INTRODUCTION TO POPULATION MODELS

J. Banasiak

Principles of mathematical modelling

By a **mathematical model** we understand an equation, or a set of equations, that describe some phenomenon that we observe in science, engineering, economics, or some other area, that provides a quantitative explanation and, ideally, prediction of observations.

Mathematical modelling we mean the process by which we formulate and analyze model equations and compare observations to the predictions that the model makes.

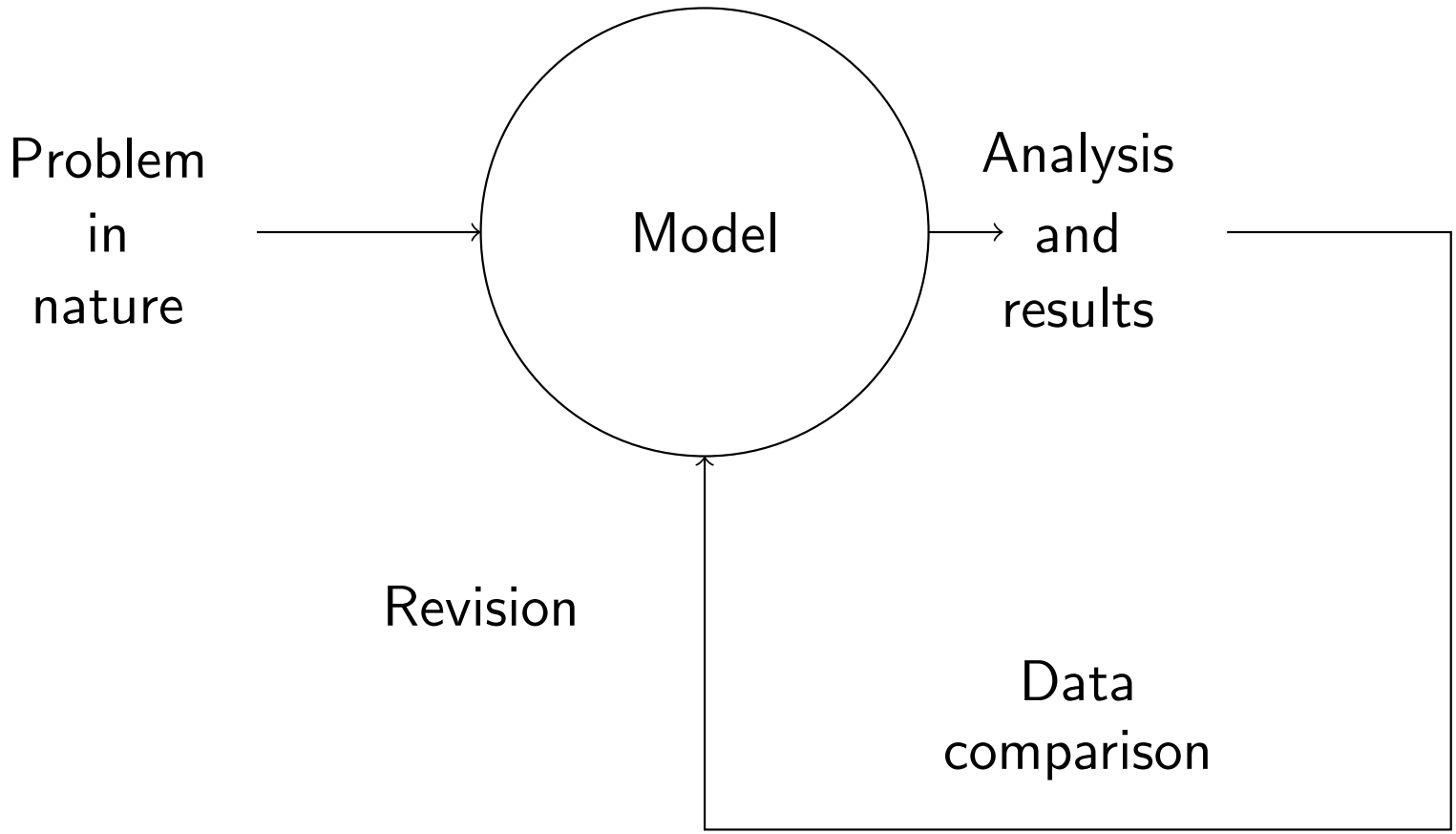


Figure: The process of mathematical modelling.

Note:

- Modelling is not mathematics – it is impossible to prove that a model is correct;
- One counterexample disproves the model.

A good model:

- has predictive powers – a model based on available observations gives correct answers in other cases:
  – General Theory of Relativity – light deflection, perihelion precession of Mercury, gravitational waves,
  – Dirac equations – existence of positrons;
- contains earlier working models as subcases:
  – Newton's mechanics is contained in Special/General Theory of Relativity for small velocities and away from large masses,
  – Quantum mechanics yields the same results as Newton's mechanics for large distances, large energies.

Descriptive versus explanatory models.

Abundance of data often leads to statistical fitting the data with formulae. One can get a variety of statistical information such as expectations, medians, variance, correlations...

Remember: do not mistake correlations for causation

Example: it has been observed that since the 1950s, both the atmospheric CO_2 levels and obesity levels in the US have increased sharply. Hence, obesity is caused by high levels of CO_2.

We shall focus on models which try to understand the underlying reasons for the phenomena we observe. Nevertheless, statistical analysis of the data is important as it separates their significant part from the noise.

Statistical (descriptive) models must not be mixed up with **stochastic models**. Stochastic modelling aims to explain the underlying mechanisms of the observed phenomena taking into account inherent(?) randomness of nature. Such models give probabilities of certain events and are indispensable in modeling small populations.

We shall focus, however, on **deterministic models** that sometimes can be thought as stochastic models averaged over many individual trajectories (Law of Large Numbers) and giving answers in terms of the evolution of the densities of the populations. Nevertheless, stochastic models are often used explicitly to derive a deterministic model.

Conservation principles and constitutive relations

Conservation principles: Mathematical biology and epidemiology must obey laws of physics; in particular the balance law. Let Q be a quantity of interest (the number of animals, mass of a pollutant, amount of heat energy, number of infected individuals) in a fixed domain Ω. Over any fixed time interval in Ω we have

$$\begin{aligned} \text{The change of } Q \quad = \quad & \text{Inflow of Q} - \text{Outflow of Q} \\ & + \text{Creation of Q} - \text{Destruction of Q}. \end{aligned} \tag{1}$$

In probabilistic approach this is the same as saying that the probability that one of all possible events occurs equals one.

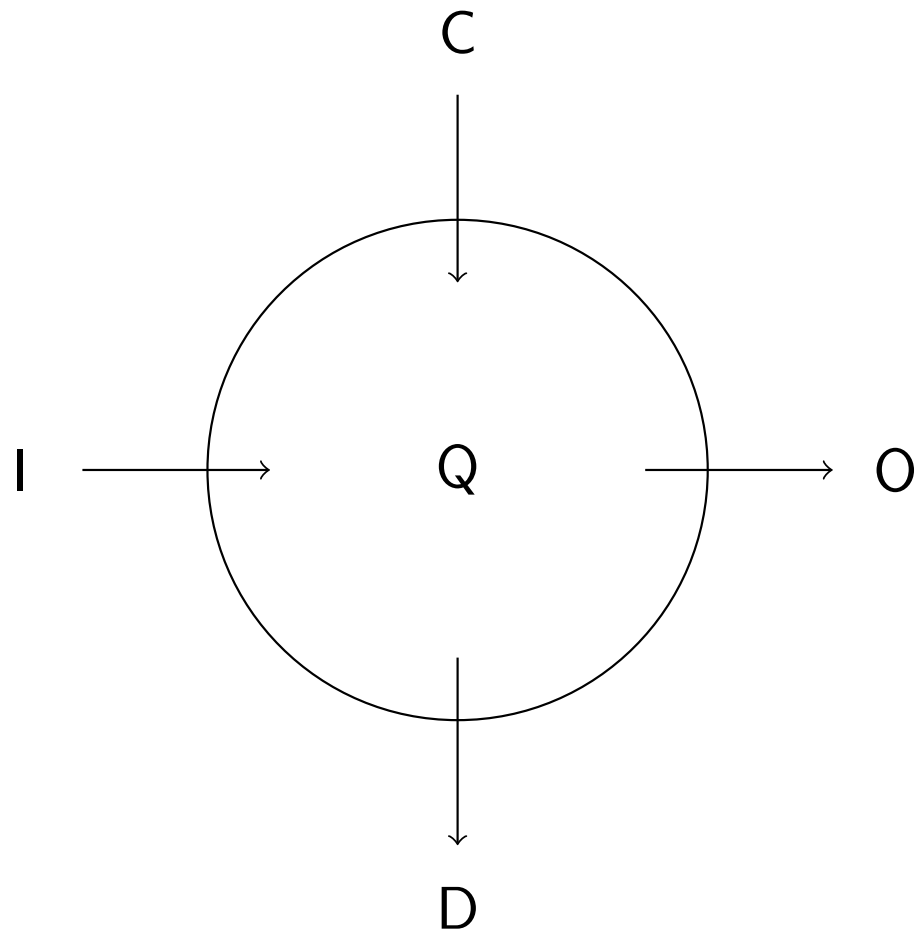


Figure: Conservation law for the substance Q.

Real modelling consists in determining the form of Q, I, O, C and D and the relations between them – these are known as **constitutive relations**.

However, before we proceed, we must decide whether we model with **continuous time**, or **discrete time**.

We use **discrete time models** if we believe that significant changes in the system only occur during evenly spaced short time intervals, or we only can observe the system at evenly spaced time instances and have a reason to believe that essential parameters of the system remain unchanged between successive observations.

Then we use the time between the events/observations as the time unit and count time using the number of elapsed events/observations and (1) can be written as

$$Q(k+1) - Q(k) = I(k) - O(k) + C(k) - D(k). \tag{2}$$

Quantities $I(k), O(k), C(k), D(k)$ are the amounts of Q, respectively, that inflows, outflows, is created and destroyed in the time interval $[k, k+1]$.

Examples. Many plants and animals breed only during a short, well-defined, breeding season. Also, often the adult population dies soon after breeding. Such populations are ideal for modelling using discrete time modelling. Let us consider a few typical examples.

(i) Monocarpic plants flower once and then die. Such plants may be annual but, for instance, bamboos grow vegetatively for 20 years and then flower and die.

(ii) Animals with such a life cycle are called semelparous.

a) Insects typically die after lying eggs but their life-cycle may range from several days (e.g. house flies) to 13–17 years (cicads).

b) Similar life cycle is observed in some species of fish, such as the Pacific salmon or European eel. The latter lives 10-15 years in freshwater lakes, migrates to the Sargasso Sea, spawns and dies.

c) Some marsupials (antechinus) ovulate once per year and produce a single litter. There occurs abrupt and total mortality of males after mating. The births are synchronized to within a day or two with a predictable 'bloom' of insects.

(iii) A species is called iteroparous if it is characterized by multiple reproductive cycles over the course of its lifetime. Such populations can be modelled by difference equations if the breeding only occurs during short, regularly spaced breeding periods. It is typical for birds. For instance, females of the Greater Snow Geese lay eggs between 8th–20th of June (peak occurs at 12th–17th of June) and practically all eggs hatch between 8th and 13th of July.

If the assumptions allowing us to use discrete time modelling are not satisfied, we use **continuous time**. This, however requires some preparation, as all quantities may change at any instance of time. Thus, I, O, D, C should be considered as the **rates** of inflow, outflow, destruction or creation, respectively; in other words, the amount of Q at a given time t will be given by

$$Q(t) = Q(t_0) + \int_{t_0}^{t} I(s)ds - \int_{t_0}^{t} O(s)ds + \int_{t_0}^{t} C(s)ds - \int_{t_0}^{t} D(s)ds,$$

where $Q(t_0)$ is the initial amount of Q.

Hence, assuming that I, O, D, C are continuous functions, so that Q is differentiable, we obtain the conservation law in differential form,

$$\frac{dQ}{dt}(t) = I(t) - O(t) + C(t) - D(t). \tag{3}$$

Note 1. The meaning of I, O, C and D (and the dimension) in (3) is different than in (2).

Note 2. If we consider populations, then the value of Q always is a nonnegative integer. Such a function can never be continuous. Thus already (3) is an approximation the validity of which requires that Q be so large that it can be considered a continuum.

Constitutive relations.

We try to build the functions I, O, D, C to encompass all we know about the process. However, this is usually impossible.

Constitutive relations.

We try to build the functions I, O, D, C to encompass all we know about the process. However, this is usually impossible.

> *There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.*
>
> *Donald Rumsfeld*

Nevertheless, let us try. The functions I, O, D, C may depend on

- other unknown quantities – this leads to systems of equations that will be discussed in the second lecture;
- space or other independent quantities – this leads to partial differential equations that will be discussed in the third lecture;
- explicitly on time – this results in non-autonomous equations which will be discussed later in this lecture;

- the unknown Q in

  a) a nonlinear way, such as $I(t) = I(Q(t)) = Q^2(t)$, or

  b) a linear way, such as $I(t) = I(Q(t)) = 2Q(t)$,

  in which case we talk, respectively, about autonomous nonlinear or linear equations.

It is important to realize that a non-autonomous equations often is derived from a larger systems of autonomous nonlinear equations in which the coefficients depend on partial solutions of this system which can be determined explicitly.

Unstructured population models

Models given by a single (scalar) autonomous equations – discrete time.

To fix attention we shall focus on Q being the number P of individuals occupying certain domain, or the population density. What assumptions do we need to be able to describe this population by a single equation

$$P(k+1) = F(P(k)), \tag{4}$$

with $F(P) = P + I(P) - O(P) + C(P) - D(P)$?

In general, a single equation means that the population is influenced by (but does not interact with) an unchanging environment (possibly including other populations). Further, any variations among individuals can be disregarded. In particular,

a) the ratio of females to males remains constant;

b) each member of the population produces on average the same number of offspring;

c) each member has an equal chance of dying;

d) there are no age differences in the population;

e) the population is spatially homogenous.

We consider a population with no migration ($I = O = 0$) and, using a), we only track females.

In the simplest case of constant birth and death rates usually we see the equation

$$P(k+1) = P(k) + \beta P(k) - \mu P(k) \tag{5}$$

where β is the average number of offspring per female in a single breeding season and μ is the probability of dying in the unit time interval. In general, this is a wrong interpretation.

Let us assume that the census is taken just before each breeding season and, to incorporate monocarpic/semelparous individuals we introduce μ_0 as the probability that an individual dies from natural causes between the breeding seasons and μ_1 as the probability of death due to giving birth. Then, the population $P(k+1)$ just before the $k+1$st breeding season consists of

- individuals present before the kth breeding season who survived giving birth at k and then survived till $k+1$, $(1-\mu_0)(1-\mu_1)P(k)$, and
- surviving offspring of the population $P(k)$, $(1-\mu_0)\beta P(k)$.

In mathematical terms

$$P(k+1) = \beta(1-\mu_0)P(k) - (1-\mu_1)(1-\mu_0)P(k). \qquad (6)$$

The coefficient $(1-\mu_0)\beta$ is the (effective) per capita birth rate It is a very important parameter from the population point of view since for the survival of the population it is not only important what is the natural fertility of a female; that is, how many offspring she can produce each season, but also how many of them survive on average till the next breeding season.

The standard interpretation only would apply if there is no mortality between breeding seasons ($\mu_0 = 0$) and the death only can occur immediately after giving birth. The case of non-overlapping generations requires $\mu_1 = 1$.

Mathematically, in any case we have the so-called **Malthus model**

$$P(k+1) = rP(k), \tag{7}$$

where r is called the net growth rate. Thus

$$P(k) = r^k P(0)$$

where $P(0)$ is the initial population, so the population either decays to 0 (for $r < 1$), or stays constant (if $r = 1$), or quickly increases to infinity (if $r > 1$).

Basic nonlinear models

In real populations, the parameters depend on the population density P. This is often the case for the survival probability, $S(P) = 1 - \mu_0(P)$.

First we consider populations with no overlapping generations. Then

$$P(k+1) = F(P(k)) = R(P(k))P(k) = \beta S(P(k))P(k), \qquad (8)$$

k=0,1,....

The survival rate S in (8) reflects the intraspecific (within-species) competition for some resource (typically, food or space). The three main (idealized) forms of intraspecific competition are

(i) *No competition*; then $S(P) = 1$ for all P.

(ii) *Contest competition*: there is a finite number of units of resource. An individual who obtains one of these units survives to breed, and produces β offspring; all others die without producing offspring. Thus $S(P) = 1$ for $P \leq P_c$ and $S(P) = P_c/P$ for $P > P_c$ for some critical value P_c.

(iii) *Scramble competition*: each individual gets an equal share of a limited resource. If this amount is sufficient for survival, then all individuals survive and produce R_0 offspring each; if its is not sufficient, then all of them die. Thus, $S(P) = 1$ for $P \leq P_C$ and $S(P) = 0$ if $P > P_C$ for a critical value P_C (P_C is different from P_c).

These ideal situations do not occur in real populations: real data are not easily classified in terms of the contest or scramble competition. Threshold density is not usually seen, zero survival is unrealistic, at least for large populations.

Thus, a similar classification is done on the basis of asymptotic behaviour of $S(P)$ (or $f(P)$) as $P \to \infty$. 1. Contest competition corresponds tothe so-called *exact compensation*:

$$S(P) \sim cP^{-1}, \quad P \to \infty, \tag{9}$$

for some constant c. This describes the situation if the increased mortality compensates exactly any increase in numbers and thus only the predetermined number of individuals in the population can survive.

2. The other case is when

$$S(P) \sim c/P^b, \quad P \to \infty. \tag{10}$$

Here we have

2a. *Under-compensation* if $0 < b < 1$; that is, when the increased mortality less than compensates for the increase in numbers;

2b. *Over-compensation* if $b > 1$.

In general, if $b \approx 1$, then we say that there is a contest, and a scramble if b is large. Indeed, in the first case, $F(P)$ eventually levels-out at a nonzero level for large populations which indicates that the population stabilizes by rejecting too many newborns. On the other hand, for $b > 1$, $F(P)$ tends to zero for large populations which indicates that the resources are over-utilized leading to eventual extinction.

Beverton-Holt type models. To exhibit compensatory behaviour, we should have $PS(P) \approx const$ for large P. At the same time, or small P, $S(P)$ should be approximately 1 as we expect very small intra-species competition so that the growth should be exponential with the effective birth rate given by fertility β. A simple model of this type is given

$$P(k+1) = \frac{\beta P(k)}{1 + aP(k)}. \tag{11}$$

A generalization of this model is called the *Hassell* or again *Beverton-Holt* model, and reads

$$P(k+1) = \frac{\beta P(k)}{(1+aP(k))^b}. \tag{12}$$

It exhibits all types of compensatory behaviour, depending on b.

For $b > 1$ the models describes *scramble* competition, while for $b = 1$ we have contest.

It is useful to introduce the per capita growth rate

$$\frac{\Delta P}{P} := \frac{P(k+1)-P(k)}{P(k)} \tag{13}$$

For the Beverton-Holt model we have

$$\frac{\Delta P}{P} = \frac{\beta - 1 + aP}{1+aP}$$

so that, provided $\beta > 1$, there is a unique value of P, denoted K,

$$K = \frac{\beta - 1}{a}$$

such that if $P(k) < K$, then $P(k+1) > P(k)$, if $P(k) > K$, then $P(k+1) < P(k)$ and if $P(k) = K$, then $P(k+1) = P(k)$.

We call K, at this moment wishfully, the carrying capacity of the environment and, since it has a biological meaning, we re-write the Beverton-Holt model as

$$P(k+1) = \frac{\beta P(k)}{1 + \frac{\beta - 1}{K} P(k)}. \tag{14}$$

This concept has a wider meaning to which we return later. Here we fix some terminology. For a difference equation

$$P(k+1) = F(P(K))$$

the carrying capacity is among solutions of the fixed point equation

$$P = F(P)$$

and can graphically be found by plotting together the lines $y = x$ and $y = F(x)$:

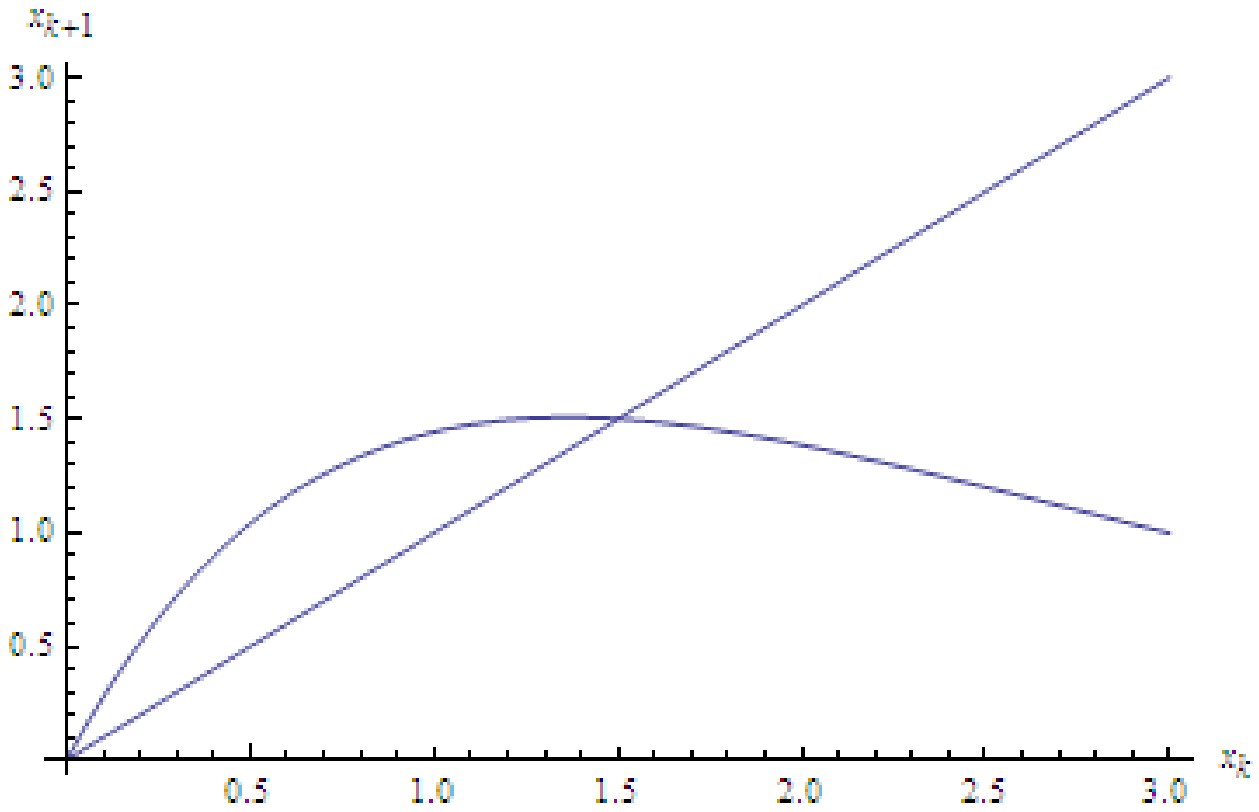


Figure: Finding the carrying capacity K.

Other nonlinear equations.

Consider a general per capita growth relation

$$\frac{\Delta P}{P} = f(P), \tag{15}$$

where f is some function. For the Malthus model

$$\frac{\Delta P}{P} = r,$$

so the graph of the per capita growth rate is a horizontal line.

If we want our model to exhibit the behaviour similar to the Beverton-Holt model; that is, to have a single positive carrying capacity, we have to replace r by a function $f(P)$ that cuts the vertical axis at a positive value, say r and the horizontal axis at $P = K$.

Discrete logistic equation.

The simplest function satisfying these requirements is a linear function

$$f(P) = r\left(1 - \frac{P}{K}\right)$$

that gives

$$P(k+1) = P(k)\left(1 + r\left(1 - \frac{P(k)}{K}\right)\right), \tag{16}$$

which is still one of the most often used discrete equations of population dynamics.

The substitution

$$x(k) = \frac{rP(k)}{(1+r)K}, \quad \beta = 1+r$$

reduces (16) to a simpler form

$$x(k+1) = \beta x(k)(1-x(k)) \tag{17}$$

which is most often used in analysis.

The problem with the discrete logistic equation is that large (close to K) populations can become negative in the next step. For instance, for

$$P(k+1) = 5P(k)(1-P(k)), \quad P(0) = 0.6,$$

we obtain $P(1) = 1.2$ and $P(2) = -1.2 < 0$.

Although we could interpret a negative population as extinct, this may not be the behaviour consistent with what we wanted to model. Indeed, if we take $P(0) = 0.99$, $P(1) = 0.0495 < 1$ and $P(2) = 0.235249$ so, despite being bigger, the latter population persists for a longer time than the former.

Ricker equation.

To remedy, we can consider other functions $f(P)$ subject to $f(P) > -1$. A possible choice is the exponential function

$$f(P) = ae^{-bP(k)} - 1.$$

If we introduce the carrying capacity K, then

$$b = \frac{\ln a}{K}$$

and, letting for simplicity $\rho = \ln a$, we obtain the *Ricker equation*

$$P(k+1) = P(k)e^{\rho(1-\frac{P(k)}{K})}. \tag{18}$$

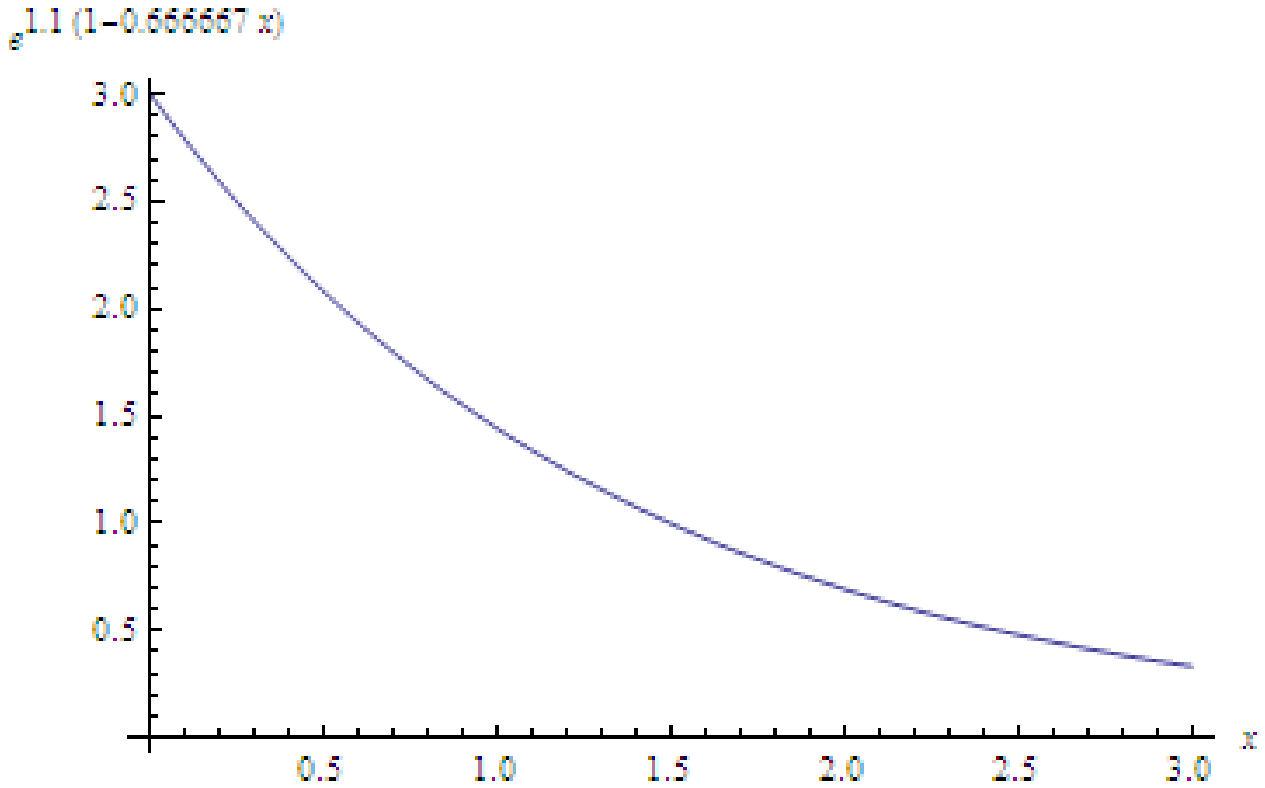


Figure: The function $f(x) = e^{\rho(1-x/K)}$ for $r = 1.1$ and $K = 1.5$.

We note that if $P(k) > K$, then $P(k+1) < P(k)$ and if $P(k) < K$, then $P(k+1) > P(k)$. The intrinsic growth rate β is given by $\beta = e^{\rho} - 1$ but, using the Maclaurin formula, for small ρ we have $\beta \approx \rho$.

Note that the Ricker model is qualitatively different from the Beverton-Holt model. For large $P(k)$ the former produces very small, but still positive, values of $P(k+1)$, while in the latter large populations stay large but only reproduce slowly.

Since $R(P)$ tends to zero faster than any power of P, we see that the Ricker model describes a scramble competition.

Allee type models.

In all previous models only the increase in the population density could slow down the growth. However, in 1931 Warder Clyde Allee noticed that in small, or dispersed, populations individual chances of survival decrease which can lead to extinction of the populations. This could be due to the difficulties of finding a mating partner or more difficult cooperation in e.g., organizing defence against predators.

