assume that recovered individuals are again susceptible. This is a reasonable approximation for many bacterial diseases (like TB, meningitis and gonorrhoea) whereas viral diseases usually confer at least some immunity to the host after recovery. The transfer diagram is

$$S \xrightarrow{\beta SI} I \xrightarrow{\gamma I} S$$

Here γ is the recovery rate. This assumes that the time spent in the infective class is exponentially distributed with mean $1/\gamma$. More realistic models assume constant infective periods, leading to integro-differential equations.
The equations are given by

$$\frac{dS}{dT} = -\beta SI + \gamma I, \qquad \frac{dI}{dT} = \beta SI - \gamma I.$$

The total size of the population is $S + I = N$, constant. Non-dimensionalising the model by introducing $u = \frac{S}{N}, v = \frac{I}{N}, t = \gamma T$ we obtain

$$\frac{du}{dt} = -(R_0 u - 1)v, \qquad \frac{dv}{dt} = (R_0 u - 1)v$$

where $R_0 = \beta N/\gamma$. By definition $u + v = 1$ and the ODEs need to be solved on the line segment (one-dimensional simplex) $S_1 = \{(u, v) \mid 0 \leq u, v \leq 1, u + v = 1\}$. In fact, we could substitute either u or v to obtain a single ODE.

The basic reproduction number R_0

The basic reproduction number R_0 is the product of (i) the expected number of secondary cases per unit time caused by a single infected individual introduced into a completely susceptible population ($S = N$) and (ii) the expected infectious lifetime. So R_0 is the expected number of secondary cases produced by the primary (index) case.

Epidemic threshold theorem
If $R_0 < 1$, the disease-free equilibrium is locally stable (an epidemic does not occur), but if $R_0 > 1$, the disease-free equilibrium is locally unstable (the disease becomes endemic in the population).

For the SIS model, we have $R_0 = \beta\, N \times \frac{1}{\gamma}$. If $R_0 < 1$, then

$$\frac{dv}{dt} = (R_0 u - 1)v < (R_0 - 1)v$$

with $R_0 - 1 < 0$. So the infected fraction decays exponentially (or faster). If $R_0 > 1$, then

$$\begin{aligned} \frac{dv}{dt} = (R_0 u - 1)v &= (R_0(1 - v) - 1)\, v \\ &= (R_0 - 1)v \left(1 - \frac{R_0 v}{R_0 - 1}\right) \end{aligned}$$

a logistic equation with intrinsic growth rate $R_0 - 1$ and carrying capacity $(R_0 - 1)/R_0 = 1 - 1/R_0$. So $v(t) \to v^* = 1 - \frac{1}{R_0}$ as $t \to \infty$. So that the disease is endemic. Hence, for the SIS model the the disease-free equilibrium is in fact globally stable if and only if $R_0 < 1$.

5.3 SIR Epidemic

We now consider disease models with a *removed* class, in which individuals may be immune, in quarantine, etc. Many childhood diseases (such as measles, chickenpox, and pertussis) have such a structure. We still ignore births and deaths in the host population. The transfer diagram is

$$S \xrightarrow{\beta SI} I \xrightarrow{\gamma I} R$$

This gives the classical model attributed to Kermack and McKendrick (1927):

$$\frac{dS}{dT} = -\beta SI, \quad \frac{dI}{dT} = \beta SI - \gamma I, \quad \frac{dR}{dT} = \gamma I.$$

The total size of the population is $S+I+R = N$, which is constant. Non-dimensionalising the model by introducing $u = \frac{S}{N}, v = \frac{I}{N}, w = \frac{R}{N}, t = \gamma T$ we obtain

