References

General

*A Concrete Approach To Mathematical Modelling. Michael Mesterton-Gibbons $85 ($50 used)
Best introduction to modeling. Covers both deterministic and stochastic. More focus on art of modeling, less on more faddish things like chaos and fractals. Best as a tutorial, not a reference. Either edition fine.

Mathematical Modeling. Mark Meerschaert. $59.95
Covers static, deterministic dynamic, and random dynamic (brief mention of fractals)

The Art of Modeling Dynamic Systems: Forecasting for Chaos, Randomness & Determinism. Foster Morrison $110
General coverage of all types of models as above. Less focus on the mechanics of model tools and more on the art of modeling.

Deterministic dynamics

*Understanding Nonlinear Dynamics. Daniel Kaplan & Leon Glass $36.95
Good introduction to deterministic dynamic models and one good chapter each on fractals and time series analysis. Very example based

Chaos – An Introduction to Dynamical Systems. Allgood, Sauer, Yorke $39.95
Very thorough somewhat mathematically rigorous introduction.

Chaos Theory Tamed. Garnett Williams $35
Written by a geologist – a how to use chaos (and fractals) cookbook in your scientific field.

Mostly fractal focused

Fractals and Chaos: An Illustrated Course. Paul Addison $39.00
Focus on fractals with some coverage of chaos

*Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. Manfred Schroeder $24.95
Light read with good coverage of the topics in title

Stochastic dynamics

The elements of Stochastic Processes (with applications to the natural sciences) Norman Bailey $89.95 ($15-$40 used)
Somewhat mathematical (but still accessible) coverage of stochastic processes, very comprehensive with strong EEB bias. Good if you just want to look up known results for EEB models or understand at a technical level.

*An Introduction to Stochastic Modeling 3rd edition. Taylor and Karlin $59.95 (older editions actually better)
Best 1st year text book on stochastic processes in use today

Introduction to Stochastic Processes. M S Bartlett $20-$45 used
Fairly good introduction and moderate EEB bias

Elements of Applied Stochastic Processes. U. Narayan Bhat $15-$40 used
Very intuitive introduction with focus on applications but mostly engineering, not biology

Schaum's Outline series – Probability, Random Variables & Random Processes - $14.95
Pretty decent coverage if you like this sort of book

Ecology/Evolution books w/ math reviews

Theoretical evolutionary ecology. Michael Bulmer, $39.95

Good introduction to optimization/evolutionary ecology/game theory

*Theory of Population Genetics and Evolutionary Ecology: an Introduction. Jonathon Roughgarden $41.20
The bible for me – both on the science with reasonably good intro to the math. At an advanced level.

Primer of Ecological Theory. Jonathon Roughgarden $52.00
Ecology is at undergraduate level, but excellent use of interwoven Matlab scripts – good way to learn Matlab

An Illustrated Guide to Theoretical Ecology. Ted Case $49.95
Covers ecology at advanced undergraduate/graduate level. As the title says, it is heavily illustrated with an emphasis on developing visual intuition.

Math packages

Excel 2000 - $104
Can do most math and graphing in this product. Sometimes easy, sometimes frustrating.

*Matlab
My preferred math package (and generally preferred by engineers). If you buy this, be sure to get the $99 student (vs. $1000 for non-students) version directly from Mathworks (www.mathworks.com), not the $79 Prentice Hall student version available in the book store as the Mathworks version has no limits on array sizes and allows you to buy mapping, statistical and other add-ins for great student prices as well (usually $59). I have a lot of cheat sheets and utilities for Matlab.

Mathematica - $139.95
The preferred math package by mathematicians (www.wolfram.com). I eventually gave up using this package as it was just too much work to get a graph up on the screen with any sort of decent formatting. Matlab was much quicker and dirtier.

Outline

References
- General
- Deterministic dynamics
- Mostly fractal focused
- Stochastic dynamics
- Ecology/Evolution books w/ math reviews
- Math packages

Outline
- A mini diversion –kinds of numbers

Overview - what a state space model is

Subject Matter

Calculus "review"
- a bestiary of functional forms - how to find a function that has the shape you want
  - Polynomials
- Rational functions
  - Power Law
  - Exponential functions
  - Trigonometric functions
  - Sigmoidal
- making functions dance – how to modify them
- the top 10 things you need to remember from 1st year calculus
- approximations - how do they do that and should they get away with that?

Linear algebra - a geometric/visual approach
- Types of transformations
- Types of linear transformations
- basis inner (dot) products (distance and angles), embeddings and projections, determinants
- How this maps to matrices
  - Matrices as basis changes
  - Matrices as representations of tranformations
- What they usually teach you in linear algebra
- the 15 minute version of 2nd year calculus
  - f: $\nabla^n \rightarrow \nabla$
  - f: $\nabla^n \rightarrow \nabla^m$
- why a Taylors series is really a linear algebra problem

Deterministic dynamical systems
- solving differential equations (the 15 minute version and why you don't need more)
- asymptotic behavior - stability analysis, chaos and everything in between
- what can we study if its chaotic – ergodic properties
- sensitivity to initial conditions and parameter values
- working with time series data
- diffusion equations

Random dynamical systems (stochastic processes)
- the 10 things you need to know about probability
- the top 10 probability distributions and how to remember them
- Markov models
  - Markov Chains
  - Counting (Poisson) and Birth Death processes
  - Stochastic Differential equations
  - Jump equations