Models having this property can be built within the considered framework by introducing two thresholds: the carrying capacity K and a parameter $0 < L < K$ at which the behaviour of the population changes so that

$$\Delta P/P < 0 \quad \text{for} \quad 0 < P < L \text{ and } P > K$$

$$\Delta P/P > 0 \quad \text{for} \quad L < P < K.$$

and the required properties can be obtained by taking $f(P) \leq 0$ for $0 < P < L$ and $P > K$ and $f(P) \geq 0$ for $L < P < K$. A simple model like that is offered by choosing $f(P) = (L - P)(P - K)$ so that

$$P(k+1) = P(k)(1 + (L - P(k))(P(k) - K)). \tag{19}$$

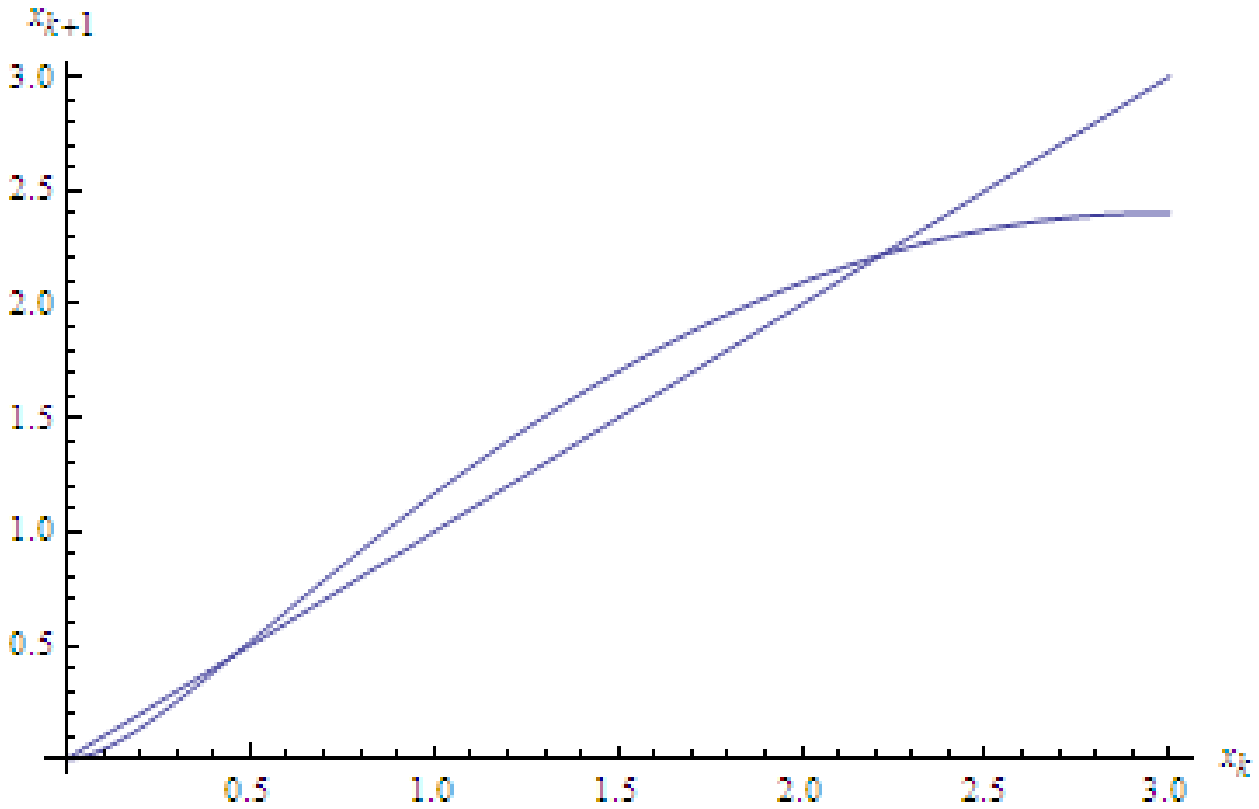


Figure: The relation $P(k+1) = P(k) + P(k)f(P(k))$ for an Allee model

Another model of this type can be justified by modelling looking of a mating partner or introducing a generalist predator (that is, preying also on other species), has the form

$$P(k+1) = P(k)\left(1+\lambda\left(1-\frac{P(k)}{C}-\frac{A}{1+BP(k)}\right)\right) \tag{20}$$

where $\lambda > 0$ and

$$1 < A < \frac{(BC+1)^2}{4CB}, BC > 1. \tag{21}$$

as well as

$$A \leq \min\left\{\frac{1+\lambda}{\lambda}, \frac{C}{K-L}\right\}. \tag{22}$$

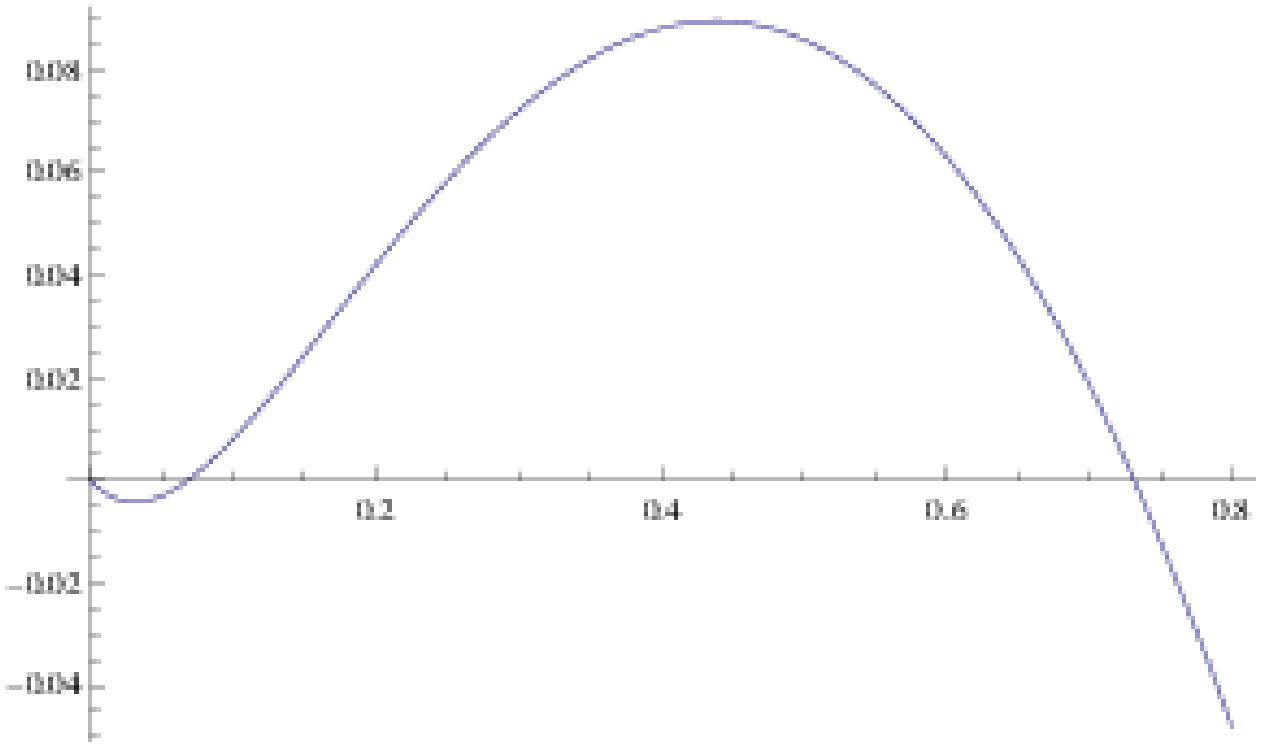


Figure: Graph of the function $1.2x(1-x) - \frac{0.03x}{0.02+0.1x}$: we see three equilibria, as required by the Allee model.

Discrete nonlinear population models from first principles.

The way we have introduced the equations may seem completely *ad hoc*. It follows, however, that it can be derived from a set of assumptions describing population in which the reproductive success of an individual is adversely affected by other individuals competing for the same resources. For this we need to introduce some probabilistic tools. The workhorse is the Poisson distribution describing the probability of exactly r events given we know their average number.

Interlude – Poisson distribution. Let us first assume that there are n individuals in a population and let p be the probability that one of them happens to be in your neighbourhood. Under usual assumption of independence of individuals occurring in the neighbourhood, the probability of having exactly r neighbours is given by the binomial formula

$$b(n,p;r) = \left(\begin{array}{c} n \\ r \end{array} \right) p^r (1-p)^{n-r}.$$

Average number of neighbours is thus $\omega = np$. This is intuitively clear and also can be checked by direct calculation

$$\begin{aligned} \omega &= \sum_{r=0}^{n} r \begin{pmatrix} n \\ r \end{pmatrix} p^r (1-p)^{n-r} \\ &= np \sum_{r=0}^{n-1} \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} p^{r-1} (1-p)^{n-r} = np. \end{aligned}$$

If we consider a potentially infinite population so that n grows to infinity in such a way that the average number of neighbours ω stays constant (so p goes to zero), then the probability of having exactly r neighbours is given by

$$\begin{aligned} p(r) &= \lim_{n\to\infty} b(n,\omega/n;r) = \lim_{n\to\infty} \frac{n!}{r!(n-r)!}\frac{\omega^r}{n^r}\left(1-\frac{\omega}{n}\right)^{n-r} \\ &= \frac{\omega^r}{r!}\lim_{n\to\infty}\frac{n(n-1)\cdot\ldots\cdot(n-r+1)}{n^r}\left(1-\frac{\omega}{n}\right)^{-r}\left(\left(1-\frac{\omega}{n}\right)^{\frac{n}{\omega}}\right)^{\omega} \\ &= \frac{e^{-\omega}\omega^r}{r!}, \end{aligned}$$

which is called the *Poisson distribution*.

The Ricker model

Returning to our problem, we consider a population, the size of which at time k is given by $P(k)$. The standard growth equation is (8)

$$P(k+1) = R(P(k))P(k), \quad k = 0, 1, \ldots, \tag{23}$$

where R gives the average number of offspring in the cycle. We assume that the number of offspring of an individual is adversely affected by the number of its neighbours living in, say, a disc D of area s. A simple possible model is that the number of offspring per capita is given by bc^r where $b \geq 0, 0 < c < 1$, where r is the number of neighbours in D.

If we assume that the population is uniformly distributed in an environment with area A, then the average number of individuals in D at time k is given by $sP(k)/A$ and, using the Poisson distribution, the probability of having r neighbours in D is

$$\frac{(sP(k))^r e^{-\frac{sP(k)}{A}}}{A^r r!}$$

and the average number of offspring per individual is

$$R(P(k)) = be^{-\frac{sP(k)}{A}} \sum_{r=0}^{\infty} \frac{(csP(k))^r}{A^r r!} = be^{-\frac{s(1-c)P(k)}{A}}.$$

Hence we obtain the Ricker model

$$P(k+1) = bP(k)e^{-\frac{s(1-c)P(k)}{A}}.$$

Comparing this expression with (18) we see that the carrying capacity K can be expressed as

$$K = \frac{A \ln b}{s(1-c)}.$$

Ricker and Beverton-Holt equations.

Consider a habitat consisting of N resource sites. At time k a population of $P(k)$ individuals is distributed and then reproduce. Let h_l be the proportion of sites with l individuals. It is a function of both $P(l)$ and N. Once on site, the individuals reproduce and the success of reproduction; that is, the number of offspring, denoted by $\phi(l)$, only depends on the number of individuals at the site. Then the difference equation governing the growth of the population is

$$P(k+1) = N \sum_{l=0}^{\infty} h_l \phi(l). \tag{24}$$

To be able to make use of this equation, we must specify the site occupation function h_l and the fecundity $\phi(l)$.

We consider two types of the site occupation function: uniform and preferential. In what follows we assume that the population is large, with a large number of sites, so that the expected occupation can be written as $\omega = P(k)/N$.

Uniform distribution. If individuals are uniformly distributed, then the probability of finding l individuals at any given site is Poisson distributed:

$$h_l = \frac{\left(\frac{P(k)}{N}\right)^l e^{-\frac{P(k)}{N}}}{l!} = \frac{\omega^l e^{-\omega}}{l!}$$

and (24) can be written as

$$P(k+1) = Ne^{-\omega} \sum_{l=0}^{\infty} \frac{\omega^l}{l!} \phi(l). \tag{25}$$

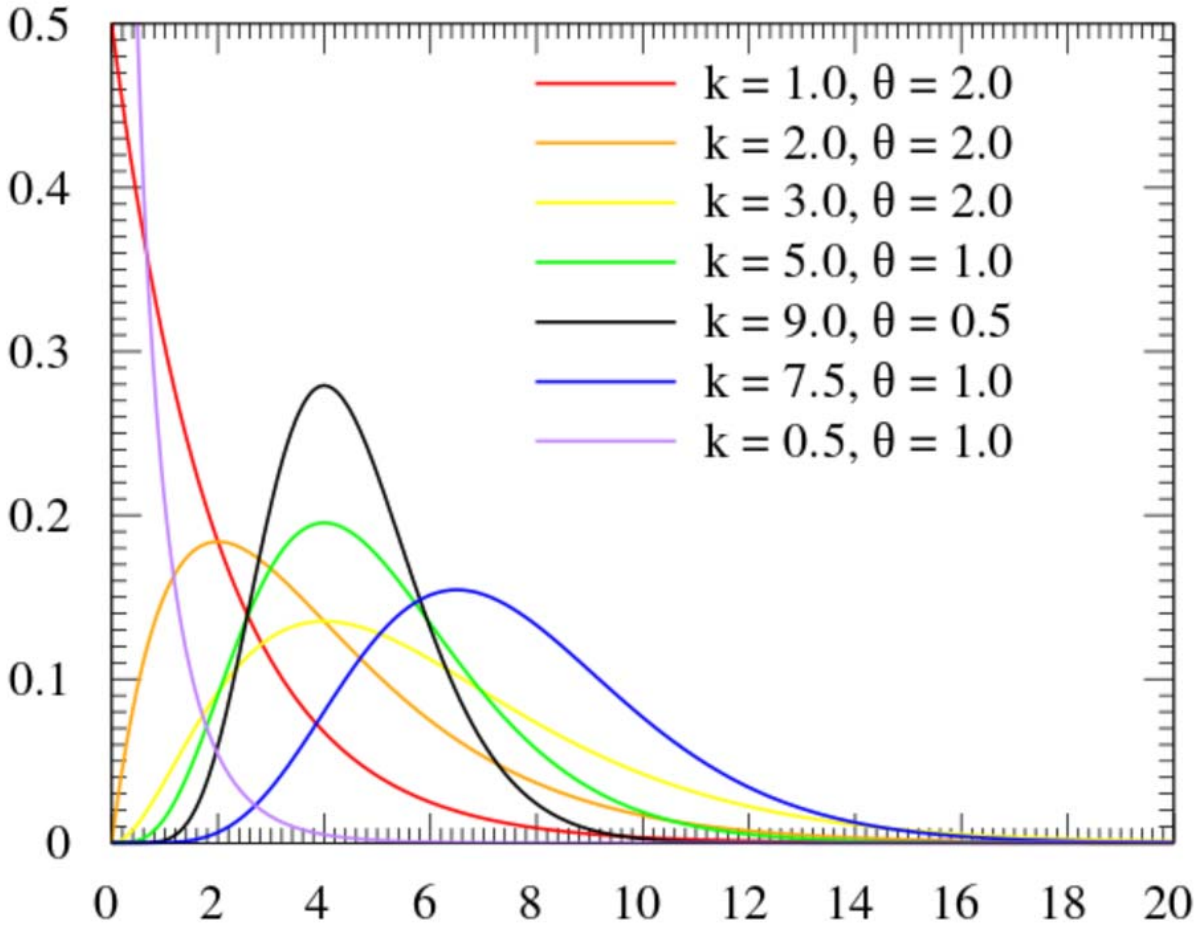
Preferential distribution. Here we assume that the sites are allocated at random a value, say $t \in \mathbb{R}_+$ which has a continuous probability density $f(t)$. Then, we assume that the average occupation of a site with value t is $t\omega$. If t was known then, as before, the number h_l of occupants of the site would be Poisson distributed: $(\omega t)^l e^{-\omega t}/l!$. Since, however, the value of t is not known, h_l will be given by

$$h_l = \int_0^\infty f(t) \frac{(\omega t)^l e^{-\omega t}}{l!} dt. \tag{26}$$

The function $f(t)$ is the probability density of the number of sites with value t and must be selected for each particular case. Quite often it is assumed to be Gamma distributed; that is,

$$f(t) = \frac{\lambda^{\lambda}}{\Gamma(\lambda)} t^{\lambda-1} e^{-\lambda t}$$

where λ is a positive parameter.


0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
k = 1.0, θ = 2.0
k = 2.0, θ = 2.0
k = 3.0, θ = 2.0
k = 5.0, θ = 1.0
k = 9.0, θ = 0.5
k = 7.5, θ = 1.0
k = 0.5, θ = 1.0

Then

$$
\begin{aligned}
h_l &= \int_0^\infty f(t)\frac{(\omega t)^l e^{-\omega t}}{l!}dt = \frac{\lambda^\lambda}{l!\Gamma(\lambda)\omega^\lambda}\int_0^\infty t^{\lambda+l-1}e^{-t\frac{\lambda+\omega}{\omega}}dt \\
&= \frac{\lambda^\lambda\omega^l}{l!\Gamma(\lambda)(\lambda+\omega)^{\lambda+l}}\int_0^\infty s^{\lambda+l-1}e^{-s}ds = \frac{\lambda^\lambda\omega^l\Gamma(l+\lambda)}{l!\Gamma(\lambda)(\lambda+\omega)^{\lambda+l}}
\end{aligned}
$$

which is the so called negative binomial distribution.

The formula for the growth of the population is then given by

$$P(k+1) = N\frac{\lambda^{\lambda}}{\Gamma(\lambda)(\lambda+\omega)^{\lambda}}\sum_{l=0}^{\infty}\frac{\omega^{l}\Gamma(l+\lambda)}{l!(\lambda+\omega)^{l}}\phi(l). \qquad (27)$$

The next step is to specify the offspring outcome at each site.

Scramble competition. Let as assume that each site contains resources to support one individual. Then

$$\phi_l = \begin{cases} b & \text{if } l = 1, \\ 0 & \text{otherwise,} \end{cases}$$

where b is the number of offspring produced by a site containing only one individual.

If the sites are distributed in a uniform way, substituting this into (25), we obtain

$$P(k+1) = bP(k)e^{-\frac{P(k)}{N}}, \tag{28}$$

which is the Ricker model.

On the other hand, the preferential distribution, resulting with the negative binomial distribution, gives

$$P(k+1) = b\frac{\lambda^{\lambda+1}P(k)}{\Gamma(\lambda)(\lambda + P(k)/N)^{\lambda+1}}, \tag{29}$$

where we used $\Gamma(\lambda+1) = \lambda\gamma(\lambda)$. Since $\lambda > 0$, this is the generalized Beverton–Holt model (14). Note, that since we have scramble competition, we cannot get here the basic Beverton-Holt model ($\lambda = 0$) which is compensatory and thus describes a contest competition.

Contest competition. Again we assume that each site can support one individual but, in contrast to the scramble competition, if there are more individuals at the site, only one emerges victorious and the others perish. Thus, the function ϕ_I is given by

$$\phi_I = \begin{cases} b & \text{if} \quad I \geq 1, \\ 0 & \text{if} \quad I = 0, \end{cases}$$

where, as before, b is the number of offspring produced by one individual.

Then the uniform distribution gives

$$P(k+1) = bNe^{-\frac{P(k)}{N}} \sum_{l=1}^{\infty} \frac{(P(k))^l}{N^l l!} = bN\left(1 - e^{-\frac{P(k)}{N}}\right), \qquad (30)$$

which is the so-called Skellam model.

Let us consider the preferential site distribution. We have

$$P(k+1) = bN\frac{\lambda^{\lambda}}{\Gamma(\lambda)(\lambda+\omega)^{\lambda}}\sum_{l=1}^{\infty}\frac{\omega^{l}\Gamma(l+\lambda)}{l!(\lambda+\omega)^{l}}. \qquad (31)$$

Now, using the fact that $\Gamma(l+\lambda) = \lambda(\lambda+1)\cdot\ldots\cdot(\lambda+l-1)\Gamma(\lambda)$ and denoting $z = \omega/(\lambda+\omega)$, we get

$$\sum_{l=0}^{\infty}\frac{\Gamma(l+\lambda)}{l!}z^{l} = \Gamma(\lambda)\sum_{l=0}^{\infty}\frac{\lambda(\lambda+1)\cdot\ldots\cdot(\lambda+l-1)}{l!}z^{l} = \Gamma(\lambda)(1-z)^{-\lambda}.$$

Now,

$$1-z=1-\frac{P(k)/N}{\lambda+P(k)/N}=\frac{\lambda}{\lambda+P(k)/N}$$

and thus (31) can be written as

$$\begin{aligned} P(k+1) &= bN\frac{\lambda^{\lambda}}{\Gamma(\lambda)(\lambda+P(k)/N)^{\lambda}}\left(\Gamma(\lambda)\lambda^{-\lambda}(\lambda+P(k)/N)^{\lambda}-\Gamma(\lambda)\right) \\ &= bN\left(1-\frac{\lambda^{\lambda}}{(\lambda+P(k)/N)^{\lambda}}\right). \end{aligned}$$

If $\lambda = 1$, the above equation corresponds to the Beverton-Holt model (14). Indeed, in this case

$$P(n+1) = bN\left(1 - \frac{1}{(1+P(n)/N)}\right) = \frac{bP(n)}{1+P(n)/N}.$$

We see that the carrying capacity is given by $K = N(b-1)$; that is, it is proportional to the number of sites as well as to the per capita birth rate above the simple reproduction.

Continuous time models.

Most discrete models introduced above have their continuous time counterparts.

Malthusian model. If births and death rates are constant then, denoting the net growth rate by r we obtain

$$\frac{dP}{dt} = rP. \tag{32}$$

which has a general solution given by

$$P(t) = P(0)e^{rt}, \tag{33}$$

where $P(0)$ is the size of the population at $t = 0$.

The U.S. Department of Commerce estimated that the Earth population in 1965 was 3.34 billion and that the population was increasing at an average rate of 2% per year during the decade 1960-1970. Thus $P(0) = 3.34 \times 10^9$ with $r = 0.02$, and

$$P(t) = 3.34 \times 10^9 e^{0.02t}. \tag{34}$$

Then the population will double in

$$T = 50 \ln 2 \approx 34.6 \text{ years},$$

which is in a good agreement with the estimated value of 6070 billion inhabitants of Earth in 2000. It also agrees relatively well with the observed data if we don't go too far into the past.

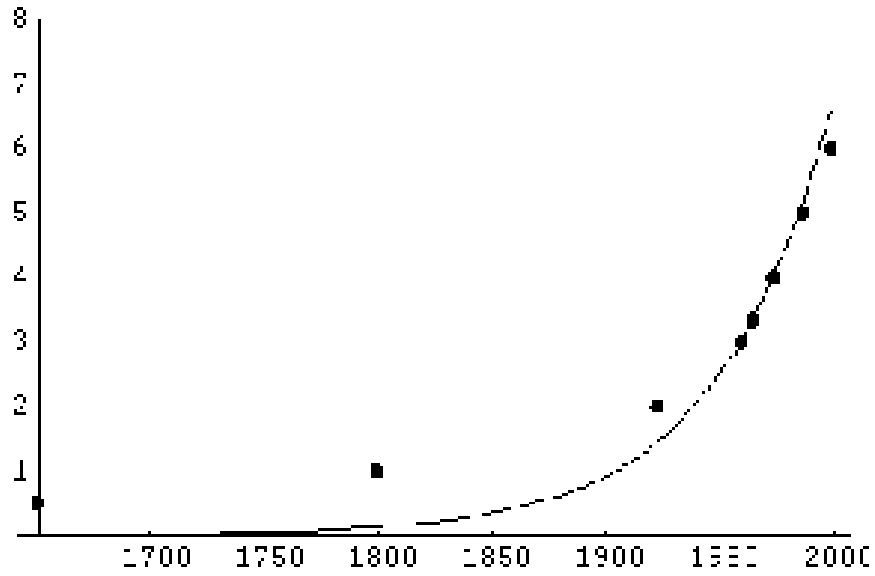


Fig 1.1. Comparison of actual population figures (points) with those obtained from equation (34).

On the other hand, if we try to extrapolate this model then in, say, 2515, the population would reach $199980 \approx 200000$ billion giving each of us area of $(86.3\,\mathrm{cm} \times 86.3\,\mathrm{cm})$ to live on.

Nevertheless, the Malthusian model has its uses for short term prediction. It also provides a useful link about the death rate and the expected life span of an individual.

Consider a population in which individuals die at a constant rate μ

$$P' = -\mu P.$$

Then the probability that an individual dies in a time interval Δt is approximately equal to $\mu \Delta t$. Let $p(t)$ be the probability that the individual is alive at time t. Then the probability $p(t + \Delta t)$ of it being alive at $t + \Delta t$ provided he/she was alive at t is $p(t + \Delta t) = (1 - \mu \Delta t)p(t)$ which, as above, yields

$$p' = -\mu p$$

with $p(0) = 1$ (expressing the fact that the individual was born, and thus alive, at $t = 0$) yielding $p(t) = e^{-\mu t}$.

The average life span is given by

$$L = \int_0^\infty sm(s)ds,$$

where $m(s)$ is the probability (density) of dying exactly at age s.

Since the probability of dying at the age between t and $t+\Delta t$ is

$$-p(t+\Delta) + p(t) = -\int_t^{t+\Delta t} \frac{d}{ds} p(s)ds$$

(one should be alive at t and dead at $t+\Delta t$, we have

$m(s) = -\frac{d}{ds}p(s)$ and

$$L = -\int_0^\infty s\frac{d}{ds}e^{-\mu s}ds = \mu\int_0^\infty se^{-\mu s}ds = \frac{1}{\mu}. \qquad (35)$$

Logistic equation. Passing to the limit in the discrete logistic equation valid between t and $t+\Delta t$,

$$P(t+\Delta t) - P(t) = r\Delta t \left(1 - \frac{P(t)}{K}\right)$$

we obtain the continuous logistic model

$$\frac{dP}{dt} = rP(t)\left(1 - \frac{P}{K}\right), \tag{36}$$

which proved to be one of the most successful models for describing a single species population.

The equation has two constant solutions, $P(t) = 0$ and $P(t) = K$, with the latter being the carrying capacity of the environment.

Other solutions can be obtained by separation of variables:

$$P(t) = \frac{P(0)K}{P(0) + (K - P(0))e^{-rt}}. \tag{37}$$

We have

$$\lim_{t\to\infty} P(t) = K, \quad P(0) > 0,$$

hence our model correctly reflects the initial assumption that K is the carrying capacity of the habitat. Next, we obtain

$$\frac{dP}{dt} > 0 \quad \text{if} \quad 0 < P(0) < K,$$
$$\frac{dP}{dt} < 0 \quad \text{if} \quad P(0) > K,$$

thus, if $P(0) < K$, the population monotonically increases, whereas if $P(0) > K$, then such a population will decrease until it reaches K.

Also, for $0 < P(0) < K$,

$$\frac{d^2P}{dt^2} > 0 \quad \text{if} \quad 0 < P(t) < K/2,$$
$$\frac{d^2P}{dt^2} < 0 \quad \text{if} \quad P(0) > K/2,$$

thus, as long as the population is small (less then half of the capacity), then the rate of growth increases, whereas for larger population the rate of growth decreases. This results in the famous *logistic* or *S-shaped* curve that describes saturation process.

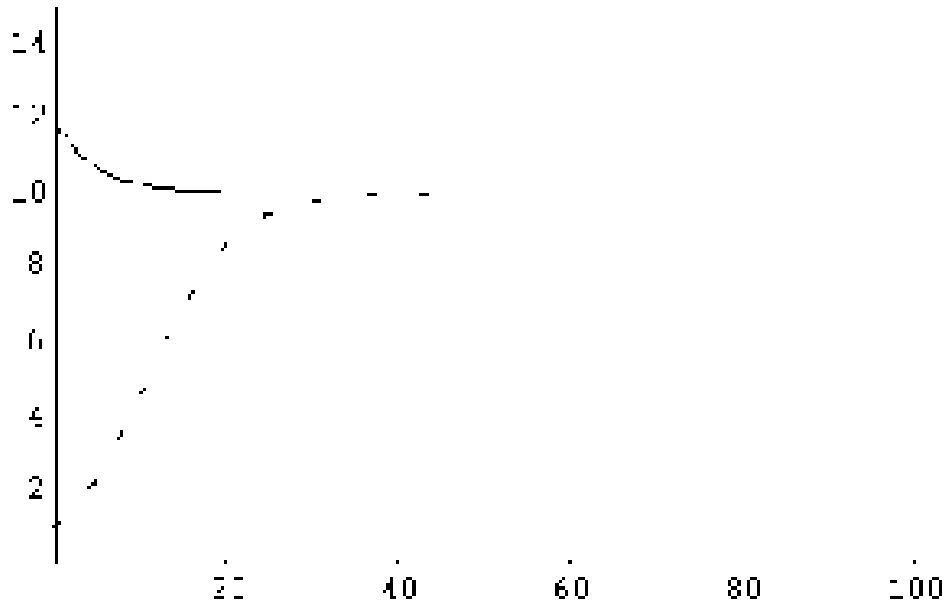


Figure: Logistic curves with $P_0 < K$ (dashed line) and $P_0 > K$ (solid line) for $K = 10$ and $r = 0.02$.

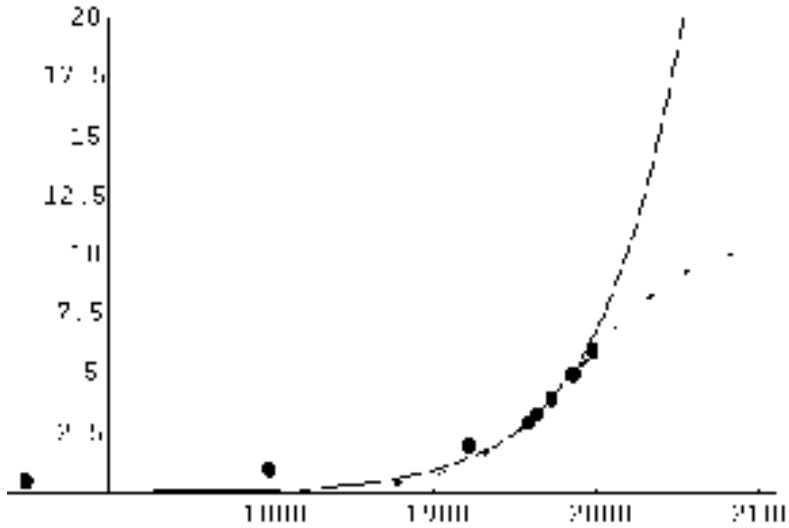


Figure: Human population on Earth with $K = 10.76$ billion and $r = 0.029$ and $P(1965) = 3.34$ billion. Observational data (points), exponential growth (solid line) and logistic growth (dashed line).

	Actual	Predicted	Error	%
1790	3,929,000	3,929,000	0	0.0
1800	5,308,000	5,336,000	28,000	0.5
1810	7,240,000	7,228,000	−12,000	−0.2
1820	9,638,000	9,757,000	119,000	1.2
1830	12,866,000	13,109,000	243,000	1.9
1840	17,069,000	17,506,000	437,000	2.6
1850	23,192,000	23,192,000	0	0.0
1860	31,443,000	30,412,000	−1,031,000	−3.3
1870	38,558,000	39,372,000	814,000	2.1
1880	50,156,000	50,177,000	21,000	0.0
1890	62,948,000	62,769,000	−179,000	−0.3
1900	75,995,000	76,870,000	875,000	1.2
1910	91,972,000	91,972,000	0	0.0
1920	105,711,000	107,559,000	1,848,000	1.7
1930	122,775,000	123,124,000	349,000	0.3
1940	131,669,000	136,653,000	4,984,000	3.8
1950	150,697,000	149,053,000	−1,644,000	−1.1

Figure: Comparison of actual and logistic model population in the United States

On the other hand, Verhulst in 1845 predicted, on the basis of the logistic equation, that the maximum population of Belgium is 6 600 000. However, already in 1930 it was close to 8 100 000. This is attributed to the global change that happened for Belgium in the XIX century - acquisition of Congo that provided resources to support increasing population (at the cost of the African population of Congo).