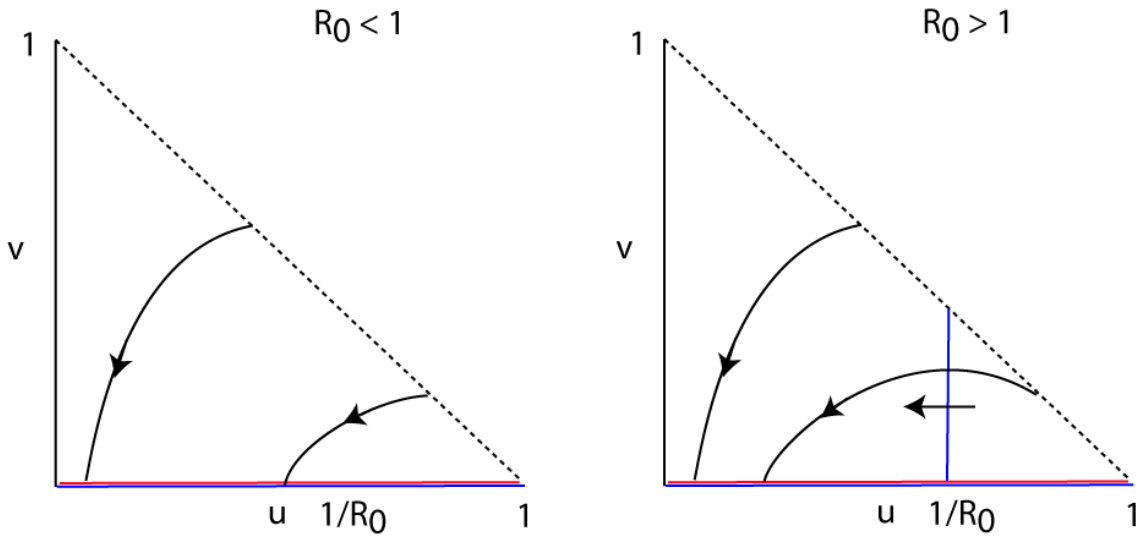
$$\frac{du}{dt} = -R_0 uv, \qquad \frac{dv}{dt} = (R_0 u - 1)v, \qquad \frac{dw}{dt} = v \tag{$\star$}$$

where $R_0 = \beta N/\gamma$.

The system has to be solved on the triangle (two-dimensional simplex) $S_2 = \{(u, v, w) \mid 0 \leq u, v, w \leq 1,\ u + v + w = 1\}$. All points in the simplex with $v = 0$ are disease-free steady states. We shall be particularly interested in the disease-free steady state (DFE) in which all individuals are susceptible, given by $S_0 = (1, 0, 0)$. Note that the first two equations do not contain w, so that we can (in principle) solve them for u and v, and then find w by $w = 1 - u - v$. Hence we look at the problem consisting of the first two equations on the triangle T in the (u, v)-plane given by $T = \{(u, v) \mid u \geq 0,\ v \geq 0,\ u + v \leq 1\}$. The triangle T is positively invariant since $du/dt = 0$ when $u = 0$, $dv/dt = 0$ when $v = 0$, and $d(u + v)/dt = -v \leq 0$ when $u + v = 1$. The two-dimensional system is given by

$$\frac{du}{dt} = -R_0 uv, \quad \frac{dv}{dt} = (R_0 u - 1)v.$$

The u-nullclines are $u = 0$ and $v = 0$. The v-nullclines are $v = 0$ and $u = 1/R_0$. All points on the horizontal line ($v = 0$) are disease-free steady states. In two dimensions the DFE with all individuals susceptible is given by $S_0 = (1, 0)$. The phase planes are as sketched below.



The Jacobian matrix J for the system is

$$J(u,v) = \begin{pmatrix} -R_0 v & -R_0 u \\ R_0 v & R_0 u - 1 \end{pmatrix}.$$

Evaluating this at the steady state $S_0 = (1, 0)$

$$J_0 = J(1,0) = \begin{pmatrix} 0 & -R_0 \\ 0 & R_0 - 1 \end{pmatrix}$$

which has eigenvalues $\lambda_1 = 0$ and $\lambda_2 = R_0 - 1$. The linearised stability analysis therefore does not determine the stability of S_0 if $R_0 < 1$, since it is then on the borderline between stability and instability, but shows that it is unstable if $R_0 > 1$. However, it can be proved that that S_0 is locally stable (but not asymptotically stable) if $R_0 < 1$, and unstable if $R_0 > 1$.

Critical host population threshold

Note that the threshold condition $R_0 = \beta N/\gamma > 1$ is equivalent to $N > \gamma/\beta =: N_\theta$. So the population threshold N_θ must be exceeded for a disease to persist. For example, it has been estimated that measles requires 300,000–500,000 people for endemic circulation.

Final size of the epidemic

For the SIR models without demography, all points on $v = 0$ are disease-free steady states and the epidemic eventually burns out, leaving the population in one of these states. The

final size of an epidemic is the total number of individuals who become infected over the course of the epidemic. If all individuals are initially susceptible, it is given by the number of individuals eventually in the removed class. We can analyse this by dividing the third equation by the first in the original system ($\star$) of three equations to obtain