Static systems - easier to solve than dynamic systems
- optimization (0 players affect the system)
- control theory (1 player affects the system)
- game theory (2+ players affect the system)

Descriptions of complex data
- Multivariate statistics
  - Regression

- Reduction of variables
- Fractals - cutting through the (very thick) hype
- Timeseries data
  - Chaos analysis
  - Fitting non-linear models
  - ARIMA
- Wavelets

A mini diversion –kinds of numbers

In general we will deal with three types of numbers:

∇ - the real numbers including irrational numbers

$\wedge$ - the integer numbers {0, ±1, ±2, …}

□ - a finite, unordered set (e.g. carnivore, omnivore, herbivore)

The first two sets are infinite while the last is finite. Notice that the "size" of the first two are different infinites. The size of $\wedge$ is called "countably" infinite, while the size of ∇ is called uncountably infinite. Believe it or not it can be mathematically proven that countably infinite is smaller than uncountably infinite

The first two sets are also ordered (3<4 means something) and metricized (the distance between 3 and 4 is well defined and in fact can be meaningfully said to be the same as the distance between 9 and 10)

We will sometimes take modifications on ∇ and $\wedge$ as follows:

∇^+ and $\wedge^+$ indicate the non-negative (≥ 0). Often the + notation means strictly positive, but not here.

∇^2 and $\wedge^2$ indicate the 2-dimensional sets — ∇^2 is the real plane of high school geometry and calculus while $\wedge^2$ is a 2-dimension lattice (grid). The obvious generalization to n-dimensions (∇^n) is also used.

Why belabor this? Because it is a critical modeling decision that drives what tools you can use.

Overview - what a state space model is

A state space model is the central concept in many areas of applied mathematics and certainly in ecology and evolutionary biology.

A state space model shows how a system changes over time.

A state space model has five parts:

1. State variables
2. Exogenous input
3. Transition rule
4. Transition parameters
5. Initial conditions

The **state variables** are a set of measurements or quantities that "fully" describe the system. It may be a single number or a vector (a set of numbers). The vector of numbers might be indexed by some index (such as spatial position).

The **exogenous input** is a set of measurements which are not affected by the system (state variables), but which affect the state variables.

The transition rule tells us what the system (state variables) will be at a future point in time given the current value of the state variables and the exogenous input.

The **transition parameters** are slight variations in the model – they are constants in the transition rule

The **initial conditions** tell us what the state variables were when they started

A model is a fully specified state space system (i.e. all 5 of the above).

A simple example is a competition model.

State variables	2 position vector (N_1,N_2) giving the population densities of species #1 and #2 at a point in time
Exogenous input	could be a 2 position vector (P,T) giving the number of predators (who mostly feed on other things and hence don't fluctuate with the N's) and the temperature. Note we need to know how these vary over time
Transition rule	A function (N_1,N_2) at time t+1 = f(N_1,N_2,u(P,T)) at time t – e.g. the Lotka Volterra equations modified to include effects of T & P
Transition parameters	parameters in the L-V equations, eg carrying capacity, K, and competition coefficients α
Initial conditions	The # of N_1 and N_2 at time t=0

Other examples of state systems include:
population genetics (the proportion of alleles is the usual state variable)
physiological models (energy reserves might be the state variable)

Note that the state variables can be: continuous or discrete, scalar or vector, indexed or not
The exogenous input can be: continuous or discrete (or absent in which case it is called an autonomous system)
The transition rule can be: discrete (map time t to t+1) or continuous (map t to t+dt - i.e. dN/dt), deterministic or stochastic

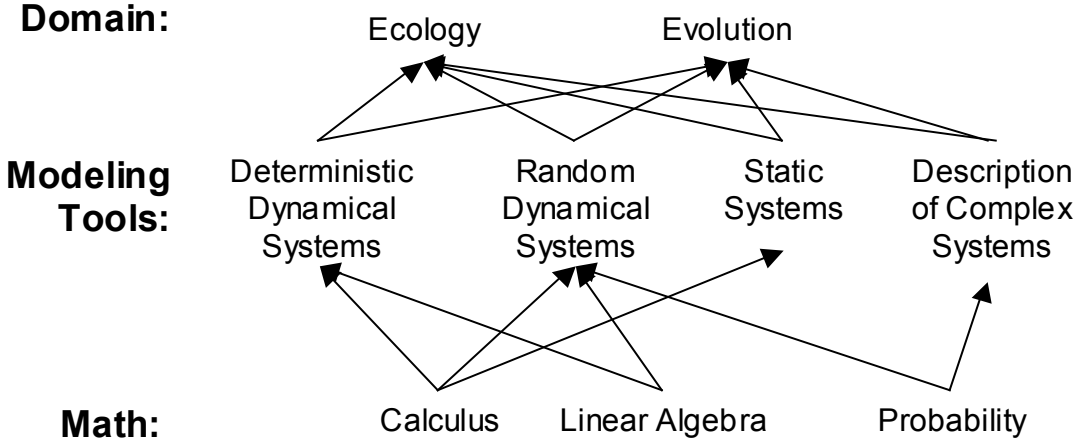
In general, models are more complicated to solve if:
there are more:
- state variables (or indexed)
- exogenous variables (none is easiest)
- transition parameters (none is easiest)

and transition rules are:
- deterministic (non-random)
- linear
- depend on the state of the system at only the immediately prior point in time (i.e. Markov, not time-delay or integro-differential)