Are discrete and continuous models essentially the same?

Malthus model. Let us consider the continuous Malthus model

$$P' = rP,$$

the solution of which is given by

$$P(t) = P(0)e^{rt},$$

and its discrete version

$$p(k+1) = \rho\, p(k)$$

with solution

$$p(k) = P(0)\rho^k.$$

It is easily seen that if we take census of the continuous process at evenly spaced time moments $t_0 = 0, t_1 = 1, \ldots, t_k = k, \ldots$, we get

$$P(k) = e^{rk} P(0) = \left(e^r\right)^k P(0)$$

which coincides with discrete model with $\rho = e^r$. Hence, discrete and continuous Malthus processes are the same, up to an adjustment of the growth rate.

Logistic model. Consider the logistic differential equation

$$P' = aP(1-P) \tag{38}$$

and the difference equation used to derive it:

$$P(t+\Delta t) = P(t) + a\Delta t(1-P(t)). \tag{39}$$

If we fix Δt and $t = 0$, then (39) generates the recurrence

$$p(k+1) = p(k) + a\Delta t p(k)(1 - p(k)), \tag{40}$$

where $p(k) = P(k\Delta t)$.

Can we claim that $P(k) = p(k)$?

If we take $\Delta t = 1$ with, say, $a = 4$, we obtain the following picture.

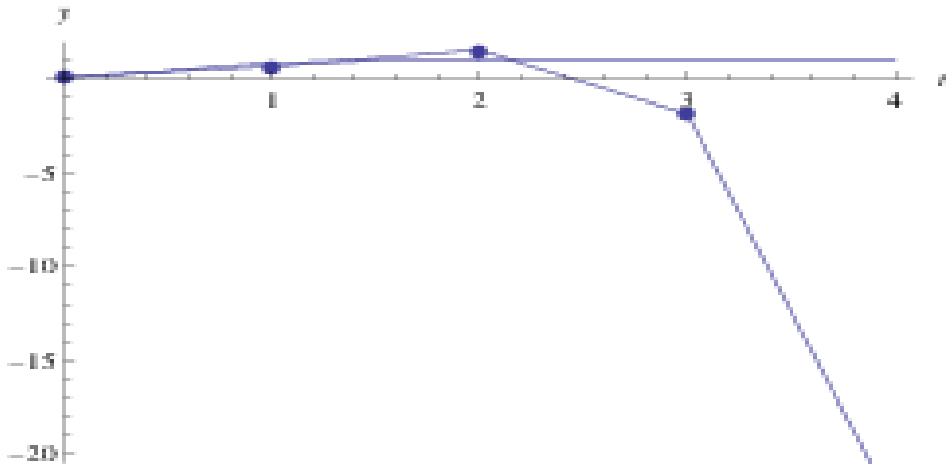


Figure: Comparison of solutions to (38) and (40) with $a = 4$ and $\Delta t = 1$.

The substitution

$$x(k) = \frac{a\Delta t}{1 + a\Delta t} p(k) \tag{41}$$

reduces (40) to

$$x(k+1) = \mu x(k)(1 - x(k)), \tag{42}$$

where $\mu = 1 + a\Delta t$.

By taking $1 + a\Delta t \leq 3$, we obtain the convergence of solutions $x(k)$ to the equilibrium $x^* = a\Delta t/(1 + a\Delta t)$ which, reverting to (41), gives the discrete approximation y_n which converges to 1, as the solution to (38). However, this convergence is not monotonic, hence the approximation is rather poor.

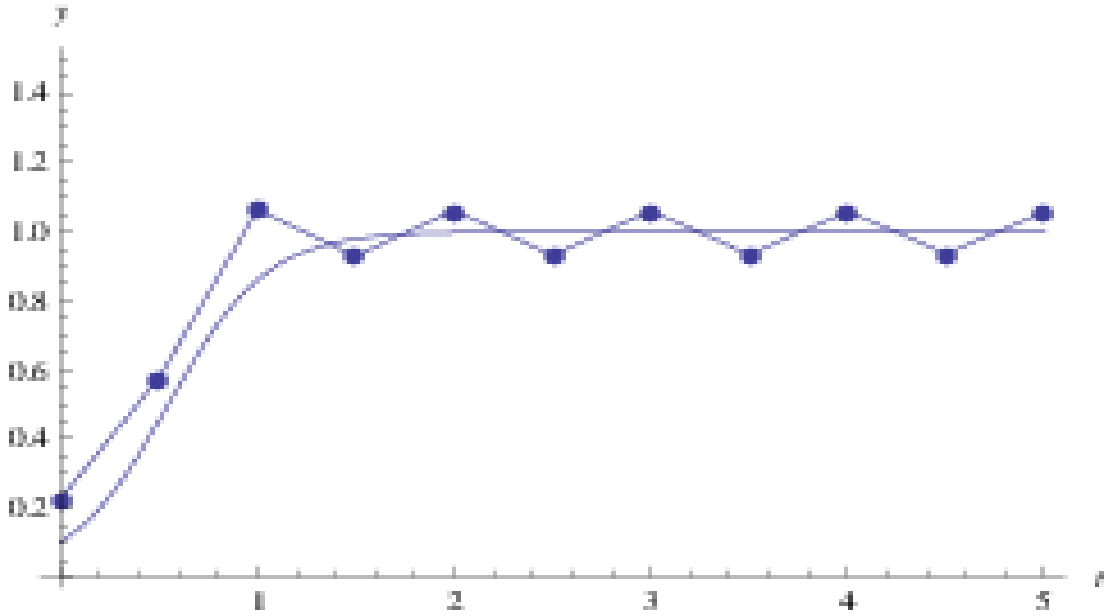


Figure: Comparison of solutions to (38) with $a = 4$ and (42) with $\mu = 3$ ($\Delta t = 0.5$).

This can be remedied by taking $1 + a\Delta t \leq 2$, in which case the qualitative features of $P(t)$ and $p(k)$ are the same.

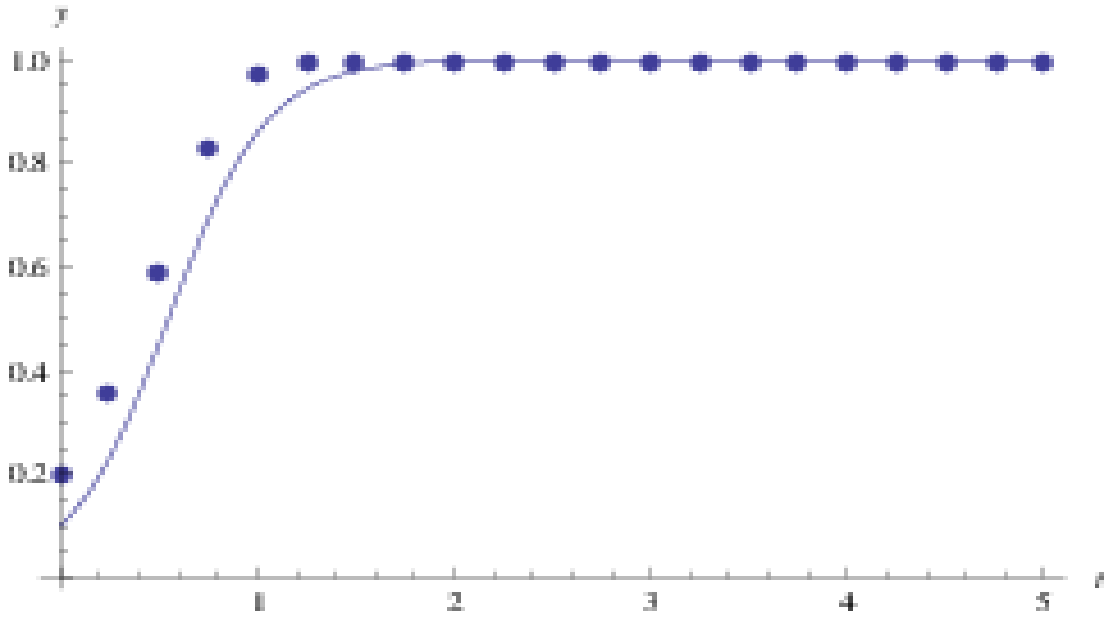


Figure: Comparison of solutions to (38) with $a = 4$ and (42) with $\mu = 2$ ($\Delta t = 0.25$).

It follows that it is the Beverton-Holt model that exactly traces the solution to the continuous logistic equation. Recall that

$$P(t) = \frac{P(0)e^{at}}{1+(e^{at}-1)P(0)}.$$

We have

$$\begin{aligned} P(k+1) &= \frac{P(0)e^{a(k+1)}}{1+(e^{a(k+1)}-1)P(0)} \\ = \frac{\frac{P(0)e^{ak}}{1+(e^{ak}-1)P(0)}e^{a}}{1+(e^{a}-1)\frac{P(0)e^{ak}}{1+(e^{ak}-1)P(0)}} &= \frac{e^{a}P(k)}{1+(e^{a}-1)P(k)} \end{aligned}$$

in which we recognize the Beverton-Holt model with the intrinsic growth rate related to the continuous logistic growth rate in the same way as in the Malthusian model.

Some explicitly solvable nonlinear population models.

Solvability of linear difference equations. We begin with the general first order difference equation

$$x(k+1) = a(k)x(k) + g(k) \tag{43}$$

with an initial condition $x(0) = x_0$. First few iterates are

$$\begin{aligned}
x(1) &= a(0)x(0) + g(0), \\
x(2) &= a(1)x(1) + g(1) = a(1)a(0)x(0) + a(1)g(0) + g(1), \\
x(3) &= a(2)x(2) + g(2) \\
&= a(2)a(1)a(0)x(0) + a(2)a(1)g(0) + a(2)g(1) + g(2).
\end{aligned}$$

We conjecture that the general form of the solution could be

$$x(k) = x(0) \prod_{r=0}^{k-1} a(r) + \sum_{r=0}^{k-1} g(r) \prod_{i=r+1}^{k-1} a(i) \tag{44}$$

where we adopted the convention that $\prod_{k}^{k-1} = 1$ and $\sum_{r=j+1}^{j} = 0$.

Then

$$
\begin{aligned}
x(k+1) &= a(k)x(k)+g(k) \\
&= a(k)\left(x(0)\prod_{r=0}^{k-1}a(r)+\sum_{r=0}^{k-1}g(r)\prod_{i=r+1}^{k-1}a(i)\right)+g(k) \\
&= x(0)\prod_{r=0}^{k}a(r)+a(k)\sum_{r=0}^{k-1}g(r)\prod_{i=r+1}^{k-1}a(i)+g(k) \\
&= x(0)\prod_{k=0}^{k}a(r)+\sum_{r=0}^{k-1}g(r)\prod_{i=r+1}^{k}a(i)+g(k)\prod_{i=k+1}^{k}a(i) \\
&= x(0)\prod_{k=0}^{k}a(r)+\sum_{r=0}^{k}g(r)\prod_{i=r+1}^{k}a(i)
\end{aligned}
$$

and induction ends the proof.

The formula simplifies if

$$x(k+1) = ax(k) + g(k). \tag{45}$$

Then

$$x(k) = a^k x(0) + \sum_{r=0}^{k-1} a^{k-r-1} g(r). \tag{46}$$

If, in particular, g is constant, we obtain

$$x(k) = \begin{cases} a^k x(0) + g\frac{a^k-1}{a-1} & \text{if} \quad a \neq 1, \\ x(0) + gk. \end{cases} \tag{47}$$

The Beverton–Holt model. We recall that the Beverton–Holt equation, Eq. (12), can be simplified to

$$x(k+1) = \frac{\beta x(k)}{(1+x(k))^b}. \tag{48}$$

While for general b this equation can display very rich dynamics, for $b = 1$ it can be solved explicitly by the so-called logistic substitution $y(k) = 1/x(k)$. Then we obtain

$$y(k+1) = \frac{1}{\beta} + \frac{1}{\beta} y(k)$$

and, by (47),

$$y(k) = \frac{1}{\beta}\frac{\beta^{-k}-1}{\beta^{-1}-1} + \beta^{-k}y_0 = \frac{1-\beta^k}{\beta^k(1-\beta)} + \beta^{-k}y_0$$

if $\beta \neq 1$ and, for $\beta = 1$,

$$y(k) = k + y_0.$$

Thus

$$x(k) = \begin{cases} \frac{\beta^k(\beta-1)x_0}{x_0(\beta^k-1)+\beta-1} & \text{if} \quad \beta \neq 1, \\ \frac{x_0}{1+x_0 k}. \end{cases}$$

We see that

$$\lim_{k\to\infty} x(k) = \beta - 1$$

if $\beta > 1$ and if $\beta \leq 1$,

$$\lim_{k\to\infty} x(k) = 0.$$

Thus, in this case, the carrying capacity is asymptotically achieved.

The logistic equation. In general, the discrete logistic equation does not admit closed form solutions and also displays a very rich dynamics. However, some special cases can be solved by an appropriate substitution. We will look at two such cases. First consider

$$x(k+1) = 2x(k)(1-x(k)) \tag{49}$$

with $x_0 \in [0,1]$. Since $f(x) = 2x(1-x)$ satisfies $0 \leq f(x) \leq 1/2$ on $[0,1]$, we see that $0 \leq x(k) \leq 1$ for any $k = 1,2,\ldots$. We use the substitution $x(k) = 1/2 - y(k)$.

Then

$$\frac{1}{2} - y(k+1) = 2\left(\frac{1}{2} - y(k)\right)\left(\frac{1}{2} + y(k)\right) = \frac{1}{2} - 2(y(k))^2,$$

so that

$$y(k+1) = 2(y(k))^2.$$

We see that if $y_0 = 0$, then $y(k) = 0$ for all $k \geq 1$. Furthermore, $y(k) > 0$ for $k \geq 1$ and we can take the logarithm of both sides getting

$$\ln y(k+1) = 2 \ln y(k) + \ln 2$$

which, upon substitution $z(k) = \ln y(k)$, becomes the inhomogeneous linear equation

$$z(k+1) = 2z(k) + \ln 2.$$

Using (47), we find the solution to be

$$z(k) = 2^k z_0 + \ln 2(2^k - 1).$$

Hence

$$y(k) = e^{z(k)} = e^{2^k \ln|y_0|} e^{\ln 2(2^k-1)} = y_0^{2^k} 2^{2^k-1},$$

where we dropped the absolute value bars as we rise y_0 to even powers. Thus

$$x(k) = \frac{1}{2} - \left(\frac{1}{2} - x_0\right)^{2^k} 2^{2^k-1}.$$

We note that for $x_0 = 1/2$ we have $x(k) = 1/2$ for all k so that we obtain a constant solution. In other words, $x = 1/2$ is an equilibrium point of (49). We observe that $x = 0$ is another equilibrium with the property that if $x_0 = 1$, then $x(k) = 0$ for $k \geq 1$. Otherwise, for $x_0 \in]0,1[$ we have

$$\lim_{k\to\infty} x(k) = \frac{1}{2}.$$

Another particular logistic equation which can be solved by substitution is

$$x(k+1) = 4x(k)(1-x(k)). \tag{50}$$

First we note that since $f(x) = 4x(1-x) \leq 1$ for $0 \leq x \leq 1$, we have $0 \leq x(k+1) \leq 1$ if $x(k)$ has this property. Thus, assuming $0 \leq x_0 \leq 1$, we can use the substitution

$$x(k) = \sin^2 y(k) \tag{51}$$

which yields

$$\begin{aligned} x(k+1) &= \sin^2 y(k+1) = 4\sin^2 y(k)(1-\sin^2 y(k)) \\ &= 4\sin^2 y(k)\cos^2 y(k) = \sin^2 2y(k). \end{aligned}$$

This gives the family of equations

$$y(k+1) = \pm 2y(k) + k\pi, \qquad k \in \mathbb{Z}.$$

Since our aim is to find $x(k)$ given by (51), the periodicity and symmetry of $\sin^2$ allows for discarding $k\pi$ and the minus sign. Thus we consider $y(k+1) = 2y(k)$ and hence

$$y(k) = C2^k,$$

where $C \in \mathbb{R}$ is arbitrary, as the general solution. Hence

$$x(k) = \sin^2 C2^k.$$

A remarkable fact is that, despite being explicitly solvable, the equation generates very irregular (chaotic) dynamics.

Equilibrium points of difference equations. Consider the autonomous first order difference equation

$$x(k+1) = f(x(k)), \qquad k \in \mathbb{N}_0, \tag{52}$$

with the initial condition x_0. In what follows we shall assume that f is at least continuous. It is clear that the solution to (52) is given by iterations

$$x(k) = f(f(\ldots f(x_0))) = f^k(x_0). \tag{53}$$

A point x^* in the domain of f is said to be an *equilibrium point* of (52) if it is a fixed point of f, that is, if

$$f(x^*) = x^*.$$

Graphically, an equilibrium is the x-coordinate of a point, where the graph of f intersects the diagonal $y = x$. This is the basis of the cobweb method of finding and analysing equilibria, described in the next subsection.

Definition

1. The equilibrium x^* is stable if for given $\varepsilon > 0$ there is $\delta > 0$ such that for any x and for any $k > 0$, $|x - x^*| < \delta$ implies $|f^k(x) - x^*| < \varepsilon$ for all $k > 0$. If x^* is not stable, then it is called unstable.
2. A point x^* is called attracting if there is $\eta > 0$ such that $|x_0 - x^*| < \eta$ implies $\lim_{k \to \infty} f^k(x_0) = x^*$. If $\eta = \infty$, then x^* is called a global attractor or globally attracting.
3. The point x^* is called an asymptotically stable equilibrium if it is stable and attracting. If $\eta = \infty$, then x^* is said to be a globally asymptotically stable equilibrium.

The cobweb diagrams. We describe an important graphical method for analysing the stability of equilibrium (and periodic) points of (52). Since $x(k+1) = f(x(k))$, we may draw a graph of f in the $(x(k), x(k+1))$ system of coordinates. Then, given $x(0) = x_0$, we pinpoint the value $x(1)$ by drawing a vertical line through $x(0)$ so that it also intersects the graph of f at $(x(0), x(1))$. Next, we draw a horizontal line from $(x(0), x(1))$ to meet the diagonal line $y = x$ at the point $(x(1), x(1))$. A vertical line drawn from the point $(x(1), x(1))$ will meet the graph of f at the point $(x(1), x(2))$. In this way we may find any $x(k)$.

This is illustrated in Fig. 13, where we presented several steps of drawing the cobweb diagram for the logistic equation

$$x(k+1) = 3x(k)(1-x(k)), \qquad x_0 = 0.2.$$

The equilibria are $x_1 = 0$ and $x_2 = 2/3$. On the basis of the diagram we can conjecture that $x_2 = 2/3$ is an asymptotically stable equilibrium as the solution converges to it as k becomes large. However, to be sure, we need to develop analytical tools for analysing stability.

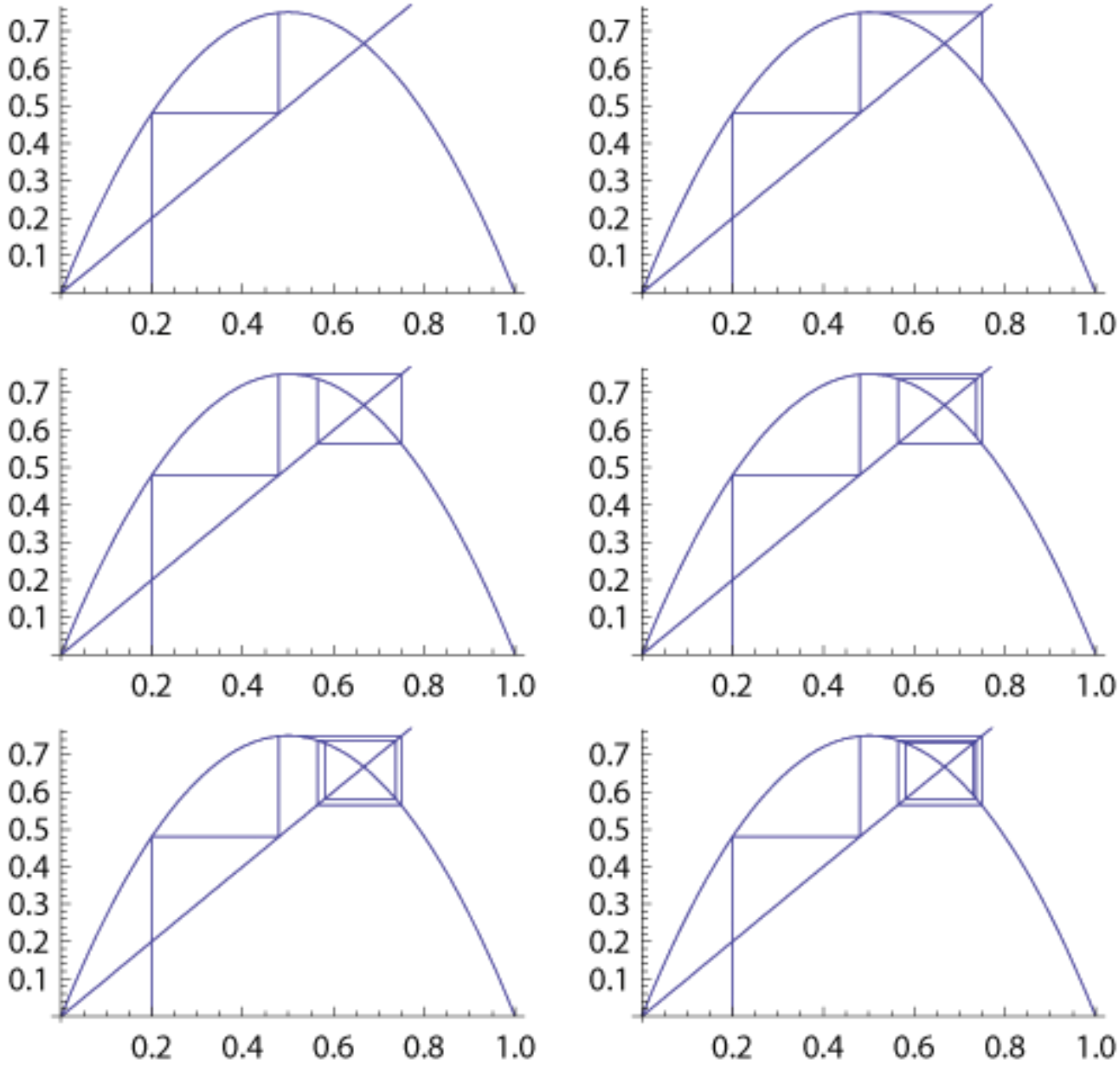


Figure: Cobweb diagram of a logistic difference equation

Analytic criterion for stability.

Theorem

Let x^* be an isolated equilibrium point of the difference equation

$$x(n+1) = f(x(n)), \tag{54}$$

where f is continuously differentiable in some neighbourhood of x^*. Then,

(i) if $|f'(x^*)| < 1$, then x^* is asymptotically stable;

(ii) if $|f'(x^*)| > 1$, then x^* is unstable.

Proof.

Suppose $|f'(x^*)| < M < 1$. Then $|f'(x)| \leq M < 1$ over some interval $J = (x^* - \gamma, x^* + \gamma)$ by the property of local preservation of sign for continuous functions. Let $x_0 \in J$. We have

$$|x(1) - x^*| = |f(x_0) - f(x^*)|$$

and, by the Mean Value Theorem, for some $\xi \in [x_0, x^*]$,

$$|f(x_0) - f(x^*)| = |f'(\xi)||x_0 - x^*|.$$

Hence,

$$|x(1) - x^*| = |f(x_0) - f(x^*)| \leq M|x_0 - x^*|.$$

Since $M < 1$, $x(1) \in J$.

By induction,

$$|x(k) - x^*| \leq M^k |x_0 - x^*|.$$

For given ε, define $\delta = \varepsilon$. Then $|x(k) - x^*| < \varepsilon$ for $n > 0$ provided $|x_0 - x^*| < \delta$ (since $M < 1$). Furthermore $x(k) \to x^*$ and $k \to \infty$ so that x^* is (locally) asymptotically stable.

To prove the converse, first observe that x^* is unstable if there is $\varepsilon > 0$ such that for any $\delta > 0$ there are x and k such that $|x - x^*| < \delta$ and $|f^k(x) - x^*| \geq \varepsilon$.

By the first part, there is $\varepsilon > 0$ such that on $J = (x^* - \varepsilon, x^* + \varepsilon)$ we have $|f'(x)| \geq M > 1$. Take an arbitrary $\delta > 0$ smaller than ε and choose x satisfying $|x - x^*| < \delta$. Using again the Mean Value Theorem, we get

$$|f(x) - x^*| \geq M|x - x^*|.$$

If $f(x)$ is outside J, then we are done.

If not, we can repeat the argument getting

$$|f^2(x) - x^*| \geq M^2|x - x^*|,$$

that is, $f^2(x)$ is further away from x^* than $f(x)$. If $f^2(x)$ is still in J, we continue the procedure. Since $M^k \to \infty$ as $k \to \infty$ there is k such that $f^{k-1}(x) \in J$ but

$$|f^k(x) - x^*| \geq M^n|x - x^*| > \varepsilon.$$

□

Equilibrium x^* with $|f'(x^*)| \neq 1$ is called hyperbolic.

What happens if the equilibrium point x^* is not hyperbolic? Let us reflect on the geometry of the situation and assume that $f'(x^*) > 0$. The equilibrium x^* is stable if the graph of f crosses the line $y = x$ from above to below as x increases. This ensures that the cobweb iterations from the left are increasing, and from the right are decreasing, while converging to x^*. In contrast, x^* is unstable if the graph of f crosses $y = x$ from below – then the cobweb iterations will move away from x^*.

If $f'(x^*) = 1$, then the graph of f is tangent to the line $y = x$ at $x = x^*$, but the stability properties follow from the geometry. If $f''(x^*) \neq 0$, then f is convex (or concave) close to x^* and the graph of f will be (locally) either entirely above or entirely below the line $y = x$. Therefore the picture is the same as in the unstable case either to the left, or to the right, of x^*. Hence, x^* is unstable in this case (remember that for instability it is sufficient to display, for any neighbourhood of x^*, only one diverging sequence of iterations emanating from this neighbourhood).

On the other hand, if $f''(x^*) = 0$, then x^* is an inflection point and the graph of f crosses the line $y = x$. This case is essentially the same as when $|f'(x^*)| \neq 1$: the equilibrium is stable if the graph of f crosses $y = x$ from above and unstable if it does it from below. If f''' exists around x^*, the former occurs when $f'''(x^*) < 0$, while the latter if $f'''(x^*)$.

Theorem

Let x^* be an isolated equilibrium with $f'(x^*) = 1$ and let f be three times continuously differentiable in a neighbourhood of x^*. Then,

(i) if $f''(x^*) \neq 0$, then x^* is unstable,

(ii) if $f''(x^*) = 0$ and $f'''(x^*) > 0$, then x^* is unstable,

(iii) if $f''(x^*) = 0$ and $f'''(x^*) < 0$, then x^* is asymptotically stable.

The case of $f'(x^*) = -1$ is more difficult. We observe more general situation. First, we note that x^* is also an equilibrium of $g(x) := f(f(x))$ and it is a stable equilibrium of f if and only if it is stable for g. This statement follows from the continuity of f: if x^* is stable for g, then $|g^k(x_0) - x^*| = |f^{2k}(x_0) - x^*|$ is small for x_0 sufficiently close to x^*. But then

$$|f^{2n+1}(x_0) - x^*| = |f(f^{2n})(x_0) - f(x^*)|$$

is also small by continuity of f. The reverse is obvious.

Then, since $g'(x^*) = 1$, we can apply the previous theorem g, getting

Theorem

Suppose that at an equilibrium point x^* we have $f'(x^*) = -1$. Define $S(x^*) = -f'''(x^*) - 3(f''(x^*))^2/2$. Then x^* is asymptotically stable if $S(x^*) < 0$ and unstable if $S(x^*) > 0$.

We can provide a fine-tuning of the notion of stability by noting that if $f'(x^*) < 0$, then the solution behaves in an oscillatory way around x^* and if $f'(x^*) > 0$, then it is monotonic. Indeed, consider (in a neighbourhood of x^* where $f'(x) < 0$)

$f(x) - f(x^*) = f(x) - x^* = f'(\xi)(x - x^*)$, $\xi \in (x^*, x)$. Since $f' < 0$, $f(x) > x^*$ if $x < x^*$ and $f(x) < x^*$ if $x > x^*$, hence each iteration moves the point to the other side of x^*. If $|f'| < 1$ over this interval, then $f^n(x)$ converges to x^* in an oscillatory way, while if $|f'| > 1$, the iterations will move away from the interval, also in an oscillatory way.

What happens if $f'(x^*) = 0$? Clearly, if there is a local extremum at x^*, then in some neighbourhood the derivative will have a fixed sign and thus the behaviour of the iterates will be as above. Let, on the other hand, $f'(x) < 0$ is some one-sided neighbourhood of x^* and $f'(x) > 0$ on the other side. Then, if the iterates start in the latter, they will stay there, converging to x^*, while if the iterates start in the former, they will begin converging to x^* in an oscillatory way until they reach the neighbourhood in which the derivative is positive and then they will converge monotonically.

Hence, we say that the equilibrium is oscillatory unstable or stable if $f'(x^*) < -1$ or $-1 < f'(x^*) < 0$, respectively, and monotonically stable or unstable depending on whether $0 \leq f'(x^*) < 1$ or $f'(x^*) > 1$, respectively.

Some applications.

Logistic model. Consider the logistic equation

$$x(k+1) = F_\beta(x(k)) := \beta x(k)(1-x(k)), \quad x \in [0,1],\ \beta > 0. \quad (55)$$

To find the equilibrium points, we solve

$$F_\beta(x^*) = x^*$$

which gives

$$x_0^* = 0, \qquad x_\beta^* = (\beta - 1)/\beta.$$

We investigate the stability of each point separately.

(a) For $x_0^* = 0$, we have $F'_\beta(0) = \beta$ and thus $x_0^* = 0$ is asymptotically stable for $0 < \beta < 1$ and unstable for $\beta > 1$. To investigate the stability for $\beta = 1$, we find $F''_\beta(0) = -2\beta \neq 0$ and thus $x_0^* = 0$ is unstable in this case. However, the instability comes from the negative values of x, which we discarded from the domain. If we restrict our attention to the domain $[0, 1]$, then $x_0^* = 0$ is stable. Such points are called *semi-stable*.