$$\frac{dw}{du} = -\frac{1}{R_0 u}.$$

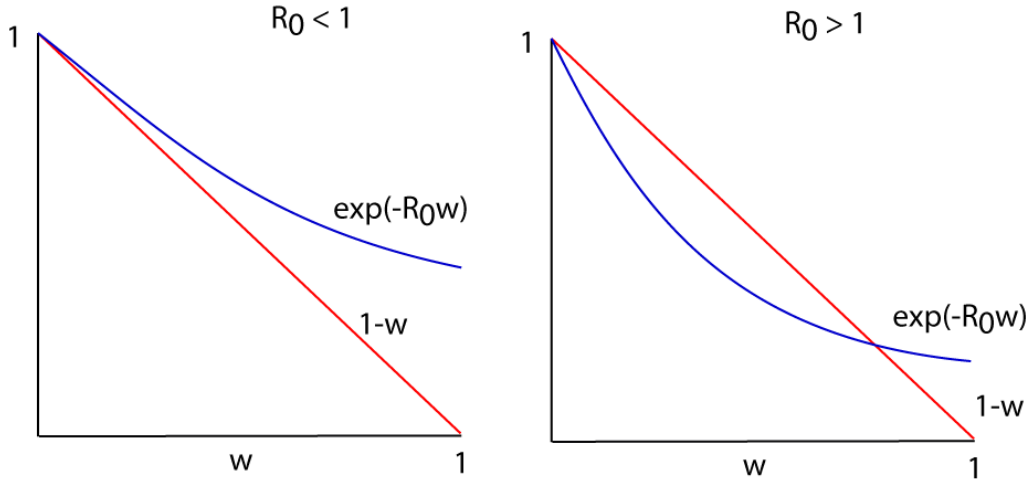
If we solve this equation for w as a function of u with initial condition $w = w_0 = 0$ when $u = u_0 = 1$ then we find a relationship between w and u that is valid everywhere on the solution trajectory that starts arbitrarily close to the DFE $S_0 = (1, 0, 0)$. We could similarly find a relationship between v and u on this trajectory, but this is not necessary. Separating variables $dw = -du/(R_0 u)$, and integrating from $(u_0, w_0) = (1, 0)$ to a general point (u, w) on the trajectory, we obtain

$$w - w_0 = -\frac{1}{R_0}(\ln u - \ln u_0)$$

from which

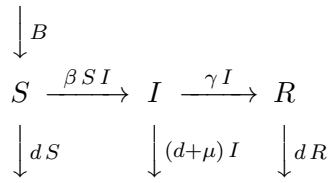
$$w = -\frac{1}{R_0} \ln u \text{ and so } u = \exp(-R_0 w).$$

Note that u and w are monotonically decreasing and increasing respectively and, since they are bounded, $u \to u_1$ and $w \to w_1$ as $t \to \infty$. Moreover, as $w \to w_1$, $\dot{w} \to 0$. But $\dot{w} = v$ so $v \to 0$ as $t \to \infty$. In consequence $(u, v, w) \to (u_1, 0, w_1) = (1 - w_1, 0, w_1)$ as $t \to \infty$. Combining this with $u = \exp(-R_0\, w)$, and taking the limit as $t \to \infty$, gives $1 - w_1 = \mathrm{e}^{-R_0 w_1}$. There is a unique positive root of this equation if $R_0 > 1$. It (implicitly) determines the final size of the epidemic. If $R_0 < 1$ the only root is zero, and the final size of the epidemic is therefore zero, as expected.



5.4 Endemic circulation

Many SIR diseases do not burn out but remain endemic in a population. They can only do so if the pool of susceptible individuals is replenished by births. To model this, we need to incorporate host demographics. In some sense we shall couple epidemic models with demographic models. The transfer diagram is



where B is the birth rate (*not* per capita birth rate) with dimension individuals $\times$ time^{-1}, d is the (natural) per capita death rate d, dimension time^{-1}, and μ is the disease-related per capita death rate, also called the virulence or pathogenicity, with dimension time^{-1}. The model assumes no vertical transmission, so all births enter the S class.

5.4.1 Constant population, no disease-related deaths

Let us assume no disease-related death, $\mu = 0$, constant per capita birth rate, b so $B = bN$, and equal birth and death rate, $b = d$, so that the population size is constant $N = S+I+R$ but we have an inflow of susceptible individuals. Then the equations are

$$\begin{aligned}\frac{dS}{dT} &= bN - \beta\, S\, I - bS\\ \frac{dI}{dT} &= \beta\, S\, I - \gamma I - bI\\ \frac{dR}{dT} &= \gamma I - bR.\end{aligned}$$

Rewriting the model in terms of fractions $u = S/N, \quad v = I/N, \quad w = R/N$ and non-dimensional time $t = (\gamma + b)T$ we obtain