	Deterministic rule		**Random Rule**	
	Discrete time	Continuous time	Discrete Time	Continuous Time
Discrete States	N/A	N/A	Markov Processes on □ and ∧	Continuous time Markov Processes on □ and ∧
Continuous States	Difference equations	Differential equations	Noise processes	Diffusion equations Stochastic differential equations (SDE's)

NB: the mathematical machinery does not exist to handle all combinations.
thus the exact specification of the state space model is *often chosen for mathematical convenience*

Subject Matter



Calculus "review"

a bestiary of functional forms - how to find a function that has the shape you want

Functions map points from one space to another. The key criteria is that from any point, it only maps to one point. If this is not true it is a relation, not a function. The converse is not true (more than one point can be mapped to a single point). If this occurs, the function is not invertible. As a notation we often write something like $f:\nabla^2 \rightarrow \nabla^3$ when it maps points from the real plane to 3-D space.

Broadly speaking we will look at two aspects of functional forms:

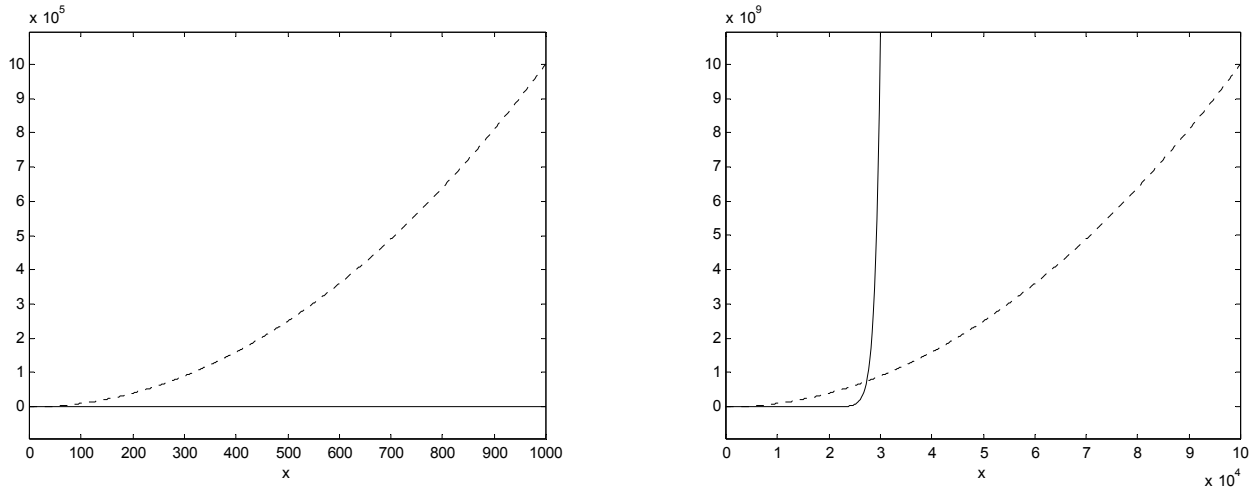
shape – the overall shape

speed – how fast it moves to where its going (∞ or asymptotes)

In particular there are two possible speeds:

- arithmetic (aka polynomial) – these functions are formed by +,-,×, ÷ and x^c
- geometric (aka exponential or non-polynomial) – these functions are often transcendental (meaning they use the exponential — e^x or log or sin or cos functions) but also c^x works

in general, geometric functions get there faster



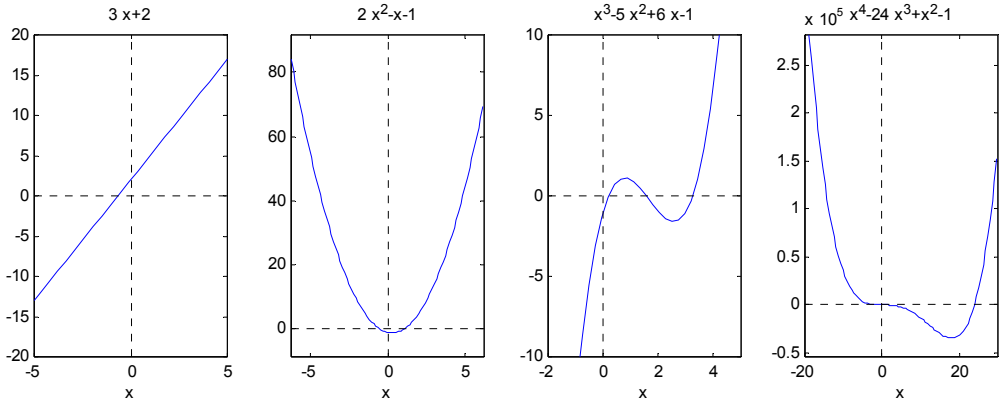
on x-scale 0-1000 on x-scale 0-10,000

Table 1- In both pictures, the arithmetic function(x^2) is the dashed line, and the exponential function (.001exp(.001x)) is the solid line. Note that even though the x^2 function is given quite an advantage in terms of parameters, the exponential catches up.