(b) The equilibrium point $x_\beta^* = (\beta - 1)/\beta$ belongs to the domain $[0,1]$ only if $\beta > 1$. Here, $F'_\beta((\beta-1)/\beta) = 2 - \beta$ and $F''_\beta((\beta-1)/\beta) = -2\beta$. Thus, using the stability theorems, we obtain that x_β^* is asymptotically stable if $1 < \beta \leq 3$ and it is unstable if $\beta > 3$.

Further, for $1 < \beta < 2$ the population approaches the carrying capacity monotonically from below. However, for $2 < \beta \leq 3$ the population can go over the carrying capacity but it will eventually stabilize around it.

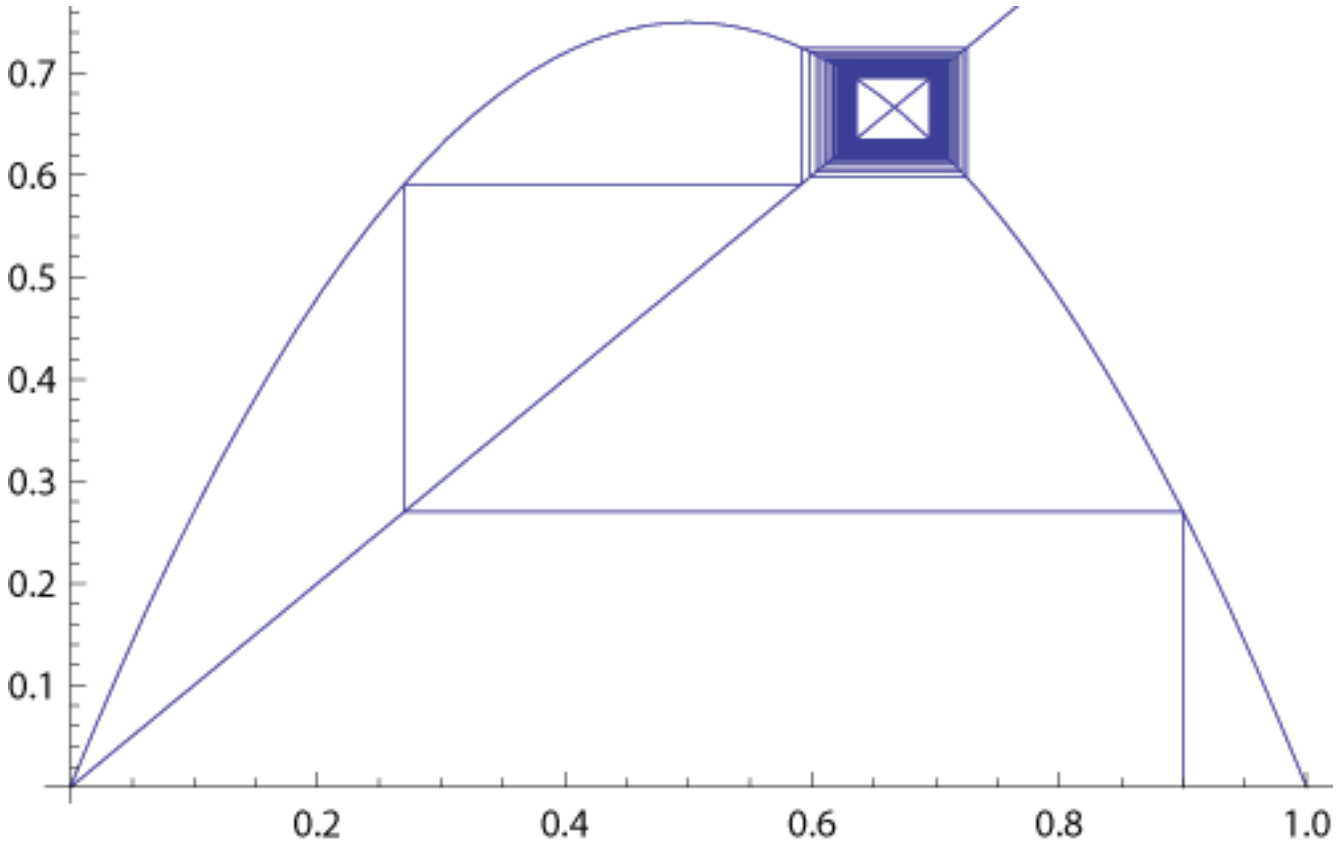


Figure: Asymptotically stable equilibrium $x_3^* = 2/3$ for $\beta = 3$.

Sustainable fishing. Let us consider a population of fish which grows according to the logistic equation with overlapping generations,

$$N(k+1) = N(k) + rN(k)\left(1 - \frac{N(k)}{K}\right).$$

This equation only can be solved in some particular cases. However, even without solving it, we can draw some conclusions relevant to the policies governing sustainable economy.

The basic idea of a sustainable economy is to find an optimal level of fishing: too much harvesting would deplete the fish population beyond recovery and too little would provide insufficient return for the community.

To maintain the population at a constant level, only the increase in the population should be harvested during any one season. The simplest model which includes fishing is

$$N(k+1) = N(k) + rN(k)\left(1 - \frac{N(k)}{K}\right) - qEN(k), \qquad (56)$$

where E is the so-called fishing effort, for instance the number of fishing boats at sea, and q is the fishing efficiency, that is, the fraction of the population caught by one boat in the unit time. Here we should find find the amount of fish that can be caught during a year to maintain the population at a constant level and hence find the population for which the yield is optimal.

To find a possible constant population level for a given level of fishing we solve

$$N = N + rN\left(1 - \frac{N}{K}\right) - qEN \tag{57}$$

This gives $N = 0$ or

$$N^* = K\left(1 - \frac{qE}{r}\right). \tag{58}$$

The first solution is trivial and not interesting. The second solution is positive if

$$qE < r.$$

Let us suppose that for the given fishing rate qE the population is kept at N^* given by (58). Then the yield is $Y(qE) = qEN^*$ or

$$Y(qE) = qEK\left(1 - \frac{qE}{r}\right).$$

Thus the yield is a quadratic function of the fishing rate qE and hence its maximum can be found as in the first part of the section. Maximum of $Y(qE)$ is attained at $E = \frac{r}{2q}$ which gives the maximum sustainable yield as

$$Y_{max}(qE) = \frac{rK}{4}.$$

Clearly, if we increase the fishing effort, $E > r/2q$, then the yield will decrease (greed does not pay!).

To understand why, let us summarize the mechanism described by the model. For a given fishing rate qE, we find that the equilibrium N^*, given by (58), is asymptotically stable provided

$$\left| \frac{d}{dN}\left(N + rN\left(1 - \frac{N}{K}\right) - qEN \right)\Big|_{N=N^*} \right| = \left| 1 + r - \frac{2rN^*}{K} - qE \right| < 1.$$

Using (58), this gives

$$r - 2 \leq qE < r.$$

Thus, if the fishing rate satisfies the above condition, then the fish population eventually will stabilize at N^* and N^* decreases if the fishing rate increases. The yield is the product of the fishing rate and the size of the population. Thus, if the fishing rate is too low, then though N^* is large, the product is not optimal. Similarly, overfishing results in the population settling at a lower N^* again resulting in a suboptimal yield.

One may wonder what is the significance of the condition $r-2 \leq qE$ which puts an upper bound on r. A common sense would suggest that the higher the net growth rate, the better the yield. To explain this, we observe that (57) can be written as

$$N(k+1) = (1+r-qE)\,N(k)\left(1-\frac{N}{\frac{K(1+r-qE)}{r}}\right)$$

and the stable equilibrium exists only if

$$1+r-qE \leq 3 \quad \text{iff} \quad r-2 \leq qE.$$

Thus, the case $r > qE+2$ does not give a stable fish population ensuring the sustainable constant yield.

Biological pest control. Assume that we have to deal with an insect population which invades our plantation. The insects reproduce according to the Beverton-Holt model (11),

$$P(k+1) = \frac{\beta P(k)}{1 + aP(k)} \tag{59}$$

where β is the natural fertility of insects and $a = (\beta - 1)/K$ where K is the capacity of the environment. As we learned, if $\beta > 1$, then the population $P(k)$ tends a nonzero equilibrium $K = (\beta - 1)/a$; that is, there is a nonzero stable population of insects.

An ecological way of eradicating the pest is to decrease the birth rate and one of the methods is to introduce a number S of sterile insects into the population. We assume that S is under our control and we can keep the number of the sterile insects constant in time. Though suppressed in the model, insects reproduce sexually and thus the effective birth rate depends on the probability of finding a mate. If, say, S individuals are sterile, and $P(k)$ is the number of fertile insects, then the probability of picking a fertile insect is $P(k)/(P(k)+S)$.

Thus, the Beverton-Holt model can be modified as

$$P(k+1) = \beta P(k) \frac{P(k)}{S + P(k)} \frac{1}{1 + aP(k)} = \beta P(k) f(P(k)). \quad (60)$$

To find the equilibria of (60), we solve

$$P = \beta P \frac{P}{S + P} \frac{1}{1 + aP}$$

which gives $P_0^* = 0$ and the simplified equation

$$1 = \beta \frac{P}{S + P} \frac{1}{1 + aP}.$$

While the above equation can be solved for P_1^*, a faster way to look at the functional relationship between P and S.

Thus, we solve the above equation for S as a function of P, getting

$$S(P) = \frac{(\beta - 1 - aP)P}{1 + aP}. \tag{61}$$

We find that $S(0) = 0$ and the derivative is given by

$$S'(P) = -1 + \beta/(1 + aP)^2.$$

Hence, the maximum in the interval $(0, (\beta - 1)/a)$ is attained at $P = (\sqrt{\beta} - 1)/a$; the maximum is $S_{\max} = (\sqrt{\beta} - 1)^2/a$.

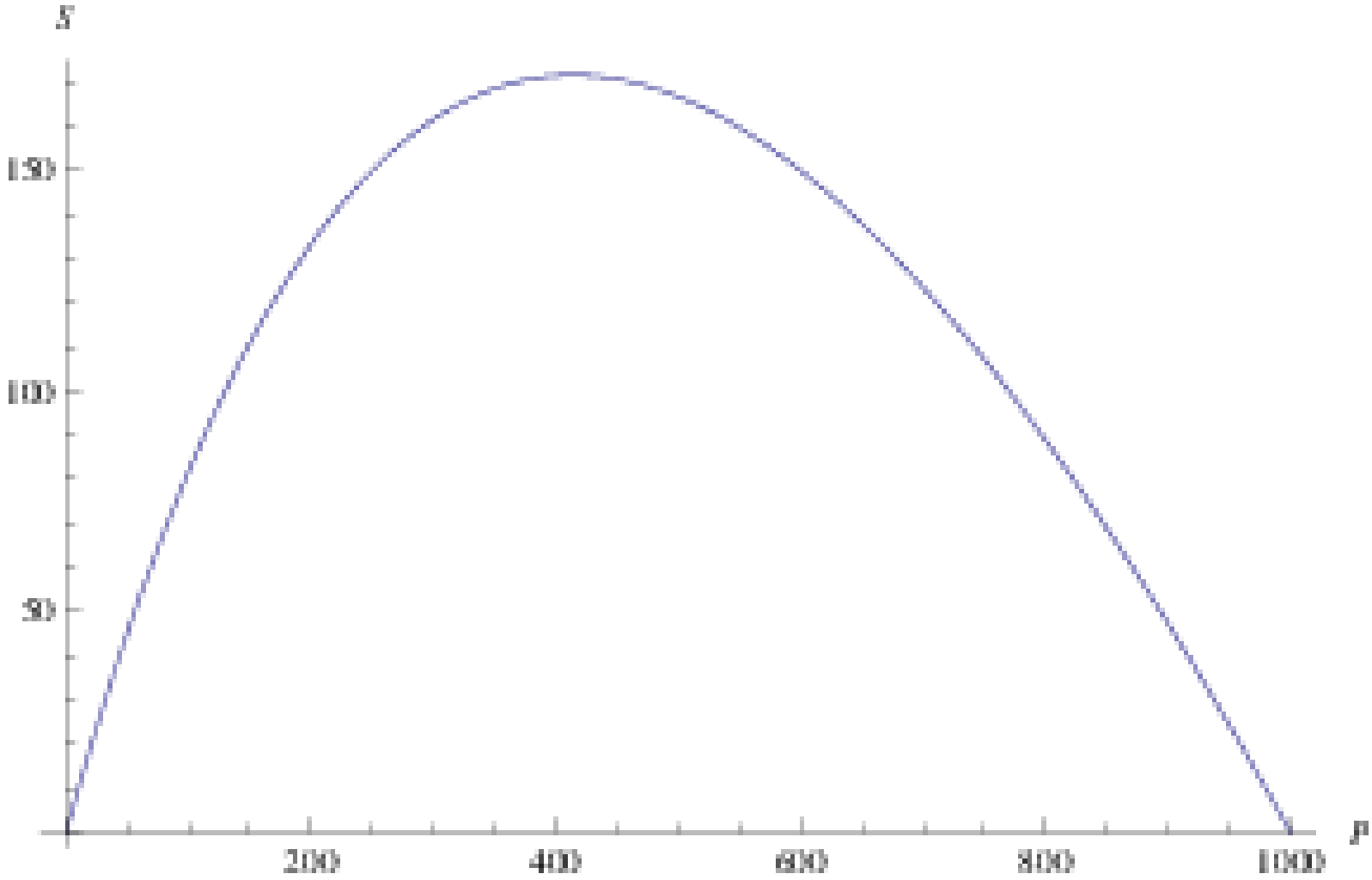


Figure: The graph of $S(P)$, given by (61),for $\beta = 2$ and $a = 0.001$.

Thus, we conclude that if $S > S_{max}$, the only equilibrium of (60) is $P_0^* = 0$. Further,

$$\beta \frac{d}{dP} Pf(P) = \beta \frac{P^2(1+aS)+2PS}{((P+S)(1+aP))^2}.$$

Hence, $\beta \frac{d}{dP} Pf(P)|_{P=0} = 0$ and the equilibrium $P_0^* = 0$ is asymptotically stable (for any S).

To determine the stability of the other two equilibria that exist for $0 < S < S_{\max}$ we use geometric considerations.

First we notice that $\beta \frac{d}{dP} Pf(P) > 0$ for all $P > 0$ (provided $S > 0$). This means that the derivative never equals -1. Let $0 < S < S_{\max}$. The curve $\beta Pf(P)$ starts at zero below the diagonal and thus at the smaller equilibrium it must cross the diagonal from below. Hence, this equilibrium is unstable. At the second equilibrium, the curve crosses the diagonal from above and, since the curve is ascending, the derivative is between 0 and 1. Thus this equilibrium is asymptotically stable. If $S = S_{\max}$, we have a tangent point which means that the (unique) positive equilibrium is semi-stable - unstable from the left and stable from the right.

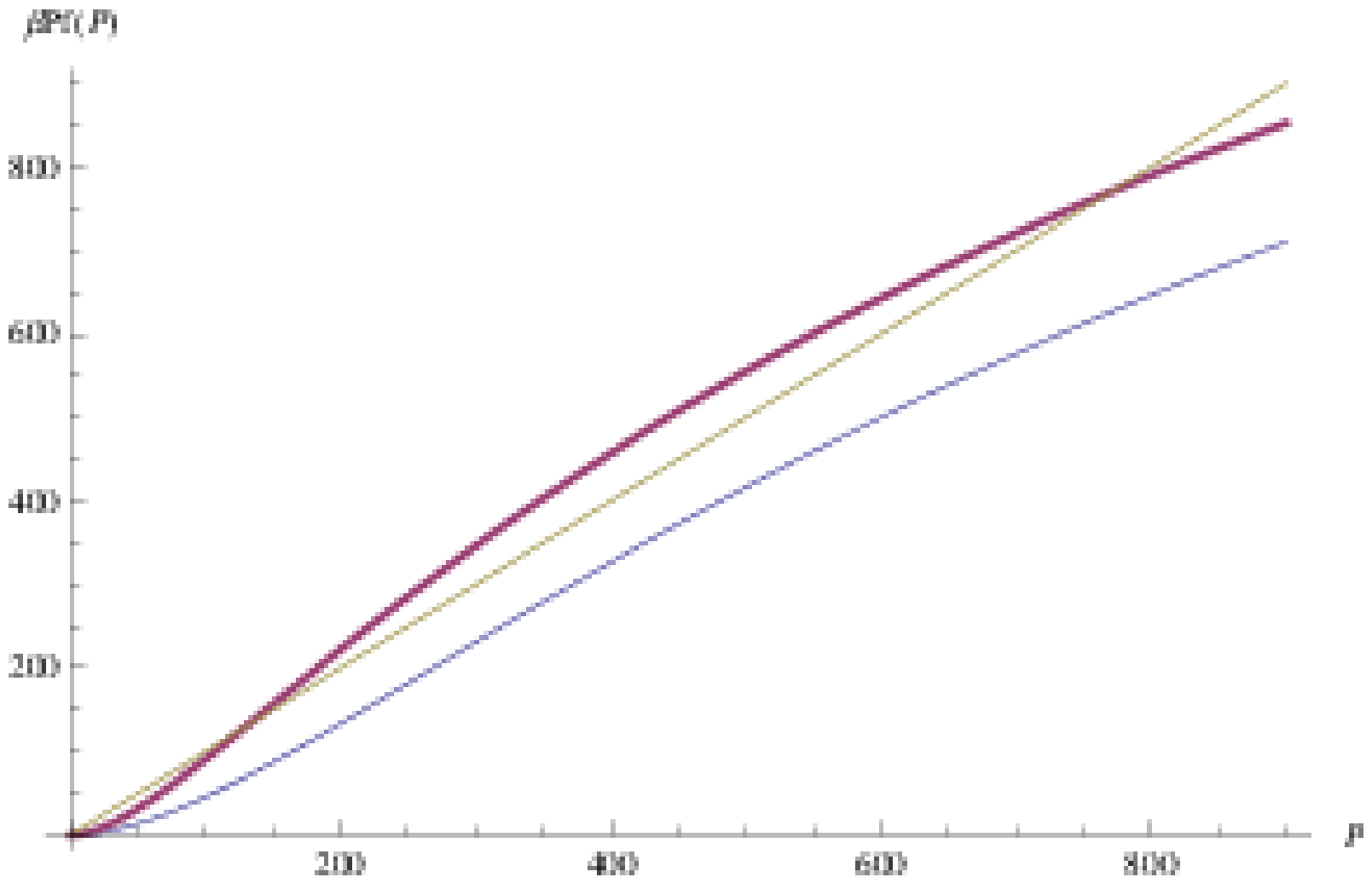


Figure: The graphs of $\beta P f(P)$, as in (60),for $\beta = 2, a = 0.001$ with $S = 300$ (blue line) and $S = 100$ (red line).

Summarizing, to eradicate the pest population we should introduce the number $S > S_{\max}$ of sterile insects. Then the population will converge to the extinction equilibrium provided the number S of sterile insects is kept above $S_{\max}$ at each cycle (this may require our intervention as the insects die of natural causes). Otherwise, we need to drive the population of pest below the smaller equilibrium – then the population will also converge to the extinction equilibrium.

Introducing structure

Projection matrices. In most applications the assumption that the analysed population consists of undistinguishable individuals is unrealistic – for instance the birth rate significantly depends on the age. Thus, we divide the population into subpopulations with regards to some feature of interest. For instance, we can consider clusters of cells divided into classes with respect to their size, cancer cells divided into classes on the basis of the number of copies of a particular gene responsible for its drug resistance, or a population divided into subpopulations depending on the geographical patch they occupy in a particular moment of time.

Let us suppose we have n states.

The state of the population at time k is described by a vector

$$\boldsymbol{v}(k) = (v_1(k), \ldots, v_n(k)),$$

where v_i gives the number of elements in the state i, $i = 1, \ldots, n$.

As before, we describe the changes of such a structured population from one generation to another by considering what can happen between subsequent censuses.

Each individual in a given state j contributes on average to, say, a_{ij} individuals in state j.

Typically, this occurs due to a state-j individual:

- migrating to i-th subpopulation with probability p_{ij};
- contributing to the birth of an individual in the i-th subpopulation at the rate b_{ij};
- dying with probability d_j,

other choices and interpretations are, however, also possible.

Then

$$\boldsymbol{v}(k+1) = \mathscr{A}\,\boldsymbol{v}(k), \tag{62}$$

where, with $a_{ij} = p_{ij} + b_{ij}$ for $i \neq j$ and $a_{ii} = p_{ii} + b_i - d_i$,

$$\mathscr{A} := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2\,n-1} & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n\,n-1} & a_{nn} \end{pmatrix} \tag{63}$$

so that

$$\boldsymbol{v}(k) = \mathscr{A}^k \boldsymbol{v}_0,$$

where $\boldsymbol{v}_0$ is the initial distribution.

Example. Any chromosome ends with a *telomer* which protects it against damage during the DNA replication process. Recurring divisions of cells can shorten the length of telomers and this process is considered to be responsible for cell's aging. If telomer is too short, the cell cannot divide which explains why many cell types can undergo only a finite number of divisions. Let us consider a simplified model of telomer shortening. The length of a telomer is a natural number from 0 to n, so cells with telomer of length i are in subpopulation i. A cell from subpopulation i can die with probability μ_i and divide (into 2 daughters).

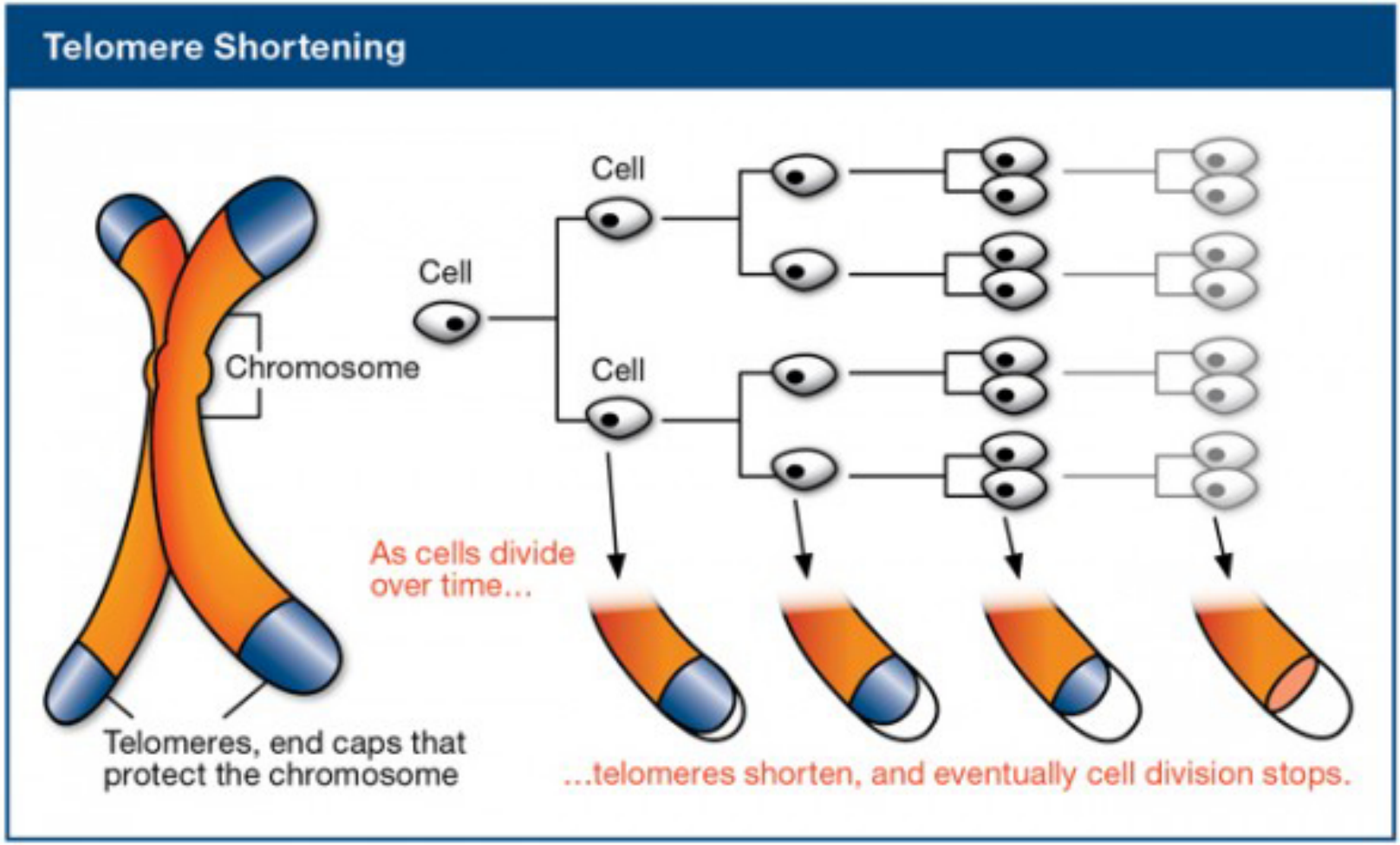


Figure: The process of telomere shortening.

Any daughter can have a telomer of length i with probability a_i and of length $i-1$ with probability $1-a_i$. Cells of 0 length telomer cannot divide and thus will die some time later. To find coefficients of the transition matrix, we see that the average production of offspring with telomer of length i by a parent of the same class is

$$2a_i^2 + 2a_i(1-a_i) = 2a_i,$$

(2 daughters with telomer of length i produced with probability a_i^2 and 1 daughter with telomer of length $i-1$ produced with probability $2a_i(1-a_i)$). Similarly, average production of daughters with length $i-1$ telomer is $2(1-a_i)$. However, to have offspring, the cell must have survived from one census to another, which happens with probability $1-\mu_i$. Hence, defining $r_i = 2a_i(1-\mu_i)$ and $d_i = 2(1-a_i)(1-\mu_i)$, we have

$$\mathscr{A} := \begin{pmatrix} 1-\mu_0 & d_1 & 0 & \cdots & 0 \\ 0 & r_1 & d_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & r_n \end{pmatrix}, \tag{64}$$

The model can be modified to make it closer to reality by allowing, for instance, for shortening of telomers by different lengthes or consider models with more telomers in a cell and with probabilities depending on the length of all of them.

Markov matrices. A particular version of (63) is obtained when we model a population that has constant size. Then $b_{ij} = d_j = 0$ for any $1 \leq i,j \leq n$ and thus $a_{ij} = p_{ij}$ is the fraction of j-th subpopulation which, on average, moves to the i-th subpopulation in one unit of time. Then $0 \leq p_{ij} \leq 1$ and

$$\sum_{1 \leq i \leq n} p_{ij} = 1,$$

or

$$p_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^{n} p_{ij}, \quad i = 1, \ldots, n. \tag{65}$$

Matrices of this form are called *Markov matrices*.

We can check that this condition ensures that the size of the population is constant. Indeed, the size of the population at time k is $P(k) = v_1(k) + \ldots + v_n(k)$ so that

$$\begin{aligned} P(k+1) &= \sum_{1\leq i\leq n} v_i(k+1) = \sum_{1\leq i\leq n} \left(\sum_{1\leq j\leq n} p_{ij} v_j(k) \right) \\ &= \sum_{1\leq j\leq n} v_j(k) \left(\sum_{1\leq i\leq n} p_{ij} \right) = \sum_{1\leq j\leq n} v_j(k) = N(k). \end{aligned}$$

Leslie matrices Assume that we have a population in which individuals only differ from each other by age: the population is divided into age groups with the age of an individual in group i is exactly i (in the chosen time units). Apart from this, we adopt all other assumption of unstructured mode. In particular, we only track females and that the census is taken immediately after the reproductive period (the length of which is negligible). Further, assume that there is an oldest age class $n-1$ and no individual can stay in an age class for more than one time period.

Finally, we introduce the s_i as the probability of survival from age i to age $i+1$ and the age dependent birth rate m_i. Thus, say in the kth breeding season, we have $v_i(k)$ individuals of age i, s_i of them survives to the $k+1$th breeding season; that is, to the age $i+1$, producing on average

$$f_i v_i(k) := s_i m_{i+1} v_i(k)$$

offspring. In this case, the evolution of the population can be described by the difference system

$$\boldsymbol{v}(k+1) = \mathscr{L}\boldsymbol{v}(k),$$

where $\mathscr{L}$ is the $n \times n$ matrix

$$\mathscr{L} := \begin{pmatrix} f_0 & f_1 & \cdots & f_{n-2} & f_{n-1} \\ s_0 & 0 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-2} & 0 \end{pmatrix}. \tag{66}$$

The matrix of the form (66) is referred to as a *Leslie matrix*.

A generalization of the Leslie matrix can be obtained by assuming that a fraction τ_i of i-th population stays in the same population. This gives the matrix

$$\mathscr{L} := \begin{pmatrix} f_0 + \tau_0 & f_1 & \cdots & f_{n-2} & f_{n-1} \\ s_0 & \tau_1 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-2} & \tau_{n-1} \end{pmatrix}. \tag{67}$$

Such matrices are called *Usher matrices*

In most cases $f_i \neq 0$ only if $\alpha \leq i \leq \beta$ where $[\alpha, \beta]$ is the fertile period. For example, for a typical mammal population we have three stages: immature (pre-breeding), breeding and post-breeding. If we perform census every year, then naturally a fraction of each class remains in the same class. Thus, the transition matrix in this case is given by

$$\mathscr{L} := \begin{pmatrix} \tau_0 & f_1 & 0 \\ s_0 & \tau_1 & 0 \\ 0 & s_1 & \tau_2 \end{pmatrix}. \tag{68}$$

On the other hand, in many insect populations, reproduction occurs only in the final stage of life and in such a case $f_i = 0$ unless $i = n$.

Long time behaviour of structured population models.

The main interest in population theory is to determine the long time structure of the population.

Before we embark on mathematical analysis, let us consider two numerical examples which indicate what we could expect from the models.

Example. Let us consider a population divided into three classes, evolution of which is modelled by the Leslie matrix

$$\mathscr{L} = \begin{pmatrix} 2 & 1 & 1 \\ 0.5 & 0 & 0 \\ 0 & 0.4 & 0 \end{pmatrix},$$

so that the population $\boldsymbol{v} = (v_1, v_2, v_3)$ evolves according to

$$\boldsymbol{v}(k+1) = \mathscr{L}\boldsymbol{v}(k), \quad k = 0,1,2\ldots,$$

or

$$\boldsymbol{v}(k) = \mathscr{L}^k \overset{\circ}{\boldsymbol{v}},$$

where $\overset{\circ}{\boldsymbol{v}}$ is an initial distribution of the population.