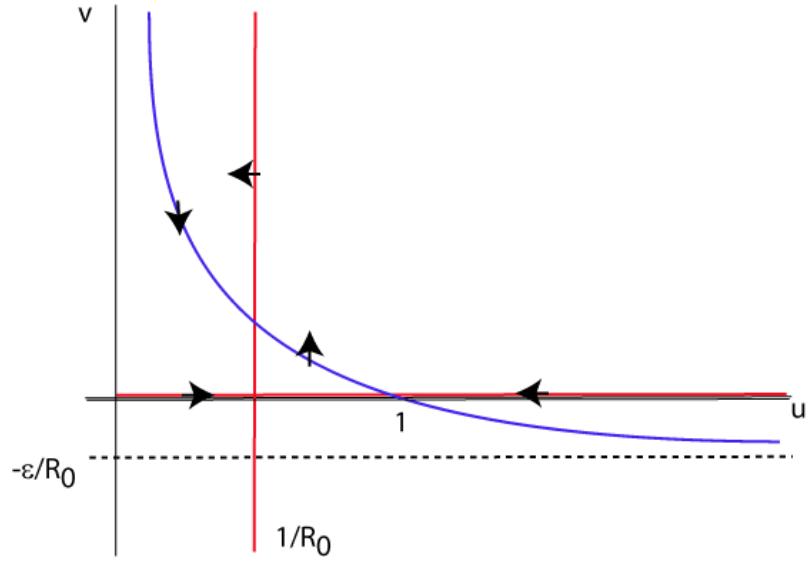
$$\begin{aligned}\frac{du}{dt} &= \epsilon(1-u) - R_0 uv\\ \frac{dv}{dt} &= (R_0 u - 1)v\\ \frac{dw}{dt} &= (1-\epsilon)v - \epsilon w\end{aligned}$$

where $\epsilon = b/(\gamma + b)$ and $R_0 = \beta N/(\gamma + b)$ is the basic reproduction number.

The model reduces to the epidemic case when $b = 0$, or $\epsilon = 0$. Recall that model was structurally unstable due because $v = 0$ was a line of steady states. Usually the life expectancy of the host $1/b$ is much larger than the recovery period $1/\gamma$, at least for acute diseases such as measles. Hence $b \ll \gamma$, and so $\epsilon \ll 1$. The demographic terms reduce the expected infectious lifetime from $1/\gamma$ to $1/(\gamma + b)$ because an infective may die of natural causes while infected. The basic reproduction number R_0 therefore decreases from $\beta N/\gamma$ to $\beta N/(\gamma + b) = (1-\epsilon)\beta N/\gamma$. Since $\epsilon \ll 1$, its value is almost unchanged (for acute diseases). However, even though the demographic terms are small, they make a crucial difference to the qualitative behaviour of the system.

The equations for u and v do not depend on w. We can therefore consider the model on the triangle T in the (u, v)-plane given by
$T = \{(u, v) \mid u \geq 0,\ v \geq 0,\ u + v \leq 1\}$. The steady states are $S_0 = (1, 0)$, the disease-free steady state (DFE) with all individuals susceptible, and $S_1 = (u^*, v^*)$, the endemic steady state (EE), given by $u^* = \frac{1}{R_0}, v^* = \epsilon\left(1 - \frac{1}{R_0}\right)$ which is biologically realistic as long as

$R_0 > 1$. The u-nullcline is $v = \epsilon(1-u)/(R_0 u)$ and the v-nullclines are $v = 0$ and $u = 1/R_0$. There is no longer a line $v = 0$ of disease-free steady states. The nullclines and steady states for $R_0 > 1$ are sketched below.



We assesses the stability of the steady states by considering the Jacobian matrix J, given at a general point (u, v) by

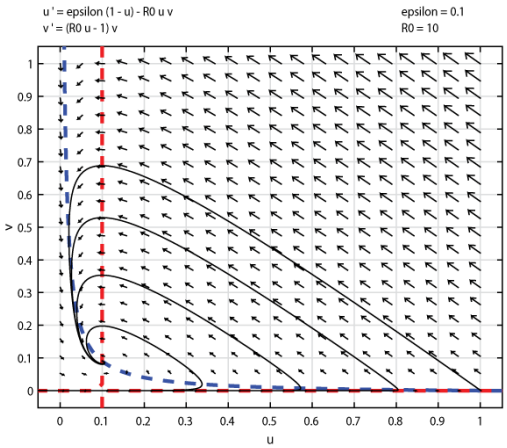
$$J(u,v) = \begin{pmatrix} -\epsilon - R_0 v & -R_0 u \\ R_0 v & R_0 u - 1 \end{pmatrix}.$$