Polynomials

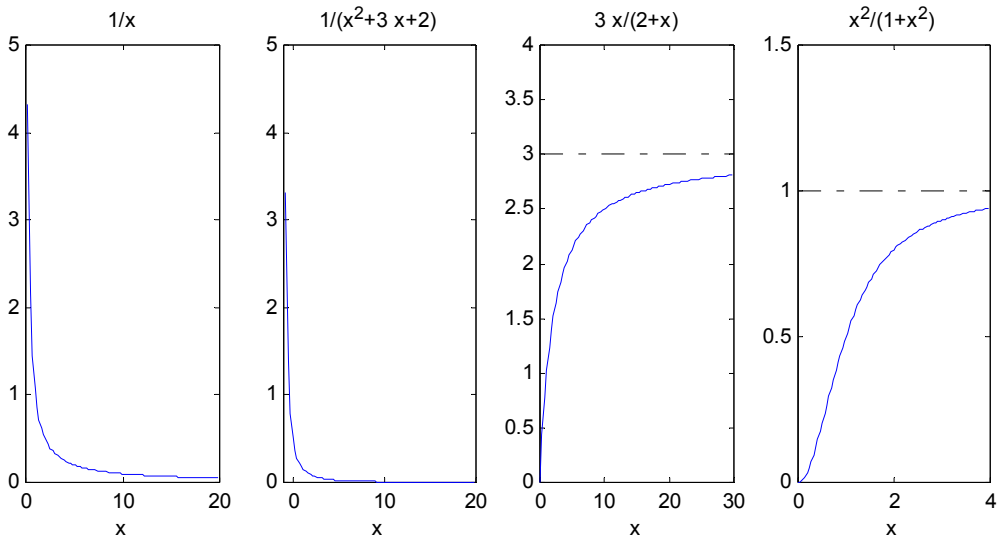
$f(x)=a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$

Polynomials are simple and well behaved. The degree of the polynomial is the highest exponent (n). An n-degree polynomial has n-1 humps or changes in direction and n roots (crosses X-axis).



Rational functions

Rational functions are the ratio of two polynomials.

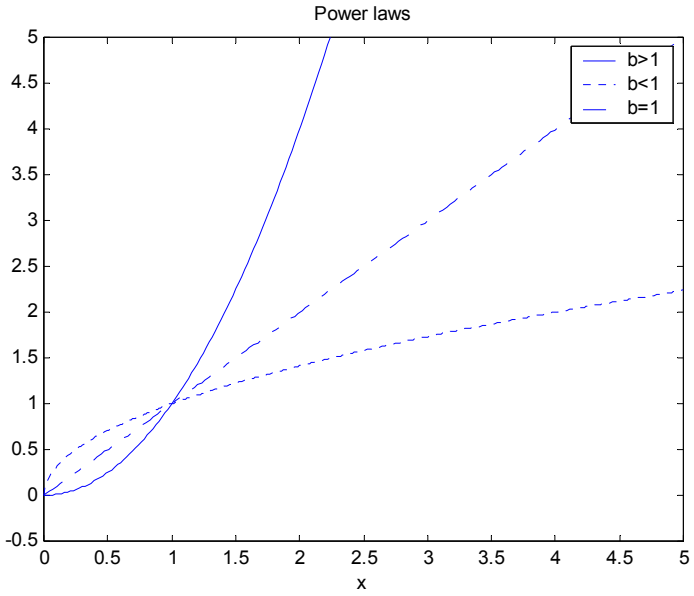


A few things to note:

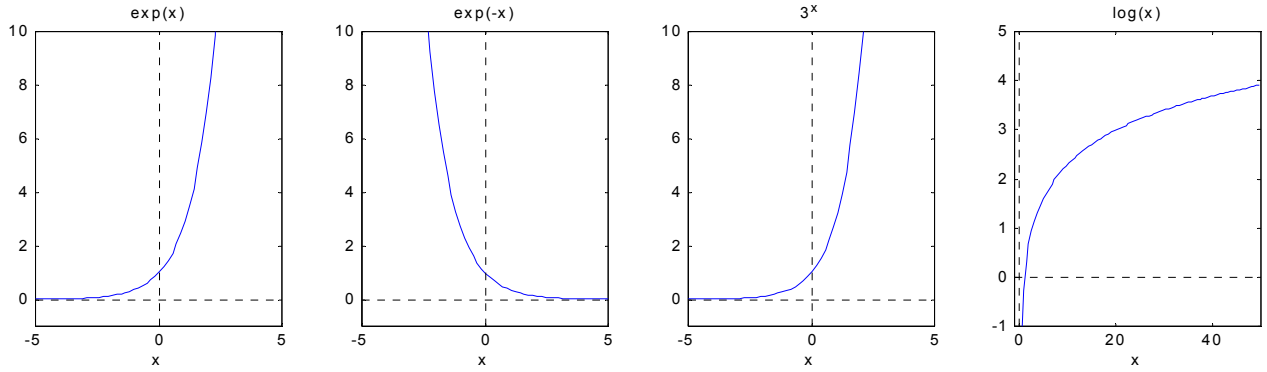
- The first two functions are a subclass of rational functions called hyperbolics where the numerator is 1 (or some other constant)
- Increasing the degree of the denominator polynomial in a hyperbolic function just makes it closer to a 90 degree turn
- The third function is the famous Michaelis-Menton equation and it asymptotes to a line. In general if a M-M equation is in the form ax/(b+cx) then the line will asymptote at a/c and it will be half way to the asymptote at b/c.
- The fourth function is sometimes called a Hill function (with the x^2 terms replaced by x^n). It is similar to the M-M but it asymptotes much faster and is now somewhat sigmoidal.