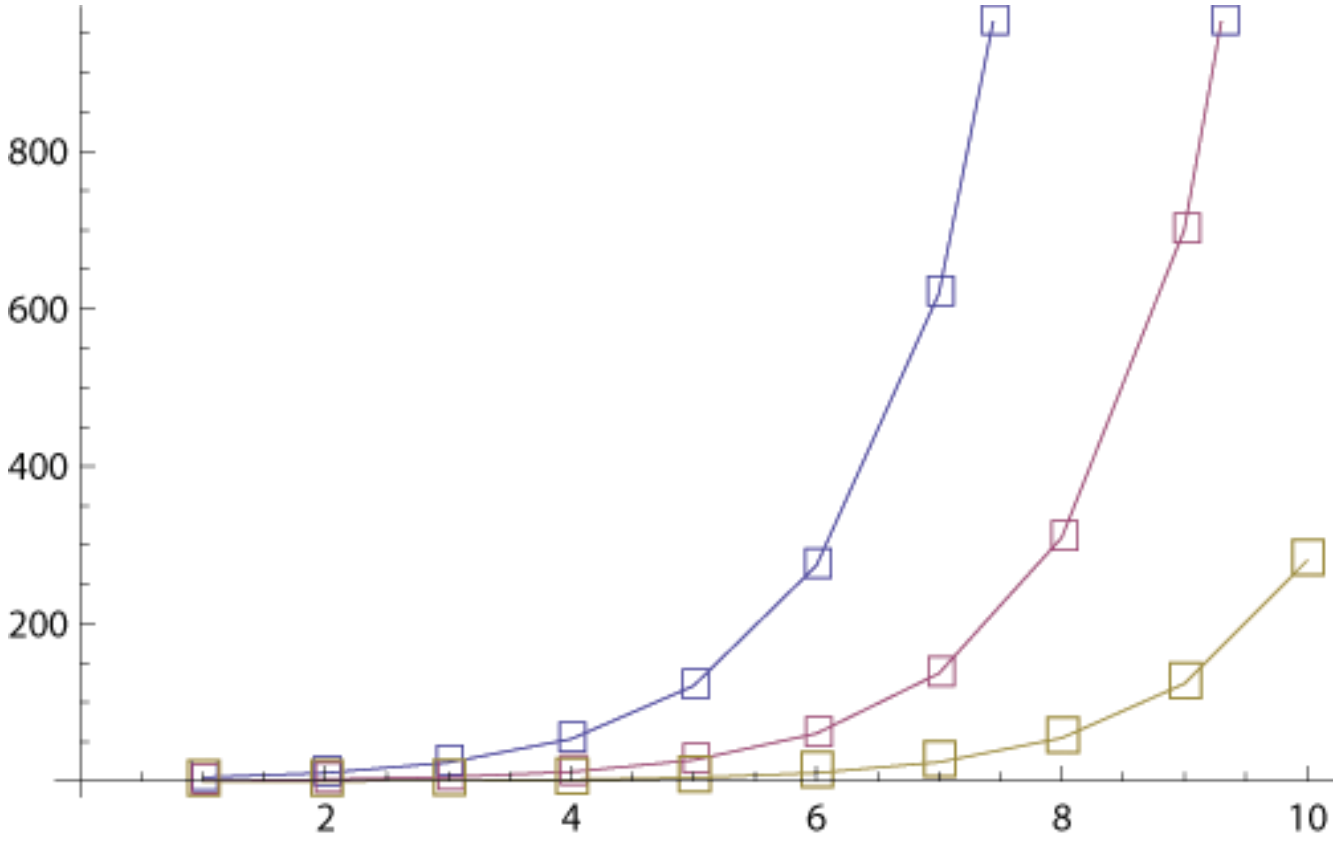


Figure: Evolution of $v_1(k)$ (blue), $v_2(k)$ (magenta) and $v_3(k)$ (brown) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (1,0,3)$ and $k = 1,\ldots,20$.

We may observe that each component grows very fast with k. However, if we compare growth of $v_1(k)$ with $v_2(k)$ and of $v_2(k)$ with $v_3(k)$ (next picture) we see that the ratios stabilize quickly around 4.5 in the first case and around 5.62 in the second case. This suggests that there is a scalar function $f(k)$ and a vector $\boldsymbol{e} = (e_1, e_2, e_3) = (25.29, 5.62, 1)$ such that for large k

$$\boldsymbol{v}(k) \approx f(k)\boldsymbol{e}. \tag{69}$$

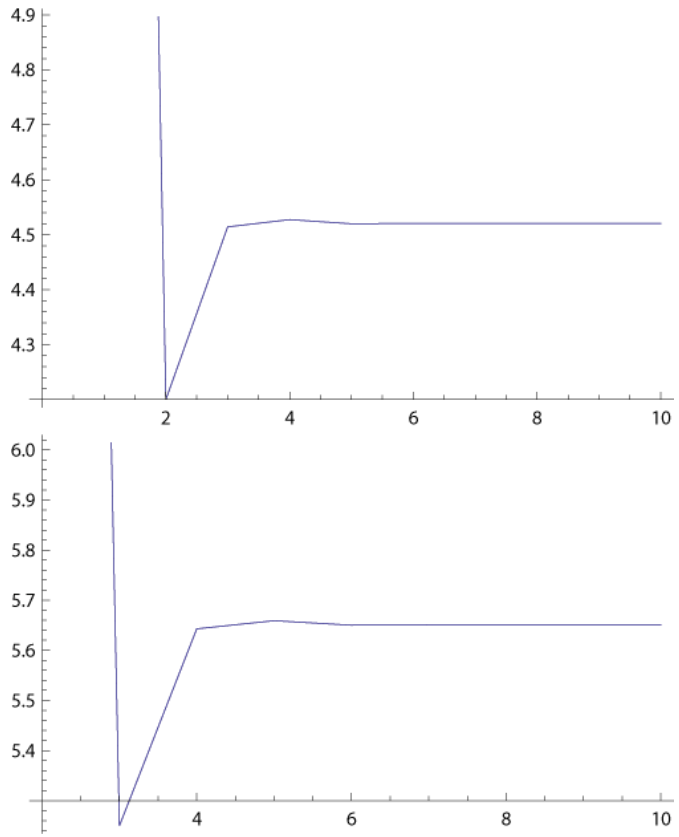


Figure: Evolution of $v_1(k)/v_2(k)$ (top) and $v_2(k)/v_3(k)$ (bottom) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (1,0,3)$ and $k = 1,\ldots,20$.

Let us consider another initial condition, say, $\overset{\circ}{v} = (2,1,4)$:

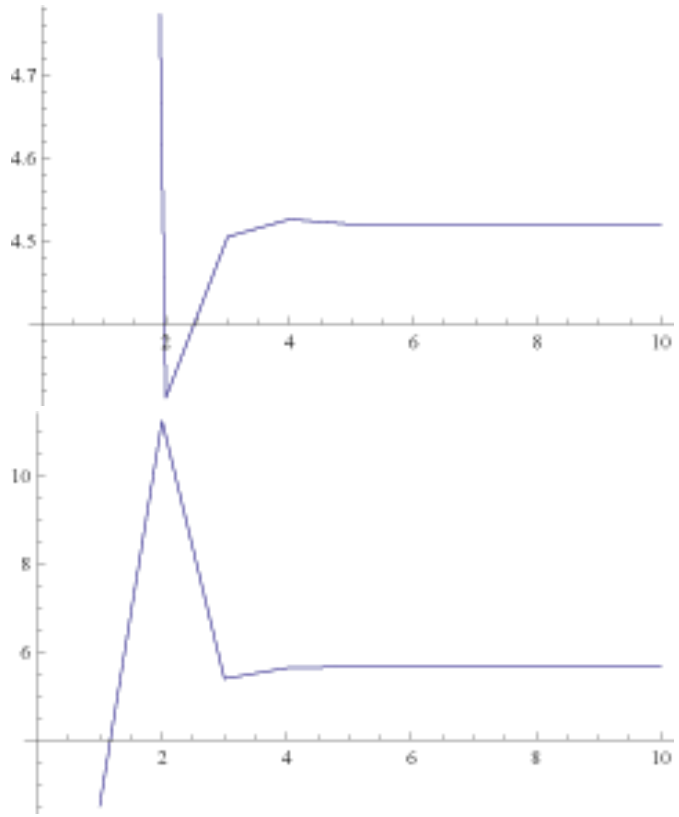


Figure: Evolution of $v_1(k)/v_2(k)$ (top) and $v_2(k)/v_3(k)$ (bottom) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (2,1,4)$.

It turns out that the ratios stabilize at the same level which further suggest that $\boldsymbol{e}$ does not depend on the initial condition so that (69) can be refined to

$$\boldsymbol{v}(k) \approx f_1(k) g(\overset{\circ}{\boldsymbol{v}}) \boldsymbol{e}, \quad k \to \infty \tag{70}$$

where g is a linear function.

Anticipating the development of the theory, it can be proved that $f_1(k) = \lambda^k$ where λ is the largest eigenvalue of $\mathscr{L}$, $\boldsymbol{e}$ is the eigenvector corresponding to λ and $g(\boldsymbol{x}) = \boldsymbol{g} \cdot \boldsymbol{x}$ with $\boldsymbol{g}$ being the eigenvector of the transpose matrix corresponding to λ. In our case, $\lambda \approx 2.26035$ and the ratios $v_i(k)/\lambda^k$ stabilize as expected from numerical simulations.

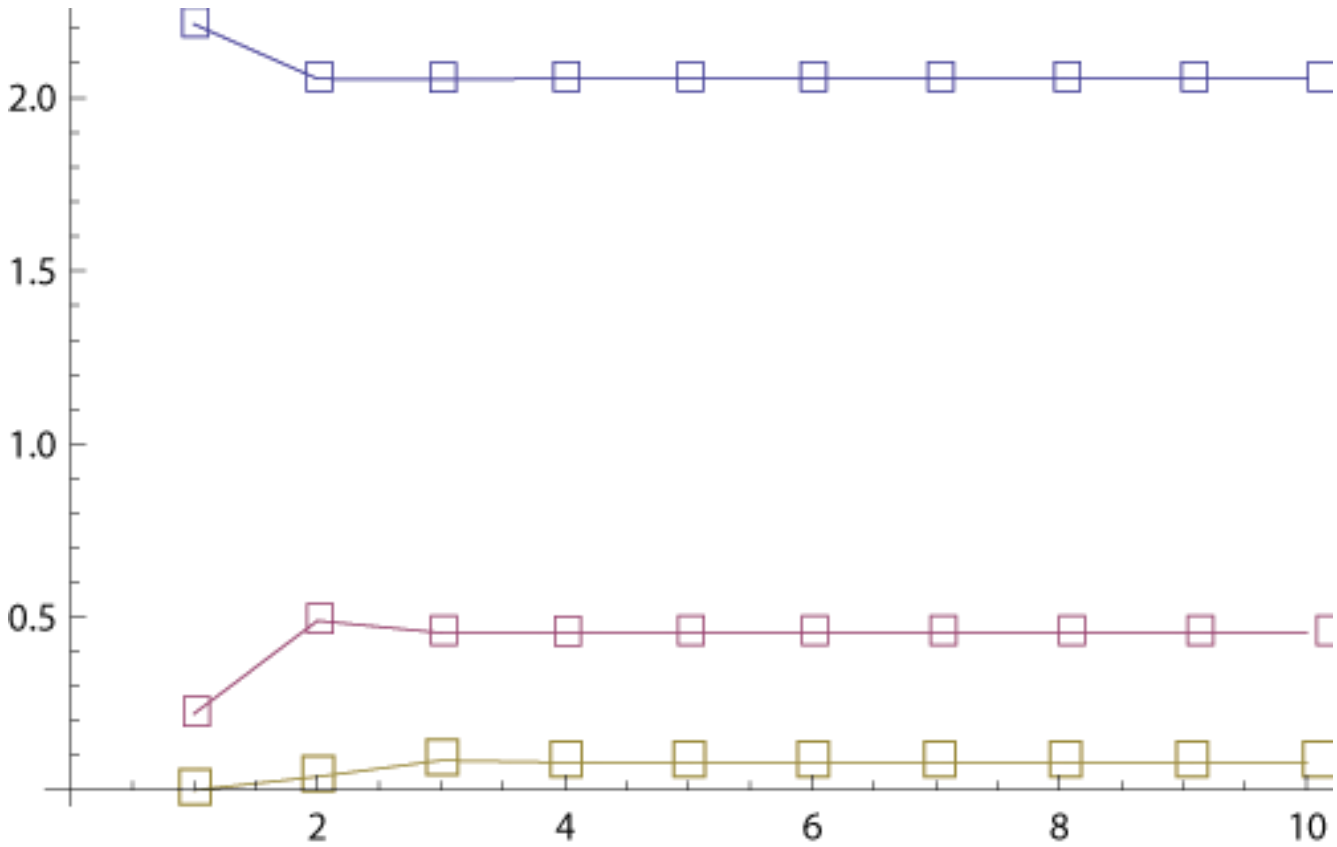


Figure: Evolution of $v_1(k)/\lambda^k$) (blue), $v_2(k)/\lambda^k$ (magenta)and $v_3(k)/\lambda^k$ (brown) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (1,0,3)$ and $k = 1,\ldots,20$.

The situation in which the structure of the population after long time does not depend on the initial condition but only on the intrinsic properties of the model (here the leading eigenvalue) is called the *asynchronous exponential growth (AEG)* property.

Unfortunately, not all Leslie matrices enjoy this property.

Example. Consider a Leslie matrix given by

$$\mathscr{L} = \begin{pmatrix} 0 & 0 & 3 \\ 0.5 & 0 & 0 \\ 0 & 0.4 & 0 \end{pmatrix}$$

and $\overset{\circ}{\boldsymbol{y}} = (2,3,4)$.

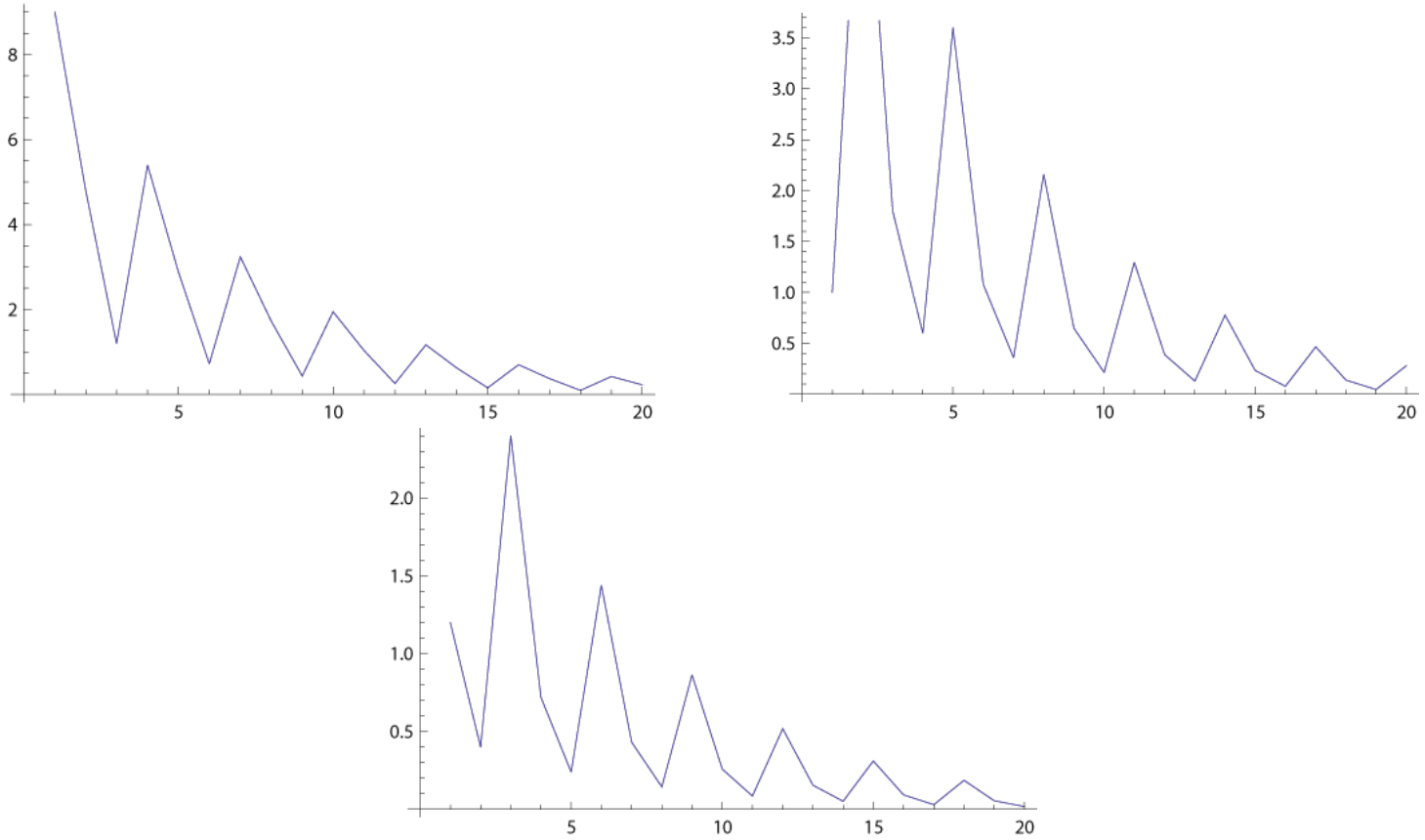


Figure: Evolution of $y_1(k)$ (top) and $y_2(k)$ (middle) and $y_3(k)$ (bottom) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (2,3,4)$ and $k = 1,\ldots,10$.

The picture is completely different from that obtained in the previous example. We observe some pattern but the ratios do not tend to a fixed limit but oscillate, as shown on the next figure.

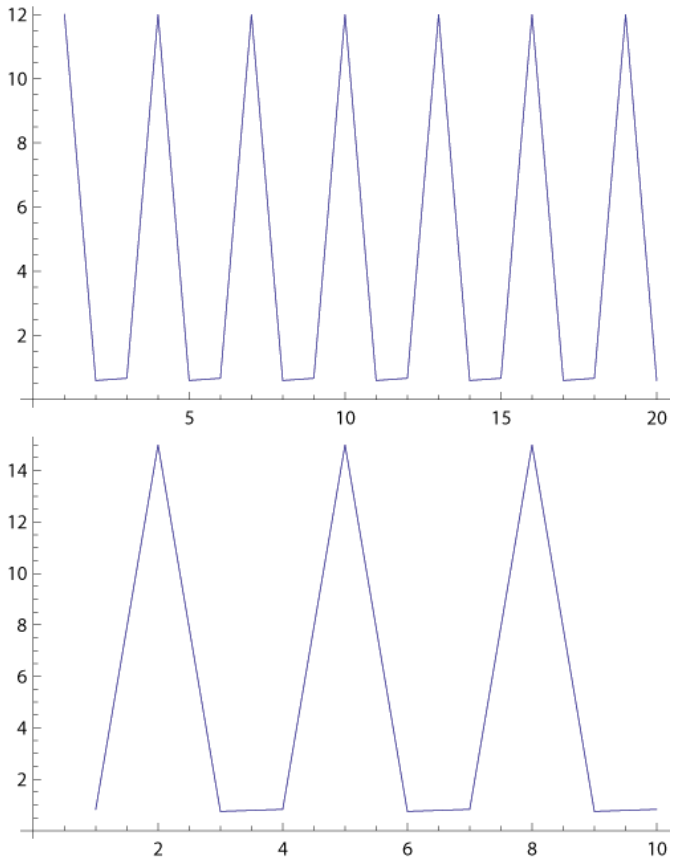


Figure: Evolution of $y_1(k)/y_2(k)$ (top) and $y_2(k)/y_3(k)$ (bottom) for the initial distribution $\overset{\circ}{\boldsymbol{v}} = (2,3,4)$ and $k = 1,\ldots,20$.

This can be explained by the spectral decomposition that is explained below in detail. The eigenvalues of $\mathscr{L}$ are given by $\lambda_1 = 0.843433, \lambda_2 = -0.421716 + 0.730434i, \lambda_2 = -0.421716 - 0.730434i$ and we can check that $|\lambda_1| = |\lambda_2| = |\lambda_3| = 0.843433$ and thus we do not have the dominant eigenvalue. The question we will try to answer in the next chapter is what features of the population are responsible for such behaviour.

Spectral decomposition of a matrix and the solution of matrix recurrences. To explain and to be able to predict similar behaviours in population models, first we discuss basic facts concerning eigenvalues and eigenvectors of a matrix.

We are interested in solving

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k), \qquad \boldsymbol{x}(0) = \mathring{\boldsymbol{x}} \tag{71}$$

where $\mathscr{A}$ is an $n \times n$ matrix $\mathscr{A} = \{a_{ij}\}_{1 \le i,j \le n}$ and $\boldsymbol{x}(k) = (x_1(k), \ldots, x_n(k))$.

It is obvious, by induction, to see that the solution to (71) is given by

$$\boldsymbol{x}(k) = \mathscr{A}^k \mathring{\boldsymbol{x}}, \quad k = 1, 2, \ldots. \tag{72}$$

The problem with (72) is that it is rather difficult to give an explicit form of $\mathscr{A}^k$.

Since $\mathbb{R}^n$ is n-dimensional, it is enough to find n linearly independent vectors $\boldsymbol{v}^{\boldsymbol{i}}$, $i = 1, \ldots, n$, for which $\mathscr{A}^k \boldsymbol{v}^{\boldsymbol{i}}$ can be easily evaluated.

Assume for a moment that such vectors have been found. Then, for arbitrary $\mathring{\boldsymbol{x}} \in \mathbb{R}^n$ we can find constants $c_1, \ldots, c_n$ such that

$$\mathring{\boldsymbol{x}} = c_1 \boldsymbol{v}^{\mathbf{1}} + \ldots + c_n \boldsymbol{v}^{\boldsymbol{n}}.$$

Precisely, let $\mathscr{V}$ be the matrix having the vectors $\boldsymbol{v}^{\boldsymbol{i}}$ as its columns

$$\mathscr{V} = \begin{pmatrix} | & \dots & | \\ \boldsymbol{v^1} & \dots & \boldsymbol{v^n} \\ | & \dots & | \end{pmatrix}. \tag{73}$$

Note, that $\mathscr{V}$ is invertible as the vectors $\boldsymbol{v^i}$ are linearly independent. Denoting $\boldsymbol{c} = (c_1, \dots, c_n)$, we obtain

$$\boldsymbol{c} = \mathscr{V}^{-1} \mathring{\boldsymbol{x}}. \tag{74}$$

Thus, for an arbitrary $\mathring{\boldsymbol{x}}$ we have

$$\mathscr{A}^n \mathring{\boldsymbol{x}} = \mathscr{A}^n(c_1 \boldsymbol{v^1} + \ldots + c_2 \boldsymbol{v^n}) = c_1 \mathscr{A}^n \boldsymbol{v^1} + \ldots + c_k \mathscr{A}^n \boldsymbol{v^n}. \quad (75)$$

Now, if

$$\mathscr{A}_k = \begin{pmatrix} | & \ldots & | \\ \mathscr{A}^k \boldsymbol{v^1} & \ldots & \mathscr{A}^k \boldsymbol{v^n} \\ | & \ldots & | \end{pmatrix}.$$

the matrix whose columns are vectors $\mathscr{A}^k \boldsymbol{v^1}, \ldots, \mathscr{A}^k \boldsymbol{v^n}$, then

$$\mathscr{A}^k \mathring{\boldsymbol{x}} = \mathscr{A}_k \boldsymbol{c} = \mathscr{A}_k \mathscr{V}^{-1} \mathring{\boldsymbol{x}}. \quad (76)$$

Hence, the problem is to find linearly independent vectors $\boldsymbol{v}^{\boldsymbol{i}}$, $i = 1, \ldots, k$, of which the powers of $\mathscr{A}$ can be easily evaluated. We shall use eigenvalues and eigenvectors for this purpose. Firstly, note that if $\boldsymbol{v}^{\mathbf{1}}$ is an eigenvector of $\mathscr{A}$ corresponding to an eigenvalue λ_1, that is, $\mathscr{A}\,\boldsymbol{v}^{\mathbf{1}} = \lambda_1 \boldsymbol{v}^{\mathbf{1}}$, then by induction

$$\mathscr{A}^k \boldsymbol{v}^{\mathbf{1}} = \lambda_1^k \boldsymbol{v}^{\mathbf{1}}.$$

Therefore, if we have n linearly independent eigenvectors $\boldsymbol{v}^{\mathbf{1}}, \ldots, \boldsymbol{v}^{\boldsymbol{n}}$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct), then from (75) we obtain

$$\mathscr{A}^k \mathring{\boldsymbol{x}} = c_1 \lambda_1^k \boldsymbol{v}^{\mathbf{1}} + \ldots + c_n \lambda_n^k \boldsymbol{v}^{\boldsymbol{n}}.$$

with $c_1, \ldots, c_n$ given by (74), or

$$\mathscr{A}^k \mathring{\boldsymbol{x}} = \begin{pmatrix} | & \ldots & | \\ \lambda_1^k \boldsymbol{v}^{\mathbf{1}} & \ldots & \lambda_n^k \boldsymbol{v}^{\boldsymbol{n}} \\ | & \ldots & | \end{pmatrix} \mathscr{V}^{-1} \mathring{\boldsymbol{x}} \tag{77}$$

Eigenvalues, eigenvectors and associated eigenvectors. Let $\mathscr{A}$ be an $n \times n$ matrix. We say that a number λ (real or complex) is an *eigenvalue* of $\mathscr{A}$ is there exist a non-zero solution of the equation

$$\mathscr{A}\,\boldsymbol{v} = \lambda\,\boldsymbol{v}. \tag{78}$$

Such a solution is called an *eigenvector* of $\mathscr{A}$. The set of eigenvalues of $\mathscr{A}$ is called the spectrum of $\mathscr{A}$ and is denoted by $\sigma(\mathscr{A})$. Eq. (78) is equivalent to the homogeneous system $(\mathscr{A} - \lambda\mathscr{I})\boldsymbol{v} = \boldsymbol{0}$, where $\mathscr{I}$ is the identity matrix.

Therefore λ is an eigenvalue of $\mathscr{A}$ if and only if the determinant of $\mathscr{A}$ satisfies

$$p_{\mathscr{A}}(\lambda) = det(\mathscr{A} - \lambda \mathscr{I}) = \begin{vmatrix} a_{11} - \lambda & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} - \lambda \end{vmatrix} = 0. \quad (79)$$

Evaluating the determinant we obtain a polynomial in λ of degree n. This polynomial is also called the *characteristic polynomial* of the matrix $\mathscr{A}$.

From algebra we know that there are exactly n, possibly complex, roots of $p_{\mathscr{A}}(\lambda)$ and, in general,

$$p_{\mathscr{A}}(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdot \ldots \cdot (\lambda_k - \lambda)^{n_k}, \tag{80}$$

with $n_1 + \ldots + n_k = n$. Since the coefficients of the polynomial are real, then complex roots appear always in conjugate pairs; that is, if $\lambda_j = \xi_j + i\omega_j$ is a characteristic root, then so is $\bar{\lambda}_j = \xi_j - i\omega_j$. Thus, eigenvalues are the roots of the characteristic polynomial of $\mathscr{A}$.

The exponent n_i appearing in the factorization (80) is called the *algebraic multiplicity* of λ_i. For each eigenvalue λ_i there corresponds an eigenvector $\boldsymbol{v^i}$ and eigenvectors corresponding to distinct eigenvalues are linearly independent. The set of all eigenvectors corresponding to λ_i spans a subspace, called the *eigenspace* corresponding to λ_i which we will denote by $\tilde{E}_{\lambda_i}$. The dimension of $\tilde{E}_{\lambda_i}$ is called the *geometric multiplicity* of λ_i. In general, algebraic and geometric multiplicities are different with geometric multiplicity being at most equal to the algebraic one.

Thus, in particular, if λ_i is a single root of the characteristic polynomial, then the eigenspace corresponding to λ_i is one-dimensional.

If the geometric multiplicities of eigenvalues add up to n; that is, if we have n linearly independent eigenvectors, then these eigenvectors form a basis for $\mathbb{R}^n$. In particular, this happens if all eigenvalues are single roots of the characteristic polynomial.

If this is not the case, then we do not have sufficiently many eigenvectors to span $\mathbb{R}^n$. A procedure that can be employed to find the 'missing' vectors here is to find solutions to equations of the form $(\mathscr{A} - \lambda_i \mathscr{I})^k \boldsymbol{v} = 0$ for $1 < k \leq n_i$, where n_i is the algebraic multiplicity of λ_i. Precisely speaking, if λ_i has algebraic multiplicity n_i and if

$$(\mathscr{A} - \lambda_i \mathscr{I})\boldsymbol{v} = 0$$

has only $\nu_i < n_i$ linearly independent solutions, then we consider the equation

$$(\mathscr{A} - \lambda_i \mathscr{I})^2 \boldsymbol{v} = 0.$$

Clearly all eigenvectors solve the latter equation but there is at least one more independent solution so that we have at least $\nu_i + 1$ independent vectors (note that these new vectors are no longer eigenvectors). If the number of independent solutions is still less than n_i, then we consider

$$(\mathscr{A} - \lambda_i \mathscr{I})^3 \boldsymbol{v} = 0,$$

and so on, till we get a sufficient number of them.

Note, that to make sure that in the step j we select solutions that are independent of the solutions obtained in step $j-1$, it is enough to find solutions to $(\mathscr{A}-\lambda_i\mathscr{I})^j\boldsymbol{v}=0$ that satisfy $(\mathscr{A}-\lambda_i\mathscr{I})^{j-1}\boldsymbol{v}\neq 0$.

Vectors $\boldsymbol{v}$ obtained in this way for a given λ_i are called *generalized* or *associated eigenvectors* corresponding to λ_i and they span an n_i dimensional subspace called a *generalized* or *associated eigenspace* corresponding to λ_i, denoted hereafter by E_{λ_i}.

Back to the difference system. Now we show how to apply the concepts discussed above to solve systems of difference equations. Let us return to the system

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k), \quad \boldsymbol{y}(0) = \overset{\circ}{\boldsymbol{y}}\,.$$

As discussed, we need to find formulae for $\mathscr{A}^k\boldsymbol{v}$ for a selected n linearly independent vectors $\boldsymbol{v^1}, \ldots, \boldsymbol{v^n}$.

Let us take as $\boldsymbol{v}$ the collection of all eigenvectors and associated eigenvectors of $\mathscr{A}$. We know that if $\boldsymbol{v^i}$ is an eigenvector associated to an eigenvalue λ^i, then $\mathscr{A}^k \boldsymbol{v^i} = \lambda_i^k \boldsymbol{v^i}$. Thus, the question is whether $\mathscr{A}^k$ can be effectively evaluated on associated eigenvectors. Let $\boldsymbol{v^j}$ be an associated eigenvector found as a solution to $(\mathscr{A} - \lambda_i \mathscr{I})^j \boldsymbol{v^j} = \boldsymbol{0}$ with $j \leq n_i$. Then, using the binomial expansion, we find

$$\begin{aligned}
\mathscr{A}^k \boldsymbol{v^j} &= (\lambda_i \mathscr{I} + \mathscr{A} - \lambda_i \mathscr{I})^k \boldsymbol{v^j} = \sum_{r=0}^{k} \lambda_i^{k-r} \begin{pmatrix} k \\ r \end{pmatrix} (\mathscr{A} - \lambda_i \mathscr{I})^r \boldsymbol{v^j} \\
&= \Big(\lambda_i^k \mathscr{I} + k\lambda_i^{k-1}(\mathscr{A} - \lambda_i \mathscr{I}) + \dots \\
&\quad + \frac{k!}{(j-1)!(k-j+1)!} \lambda_i^{k-j+1} (\mathscr{A} - \lambda_i \mathscr{I})^{j-1} \Big) \boldsymbol{v^j}, \qquad (81)
\end{aligned}$$

where

$$\begin{pmatrix} k \\ r \end{pmatrix} = \frac{k!}{r!(k-r)!} = \frac{1}{r!} k(k-1) \cdot \ldots \cdot (k-r+1)$$

is the Newton symbol; it is is a polynomial in k of order smaller than j.