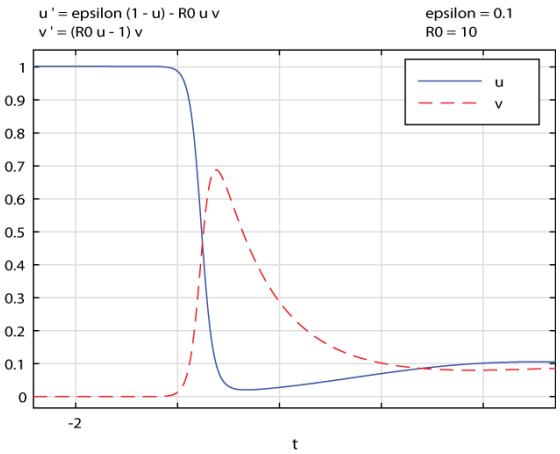
At the disease-free equilibrium $S_0 = (1, 0)$, the eigenvalues are $\lambda_1 = -\epsilon < 0, \quad \lambda_2 = R_0 - 1$. Hence S_0 is a stable node if $R_0 < 1$ and a saddle point otherwise.

The Jacobian J_1 evaluated at the endemic equilibrium $S_1 = (u^*, v^*)$ is given by

$$J_1 = J(u^*, v^*) \begin{pmatrix} -\epsilon R_0 & -1 \\ \epsilon(R_0 - 1) & 0 \end{pmatrix},$$

so that $\text{tr}\, J_1 = -\epsilon R_0 < 0, \det J_1 = \epsilon(R_0 - 1)$ and by the Routh–Hurwitz criteria S_1 is locally asymptotically stable if and only if it is biologically realistic ($R_0 > 1$). Since $\text{disc}\, J_1 = (\text{tr}\, J_1)^2 - 4 \det J_1 \approx -4\epsilon(R_0 - 1)$ then we expect S_1 to be a stable focus, so that the EE is approached via damped oscillations.





5.4.2 Constant birth rate, disease-related deaths

We now assume that the birth rate B (not the per capita birth rate b) is constant, and that infection causes additional mortality at rate $\mu > 0$. The population size may vary, but approaches a constant size in the absence of disease. The equations are

$$\begin{aligned} \frac{dS}{dT} &= B - \beta\, S\, I - dS, \\ \frac{dI}{dT} &= \beta SI - \gamma I - dI - \mu I, \\ \frac{dR}{dT} &= \gamma I - dR. \end{aligned}$$

Adding all three ODEs gives

$$\frac{dN}{dT} = B - dN - \mu I.$$

In the absence of disease, the population will grow to B/d, and there is a disease-free steady state at $(S, I, R) = (B/d, 0, 0)$. For an index case introduced into this disease-free steady state, the basic reproduction number is

$$R_0 = \beta \frac{B}{d} \times \frac{1}{\gamma + \mu + d}.$$

The mean infectious lifetime is determined by recovery, natural mortality and disease-induced mortality.

We may study any three of the four ODEs, and choose the (N, S, I) equations. There are two steady states $S_0 = (B/d, B/d, 0)$ is the disease-free equilibrium DFE, and $S_1 = (N^*, S^*, I^*)$ is the endemic equilibrium EE with $S^* = (\gamma + \mu + d)/\beta = B/(dR_0)$, $I^* = (B - dS^*)/(\beta S^*) = d(R_0 - 1)/\beta$, $N^* = (B - \mu I^*)/d$ which is realistic as long as $R_0 > 1$. One can show the usual threshold behaviour. If $R_0 < 1$ the DFE is stable. At $R_0 = 1$ the DFE exchanges stability with the emerging EE. If $R_0 > 1$, the DFE is unstable, the EE is stable. The EE is again a stable focus for acute diseases, and is approached through damped oscillations.

5.5 Eradication and control

In the disease models considered so far, the basic reproduction number was a crucial quantity determining whether a disease will die out or spread. In the simplest models, the basic reproduction number is of the form $R_0 = \frac{\beta N}{\gamma}$. Control measures usually aim at reducing R_0 below one. Most generally, this could be achieved by decreasing β, the transmission rate, increasing γ, the 'removal' rate of infective individuals, or decreasing N, which can be interpreted as the initial susceptible population size.

5.5.1 Example: SIR, no demography, one-off vaccination

The equations are

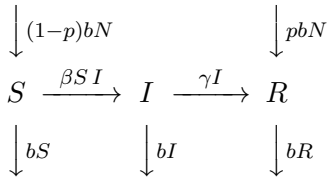
$$\frac{dS}{dT} = -\beta SI, \qquad \frac{dI}{dT} = \beta SI - \gamma I, \qquad \frac{dR}{dT} = \gamma I.$$

Suppose that (i) there is a perfect vaccine against the disease, and (ii) we can immediately vaccinate a proportion p of the population. So the initial susceptible population is reduced

from N to $(1-p)N$ or, as a fraction of the total population, from 1 to $1-p$. The model is exactly the one we have considered before, but with N replaced by $(1-p)N$. The analysis is therefore exactly as before but with $R_0 = \beta N/\gamma$ replaced by a new quantity, the *effective reproduction number* R_e given by $R_e = \beta(1-p)N/\gamma = (1-p)R_0$. In other words, the disease-free steady state (DFE) S_0, now given by $S_0 = (1-p, 0)$ is stable (but not asymptotically stable) if $R_e < 1$, which is equivalent to $(1-p)R_0 < 1$ or $p > \hat{p} = 1 - \frac{1}{R_0}$. This is the fraction of the population that needs to be vaccinated to eliminate the threat of an epidemic.