Power Law

The power law $f(x)=ax^b$ is central in biology. The coefficient a determines the behavior near 0, but the exponent b, determines the long term behavior and whether it is convex or concave.



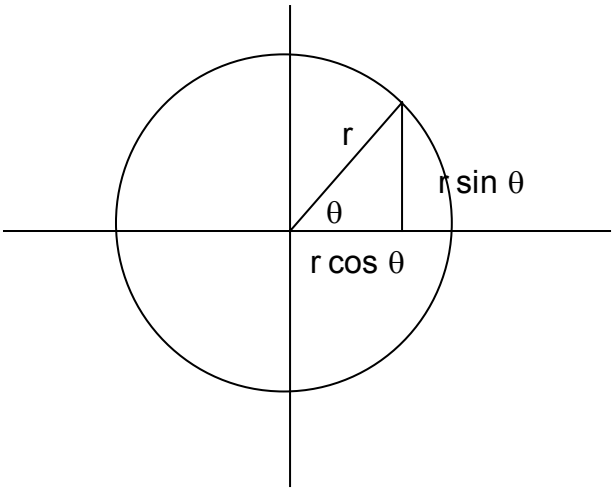
Exponential functions

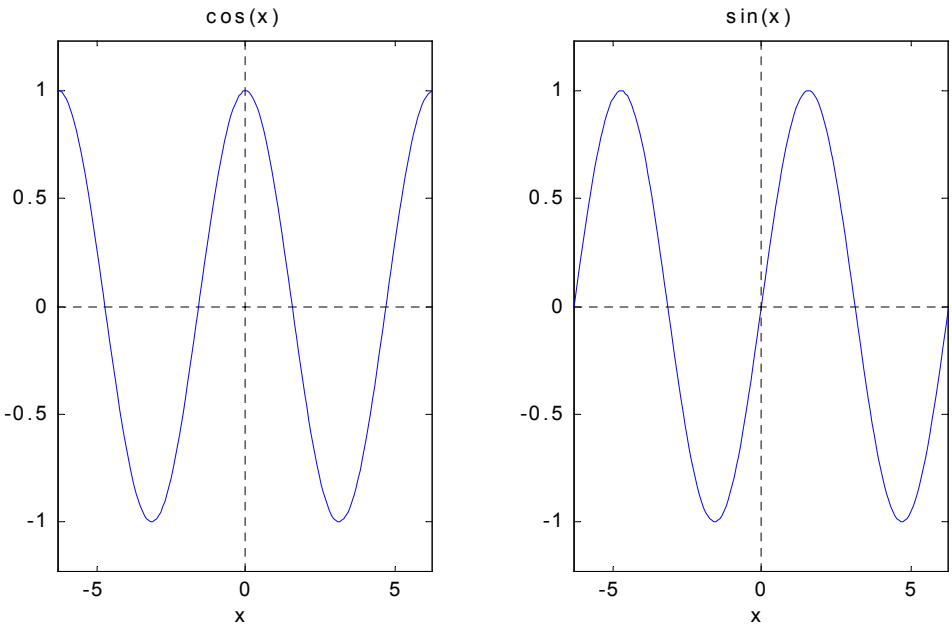


- Recall that the exponential function (1st picture) grows faster than any polynomial function, but that the geometric function (3rd picture) (c^x) behaves just like the exponential function (e^x)
- The exponential function always has an asymptote at the X-axis – whether at asymptotes at $+\infty$ or $-\infty$ depends on whether there is a –x or +x
- The log function is the inverse of the exponential function. It has a vertical asymptote at the Y-axis but no horizontal asymptote (it ends up at (∞,∞))