It is important to note that (81) is a finite sum for any k; it always terminates at most at the term $(\mathscr{A}-\lambda_1\mathscr{I})^{n_i-1}$, where n_i is the algebraic multiplicity of λ_i. Thus

$$\begin{aligned}
\mathscr{A}^k\boldsymbol{v^j} &= \lambda_i^k\sum_{r=0}^{k}\lambda_i^{-r}\begin{pmatrix}k\\ r\end{pmatrix}(\mathscr{A}-\lambda_i\mathscr{I})^r\boldsymbol{v^j}\\
&= \lambda_i^k\Big(\mathscr{I}+k\lambda_i^{-1}(\mathscr{A}-\lambda_i\mathscr{I})\boldsymbol{v^j}+\ldots\\
&\quad +\frac{k!}{(j-1)!(k-j+1)!}\lambda_i^{-j+1}(\mathscr{A}-\lambda_i\mathscr{I})^{j-1}\boldsymbol{v^j}\Big)\\
&= \lambda_i^k p_j(k,\boldsymbol{v}^{\lambda_i})
\end{aligned}$$

where p_j is a polynomial in k of order $j<n_i$.

Above we used the following result

Theorem

Each generalized eigenspace E_{λ_i} of $\mathscr{A}$ is invariant under $\mathscr{A}$; that is, for any $\boldsymbol{v} \in E_{\lambda_i}$ we have $\mathscr{A}\boldsymbol{v} \in E_{\lambda_i}$. It is also invariant under $\mathscr{A}^k, k = 1, 2, \ldots$.

This result suggests that the the evolution governed by $\mathscr{A}$ in both discrete and continuous case can be broken into several simpler and independent pieces occurring in each generalized eigenspace.

We will return to this idea after an example.

Example. Find $\mathscr{A}^k$ for

$$\mathscr{A} = \begin{pmatrix} 4 & 1 & 2 \\ 0 & 2 & -4 \\ 0 & 1 & 6 \end{pmatrix}.$$

We start with finding eigenvalues of $\mathscr{A}$:

$$p(\lambda) = \begin{vmatrix} 4-\lambda & 1 & 2 \\ 0 & 2-\lambda & -4 \\ 0 & 1 & 6-\lambda \end{vmatrix} = (4-\lambda)^3 = 0$$

gives the eigenvalue $\lambda = 4$ of algebraic multiplicity 3.

To find eigenvectors corresponding to $\lambda = 3$, we solve

$$(\mathscr{A} - 4\mathscr{I})\boldsymbol{v} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus, v_1 is arbitrary and $v_2 = -2v_3$ so that the eigenspace is two dimensional, spanned by

$$\boldsymbol{v^1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{v^2} = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}.$$

Therefore

$$\mathscr{A}^k \boldsymbol{v^1} = 4^k \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathscr{A}^k \boldsymbol{v^2} = 4^k \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}.$$

To find the last vector we consider

$$
\begin{aligned}
(\mathscr{A}-4\mathscr{I})^2\boldsymbol{v} &= \begin{pmatrix} 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}
$$

Any vector solves this equation so that we have to take a vector that is not an eigenvalue. Possibly the simplest choice is

$$\boldsymbol{v}^{\mathbf{3}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Thus, by (81)

$$\mathscr{A}^k \boldsymbol{v^3} = \left(4^k \mathscr{I} + k4^{k-1}(\mathscr{A} - 4\mathscr{I})\right) \boldsymbol{v^3}$$

$$= \left(\left(\begin{array}{ccc} 4^k & 0 & 0 \\ 0 & 4^k & 0 \\ 0 & 0 & 4^k \end{array} \right) + k4^{k-1} \left(\begin{array}{ccc} 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & 1 & 2 \end{array} \right) \right) \left(\begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right)$$

$$= \left(\begin{array}{c} 2k4^{k-1} \\ -k4^k \\ 4^k + 2k4^{k-1} \end{array} \right).$$

To find explicit expression for $\mathscr{A}^k$ we use (76). In our case

$$\mathscr{A}_k = \begin{pmatrix} | & | & | \\ \mathscr{A}^k \boldsymbol{v^1} & \mathscr{A}^k \boldsymbol{v^2} & \mathscr{A}^k \boldsymbol{v^3} \\ | & | & | \end{pmatrix} = \begin{pmatrix} 4^k & 0 & 2k4^{k-1} \\ 0 & -2\cdot 4^k & -k4^k \\ 0 & 4^k & 4^k + 2k4^{k-1} \end{pmatrix},$$

further

$$\mathscr{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \mathscr{V}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix}.$$

Therefore

$$\begin{aligned}
\mathscr{A}^k = \mathscr{A}_k \mathscr{V}^{-1} &= \begin{pmatrix} 4^k & k4^{k-1} & 2k4^{k-1} \\ 0 & 4^k - 2k4^{k-1} & -k4^k \\ 0 & k4^{k-1} & 4^k + 2k4^{k-1} \end{pmatrix} \\
&= 4^k \begin{pmatrix} 1 & \frac{k}{4} & \frac{k}{2} \\ 0 & 1 - \frac{k}{2} & -k \\ 0 & \frac{k}{4} & 1 + \frac{k}{2} \end{pmatrix} \\
&= 4^k \left(k \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & -\frac{1}{2} & -1 \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right).
\end{aligned}$$

On the basis of previous considerations we can write

$$\mathscr{A}^k \boldsymbol{x} = \sum_{\lambda \in \sigma(\mathscr{A})} \lambda^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}), \tag{82}$$

where $\boldsymbol{p}_\lambda$ are polynomials in k of degree strictly smaller than the algebraic multiplicity of λ, and with vector coefficients being linear combinations of eigenvectors and associated eigenvectors $\boldsymbol{v}^\lambda$ corresponding to λ and depending, in a linear way, on $\boldsymbol{x}$.

A special role is played by the spectral radius of $\mathscr{A}$ defined by

$$\rho(\mathscr{A}) = \max_{\lambda \in \sigma(\mathscr{A})} |\lambda|$$

and the peripheral spectrum

$$\sigma_p(\mathscr{A}) = \{\lambda \in \sigma(\mathscr{A});\ |\lambda| = \rho(\mathscr{A})\}.$$

It is clear that if $\rho(\mathscr{A}) < 1$, then $\mathscr{A}^k \to 0$ as $k \to \infty$. In what follows we assume $\rho(\mathscr{A}) \geq 1$ and re-write (82) as

$$\begin{aligned}\left(\frac{\mathscr{A}}{\rho(\mathscr{A})}\right)^k \boldsymbol{x} &= \sum_{\lambda \in \sigma_p(\mathscr{A})} \left(\frac{\lambda}{\rho(\mathscr{A})}\right)^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}) \\ &\quad + \sum_{\lambda \in \sigma(\mathscr{A}) \setminus \sigma_p(\mathscr{A})} \left(\frac{\lambda}{\rho(\mathscr{A})}\right)^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}),\end{aligned}$$

Since $\lambda/\rho(\mathscr{A}) < 1$ for $\lambda \in \sigma(\mathscr{A}) \setminus \sigma_p(\mathscr{A})$, we see that the long term behaviour of $\mathscr{A}^k$ is fully determined by the points on the peripheral spectrum.

Since $|\lambda/\rho(\mathscr{A})| = 1$ for $\lambda \in \sigma_p(\mathscr{A})$, $(\mathscr{A}/\rho(\mathscr{A}))^k$ will not have a limit as $k \to \infty$ if at least one $\lambda \in \sigma_p(\mathscr{A})$ has algebraic multiplicity greater than geometric multiplicity.

Otherwise, if all $\lambda \in \sigma_p(\mathscr{A})$ are semisimple, then

$$\left(\frac{\mathscr{A}}{\rho(\mathscr{A})}\right)^k \boldsymbol{x} = \sum_{\lambda \in \sigma_p(\mathscr{A})} \left(\frac{\lambda}{\rho(\mathscr{A})}\right)^k c_\lambda(\boldsymbol{x}) \boldsymbol{v}_\lambda + \sum_{\lambda \in \sigma(\mathscr{A}) \setminus \sigma_p(\mathscr{A})} \left(\frac{\lambda}{\rho(\mathscr{A})}\right)^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}),$$

where $\boldsymbol{v}_\lambda$ is an eigenvector corresponding to $\lambda \in \sigma(\mathscr{A})$.

We see that for $(\mathscr{A}/\rho(\mathscr{A}))^k$ to be convergent, we must have

$$\sigma_p(\mathscr{A}) = \{\rho(\mathscr{A})\}.$$

The eigenvalue, say $\lambda_1 = \rho(\mathscr{A})$, satisfying $\lambda_1 > |\lambda|$ for any other eigenvalue λ is called the *dominant* eigenvalue. Since λ_1 is simple, there corresponds to it a unique (up to a scalar multiple) eigenvector, say $\boldsymbol{v}_1$ so that we can informally write

$$\mathscr{A}^k \boldsymbol{x} \approx c_1(\boldsymbol{x})\lambda_1^k \boldsymbol{v_1}$$

for large k, provided $c_1(\boldsymbol{x}) \neq 0$. We observed this situation in the first numerical example.

If, on the other hand, $\sigma_p(\mathscr{A})$ consists of more than one element, asymptotically we have

$$\left(\frac{\mathscr{A}}{\rho(\mathscr{A})}\right)^k \boldsymbol{x} \approx \sum_{\lambda \in \sigma_p(\mathscr{A})} \left(\frac{\lambda}{\rho(\mathscr{A})}\right)^k c_\lambda(\boldsymbol{x}) \boldsymbol{v}_\lambda$$

and we see that the solution keeps oscillating, as in the second example.

Finding c_1 from the definition requires knowing all eigenvectors and associated eigenvectors of $\mathscr{A}$ and thus is not particularly handy. Here we shall describe a simpler method.

Let us recall that the transposed matrix $\mathscr{A}^T$ satisfies

$$< \mathscr{A}^T \boldsymbol{x}^*, \boldsymbol{y} > = < \boldsymbol{x}^*, \mathscr{A} \boldsymbol{y} >$$

where $< \boldsymbol{x}^*, \boldsymbol{y} > = \boldsymbol{x}^* \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i^* y_i$.

Matrices $\mathscr{A}$ and $\mathscr{A}^T$ have the same eigenvalues and, though eigenvectors and associated eigenvectors are different (unless $\mathscr{A}$ is symmetric), the structure of the generalized eigenspaces corresponding to the same eigenvalue is identical (that is, the geometric multiplicities of λ are equal and we have the same number of associated eigenvectors solving $(\mathscr{A} - \lambda \mathscr{I})^\nu \boldsymbol{v} = 0$ and $(\mathscr{A}^* - \lambda \mathscr{I})^\nu \boldsymbol{v}^* = 0$). This follows from the fact that determinant, nullity and rank of a matrix and its transpose are the same.

Theorem

Let E_λ and $E^*_{\lambda^*}$ be generalized eigenspaces of, respectively, $\mathscr{A}$ and $\mathscr{A}^T$, corresponding to different eigenvalues: $\lambda \neq \lambda^*$. If $\boldsymbol{v}^* \in E^*_{\lambda^*}$ and $\boldsymbol{v} \in E_\lambda$, then

$$< \boldsymbol{v}^*, \boldsymbol{v} >= 0. \tag{83}$$

Thus, to determine a long time behaviour of a population described by the discrete system $\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k)$ we have to

1. Find eigenvalues of $\mathscr{A}$ and determine whether there is a dominant eigenvalue; that is, a simple real eigenvalue λ_1 satisfying $\lambda_1 > |\lambda|$ for any other λ;
2. If this is the case, find the eigenvectors $\boldsymbol{v}$ of $\mathscr{A}$ and $\boldsymbol{v}^*$ of $\mathscr{A}^T$ corresponding to λ_1 normalized so as $< \boldsymbol{v}^*, \boldsymbol{v} >= 1$.
3. The long time behaviour of the population is then described by

$$\mathscr{A}^k \boldsymbol{x} \approx \lambda_1^k < \boldsymbol{v}^*, \boldsymbol{x} > \boldsymbol{v} \tag{84}$$

for large k for any initial distribution of the population satisfying $< \boldsymbol{v}^*, \boldsymbol{x} > \neq 0$.

Example. Find the long term behaviour of the process described by

$$\boldsymbol{x}(k+1) = \begin{pmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{pmatrix} \boldsymbol{x}(k).$$

To obtain the eigenvalues we calculate the characteristic polynomial

$$
\begin{aligned}
p(\lambda) &= det(\mathscr{A}-\lambda\mathscr{I}) = \begin{vmatrix} 1-\lambda & -1 & 4 \\ 3 & 2-\lambda & -1 \\ 2 & 1 & -1-\lambda \end{vmatrix} \\
&= -(1+\lambda)(1-\lambda)(2-\lambda)+12+2-8(2-\lambda)+(1-\lambda)-3(1+\lambda) \\
&= -(1+\lambda)(1-\lambda)(2-\lambda)+4\lambda-4 = (1-\lambda)(\lambda-3)(\lambda+2),
\end{aligned}
$$

so that the eigenvalues of $\mathscr{A}$ are $\lambda_1 = 1$, $\lambda_2 = 3$ and $\lambda_3 = -2$. All the eigenvalues have algebraic multiplicity 1 so that they should give rise to 3 linearly independent eigenvectors.

(i) $\lambda_1 = 1$: we seek a nonzero vector $\boldsymbol{v}$ such that

$$(\mathscr{A} - 1\mathscr{I})\boldsymbol{v} = \begin{pmatrix} 0 & -1 & 4 \\ 3 & 1 & -1 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus

$$-v_2 + 4v_3 = 0, \qquad 3v_1 + v_2 - v_3 = 0, \qquad 2v_1 + v_2 - 2v_3 = 0$$

and we get $v_2 = 4v_3$ and $v_1 = -v_3$ from the first two equations and the third is automatically satisfied.

Thus we obtain the eigenspace corresponding to $\lambda_1 = 1$ containing all the vectors of the form

$$\boldsymbol{v^1} = C_1 \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$$

where C_1 is any constant.

(ii) $\lambda_2 = 3$: we seek a nonzero vector $\boldsymbol{v}$ such that

$$(\mathscr{A} - 3\mathscr{I})\boldsymbol{v} = \begin{pmatrix} -2 & -1 & 4 \\ 3 & -1 & -1 \\ 2 & 1 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hence

$$-2v_1 - v_2 + 4v_3 = 0, \qquad 3v_1 - v_2 - v_3 = 0, \qquad 2v_1 + v_2 - 4v_3 = 0.$$

Solving for v_1 and v_2 in terms of v_3 from the first two equations gives $v_1 = v_3$ and $v_2 = 2v_3$.

Consequently, vectors of the form

$$\boldsymbol{v}^{\mathbf{2}} = C_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

are eigenvectors corresponding to the eigenvalue $\lambda_2 = 3$.

(iii) $\lambda_3 = -2$: We have to solve

$$(\mathscr{A} + 2\mathscr{I})\boldsymbol{v} = \begin{pmatrix} 3 & -1 & 4 \\ 3 & 4 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Thus

$$3v_1 - v_2 + 4v_3 = 0, \qquad 3v_1 + 4v_2 - v_3 = 0, \qquad 2v_1 + v_2 + v_3 = 0.$$

Again, solving for v_1 and v_2 in terms of v_3 from the first two equations gives $v_1 = -v_3$ and $v_2 = v_3$ so that each vector

$$\boldsymbol{v^3} = C_3 \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to the eigenvalue $\lambda_3 = -2$.

The general solution is given by

$$\begin{aligned}
\boldsymbol{x}(k) &= C_1 1^k \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix} + C_2 3^k \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + C_3 (-2)^k \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \\
&= 3^k \left(C_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + C_1 \left(\frac{1}{3}\right)^k \begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix} + C_3 \left(\frac{-2}{3}\right)^k \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \right)
\end{aligned}$$

The dominant eigenvalue is $\lambda_2 = 3$ and for large time

$$\boldsymbol{x}(k) \approx 3^k C_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \tag{85}$$

where C_2 depends on the initial condition.

The transposed matrix is given by

$$\mathscr{A}^* = \begin{pmatrix} 1 & 3 & 2 \\ -1 & 2 & 1 \\ 4 & -1 & -1 \end{pmatrix}$$

and the eigenvector $\boldsymbol{v}^*$ corresponding to $\lambda = 3$ can be calculated by

$$(\mathscr{A}^* - 3\mathscr{I})\boldsymbol{v} = \begin{pmatrix} -2 & 3 & 2 \\ -1 & -1 & 1 \\ 4 & -1 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We get $v_2 = 0$ and $v_1 = v_3$. Thus,

$$\boldsymbol{v_2}^* = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

and we can check that, indeed, $< \boldsymbol{v_2}^*, \boldsymbol{v^1} > = < \boldsymbol{v_2}^*, \boldsymbol{v^2} > = 0$.

Then, multiplying (85) by $\boldsymbol{v_2}^*$ we obtain

$$< \boldsymbol{v_2}^*, \boldsymbol{x}(k) >= C_2 3^k < \boldsymbol{v_2}^*, \boldsymbol{v^2} >$$

and, taking $k = 0$ we have

$$< \boldsymbol{v_2}^*, \overset{\circ}{\boldsymbol{x}} >= C_2 < \boldsymbol{v_2}^*, \boldsymbol{v^2} >,$$

hence $C_2 = \frac{1}{2}\left(\overset{\circ}{x}_1 + \overset{\circ}{x}_3\right)$. Clearly, long time picture of evolution given by (85) will not be realized if $\overset{\circ}{\boldsymbol{x}}$ is orthogonal to $\boldsymbol{v_2}^*$.

Frobenius-Perron theorem. The question whether any matrix with nonnegative entries gives rise to such a behaviour and, if not, what models exhibit AEG, is much more delicate and requires invoking the Frobenius-Perron theorem which will be now discussed.

To make further progress, we have to formalize a number of statements made in the previous sections and, in particular, the meaning of the approximate equality (84). For this, we have to set the problem in an appropriate mathematical framework.

First, we note that in the context of population theory, if a given difference equation/system of equations is to describe evolution of a population; that is, if the solution is the population size or density, then clearly solutions emanating from non-negative data must stay non-negative. Thus we have to extend the notion of positivity to vectors. We say that a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ is non-negative (positive), if for all $i = 1, \ldots, n$, $x_i \geq 0$ ($x_i > 0$), denoted as $\boldsymbol{x} \geq 0$ ($\boldsymbol{x} > 0$).

Similarly, we say that the matrix $\mathscr{A} = \{a_{ij}\}_{1 \leq i,j \leq n}$ is non-negative (positive), and write $\mathscr{A} \geq 0 (\mathscr{A} > 0)$, if $a_{ij} \geq 0$ $(a_{ij} > 0)$ for all $i, j = 1, \ldots, n$.

It is easy to prove that

Proposition

The solution $\boldsymbol{x}(k)$ *of*

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k), \quad \boldsymbol{x}(0) = \overset{\circ}{\boldsymbol{x}}$$

satisfies $\boldsymbol{x}(k) \geq 0$ *for any* $k = 1, \ldots,$ *for arbitrary* $\overset{\circ}{\boldsymbol{x}} \geq 0$ *if and only if* $\mathscr{A} \geq 0$.

The sequence $(\mathscr{A}^k)_{k\geq 1}$ is a dynamical system in the state space $X = \mathbb{R}^n$ (and in X_+ if $\mathscr{A} \geq 0$). Essentially, (84) is a statement about the limit of $\mathscr{A}^k \overset{\circ}{\boldsymbol{x}}$ as $k \to \infty$ so we must introduce a metric structure on X. To make the metric consistent with the linear structure of $\mathbb{R}^n$, it is typically defined by a norm, that is, a functional $\|\cdot\| : X \to \mathbb{R}_+$ satisfying, for any $\boldsymbol{x}, \boldsymbol{y} \in X, \alpha \in \mathbb{R}$,

$$\|\boldsymbol{x}\| = 0 \text{ iff } \boldsymbol{x} = 0, \quad \|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|, \quad \|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|.$$

There is a variety norms in $\mathbb{R}^n$ (all defining the same topology), the most common being the Euclidean metric.

However, bearing in mind the interpretation of our problems; that is, that the solution vector of

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k),$$

$\boldsymbol{x}(k) = (x_1(k), \ldots, x_n(k))$, defines the distribution of a population among the states, we see that the most natural norm is

$$\|\boldsymbol{x}\| = \sum_{i=1}^{n} |x_i| \tag{86}$$

which, for $\boldsymbol{x} \geq 0$ simplifies to

$$\|\boldsymbol{x}\| = \sum_{i=1}^{n} x_i \tag{87}$$

which is the total population of the ensemble.

Classification of projection matrices. The long time behaviour of $(\mathscr{A}^k)_{k\geq 1}$ is fully determined by whether $\mathscr{A}$ is a primitive irreducible, imprimitive irreducible or a reducible matrix. These concepts best can be explained in terms of graphs.

A *graph* is a nonempty finite set of vertices and (possibly empty) set of edges (edge can be interpreted as a unordered pair of vertices). An directed graph or a *digraph* is a graph with directed edges (a directed edge is then an ordered pair of vertices).

A *path* is a finite sequence of directed edges $((i_1, i_2), (i_2, i_3), \ldots, (i_{k-1}, i_k))$ in which no vertex is repeated apart from possibly $i_1 = i_k$; in the latter case the path is called a cycle.

We say that a diagraph is strongly connected if there is a path between any two vertices of it.

By the *incidence matrix* associated to a nonnegative matrix $\mathscr{A}$ we understand the matrix $\mathscr{D} = (d_{ij})_{1 \leq i,j \leq n}$ where $d_{ij} = 1$ if $a_{ij} > 0$ and $d_{ij} = 0$ otherwise. There is a one to one correspondence between diagraphs and incidence matrices (up to a permutation). For a given $\mathscr{D}$ we take $\{1, \ldots, n\}$ as the set of vertices and we draw a directed from j to i whenever $d_{ij} > 0$. Conversely, given a diagraph with n vertices, we number them $\{1, \ldots, n\}$ and set $d_{ij} = 1$ whenever there is an edge from j to i.

We say that $\mathscr{A}$ is irreducible if and only if the digraph associated with the incidence matrix of $\mathscr{A}$ is strongly connected.

An equivalent, but more algebraic, condition must be preceded by some notation. We write

$$\mathscr{A}^k = (a_{i,j}^{(k)})_{1 \leq i,j \leq n}.$$

It is easy to see that

$$a_{i,j}^{(k)} = \sum_{1 \leq i_r \leq n, r=1,\ldots,k-1} a_{i,i_1} a_{i_1,i_2} \cdot \ldots \cdot a_{i_{k-1},j}$$

If some $a_{i,i_1}a_{i_1,i_2}\cdot\ldots\cdot a_{i_{k-1},j}\neq 0$ then there is a path starting from j and passing through $i_{k-1},\ldots,i_1$ to i. Since the matrix elements are nonnegative, for $a_{i,j}^{(k)}$ to be non-zero it is enough that there exists at least one such path. Thus, $\mathscr{A}$ is irreducible if for each pair (i,j) there is k such that $a_{i,j}^{(k)}>0$.

If the matrix $\mathscr{A}$ is not irreducible, then we say that it is *reducible*. Thus, a matrix is reducible if the associated graph is not strongly connected, that is, if there are vertices i and j such that i is not accessible from j. An equivalent definition is that $\mathscr{A}$ is reducible if, by simultaneous permutation of rows and columns, it can be brought to the form

$$\begin{pmatrix} A & \mathbf{0} \\ B & C \end{pmatrix},$$

where A and C are square matrices.

In terms of age-structured population dynamics, a matrix is irreducible if each stage i can contribute to any other stage j. E.g., the Usher matrix

$$\left(\begin{array}{cccc} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \end{array}\right)$$

is reducible as the last state cannot contribute to any other state and fertility is only concentrated in one state.

Irreducible matrices are subdivided into two further classes. An irreducible matrix $\mathscr{A}$ is called *primitive* if

$$\mathscr{A}^k > 0$$

for some $k > 0$; otherwise it is called *imprimitive*.

Note the difference between irreducibility and primitivity. For irreducibility we require that for each (i,j) there is k such that $a_{i,j}^{(k)} > 0$ but for primitivity there must be k such that $a_{i,j}^{(k)} > 0$ for all (i,j).

In population dynamics, if the population has a single reproductive stage, then its projection matrix is imprimitive. E.g., the matrix

$$\mathscr{A} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

describing a semelparous population is imprimitive. Indeed

$$\mathscr{A}^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and thus none $\mathscr{A}^k > 0$.

Perron-Frobenius theorem.

Let $\mathscr{A}$ be a nonnegative matrix.

(a) There exists a real nonnegative eigenvalue $\lambda_{max} = \rho(\mathscr{A})$ such that $\lambda_{max} \geq |\lambda|$ for any $\lambda \in \sigma(\mathscr{A})$. There is an eigenvector (called the Perron eigenvector) corresponding to λ_{max} which is real and nonnegative.

(b) If, in addition, $\mathscr{A}$ is irreducible, then λ_{max} is simple and strictly positive, $\lambda_{max} \geq |\lambda|$ for $\lambda \in \sigma(\mathscr{A})$. The eigenvector corresponding to λ_{max} may be chosen to be strictly positive.

(i) If $\mathscr{A}$ additionally is primitive, then $\lambda_{max} > |\lambda|$;

(ii) If $\mathscr{A}$ is imprimitive, then there is $d > 1$ such that

$$\lambda_j = \lambda_{max} e^{2\pi i \frac{j}{d}}, \; j = 0, \ldots, d-1,$$

are simple eigenvalues of $\mathscr{A}$.

Let us apply the Perron-Frobenius theorem in a population context. Suppose that our population is divided into n classes and the state of the population is given by the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ giving the number of individuals (or density) in each class. Let $\boldsymbol{x}_0 \geq 0$ denote the initial distribution of the population among the classes. Then

$$\boldsymbol{x}(k) = \mathscr{A}^k \boldsymbol{x}_0$$

is the distribution after k periods and

$$P(k, \boldsymbol{x}) = \|\mathscr{A}^k \boldsymbol{x}_0\| = \sum_{i=1}^{n} (\mathscr{A}^k \boldsymbol{x}_0)_i = \sum_{i=1}^{n} x_i(k) = \|\boldsymbol{x}(k)\|$$

is the total population at time k evolving from the initial distribution $\boldsymbol{x}_0$.

If $\mathscr{A}$ is nonnegative, irreducible or primitive, then the transpose $\mathscr{A}^T$ has the same property. Let λ_{max} be the dominant eigenvalue of both matrices and $\boldsymbol{v}$ and $\boldsymbol{v}^*$ be the corresponding strictly positive eigenvectors of, respectively, $\mathscr{A}$ and $\mathscr{A}^T$, corresponding to λ_{max}. We normalize $\boldsymbol{v}$ so that $\|\boldsymbol{v}\| = 1$ and $\boldsymbol{v}^*$ so that $< \boldsymbol{v}^*, \boldsymbol{v} >= 1$.

Combining the Perron-Frobenius theorem with the spectral decomposition we arrive at the following result.

Fundamental Theorem of Demography. *Suppose that the projection matrix $\mathscr{A}$ is irreducible and primitive and let λ_{max} be the strictly positive dominant eigenvalue of $\mathscr{A}$, $\boldsymbol{v}$ the strictly positive eigenvector of $\mathscr{A}$ and $\boldsymbol{v}^*$ strictly positive eigenvector of $\mathscr{A}^T$ corresponding to λ_{max}. Then, for any $\boldsymbol{x}_0 \geq 0$,*

(a) $\mathscr{A}$ has the AEG property

$$\lim_{k\to\infty} \lambda_{max}^{-k} \mathscr{A}^k \boldsymbol{x}_0 =< \boldsymbol{v}^*, \boldsymbol{x}_0 > \boldsymbol{v}. \tag{88}$$

(b)

$$\lim_{k\to\infty} \frac{\boldsymbol{x}(k)}{P(k,\boldsymbol{x}_0)} = \lim_{k\to\infty} \frac{\mathscr{A}^k \boldsymbol{x}_0}{P(k,\boldsymbol{x}_0)} = \boldsymbol{v}. \tag{89}$$

(c) If $\lambda_{max} < 1$, *then*

$$\lim_{k\to\infty} P(k,\boldsymbol{x}_0) = 0$$

and

$$\lim_{k\to\infty} P(k,\boldsymbol{x}_0) = \infty$$

if $\lambda_{max} > 1$.

Proof.

(a) We use (82), (84) and Perron-Frobenius Theorem 1.(i):

$$\begin{aligned} \mathscr{A}^k \boldsymbol{x}_0 &= \sum_{\lambda \in \sigma(\mathscr{A})} \lambda^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}_0) = \lambda_{max}^k < \boldsymbol{v}^*, \boldsymbol{v} > \boldsymbol{v} \\ &+ \sum_{\lambda \in \sigma(\mathscr{A}) \setminus \{\lambda_{max}\}} \lambda^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}_0). \end{aligned} \tag{90}$$

By primitivity of $\mathscr{A}$, $\lambda_{max} > |\lambda|$ for any $\lambda \in \sigma(\mathscr{A} \setminus \{\lambda_{max}\})$ and $p_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}_0)$ are polynomials in k, we have

$$\left\| \left(\frac{\lambda}{\lambda_{max}} \right)^k \boldsymbol{p}_\lambda(k, \boldsymbol{v}^\lambda, \boldsymbol{x}_0) \right\| \to 0, \qquad k \to \infty$$

and (a) is proved.