5.5.2 Example: SIR model with demography and paediatric vaccination

If there is continuous replenishment of the susceptible population due to births the previous one–off vaccination programme does not seem suitable. For many potentially dangerous human infections (such as measles, mumps, rubella, pertussis, polio), there has been a focus on vaccinating newborns or very young infants. We model this by vaccinating a fraction p of the individuals entering the population, as in the following transfer diagram.



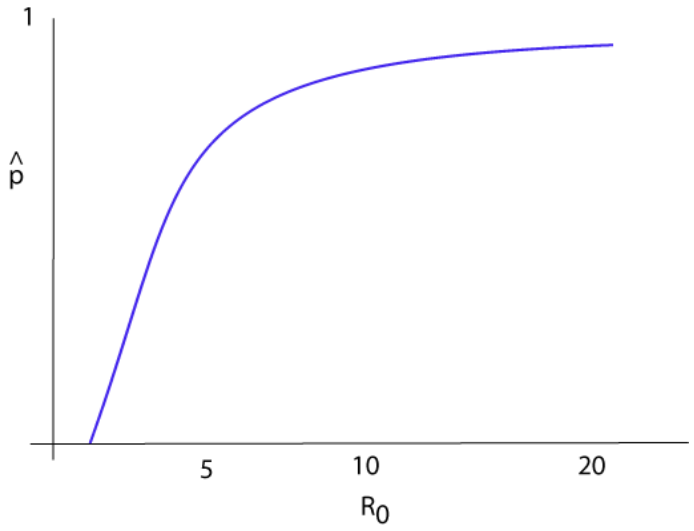
The equations are

$$\begin{aligned}
\frac{dS}{dT} &= (1-p)bN - \beta SI - bS, \\
\frac{dI}{dT} &= (\beta S - \gamma - b)I, \\
\frac{dR}{dT} &= pbN + \gamma I - bR.
\end{aligned}$$

Adding the equations, $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$, so the population size is constant. The R equation is decoupled from the system. The problem for the S and I equations is as before with N replaced by $(1-p)N$, and hence R_0 replaced by $R_e = (1-p)R_0$. We can immediately conclude that the threat of an epidemic can be eliminated if $R_e < 1$ which is equivalent to $p > \hat{p} = 1 - 1/R_0$. Otherwise, the disease will become endemic in the population.

5.5.3 Herd immunity

Not all individuals need to be vaccinated in order to eradicate an infection. Only a critical proportion $\hat{p} = \frac{R_0 - 1}{R_0}$ (solely determined by the reproduction number of the infection) must be protected. This phenomenon is referred to as herd immunity. The relationship between R_0 and $\hat{p}$ is shown below. Diseases with high R_0 (such as measles and pertussis) need a vaccination coverage of newborns (or young infants) well above 90 %. To date, smallpox is the only high-profile example of global eradication of a disease which has been attributed to systematic vaccination, but polio is coming close.



5.6 Malaria and other vector-borne diseases

There are many infectious agents that are transmitted by blood-sucking arthropods known as vectors but, in general, cannot be passed directly between primary hosts (i.e. human to human or animal to animal). Malaria is undoubtedly the most significant vector-borne disease in the world. Others include dengue, yellow fever, Chagas, leishmaniasis and schistosomiasis. Here we will study mathematical models for malaria epidemiology, but most of the theory applies equally well to other diseases.