Trigonometric functions

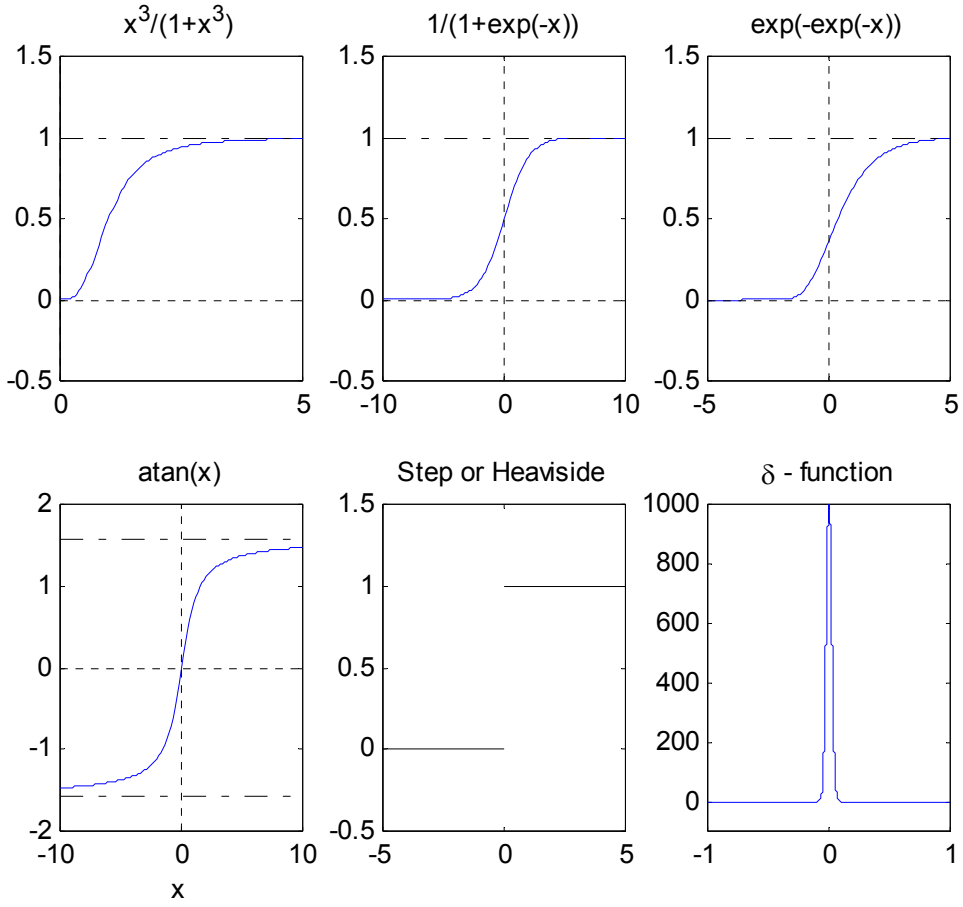
Aside from exponentials, trigonometric functions are the other commonly encountered transcendental functions (can't be written as ratios of polynomials).





Sigmoidal

Sigmoidal or S-shaped functions have two horizontal asymptotes. They can be obtained using either rational or several types of transcendental functions.

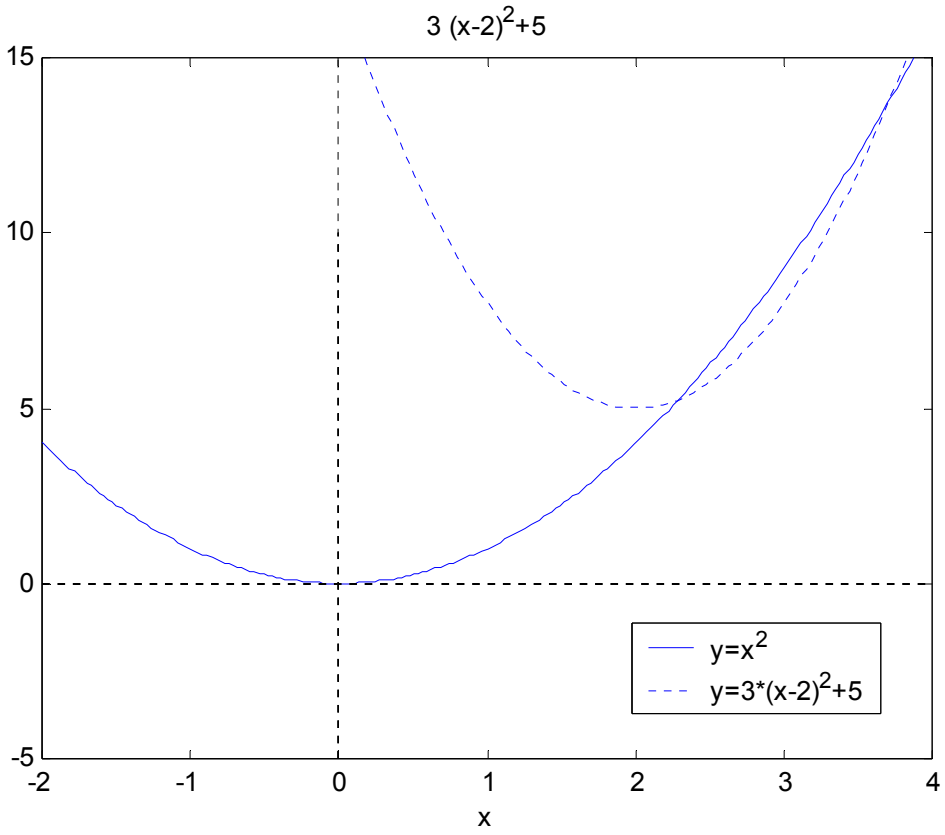


- The first is rational and sometimes called the Hill function, the 2nd and 3rd are transcendental based on the exponential function and called the logistic and Gompertz functions respectively, the 4th is transcendental based on trig functions. The 5th is trivial but is the limit of the other functions as they get steeper.
- The Hill and Gompertz functions are asymmetric about the inflection point, while the logistic and arctan functions are symmetric.
- The 6th is of course not a sigmoidal function but is sometimes called a delta function – it is the derivative of the step function. It also has the property that $\int f(x)\delta(x-x_0)dx=f(x_0)$.

making functions dance – how to modify them

To stretch along x axis	y=f(a*x)
To shrink along x axis	y=f(a/x)
To shift along x axis	y=f(x-a) (a>0 shifts right)
To stretch/shrink along y axis	y=a*f(x)
To shift along y axis	y=a+f(x)
to flip about y=x line	x=f(y) or $y=f^{-1}(x)$
to flip about x-axis	y=f(-x)
to flip about y-axis	y=-f(x)