For (b) we see that, by (a),

$$\begin{aligned} \lim_{k\to\infty} \frac{P(k,\boldsymbol{x}_0)}{\lambda_{max}^k} &= \lim_{k\to\infty} \frac{\|\mathscr{A}^k \boldsymbol{x}_0\|}{\lambda_{max}^k} = \lim_{k\to\infty} \left\| \frac{\mathscr{A}^k \boldsymbol{x}_0}{\lambda_{max}^k} \right\| \\ &= |<\boldsymbol{v}^*, \boldsymbol{v}>| > 0. \end{aligned} \tag{91}$$

Hence, by (a) and (91),

$$\lim_{k\to\infty} \frac{\mathscr{A}^k \boldsymbol{x}_0}{P(k,\boldsymbol{x}_0)} = \lim_{k\to\infty} \frac{\lambda_{max}^{-k} \mathscr{A}^k \boldsymbol{x}_0}{\lambda_{max}^{-k} P(k,\boldsymbol{x}_0)} = \boldsymbol{v},$$

which gives (89).

To prove (c) we observe that

$$\begin{aligned} P(k,\boldsymbol{x}_0) &= \left\| \mathscr{A}^k \left(\frac{\boldsymbol{x}_0}{P(k-1,\boldsymbol{x}_0)} \right) \right\| P(k-1,\boldsymbol{x}_0) \\ &= \lambda_{k-1} P(k-1,\boldsymbol{x}_0) \end{aligned}$$

where

$$\begin{aligned} \lambda_{k-1} &= \left\| \mathscr{A}^k \left(\frac{\boldsymbol{x}_0}{P(k-1,\boldsymbol{x}_0)} \right) \right\| = \lambda_{max} \left\| \frac{\lambda_{max}^{-k} \mathscr{A}^k \boldsymbol{x}_0}{\lambda_{max}^{-(k-1)} P(k-1,\boldsymbol{x}_0)} \right\| \\ &\to \lambda_{max} \frac{\|\boldsymbol{v}\|}{|<\boldsymbol{v}^*,\boldsymbol{v}>|} = \lambda_{max}, \quad \text{as} \quad k \to \infty \end{aligned}$$

by (a), (b) and the normalization of $\boldsymbol{v}, \boldsymbol{v}^*$.

Thus, if $\lambda_{max} < 1$, then we can pick $\bar{\lambda} < 1$ such that $\lambda_{k-1} \leq \bar{\lambda}$ for all k larger than some k_0 and

$$P(k_0 + i, \boldsymbol{x}_0) \leq \bar{\lambda}^i P(k_0, \boldsymbol{x}_0) \to 0 \quad \text{as} \quad i \to \infty.$$

Similarly, if $\lambda_{max} > 1$, then we can pick $\tilde{\lambda} > 1$ such that $\lambda_{k-1} \geq \tilde{\lambda}$ for all k larger than some k_0 and

$$P(k_0 + i, \boldsymbol{x}_0) \geq \tilde{\lambda}^i P(k_0, \boldsymbol{x}_0) \to \infty \quad \text{as} \quad i \to \infty,$$

as $P(k_0, \boldsymbol{x}_0) \neq 0$ for any finite k_0. Indeed, otherwise from nonnegativity we would have $\boldsymbol{x}(k_0) = 0$ and thus $\boldsymbol{x}(k) = 0$ for $k \geq k_0$, contradicting (a). □

If we note that for each $1 \leq i \leq n$

$$\frac{(\mathscr{A}^k \boldsymbol{x})_i}{P(k, \boldsymbol{x})}$$

is the fraction of the population in the state i at time k, then the result above states that for large times the fraction of the population in the state i approximately is given by the i coordinate of the Perron eigenvector and is independent of the initial distribution $\boldsymbol{x}$. Moreover $\boldsymbol{v}$ is approached (or departed from) at an exponential rate, hence the name *asynchronous exponential growth*.

Population model with transition matrix of indefinite sign

Natchez social structure. Many societies are divided into classes membership of which is largely hereditary. A way to prevent watering down of the elite is endogamy; that is, marrying within one's own class. Some societies, however, practice an open class system to prevent stagnation of the structure. An example is offered by the civilization of Natchez.

Natchez were Native Americans who lived in the lower Mississipi in North America. The civilisation ceased to exist after the so-called Natchez massacre in 1731 when they were defeated and then dispersed or were enslaved.

Natchez created a complex system of open class structure based on the exogamous marriages so that the power is passed between people born to different social classes. The society was divided into two main classes – nobility and commoners (so-called Stinkards). The nobility was further divided into subclasses (casts): Suns, Nobles and Honoured People. A member of nobility only could marry a Stinkard.

We simplify the analysis by merging Nobles and Honoured into one class – say Nobles.

The table below summarizes the permissible marriages and inherited statuses.

Mother/Father	**Sun**	**Noble**	**Stinkard**
Sun			Sun
Noble			Noble
Stinkard	Noble	Stinkard	Stinkard

Table: Possible marriages in the Natchez population and the status of their offspring

We adopt the following simplifying assumptions.

1. There is the same number of males and females in each class in each generation – we track only males.
2. Each person marries only once and the spouse is from the same generation.
3. Each pair has exactly one son and one daughter.

Let the population (of males) in the kth generation be described by

$$\boldsymbol{x}(k) = (x_1(k), x_2(k), x_3(k))$$

with the classes numbered as follows: 1 -Sun, 2 -Noble, 3 -Stinkard.

Since a Sun son only can be born to a Sun mother and there is no other way to become a Sun, using the fact that the number of female Suns equals the number of male Suns we can write

$$x_1(k+1) = x_1(k).$$

A Noble son only can be born to Sun father or to Noble mother, using the parity of males and females in the Noble class we get

$$x_2(k+1) = x_1(k) + x_2(k).$$

Finally, the number of male offspring in the Stinkard class is equal to the number of Stinkard males who are not married to females from the nobility plus the number of sons of Stinkard mothers and Noble fathers (remember that the son of a Stinkard father and a Noble mother is a Noble but then the son of a Stinkard mother and a Noble father is a Stinkard). Hence

$$x_3(k+1) = -x_1(k) - x_2(k) + x_2(k) + x_3(k) = -x_1(k) + x_3(k).$$

Writing the model in matrix form, we have

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k) = (\mathscr{I} + \mathscr{B})\boldsymbol{x}(k),$$

where

$$\mathscr{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \qquad \mathscr{B} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

We see that $\mathscr{B}^2 = 0$. Hence, $\mathscr{A}^k = \mathscr{I} + k\mathscr{B}$, or

$$\mathscr{A}^k = \begin{pmatrix} 1 & 0 & 0 \\ k & 1 & 0 \\ -k & 0 & 1 \end{pmatrix}.$$

If $x_1(0) = 0$, the structure of the population does not change since

$$\boldsymbol{x}(k) = (0, x_2(0), x_3(0)).$$

If, however, originally we have some members of the Sun class, then the size of the Stinkard becomes negative in finite time. Indeed $x_3(k) = -kx_1(0) + x_3(0)$ and thus $x_1(0) > 0$ yields $x_3(k) < 0$ for $k > x_3(0)/x_1(0)$. For instance, if the ratio of Stinkards and Suns is 5 at the beginning, Stinkards will become extinct after four generations. Thus, nobody from the nobility will be able to marry and thus the population will become extinct. Since, according to the historical records, the Natchez civilisation survived for several hundred years, the model in the presented form cannot be correct.

Luenberger solution. In many societies it is observed that the birth rate depends on the status of the parents. Can we find a birth rate for each combination of parents in the Natchez community which will result in its a stable class distribution? Consider the following the intermarriage/fecundity scheme,

Mother/Father	Sun	Noble	Stinkard
Sun			Sun α_1
Noble			Noble α_3
Stinkard	Noble α_2	Stinkard α_4	Stinkard α_5

where α_i gives the average number of male offspring from such a marriage.

Modelling as before leads to

$$\boldsymbol{x}(k+1) = \mathscr{A}\boldsymbol{x}(k),$$

where

$$\mathscr{A} = \begin{pmatrix} \alpha_1 & 0 & 0 \\ \alpha_2 & \alpha_3 & 0 \\ -\alpha_5 & (\alpha_4 - \alpha_5) & \alpha_5 \end{pmatrix} \tag{92}$$

As in the Perron-Frobenius theorem, we will try to find a positive eigenvector of $\mathscr{A}$ associated with the largest positive eigenvalue. Such a vector would correspond to a stable population structure to which the system would converge and, once attained, would not change. The total population would change as powers of the eigenvalue.

Since the matrix $\mathscr{A}$ is triangular, its eigenvalues are given by α_1, α_3 and α_5. If α_5 is the dominant eigenvalue, then the long term structure is given by

$$(0,0,1);$$

that is, the population will only consists of Stinkards. While it is certainly possible, we are interested in survival of the community as a whole which is not possible if α_5 is the dominant eigenvalue. A similar outcome is obtained if we assume that α_3 is the dominant eigenvalue. Then the stable population structure is

$$\boldsymbol{e} = (0,1,(\alpha_4-\alpha_5)/(\alpha_3-\alpha_5)),$$

provided $\alpha_4 > \alpha_5$.

If α_1 is the dominant eigenvalue, then the stable population structure is given by

$$\left(1, \frac{\alpha_2}{(\alpha_1 - \alpha_3)}, \frac{1}{\alpha_1 - \alpha_5}\left(-\alpha_5 + \alpha_2 \frac{\alpha_4 - \alpha_5}{\alpha_1 - \alpha_3}\right)\right)$$

and it is positive if and only if

$$\alpha_2(\alpha_4 - \alpha_5) > \alpha_5(\alpha_1 - \alpha_3). \tag{93}$$

To complete our considerations, we must show that in the stable structure we have sufficiently many Stinkards for Suns and Nobles to marry. Thus we require that the components of the stable distribution vector additionally satisfy

$$x_3 \geq x_1 + x_2. \tag{94}$$

Substituting here the formulae for x_2 and x_3, which were derived earlier, we obtain

$$\frac{1}{\alpha_1 - \alpha_5}\left(-\alpha_5 + \alpha_2 \frac{\alpha_4 - \alpha_5}{\alpha_1 - \alpha_3}\right) \geq 1 + \frac{\alpha_2}{(\alpha_1 - \alpha_3)}. \qquad (95)$$

This inequality will be satisfied if α_4 is sufficiently large. In other words, to ensure that in the stable population distribution we have sufficiently many Stinkards, the fecundity in marriages of Stinkard mothers and Noble fathers should be sufficiently large.

Full solution. We observe that the provided analysis is not complete. It only shows that there is a structure of the society which can persist in a stable way. However, we do not know whether any positive initial population satisfying (94) will be be positive and satisfy (94) in each generation and will eventually stabilize at the structure determined by the eigenvector found above. Fortunately, the problem allows for a more detailed solution which confirms the results obtained by the analysis of the dominant eigenvalue and the corresponding eigenvector. For the solution with $\overset{\circ}{x}_1, \overset{\circ}{x}_2 > 0$ we have

$$x_1(k) = \alpha_1^k \overset{\circ}{x}_1 > 0 \qquad x_2(k) = \alpha_3^k \overset{\circ}{x}_2 + \alpha_2 \overset{\circ}{x}_1 \frac{\alpha_1^k - \alpha_3^k}{\alpha_1 - \alpha_3} > 0.$$

Further

$$x_3(k) = \alpha_5^k \overset{\circ}{x}_3 + \frac{\alpha_3^k - \alpha_5^k}{\alpha_3 - \alpha_5}(\alpha_4 - \alpha_5)\left(\overset{\circ}{x}_2 - \overset{\circ}{x}_1 \frac{\alpha_2}{\alpha_1 - \alpha_3}\right) + \frac{\alpha_1^k - \alpha_5^k}{\alpha_1 - \alpha_5}\left(\frac{\alpha_2(\alpha_4 - \alpha_5)}{\alpha_1 - \alpha_3} - \alpha_5\right)\overset{\circ}{x}_1 . \quad (96)$$

Hence $x_3(k)$ is positive provided (93) is satisfied and

$$\frac{\overset{\circ}{x}_1}{\overset{\circ}{x}_2} \leq \frac{\alpha_1 - \alpha_3}{\alpha_2}. \quad (97)$$

In other words, for the solution to remain positive, the ratio of the initial population of Suns and Nobles must be smaller that the ratio of these populations in the stable population distribution vector.

One can prove from (96) that

$$x_3(k) \geq x_1(k) + x_2(k) \tag{98}$$

is satisfied for any k but there is a smarter way.

Polar sets. The polar set S^* of a non-empty set S in $\mathbb{R}^n$ is defined to be

$$S^* = \{\boldsymbol{z} \in \mathbb{R}^n;\ (\boldsymbol{z}, \boldsymbol{y}) \geq 0 \text{ for all } \boldsymbol{y} \in S\}.$$

It follows that if C is a closed, convex and generating cone in $\mathbb{R}^n$, then so is C^*. Furthermore, $C^{**} = C$.

Proposition

Let C be a cone in $\mathbb{R}^n$ and A be an $n \times n$ matrix. Then

$$\mathscr{A}C \subseteq C \text{ if and only if } (\boldsymbol{z}, \mathscr{A}\boldsymbol{y}) \geq 0 \text{ for all } \boldsymbol{y} \in C, \boldsymbol{z} \in C^*.$$

Polyhedral cones. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ be some vectors in $\mathbb{R}^n$. We say that a cone $C \subset \mathbb{R}^n$ is polyhedral if

$$C = \{\boldsymbol{x};\ (\boldsymbol{a}_j, \boldsymbol{x}) \geq 0,\ j = 1, \ldots, r\}$$

and that it is finitely generated if

$$C = \{\boldsymbol{x};\ \boldsymbol{x} = \sum_{j=1}^{r} \mu_j \boldsymbol{a}_j,\ \mu_j \geq 0, j = 1,\ \ldots, r\} =: \mathit{cone}(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}).$$

Farkas–Minkowski–Weyl theorem. If $C = \mathit{cone}(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\})$, then

$$C^* = \{\boldsymbol{x};\ (\boldsymbol{a}_j, \boldsymbol{x}) \geq 0,\ j = 1, \ldots, r\};$$

$$\{\boldsymbol{x};\ (\boldsymbol{a}_j, \boldsymbol{x}) \geq 0,\ j = 1, \ldots, r\}^* = C.$$

Furthermore, a cone is polyhedral if and only if it is finitely generated.

In our case we consider

$$C = \{(x_1, x_2, x_3);\ x_1 \geq 0, -\frac{\alpha_2}{\alpha_1 - \alpha_3} x_1 + x_2 \geq 0, x_3 \geq x_1 + x_2\};$$

equivalently

$$C = cone\left(\left\{\left(1, \frac{\alpha_2}{\alpha_1 - \alpha_3}, 1 + \frac{\alpha_2}{\alpha_1 - \alpha_3}\right), (0,1,1), (0,0,1)\right\}\right).$$

Then

$$\begin{aligned} C^* \quad = \quad & \left\{(y_1, y_2, y_3);\ y_1 + \frac{\alpha_2}{\alpha_1 - \alpha_3} y_2 + \left(1 + \frac{\alpha_2}{\alpha_1 - \alpha_3}\right) y_3 \geq 0, \right. \\ & \left. y_2 + y_3 \geq 0, y_3 \geq 0\right\}. \end{aligned}$$

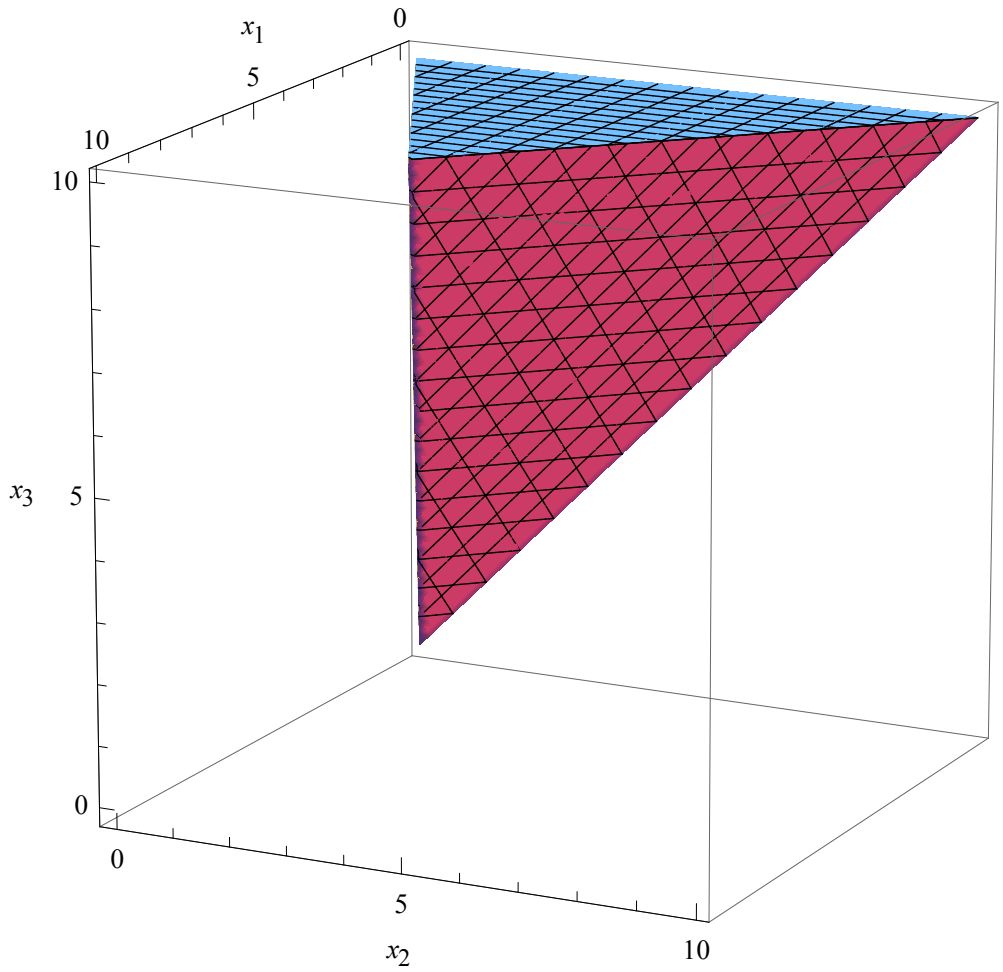


Figure: The viability cone C

We need to prove $(\boldsymbol{y}, \mathscr{A}\boldsymbol{x}) \geq 0$ for all $\boldsymbol{y} \in C^*, \boldsymbol{x} \in C$. It is equivalent to

$$\begin{aligned} \alpha_1 x_1 &\geq 0 \\ -\frac{\alpha_2}{\alpha_1 - \alpha_3} x_1 + \alpha_2 x_1 + \alpha_3 x_3 &\geq 0 \\ -\alpha_1 x_1 - \alpha_2 x_1 - \alpha_3 x_2 - \alpha_5 x_1 + (\alpha_4 - \alpha_5) x_2 + \alpha_5 x_3 &\geq 0 \end{aligned}$$

for all $(x_1, x_2, x_3) \in C$. In particular, it is sufficient to consider vectors that span C.

- substituting $(x_1, x_2, x_3) = (0, 0, 1)$ gives $\alpha_5 \geq 0$;
- substituting $(x_1, x_2, x_3) = (0, 1, 1)$ gives $\alpha_3 \geq 0$ and $\alpha_4 \geq \alpha_5$;
- substituting $(x_1, x_2, x_3) = \left(1, \frac{\alpha_2}{\alpha_1 - \alpha_3}, 1 + \frac{\alpha_2}{\alpha_1 - \alpha_3}\right)$ gives

$$\begin{aligned} \alpha_1 &\geq 0, \\ \alpha_4 &\geq \frac{\alpha_1(\alpha_1 - \alpha_3 + \alpha_2)}{\alpha_2} \end{aligned}$$

We see that (93) is automatically satisfied. Indeed

$$\begin{aligned}
&\frac{1}{\alpha_1-\alpha_5}\left(-\alpha_5+\alpha_2\frac{\alpha_4-\alpha_5}{\alpha_1-\alpha_3}\right)\\
&\geq \frac{1}{\alpha_1-\alpha_5}\left(-\alpha_5+\frac{\alpha_2}{\alpha_1-\alpha_3}\left(\frac{\alpha_1(\alpha_1-\alpha_3+\alpha_2}{\alpha_2}-\alpha_5\right)\right)\\
&= \frac{1}{\alpha_1-\alpha_5}\left(\frac{-\alpha_5(\alpha_1+\alpha_2-\alpha_3)}{\alpha_1-\alpha_3}+\frac{\alpha_1(\alpha_1+\alpha_2-\alpha_3)}{\alpha_1-\alpha_3}\right)\\
&= \frac{\alpha_1+\alpha_2-\alpha_3}{\alpha_1-\alpha_3}=1+\frac{\alpha_2}{\alpha_1-\alpha_3}\geq 0.
\end{aligned}$$

Summarizing

Theorem

Let $\alpha_i, i = 1, \ldots, 5$, *be positive,* $q := \max\{\alpha_3, \alpha_5\}/\alpha_1 < 1$ *and*

$$\alpha_4 \geq \frac{\alpha_1(\alpha_1 - \alpha_3 + \alpha_2)}{\alpha_2}.$$

Then the cone

$$C = \{(x_1, x_2, x_3);\ x_1 \geq 0, -\frac{\alpha_2}{\alpha_1 - \alpha_3} x_1 + x_2 \geq 0, x_3 \geq x_1 + x_2\};$$

is invariant under $\mathscr{A}$ *and, for* $\mathring{\boldsymbol{x}} \in IntC$,

$$\frac{\mathscr{A}^k \overset{\circ}{\boldsymbol{x}}}{\alpha_1^5} = \mathring{x}_1 \left(1, \frac{\alpha_2}{(\alpha_1 - \alpha_3)}, \frac{1}{\alpha_1 - \alpha_5}\left(-\alpha_5 + \alpha_2 \frac{\alpha_4 - \alpha_5}{\alpha_1 - \alpha_3}\right)\right) + O(q^k).$$

Leslie matrices. Let us consider the Leslie matrix

$$\mathscr{L} := \begin{pmatrix} f_0 & f_1 & \cdots & f_{n-2} & f_{n-1} \\ s_0 & 0 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-2} & 0 \end{pmatrix}, \tag{99}$$

and find under what conditions the population described by $\mathscr{L}$ exhibits asynchronous exponential growth.

Irreducible case.

First we observe that for irreducibility we need all $s_i \neq 0, 0 \leq i \leq n-2$. Indeed, if for some i the coefficient $s_i = 0$, then there would be no path from $k \leq i$ to $k > i$. In other words, there would be no way of reaching the age $k > i$. Assuming this, $\mathscr{L}$ is irreducible if and only if $f_{n-1} > 0$. Clearly, if $f_{n-1} = 0$, then there is no communication from class $n-1$ to any other class and thus $\mathscr{L}$ is reducible.

Now, let $f_{n-1} > 0$ and pick a (i,j). If $j < i$, then there is a path $(j,j+1)...(i-1,i)$ ensured by the survival coefficients $s_j, s_{j+1}, \ldots s_{i-1}$. If $j \geq i$, then the survival coefficients ensure that we reach the last class $n-1$, then since $f_{n-1} > 0$ we reach the class 0 and then we arrive at i by aging, that is $(j,j+1)\ldots(n-2,n-1)(n-1,0)(0,1),\ldots(i-1,i)$.

The question of primitivity is more complicated. Let us first assume that $f_j > 0$ for $j = 0, \ldots, n-1$, that is, that any age group is capable of reproduction. Let us consider arbitrary initial state j. Then there is an arc between j and 0 ($a_{0j} = f_j > 0$) and then from state 0 one can reach any state i in exactly i steps ($s_0 s_1 \cdot \ldots \cdot s_i$). Thus, there is a path joining j and i of length $i+1$ which still depends on the target state.

However, there is an arc from 0 to itself, so we can wait at 0 for any number of steps. In particular we can wait for $n-i$ steps so that j can be connected with i is $n+1$ steps. In other words

$$s_{i-1} \cdot \cdots \cdot s_1 s_0 f_0 \cdot \ldots \cdot f_0 f_j > 0$$

where f_1 occurs $n-i$ times. Hence $\mathscr{L}^n > 0$.

Remark. The above argument shows that any irreducible matrix in which at least one diagonal entry is not equal to zero is primitive.

This result assumes too much - typically young individuals cannot reproduce. We will strengthen this result. Let

$$0 = \det(\mathscr{L} - \lambda I) = \lambda^n + a_{n_1}\lambda^{n_1} + \ldots a_{n_i}\lambda^{n_i} \tag{100}$$

with $n > n_1 > \ldots > n_i$, $a_{n_k} \neq 0, k = 1, \ldots, i$ be the characteristic equation of $\mathscr{L}$.

It follows that if $\mathscr{L}$ is imprimitive of index d, then d is the greatest common divisor of $n - n_1, n_1 - n_2, \ldots, n_{i-1} - n_i$. This is related to the fact that $\lambda^d - r^d$ is a factor of the characteristic polynomial but full proof requires more subtle characterization of the spectrum of imprimitive matrices.

For Leslie matrices the characteristic equation can be calculated explicitly and it has its own biological interpretation. Also the above criterion can be proved directly.

Recall that $s_i = l_{i+1}/l_i$ with $l_0 = 1$, where l_i is the probability of surviving from birth to age i and $f_i = m_{i+1}s_i$. Consider the eigenvalue-eigenvector equation for a Leslie matrix

$$\mathscr{L}\boldsymbol{v} = \begin{pmatrix} f_0 & f_1 & \cdots & f_{n-2} & f_{n-1} \\ s_0 & 0 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-2} & 0 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_{n-1} \end{pmatrix} = \lambda \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_{n-1} \end{pmatrix}.$$

The equations from the second row down read

$$s_0 v_0 = \lambda v_1, \quad s_1 v_1 = \lambda v_2, \quad \ldots s_{n-2} v_{n-2} = \lambda v_{n-1}.$$

Taking $v_0 = 1$, we obtain

$$v_1 = \frac{s_0}{\lambda}, \quad v_2 = \frac{s_0 s_1}{\lambda^2}, \quad \ldots v_{n-1} = \frac{s_0 s_1 \ldots s_{n-2}}{\lambda^{n-1}}.$$

Now, the first row gives the equation

$$\lambda = \left(f_0 + \frac{f_1 s_0}{\lambda} + \frac{f_2 s_0 s_1}{\lambda^2} + \ldots + \frac{f_{n-1} s_0 s_1 \ldots s_{n-2}}{\lambda^{n-1}} \right).$$

We use $s_i = l_{i+1}/l_i$ where l_i is probability of survival till the $i+1$st reproductive cycle from birth (thus s_i is conditional probability of survival to the next reproductive cycle if one survived till i from birth) and $f_i = m_{i+1}s_i$ to rewrite the above as

$$1 = \left(\frac{m_1 l_1}{\lambda} + \frac{m_2 l_2}{\lambda^2} + \frac{m_3 l_3}{\lambda^3} + \ldots + \frac{m_n l_n}{\lambda^n} \right),$$

where we used $l_0 = 1$. Using the criterion mentioned above, a Leslie matrix is irreducible and primitive if e.g. the fertility of the oldest generation m_n is not zero and two subsequent generations have nonzero fertility.

Relative simplicity of the characteristic equation allows to strengthen this result even more. In fact, $\mathscr{L}$ is imprimtive if and only if the maternity function is periodic, that is, if the greatest common divisor of ages of positive reproduction, called the period, is greater than 1. For instance, $m_2, m_4, m_6 \ldots$ has period 2. In particular, the period is equal to the imprimitivity index. Indeed, suppose that

$$\lambda_j = re^{i\theta}, \quad \theta \neq 2\pi n,$$

is a negative or complex root to

$$\begin{aligned} \psi(\lambda) &= \lambda^{-1} m_1 l_1 + \ldots + \lambda^{n-1} m_{n-1} l_{n-1} + \lambda^n m_n l_n \\ &= \sum_{k=1}^{n} \lambda^{-k} m_k l_k = 1. \end{aligned} \tag{101}$$

Then

$$\sum_{k=1}^{n} r^{-k} e^{-ik\theta} l_k m_k = 1 \tag{102}$$

or, taking real and imaginary parts,

$$\sum_{k=1}^{n} r^{-k} \cos(k\theta) l_k m_k = 1, \tag{103}$$

$$\sum_{k=1}^{n} r^{-k} \sin(k\theta) l_k m_k = 0. \tag{104}$$

If m_k is periodic, then the only nonzero terms correspond to multiples of d, $m_{k_1 d}, m_{k_2 d}, m_{k_3 d}, \ldots$. Taking $\theta_j = 2\pi j/d$, $j = 0, 1, \ldots, d-1$, we see $\cos k_l d\theta_j = 1$, $\sin k_l d\theta_j = 0$ and so, if the above equations are satisfied by r, they are also satisfied for any $\lambda_j = re^{i\theta_j}$.

If m_k is periodic, then the only nonzero terms correspond to multiples of d, $m_{k_1 d}, m_{k_2 d}, m_{k_3 d}, \ldots$. Taking $\theta_j = 2\pi j/d$, $j = 0, 1, \ldots, d-1$, we see $\cos k_l d\theta_j = 1$, $\sin k_l d\theta_j = 0$ and so, if the above equations are satisfied by r, they are also satisfied for any $\lambda_j = re^{i\theta_j}$.

If m_k is aperiodic, then for some k we have $\cos k\theta < 1$. But then, if (103) is satisfied, we must have

$$\sum_{k=1}^{n} |\lambda_j|^{-k} l_k m_k > 1.$$

On the other hand, since

$$\sum_{k=1}^{n} r^{-k} l_k m_k = 1,$$

we obtain $|\lambda_j| < r$.

Reducible case. Let us consider a more complicated case where the fertility is restricted to some interval $[n_1, n_2]$, that is, when $f_j > 0$ for $j \in [n_1, n_2]$. As we noted earlier, if $n_2 < n$, the matrix cannot be irreducible as there is no communication between postreproductive stages and the reproductive ones. Consequently, if we start only with individuals in postreproductive age, the population will die out in finite time. Nevertheless, if $n_1 < n_2$ then the population still displays asynchronous exponential growth, albeit with a slight modification, as explained below.