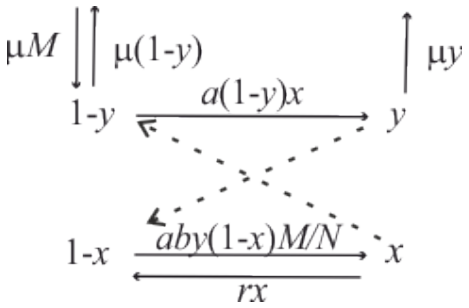
The first mathematical models for the transmission of malaria were written down by Ross in the early 20th century. These models have been adapted, adjusted and re-evaluated many times over the last hundred years, but remain at the core of much of the current work on the theory of malaria epidemiology. Ross initially produced a discrete-time model. However, it is the continuous-time formulation, improved by MacDonald in the mid 20th century and now known as the Ross-MacDonald model, that has endured. The model considers the coupled infection dynamics in populations of humans and mosquitoes (the vectors). Transmission is criss-cross - a mosquito infects a human and this human infects another mosquito. The human population size is constant N, there are no births and deaths, infected humans recover at rate r but do not acquire immunity. The mosquito population size is constant M, the per capita birth and death rates are both μ, and there is no recovery. Mosquitoes bite at frequency dependent rate a. The probability that a bite by an infected mosquito on an uninfected human leads to the human becoming infected is b. The probability that a bite by a susceptible mosquito on an infected human leads to the mosquito becoming infected is 1. Let x be the proportion of the human population that is infected, and y be the infected proportion of the mosquito population. Constant populations sizes mean the corresponding susceptible proportions are $1-x$ and $1-y$ respectively. Then the incidence function in humans is

$$f_H(x,y) = ab\frac{M}{N}y(1-x)$$

The incidence function in mosquitoes is

$$f_M(x,y) = a(1-y)x$$

The transfer diagram is



The complete model is

$$\begin{aligned} \frac{dx}{dt} &= \tfrac{abM}{N} y(1-x) - rx \\ \frac{dy}{dt} &= ax(1-y) - \mu y \end{aligned}$$

Basic reproduction number
If we define R_0^2 as the number of secondary cases *in humans* produced by a single primary human case introduced into an entirely susceptible population (of both humans and mosquitoes)

$$R_0^2 = \underbrace{\frac{abM}{N}\frac{1}{r}}_{\#\text{ cases } H\to M} \times \underbrace{a\frac{1}{\mu}}_{\#\text{ cases } M\to H} = \frac{a^2 bM}{r\mu N}.$$

Steady states
There are two steady states $S_0 = (0,0)$ is the disease-free equilibrium (DFE) and $S^* = (x^*, y^*)$ is the endemic equilibrium (EE) with $x^* = \frac{R_0^2-1}{R_0^2+\frac{a}{\mu}}$ and $y^* = \frac{R_0^2-1}{R_0^2}\left(\frac{\frac{a}{\mu}}{1+\frac{a}{\mu}}\right)$, which is biologically meaningful if $R_0 > 1$. The term a/μ is the expected number of bites a mosquito makes in its lifetime. It is sometimes called MacDonald's stability index, although this does not refer to stability in the mathematical sense. In the 1950s MacDonald observed that in geographical regions where this index was high, malaria was often endemic. He called this stable malaria. Meanwhile, in regions where this index was lower, there might be malaria epidemics but it was not endemic. He called this unstable malaria. Inspection of S^* suggests that the infected mosquito proportion y^* is less sensitive to changes in a/μ when a/μ is large than when it is small. The phase-plane analysis below offers some further insight.

Linear stability
The Jacobian matrix J at a general point (x, y) is

$$J(x,y) = \begin{pmatrix} -\frac{abM}{N}y - r & \frac{abM}{N}(1-x) \\ a(1-y) & -ax - \mu \end{pmatrix}$$

The Jacobian J_0 at the DFE S_0 is

$$J_0 = \begin{pmatrix} -r & \frac{abM}{N} \\ a & -\mu \end{pmatrix},$$

and tr $J_0 = -r - \mu < 0$, det $J_0 = r\mu - \frac{a^2bM}{N} = r\mu(1 - R_0^2)$. So S_0 is asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.

The Jacobian matrix J^* at the EE S^* is

$$J^* = \begin{pmatrix} -\frac{abM}{N}y^* - r & \frac{abM}{N}(1 - x^*) \\ a(1 - y^*) & -(ax^* + \mu) \end{pmatrix} = \begin{pmatrix} -\frac{abM}{N}\frac{y^*}{x^*} & r\frac{x^*}{y^*} \\ \mu\frac{y^*}{x^*} & -a\frac{x^*}{y^*} \end{pmatrix}.$$