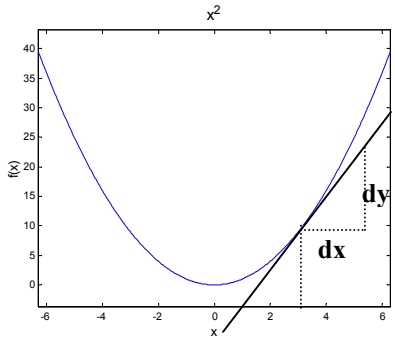
For example compare:
$y=x^2$ vs. $y=3(x-2)^2+5$
The second curve should be shifted right by 2, stretched up by a factor of 3, and shifted up by a factor of 5



also note that the order matters – reflecting about the x axis and then shifting (f(-x-a)) is not the same as shifting and then reflecting about the x axis (f(a-x))

the top 10 things you need to remember from 1st year calculus

1. Derivative=slope - A derivative is the slope of the line tangent to the function $\frac{dy}{dx}$



2. Differential=rate of change in variables – A differential represents how much y changes with a change in x: dy=df(x)dx. Note that for infinitesimally small dx, this is completely accurate, as dx gets bigger, this becomes less accurate (in fact it is a first order or linear approximation).
3. Derivatives and differentials are defined at a single point, but if we use this definition at every point, we get a new function. This is the derivative of f(x) and is denoted f'(x)
4. Simple rules for taking derivatives are:

f(x)=c	f'(x)=0
f(x)=ax	f(x)=a
$f(x)=x^n$	$f'(x)=nx^{n-1}$
f(x)=exp(x)	f(x)=exp(x)
f(x)=ln(x)	f(x)=1/x
f(x)=g(x)*h(x)	f(x)=g'(x)h(x)+g(x)h'(x)
f(x)=g(h(x))	f'(x)=g'(h(x))h'(x)
f(x)=1/g(x)	$f'(x)=-g'(x)/g(x)^2$

5. Derivatives can be taken of derivatives. This yields second derivatives and so on. A second derivative tells how fast the slope is changing. A positive second derivative means the slope is increasing (like the right half of the function above) and hence is concave up.
6. An integral is an antiderivative
7. An integral is an area under a curve
8. An integral is an infinite sum (uncountable)
9. The geometric series:

$$(1-r)(1+r+r^2+r^3+\ldots+r^n) = 1-r^{n+1} \Rightarrow \sum_{i=m}^{n} cr^i = \frac{cr^m(1-r^{n+1})}{1-r} = \frac{cr^m}{1-r} \text{ if } n=\infty \text{ and } |r|<1$$

10. The Taylor expansion:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x-a)^n \approx f(a) + f'(a)(x-a) + f''(a)(x-a)^2/2! + \ldots$$

approximations - how do they do that and should they get away with that?

The Taylors series provides the means for common approximations. For example:
$e^x \approx 1+x+x^2/2+x^3/6+\ldots \approx 1+x$ for $x \approx 0$
$\ln(1+x) = x - x^2/2 + x^3/3 - x^4/12 + \ldots \approx x$ for $x \approx 0$

in a more complicated vein:

$$\prod_{i=1}^{n}(1+x_i) = e^{\sum \ln(1+x_i)} \approx e^{\sum x_i} \approx 1+\sum_{i=1}^{n} x_i$$

why do people do this? – as always to solve problems that otherwise wouldn't be solveable
is it justified – it depends a lot on two things:

- how close x is to a (if a is 0 then it often works for x's of .1 or .2 (.1*.1=.01 which is 1/10 the size of .1)
- how many terms are included – often just the constant and first order (x) term are included but it becomes more and more accurate as more terms are included

Linear algebra - a geometric/visual approach

A study of a certain subset of transformations.
Transformations are functions but conceived of as mapping one set to another set (e.g. to)

Types of transformations

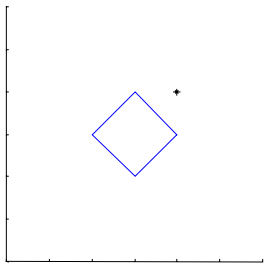
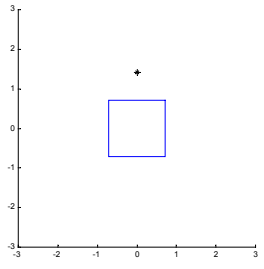
There are a number of categories of transformations of interest to mathematicians:

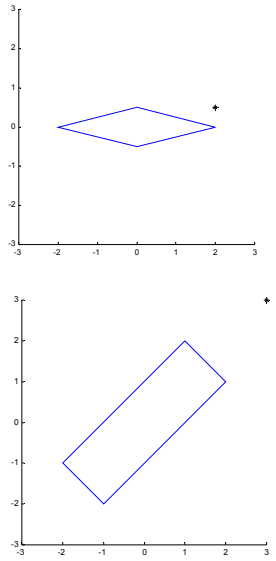
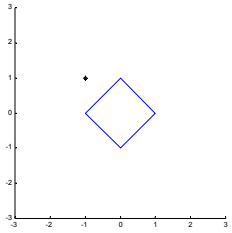
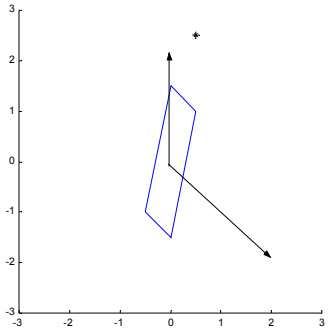
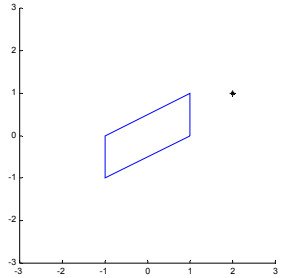
Category	**Functional property**	**Preserves**
Rubber-sheet topology	Continuous function	Dimensionality + # of holes (donut=coffee cup)
Diffeomorophisms	" + differentiable	"+ corners and kinks ($\|x\| \neq x^2$)
Affine maps	AX+b	"+straight lines Preserves dimensionality if A non-singular
Linear maps	AX	" + origin
Orthogonal maps	AX, A orthogonal	" + distance + angles