To analyse this model, we note that since we cannot move from stages with $j > n_2$ to earlier stages, the part of the population with $j \leq n_2$ evolves independently from postreproductive part (but feeds into it.) Assume that $n_1 < n_2$ and introduce the restricted matrix

$$\tilde{\mathscr{L}} = \begin{pmatrix} f_0 & f_1 & \cdots & f_{n_2-1} & f_{n_2} \\ s_0 & 0 & \cdots & 0 & 0 \\ 0 & s_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n_2-1} & 0 \end{pmatrix}$$

and the matrix providing (one-way) link from reproductive to postreproductive stages is given by

$$\mathscr{R} = \begin{pmatrix} 0 & \cdots & s_{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & s_{n-2} & 0 \end{pmatrix}$$

For the matrix $\tilde{\mathscr{L}}$, $f_{n_2} > 0$ and $f_{n_2-1} > 0$ and we can apply the considerations of the previous section and the Fundamental Theorem of Demography. Thus, there is $r > 0$ there are vectors $\boldsymbol{v} = (v_0, \dots v_{n_2})$ and $\boldsymbol{v}^* = (v_0^*, \dots v_{n_2}^*)$ such that $\tilde{\mathscr{L}}\boldsymbol{v} = r\boldsymbol{v}$ and

$$\lim_{k\to\infty} r^{-k}\boldsymbol{x}(k+1) = \lim_{k\to\infty} r^{-k}\tilde{\mathscr{L}}^k\boldsymbol{x}_0 = \boldsymbol{v} < \boldsymbol{v}^*, \boldsymbol{x}_0 >,\ 0 \le \boldsymbol{x}_0 \in \mathbb{R}^{n_2}. \tag{105}$$

For $n_2 \leq j < n, k \geq 0$, we have $x_{j+1}(k+1) = s_j x_j(k)$. Hence, starting from $x_{n_2}(k)$ we get $x_{n_2+i}(k+i) = c_i x_{n_2}(k)$, where $c_i = s_{n_2+i-1} \cdot \ldots \cdot s_{n_2}$, as long as $i \leq n - n_2 - 1$. So

$$\lim_{k\to\infty} \lambda^{-k} x_{n_2+i}(k+i) = c_i v_{n_2} < \boldsymbol{v}^*, \boldsymbol{x}_0 >, \quad 0 \leq \boldsymbol{x}_0 \in \mathbb{R}^{n_2},$$

and hence, changing $k+i$ into k

$$\lim_{k\to\infty} \lambda^{-k} x_{n_2+i}(k) = c_i \lambda^{-i} v_{n_2} < \boldsymbol{v}^*, \boldsymbol{x}_0 >, \quad 0 \leq \boldsymbol{x}_0 \in \mathbb{R}^{n_2},$$

for any $i = 1, \ldots, n - n_2 - 1$.

Hence, we see that the formula (84) is satisfied if we take

$$\begin{aligned} \boldsymbol{v} &= (v_0, \ldots v_{n_2}, c_1 \lambda^{-1} v_{n_2}, \ldots, c_{n-n_2-1} \lambda^{-(n-n_2-1)} v_{n_2}) \\ \boldsymbol{v}^* &= (v_0^*, \ldots v_{n_2}^*, 0, \ldots, 0). \end{aligned}$$

Finally, we observe that if only one f_j is positive (semelparous population), then we do not have asynchronous exponential growth. Indeed, in this case starting from initial population in one class we will have a cohort of individuals in the same age group moving through the system. We have observed such a behaviour in the second numerical example.

McKendrick model.

From discrete Leslie model to continuous McKendrick problem. In the classical Leslie model the census is taken in equal intervals equal, for convenience, to the unit of time. If the time between censuses and the length of each age class are instead taken to be $h > 0$ then, starting from some time t the Leslie model would take the form

$$\begin{pmatrix} x_0(t+h) \\ x_h(t+h) \\ x_{2h}(t+h) \\ \vdots \\ x_{(n-1)h}(t+h) \end{pmatrix} \tag{106}$$

$$= \begin{pmatrix} f_0(h) & f_h & \cdots & f_{(n-2)h(h)} & f_{(n-1)h}(h) \\ s_0(h) & 0 & \cdots & 0 & 0 \\ 0 & s_h(h) & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{(n-2)h}(h) & 0 \end{pmatrix} \begin{pmatrix} x_0(t) \\ x_h(t) \\ x_{2h}(t) \\ \vdots \\ x_{(n-1)h}(t) \end{pmatrix}.$$

The maximum age of individuals $\omega = nh$ is thus divided into n age intervals $[0,h),[h,2h)\ldots[(n-1)h,nh)$ with the convention that if the age a of an individual is in $[kh,(k+1)h)$ is considered to be kh. In this definition, as in the discrete case, nobody actually lives till exactly ω. Thus, $x_a(t)$ denotes the number of individuals of age a, $s_a = l_{a+h}/l_a$ is the probability of survival to the age of $a+h$ conditioned upon surviving up to age a with $l_0 = 1$ and $f_a = m_{a+h}s_h$ is the effective fecundity with m_{a+h} being the average fertility of females of age $a+h$. We note that $(1-s_a)x_a(t)$ is the number of individuals who do not survive from a to $a+h$.

We make the following assumptions and notation: for any $a \geq 0$

$$\lim_{h \to 0^+} s_a(h) = s_a(0) = 1, \tag{107}$$

$$\lim_{h \to 0^+} \frac{1 - s_a(h)}{h} = \mu(a), \tag{108}$$

$$\lim_{h \to 0^+} \frac{f_a(h)}{h} = \beta(a). \tag{109}$$

To explain these notation, we note that probability of survival over a very short period of time should be close to 1, as in Eq. (107).

Further, using subsection on the average life span, we note that if death rate μ is constant, then the probability of surviving over a short time interval h approximately is $s_a(h) = 1 - \mu h$ for any a and thus the limit in Eq. (108) can serve as a more general definition of the age dependant death rate. Similarly, if the average number of births per female over a unit time is a constant β then the number of births over h will be βh and the last equation gives the general definition of the age dependent birth rate which, moreover, is independent of the survival rate by Eq. (107).

Finally, we assume that there is a density function $n(a,t)$

$$x_a(t) = \int_a^{a+h} n(\alpha,t)d\alpha. \tag{110}$$

We are going to derive a differential equation for n. Consider a fixed age $a = ih > 0$. From (106) we see that

$$x_{a+h}(t+h) = s_a(h)x_a(t), \qquad a = 0,h,\ldots,(n-2)h. \tag{111}$$

Using (110)

$$x_{a+h}(t+h) = \int_{a+h}^{a+2h} n(\alpha, t+h)d\alpha = \int_{a}^{a+h} n(\alpha+h, t+h)d\alpha,$$

thus (111) can be written as

$$\int_{a}^{a+h} n(\alpha+h, t+h)d\alpha = s_a(h) \int_{a}^{a+h} n(\alpha, t)d\alpha.$$

We re-write it as

$$\frac{1}{h^2}\left(\int_a^{a+h} n(\alpha+h,t+h)d\alpha - \int_a^{a+h} n(\alpha,t)d\alpha\right)$$

$$= -\frac{1-s_a(h)}{h^2}\int_a^{a+h} n(\alpha,t)d\alpha.$$

Assuming that the directional derivative

$$Dn(a,t) = \lim_{h\to 0^+} \frac{n(a+h,t+h)-n(a,t)}{h}$$

exists, under some technical assumptions we can pass to the limit above arriving, by (108), at

$$Dn(a,t) = -\mu(a)n(a,t), \qquad a>0, t>0.$$

Assuming that the partial derivatives $\partial_t n, \partial_a n$ at (a,t) exist, we can further transform the last equation to

$$\partial_t n(a,t) + \partial_a n(a,t) = -\mu(a) n(a,t), \qquad a > 0, t > 0.$$

This is the most commonly used form of the equation for n though, as we shall see later, not the best for its analysis and, in fact, false in many cases as the differentiability assumptions are often not satisfied.

Now consider the class of neonates in the Leslie formulation:

$$x_0(t+h) = \sum_{j=0}^{n-1} f_{jh} x_{jh}(t)$$

which can be rewritten as

$$\frac{1}{h} x_0(t+h) = \sum_{j=0}^{n-1} \frac{1}{h} f_{jh}(h) \frac{1}{h} x_{jh}(t) h.$$

Now, if n is continuous and f is differentiable at 0, then

$$\frac{1}{h} x_{jh}(t) = \frac{1}{h} \int_{jh}^{(j+1)h} n(\alpha, t) d\alpha = n(jh + \theta_j h)$$

and

$$\frac{f_{jh}(h)}{h} = \beta(jh + \theta_j' h).$$

for some $0 < \theta_j, \theta_j' < 1$. Thus

$$n(\theta_j h, t) = \sum_{j=0}^{n-1} n(jh + \theta_j h)\beta(jh + \theta_j' h)h.$$

If we further assume that β is a continuous function, then the right hand side is the Riemann sum and we can pass to the limit as $h \to 0^+$ getting

$$n(0,t) = \int_0^\omega \beta(\alpha) n(\alpha,t) d\alpha.$$

Thus, we arrived at the classical formulation of the McKendrick model

$$\begin{aligned} \partial_t n(a,t) + \partial_a n(a,t) &= -\mu(a) n(a,t), \qquad a > 0, t > 0, & (112) \\ n(0,t) &= \int_0^\omega \beta(\alpha) n(\alpha,t) d\alpha, t > 0, & (113) \\ n(a,0) &= n_0(a), & (114) \end{aligned}$$

where the last equation provides the initial distribution of the

If $\omega < +\infty$ then we have to ensure that $n(a,t) = 0$ for $t \geq 0, a \geq \omega$, which can be done either by imposing an additional boundary condition on n or by introducing assumptions on the coefficients which ensure that no individual survives beyond ω. If $\omega = \infty$ then, instead of such an additional condition, we impose some requirements on the behaviour of the solution at ∞, e.g. that they are integrable over $[0, \infty)$.

Linear constant coefficient case

Before we embark on more advanced analysis of (112)–(114) let us get a taste of the structure of the problem by solving the simplest case with $\mu(a) = \mu$ and $\beta(a) = \beta$:

$$\partial_t n(a,t) + \partial_a n(a,t) = -\mu n(a,t). \tag{1}$$

coupled with the boundary condition

$$n(0,t) = \beta \int_0^\infty n(a,t)da,$$

and the initial condition

$$n(a,0) = n_0(a).$$

Interlude - solving first order partial differential equations

Let us consider more general linear first order partial differential equation (PDE) of the form:

$$au_t + bu_x = 0, \quad t, x \in \mathbb{R} \tag{2}$$

where a and b are constants. This equation can be written as

$$D_{\boldsymbol{v}} u = 0, \tag{3}$$

where $\boldsymbol{v} = a\boldsymbol{j} + b\boldsymbol{i}$ ($\boldsymbol{j}$ and $\boldsymbol{i}$ are the unit vectors in, respectively, t and x directions), and $D_{\boldsymbol{v}} = \nabla u \cdot \boldsymbol{v}$ denotes the directional derivative in the direction of $\boldsymbol{v}$.

This means that the solution u is a constant function along each line having direction $\boldsymbol{v}$; that is, along each line of equation $bt - ax = \xi$. Along each such a line the value of the parameter ξ remains constant. However, the solution can change from one line to another, therefore the solution is a function of ξ, that is the solution to Eq. (2) is given by

$$u(x,t) = f(bt - ax), \tag{4}$$

where f is an arbitrary differentiable function. Such lines are called the *characteristic lines* of the equation.

Example. To obtain a unique solution we must specify the initial value for u. Hence, let us consider the initial value problem for Eq. (2): find u satisfying both

$$\begin{aligned} au_t + bu_x &= 0 \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= g(x) \quad x \in \mathbb{R}, \end{aligned} \tag{5}$$

where g is an arbitrary given function. From Eq. (4) we find that

$$u(x,t) = g\left(-\frac{bt - ax}{a}\right). \tag{6}$$

We note that the initial shape propagates without any change along the characteristic lines.

Example Let us consider a variation of this problem and try to solve the initial- boundary value problem

$$\begin{aligned} au_t + bu_x &= 0 \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= g(x) \quad x > 0, \qquad (7) \\ u(0,t) &= h(t) \quad t > 0, \qquad (8) \end{aligned}$$

for $a, b > 0$ From the previous example we have the general solution of the equation in the form

$$u(x,t) = f(bt - ax).$$

Putting $t = 0$ we get $f(-ax) = g(x)$ for $x > 0$, hence $f(x) = g(-x/a)$ for $x < 0$. Next, for $x = 0$ we obtain $f(bt) = h(t)$ for $t > 0$, hence $f(x) = h(x/b)$ for $x > 0$. Combining these two equations we obtain

$$u(x,t) = \begin{cases} g(-\frac{bt-ax}{a}) & \text{for} \quad x > bt/a \\ h(\frac{bt-ax}{b}) & \text{for} \quad x < bt/a \end{cases}$$

Solution of the McKendrick equation

First, let us simplify the equation (1) by introducing the integrating factor

$$\partial_t(e^{\mu a}n(a,t)) = -\partial_a(e^{\mu a}n(a,t))$$

and denote $u(a,t) = e^{\mu a}n(a,t)$. Then

$$u(0,t) = n(0,t) = \beta \int_0^\infty e^{-\mu a}u(a,t)da$$

with $u(a,0) = e^{\mu a}n_0(a) =: u_0(a)$. If we knew $\psi(t) = u(0,t)$, then

$$u(a,t) = \begin{cases} u_0(a-t), & t < a, \\ \psi(t-a), & a < t. \end{cases} \tag{9}$$

The boundary condition can be rewritten as

$$\begin{aligned}\psi(t) &= \beta\int_0^\infty e^{-\mu a}u(a,t)da = \beta\int_0^t e^{-\mu a}\psi(t-a)da+\beta\int_t^\infty e^{-\mu a}u_0(a-t)da\\ &= \beta e^{-\mu t}\int_0^t e^{\mu\sigma}\psi(\sigma)d\sigma+\beta e^{-\mu t}\int_0^\infty e^{-\mu r}u_0(r)dr\end{aligned}$$

which, upon denoting $\phi(t)=\psi(t)e^{\mu t}$ and using the original initial value, can be written as

$$\phi(t)=\beta\int_0^t \phi(\sigma)d\sigma+\beta\int_0^\infty n_0(r)dr. \qquad (10)$$

Now, if we differentiate both sides, we get

$$\phi' = \beta\phi$$

which is just a first order linear equation. Letting $t = 0$ in (10), we obtain the initial value for ϕ: $\phi(0) = \beta \int_0^\infty n_0(r)dr$. Then

$$\phi(t) = \beta e^{\beta t} \int_0^\infty n_0(r)dr$$

and

$$\psi(t) = \beta e^{(\beta-\mu)t} \int_0^\infty n_0(r)dr.$$

Then

$$n(a,t) = e^{-\mu a}u(a,t) = e^{-\mu t}\begin{cases} n_0(a-t), & t < a, \\ \beta e^{\beta(t-a)}\int_0^\infty n_0(r)dr, & a < t. \end{cases}$$

Observe that

$$\lim_{a \to t^+} n(a,t) = e^{-\mu t} n_0(0)$$

and

$$\lim_{a \to t^-} n(a,t) = \beta e^{-\mu t} \int_0^\infty n_0(r)$$

so that the solution is continuous only if the initial condition satisfies the following compatibility condition

$$n_0(0) = \beta \int_0^\infty n_0(r) dr. \tag{11}$$

Thus, as we noted earlier, we must be very careful with using (112)-(114) in the differential form and interpreting the solution.

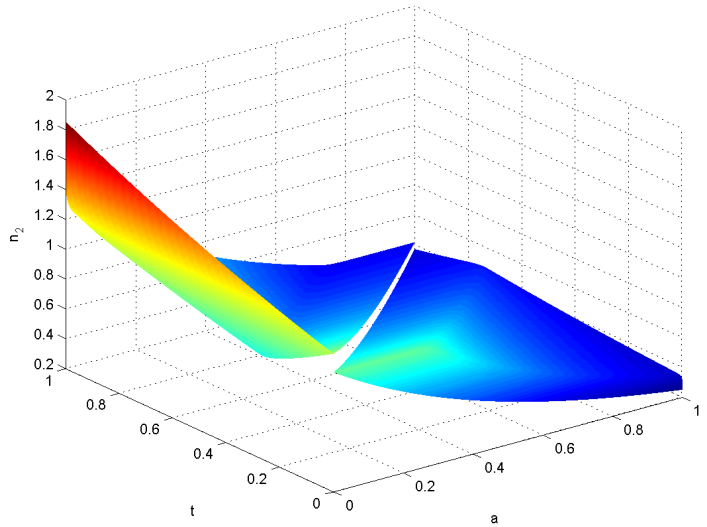


Figure: Discontinuity of the population density $n(a, t)$.

General linear McKendrick problem The ideas used to solve the McKendrick case in this simple case also is used in more general situations but, unfortunately, the resulting integral equation (10) cannot be explicitly solved. Before, however, we discuss solvability of more general cases, let us introduce certain functions related to (112)-(114) which are relevant to the population dynamics.

Demographic parameters of the McKendrick problems

Consider again the general McKendrick problem

$$\partial_t n(a,t) + \partial_a n(a,t) = -\mu(a) n(a,t)$$

$$n(0,t) = \int_0^\omega \beta(\alpha) n(\alpha,t) d\alpha,$$

$$n(a,0) = n_0(a)$$

We recall that $\beta(a)$ is the *age specific fertility* which is the number of newborns, in one time unit, coming from a single individual whose age is in a small age interval $[a, a+da)$. So, the number of births coming from all individuals in the population aged between a_1 and a_2 in a one time unit is

$$\int_{a_1}^{a_2} \beta(\alpha) n(\alpha, t) da$$

and we can define the *total birth rate* as

$$B(t) = \int_0^{\omega} \beta(\alpha) n(\alpha, t) da$$

which gives the total number of newborns in a unit time.

Let us consider the death rate $\mu(a)$ which is average number of deaths per unit of population aged a. We can relate $\mu(a)$ to a number of vital characteristics of the population. Similarly to the discrete case, we introduce the *survival probability* $S(a)$ as the proportion of the initial population surviving to age a. We can relate μ and S by the following argument.

Consider a population beginning with n_0 individuals of age 0. Then $n_0(a)S(a)(= n(a))$ is the average number of individuals that survived to age a. The decline in the population over a short age period $[a, a+da]$ is $n_0(a)S(a) - n_0(a)S(a+da)$. On the other hand, this decline can only be attributed to deaths: if the death rate is μ, then in this age interval we will have approximately $n_0(a)S(a)\mu(a)da$ deaths.

Equating and passing to the limit as $da \to 0$ yields

$$\frac{dS}{da} = -S\mu$$

or

$$S(a) = S(0)e^{-\int_0^a \mu(\sigma)d\sigma}.$$

Since, however, the probability of surviving to age 0 is 1, we have

$$S(a) = e^{-\int_0^a \mu(\sigma)d\sigma}. \tag{12}$$

We note that if no individuals can survive beyond ω, we must have $S(\omega) = 0$ or, equivalently,

$$\int_0^\omega \mu(\sigma)d\sigma = \infty. \tag{13}$$

These considerations can be used to find the average life span of individuals in the population. In fact, the average life span is the mean value of the length of life in the population, which can be expressed as

$$L = \int_0^{\omega} ap(a)da$$

where $p(a)$ is the probability (density) of an individual dying at age a. The probability of dying in the age interval $[a_1, a_2)$ is

$$\int_{a_1}^{a_2} p(a)da = S(a_1) - S(a_2) = -\int_{a_1}^{a_2} \frac{dS}{da}(a)da$$

and hence $p(a) = S(a)\mu(a)$.

Thus

$$L = \int_0^{\omega} a\mu(a) e^{-\int_0^a \mu(s)ds} da = -\int_0^{\omega} a \frac{d}{da} e^{-\int_0^a \mu(s)ds} da = \int_0^{\omega} S(a) da$$

where we used integration by parts and $S(\omega) = 0$.

Further, we introduce

$$K(a) = \beta(a)S(a) \tag{14}$$

which is called the *maternity function* and describes the rate of birth relative to the surviving fraction of the population and is the continuous equivalent to the coefficients $f_0, f_1, \ldots, f_{n-1}$. Further,

$$R = \int_0^\omega \beta(a)S(a)da \tag{15}$$

and call it *net reproduction rate* of the population. It is the expected number of offspring produced by an individual during her reproductive life.

Solution of the general linear McKendrick problem

One of the easiest way of analysing the general McKendrick model

$$\begin{aligned} &\partial_t n(a,t) + \partial_a n(a,t) = -\mu(a) n(a,t) \\ &n(0,t) = \int_0^\omega \beta(a) n(a,t) da, \\ &n(a,0) = n_0(a) \end{aligned} \tag{16}$$

is to reduce it to an integral equation in the same way as we proceeded previously, though the technicalities are slightly more involved due to age dependence of the mortality and maternity functions.

First, we simplify (16) by introducing the integrating factor

$$\partial_t \left(e^{\int_0^a \mu(\sigma)d\sigma} n(a,t) \right) = -\partial_a \left(e^{\int_0^a \mu(\sigma)d\sigma} n(a,t) \right) \qquad (17)$$

and denote $u(a,t) = e^{\int_0^a \mu(\sigma)d\sigma} n(a,t)$. Then

$$u(0,t) = n(0,t) = \int_0^\omega \beta(a) e^{-\int_0^a \mu(\sigma)d\sigma} u(a,t)da = \int_0^\omega K(a)u(a,t)da,$$

where we recognized that the kernel in the integral above is the maternity function introduced in (14). Further, $u(a,0) = e^{\int_0^a \mu(s)ds} n_0(a) =: u_0(a)$. Also, the right hand side defines the total birth rate $B(t)$.

Now, if we knew $B(t) = u(0,t)$, then

$$u(a,t) = \begin{cases} u_0(a-t), & t < a, \\ B(t-a), & a < t. \end{cases} \tag{18}$$

The boundary condition can be rewritten as

$$\begin{aligned} B(t) &= \int_0^\infty \beta(a) e^{-\int_0^a \mu(\sigma)d\sigma} u(a,t)da \\ &= \int_0^t \beta(a) e^{-\int_0^a \mu(\sigma)d\sigma} B(t-a)da + \int_t^\infty \beta(a) e^{-\int_0^a \mu(\sigma)d\sigma} u_0(a-t)da \\ &= \int_0^t K(t-a)B(a)da + \int_0^\infty \beta(a+t) e^{-\int_0^{a+t} \mu(\sigma)d\sigma} e^{\int_0^a \mu(s)ds} n_0(a)da, \end{aligned}$$

where to shorten notation we extended coefficients by zero beyond $a = \omega$.

Summarizing, we arrived at the integral equation for the total birth rate

$$B(t) = \int_0^t K(t-a)B(a)da + G(t) \tag{19}$$

where

$$G(t) = \int_0^\infty \beta(a+t)\frac{S(a+t)}{S(a)}n_0(a)da, \tag{20}$$

is a known function.

Explicitly, we have

$$\begin{aligned} B(t) &= \int_0^t K(t-a)B(a)da + \int_0^{\omega-t} \beta(a+t)\frac{S(a+t)}{S(a)}n_0(a)da \\ &= \int_0^t K(t-a)B(a)da + \int_t^{\omega} \beta(a)\frac{S(a)}{S(a-t)}n_0(a-t)da \quad (21) \end{aligned}$$

for $0 \leq t \leq \omega$ and

$$B(t) = \int_0^{\omega} K(t-a)B(a)da \quad (22)$$

for $t > \omega$.

Unlike in the constant coefficients case, this equation cannot be solved explicitly and we have to use more abstract approach. For this we have to introduce a proper mathematical framework. As in the discrete case, the natural norm will be

$$\|n\|_1 = \int_0^\omega |n(\alpha)| d\alpha$$

which in the current context, with $n \geq 0$ being the density of the population distribution with respect to age, is the total population. Thus, the state space is the space $X_0 = L_1([0, \omega))$ of Lebesgue integrable functions on $[0, \omega)$.

Since we are dealing with functions of two variables, we often consider $(a,t) \to n(a,t)$ as a function $t \to u(t,\cdot)$, that is, for each t the value of this function is a function with a argument. For such functions, we consider the space $C([0,T],L_1([0,\omega]))$ of $L_1([0,\omega])$-valued continuous functions. For functions f bounded on $[0,\omega]$ we introduce $\|f\|_\infty = \sup_{0\le a\le\omega}|f(a)|$. We make the following assumptions.

(i)

$$\beta \geq 0 \text{ is bounded on } [0, \omega], \tag{23}$$

(ii)

$$0 \leq \mu \in L_1([0, \omega']) \text{ for any } \omega' < \omega \tag{24}$$

with

$$\int_0^\omega \mu(\alpha) d\alpha = \infty, \tag{25}$$

(iii)

$$0 \leq n_0 \in L_1([0, \omega]). \tag{26}$$

Now, if (23)-(26) are satisfied, then we can show that K is a non-negative bounded function which is zero for $t \geq \omega$ and G is a continuous function which also is zero for $t \geq \omega$. If, additionally

$$n_0 \in W^{1,1}([0,\omega]) \quad \text{and} \quad \mu n_0 \in L_1([0,\omega]), \tag{27}$$

(here by W_1^1 we denote the Sobolev space of functions from L_1 with generalized derivatives in L_1), then G is differentiable with bounded derivative.

Indeed, let us look at G for $t < \omega$

$$G(t) = \int_t^\omega \beta(a) \frac{S(a)}{S(a-t)} n_0(a-t)da = \int_t^\omega \beta(a) e^{-\int_{a-t}^{a} \mu(s)ds} n_0(a-t)da$$

If we formally differentiate using the Leibniz rule, we get

$$\begin{aligned} G'(t) &= -\beta(t)S(t)n_0(0) + \int_t^\omega \beta(a)e^{-\int_{a-t}^{a}\mu(s)ds}\mu(a-t)n_0(a-t)da \\ &\quad + \int_t^\omega \beta(a)e^{-\int_{a-t}^{a}\mu(s)ds}n_0'(a-t)da \end{aligned}$$

so we see that for existence of the integrals we need integrability of μn_0 and differentiability of n_0. Then we can prove the main result

Theorem

If (23)-(26) are satisfied, then (19) has a unique continuous and nonnegative solution. If, additionally, (27) is satisfied, then B is differentiable with B' bounded on bounded intervals.

Proof.

We define iterates

$$\begin{aligned} B_0(t) &= G(t), \\ B^{k+1}(t) &= G(t) + \int_0^t K(t-s)B^k(s)ds. \end{aligned} \tag{28}$$

Take $T > 0$. Then, for any $t \in [0, T]$ we have

$$|B^1(t) - B^0(t)| = \int_0^t |K(t-s)F(s)|ds \leq tK_m F_m$$

where $K_m = \sup_{0 \leq t \leq T} |K(s)|$ and $L_m = \sup_{0 \leq t \leq T} |F(s)|$. Then

$$|B^2(t) - B^1(t)| \leq K_m \int_0^t |B^1(s) - B^0(s)|ds \leq \frac{K_m^2 F_m}{2} t^2$$

and, by induction,

$$|B^{k+1}(t) - B^k(t)| \leq K_m \int_0^t |B^k(s) - B^{k-1}(s)| ds \leq \frac{K_m^{k+1} F_m}{(k+1)!} t^{k+1}. \tag{29}$$

Further

$$\lim_{k\to\infty} B^{k+1}(t) = G(t) + \lim_{k\to\infty} \sum_{i=0}^{k} (B^{i+1}(t) - B^{i}(t))$$

with

$$\begin{aligned} \sup_{0\le t\le T} \left| \sum_{i=0}^{k} (B^{i+1}(t) - B^{i}(t)) \right| &\le \sum_{i=0}^{k} \sup_{0\le t\le T} \left| B^{i+1}(t) - B^{i}(t) \right| \\ &\le F_m \sum_{i=0}^{k} \frac{(TK_m)^{k+1}}{(k+1)!}. \end{aligned}$$

The series on the right hand side converges to $F_m e^{TK_m}$ and thus $(B^k(t))_{k\ge 0}$ converges uniformly to a continuous solution B of (19). Uniqueness follows by the Gronwall inequality.

If, in addition, (27) is satisfied, then B^k can be differentiated with respect to t and

$$V^k := \frac{d}{dt} B^k$$

satisfy the recurrence

$$V^{k+1}(t) = F'(t) + K(t)F(0) + \int_0^t K(t-s)V^k(s)ds$$

which converges uniformly to some continuous function V which, by the theorem of uniform convergence of derivatives, must be the derivative of B. □

Once we have B, we can recover n by (18) and back substitution

$$n(a,t) = e^{-\int_0^a \mu(\sigma)d\sigma} u(a,t) = \begin{cases} \frac{S(a)}{S(a-t)} n_0(a-t), & t < a, \\ S(a)B(t-a), & a < t. \end{cases} \tag{30}$$

Thus, if (27) is satisfied in addition to (23)-(26), then it is easy to see that n defined above satisfies the equation (112) everywhere except the line $a = t$. Along this line we have, as before

$$\lim_{a \to t^+} n(a,t) = S(0)n_0(0) = n_0(0)$$

and

$$\lim_{a \to t^-} n(a,t) = S(0)B(0) = \int_0^\omega \beta(a)n_0(a)da$$

and, to ensure at least continuity of the solution we need to assume the compatibility condition

$$n_0(0) = \int_0^\omega \beta(a)n_0(a)da. \tag{31}$$

We note that if a function is continuous at a point and differentiable in both one sided neighbourhoods, then it is a Lipschitz function and it is in fact differentiable almost everywhere (in the sense that the function can be recovered from its derivative). On the other hand, if a function has a jump at a point, then its derivative at this point is of a Dirac delta type. Thus, we can state that if (31) is satisfied, then the solution is continuous and satisfies (112) almost everywhere.

If we do not assume (31) then we can still claim that the solution satisfies

$$Dn(a,t) = \lim_{h \to 0^+} \frac{n(a+h, t+h) - n(a,t)}{h} = -\mu(a) n(a,t), \qquad a > 0, t > 0.$$

Furthermore, both the birth rate B and the solution n itself grow at most at an exponential rate. Consider again (19)

$$B(t) = \int_0^t K(t-a)B(a)da + G(t).$$

with G given by (20).

$$S(a) = e^{-\int_0^a \mu(\sigma)d\sigma}.$$

and $K(a) = \beta(a)S(a)$

We see that $K(t) \leq \|\beta\|_\infty$ and $G(t) \leq \|\beta\|_\infty \|n_0\|_1$ so that

$$\begin{aligned} B(t) &\leq \max_{0\leq a\leq \omega} \beta(a) \int_0^t B(s)ds + \max_{0\leq a\leq \omega} \beta(a) \int_0^\omega n_0(s)ds \\ &=: \|\beta\|_\infty \int_0^t B(s)ds + \|\beta\|_\infty \|n_0\|_1, \end{aligned}$$

which, by Gronwall's inequality, yields

$$B(t) \leq \|\beta\|_\infty \|n_0\|_1 e^{t\|\beta\|_\infty}. \tag{32}$$

This gives the estimate for n:

$$\begin{aligned}\|n(\cdot,t)\|_1 &\leq \int_0^t B(t-s)S(s)ds + \int_t^\infty \frac{S(s)}{S(s-t)} n_0(s-t)ds \\ &\leq \|\beta\|_\infty \|n_0\|_1 \left(\int_0^t e^{(t-s)\|\beta\|_\infty} ds + 1 \right),\end{aligned}$$

where we used $S(s)/S(s-t) \leq 1$. Then, by integration

$$\|n(\cdot,t)\|_1 \leq \|n_0\|_1 + \|n_0\|_1 e^{t\|\beta\|_\infty}(1 - e^{-t\|\beta\|_\infty}) = \|n_0\|_1 e^{t\|\beta\|_\infty}. \quad (33)$$