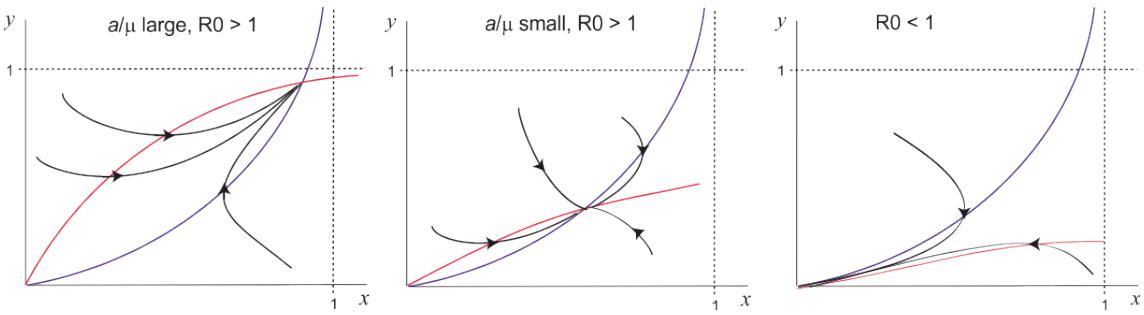
It is easy to see that tr $J^* < 0$ (if S^* is realistic) while det $J^* = \frac{a^2bM}{N} - r\mu = r\mu(R_0^2 - 1)$. Hence S^* is asymptotically stable if $R_0 > 1$, and not biologically meaningful if $R_0 < 1$.

Eradication and control

A population will be protected against epidemics if $R_0 < 1$. $R_0^2 = \frac{a^2bM}{r\mu N}$ shows that this threshold may be achieved by reducing M, a or b, or by increasing N, r or μ. Increasing the human population size N works because, when N is large compared to M, the chance of an infected human being bitten a second time (to transmit the infection) is low - the mosquito bites are 'wasted' in a *dilution effect*, and the disease dies out. This has inspired various control attempts. Of the remaining parameters, the quadratic dependence of R_0^2 on a suggests it might be most effective to focus on decreasing the biting rate.

Phase-plane

The x-nullcline is $y = \frac{rN}{abM}\left(\frac{x}{1-x}\right)$, which has initial gradient $\frac{rN}{abM}$. The y-nullcline is $y = \frac{ax}{ax+\mu}$, which has initial gradient $\frac{a}{\mu}$. Phase plane portraits for $R_0 > 1$ with $\frac{a}{\mu}$ large, $R_0 > 1$ with $\frac{a}{\mu}$ small and $R_0 < 1$ are shown below. In the case of $R_0 > 1$ and $\frac{a}{\mu}$ large, the endemic equilibrium is a strong attractor and, due to the configuration of the nullclines, insensitive to parameter variation. In the case of $R_0 > 1$ and $\frac{a}{\mu}$ large, the endemic equilibrium is a weak attractor and its location and existence is sensitive to parameter variation.



Incubation in the mosquito

The version of the Ross-MacDonald model we have been looking at assumes that an infected human or mosquito is immediately infectious. However, the reality is that there can be an incubation period of several days between infection and infectiousness. MacDonald modified the model to consider the implications of this incubation period in the mosquito. One approach is to introduce an incubating state for the mosquito z, assume that the incubation period lasts exactly time τ, and write the model in terms of delay differential

equations with delayed state variables $\hat{x} = x(t-\tau), \hat{y} = y(t-\tau), \hat{z} = z(t-\tau)$. Then

$$\begin{aligned}
\frac{dx}{dt} &= \frac{abM}{N} y(1-x) - rx \\
\frac{dy}{dt} &= a\hat{x}(1-\hat{y}-\hat{z})e^{-\mu\tau} - \mu y \\
\frac{dz}{dt} &= ax(1-y-z) - a\hat{x}(1-\hat{y}-\hat{z})e^{-\mu\tau} - \mu z
\end{aligned}$$

The basic reproduction number can be constructed by considering a single infected human in an otherwise uninfected population. The transmission pathway is the same as in the Ross-MacDonald model, except that the probability an infected mosquito survives the incubation period, and so becomes infectious, is $e^{-\mu\tau}$. Hence the basic reproduction number is

$$R_0^2 = \frac{a^2 bM}{r\mu N} e^{-\mu\tau}.$$

At an equilibrium $x^*(t) = x^*(t-\tau)$, and similarly for $y(t), z(t)$. So, after some algebra,

$$y^* = \frac{R_0^2 - 1}{R_0^2} \left(\frac{\frac{a}{\mu}}{1 + \frac{a}{\mu}} \right) e^{-\mu\tau}.$$

It may be that the expected mosquito incubation period τ is similar to, or even greater than, the expected lifespan ($\tau > 1/\mu$). In this case, $\tau\mu$ is around 1 and the proportion of the mosquito population that is infectious y^* may be small, even when the infection proportion of the human population x^* is high, and indicators of epidemic risk and endemic stability, R_0 and a/μ, are large.

This model also shows that R_0 decreases linearly with the size of the adult mosquito population M but exponentially with adult mosquito mortality rate μ. Hence control strategies that reduce adult lifespan are likely to be more effective than those that reduce recruitment into the adult population e.g. by targeting larval stages. This observation has informed practical malaria control strategies.