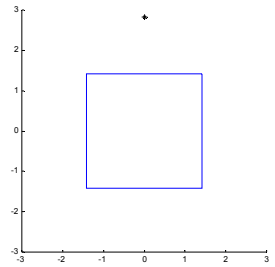
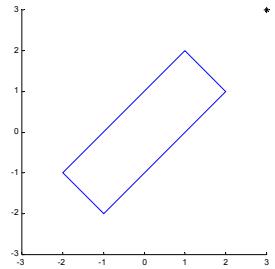
Linear algebra is the study of, no suprirse, linear maps.

Types of linear transformations

For the moment forget matrices. We're just going to look at maps of the real plane to itself. We'll look at the types of visual transformations that are possible. To do this, we are going to follow the fate of a figure and see what it looks like under different transformations:

Identity	
Rotation	

Orthogonal-Stretch	
Reflection	
Non-Orthognal stretch	
Shear	

Combinations: e.g. rotate-stretch	
Combination: e.g. orthostretch-reflection	

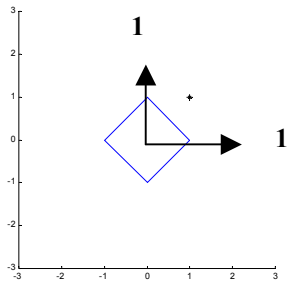
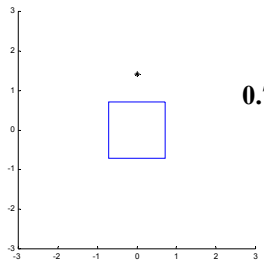
It turns out no other type of transformation is available which preserves lines and maps the origin to the origin.

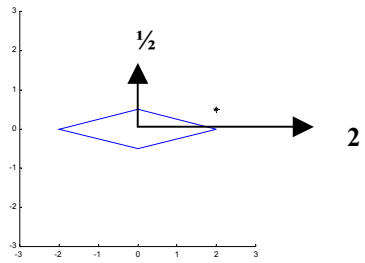
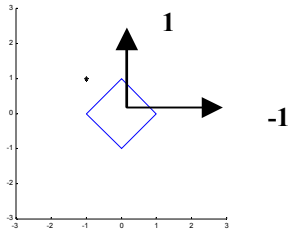
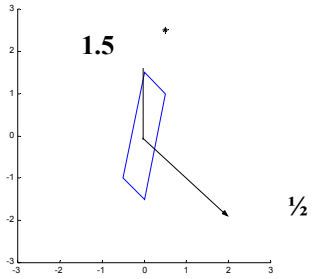
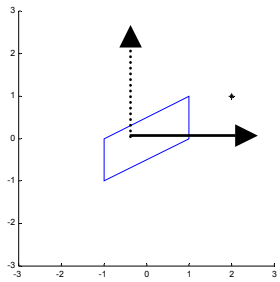
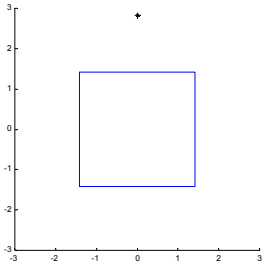
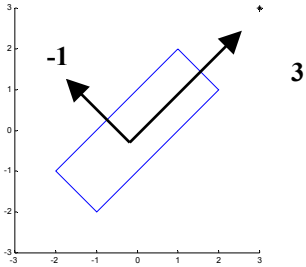
It is now time to introduce the concept of eigenvalues and eigenvectors. Its quite simple:

- Eigenvectors are the axes of the transformation (stretching, reflecting)
- Eigenvalues are the extent of the transformation along the corresponding axis (e.g. >1=stretch, <1=shrink, -1 = reflection)

It is a little more tricky to imagine what the axes are for rotations and shears.
For rotations, you get conjugate pairs. The eigenvalues are: a±bi (a represents the degree of shrinking/expanding that happens) and the eigenvectors are [a ±bi];
For shears, you get one eigenvector in the direction of stretch. You get another non-eigenvector perpendicular to this. This leads to a redundant eigenvalue which.

Lets look at the above transformations with their eigenvectors mapped on. Recall that eigenvectors are only a direction, they can be of any length.

Identity	
Rotation	

Orthogonal-Stretch	
Reflection	
Non-Orthognal stretch	
Shear	
Combinations: e.g. rotate-stretch	
Combination: e.g. orthostretch-reflection	

basis inner (dot) products (distance and angles), embeddings and projections, determinants

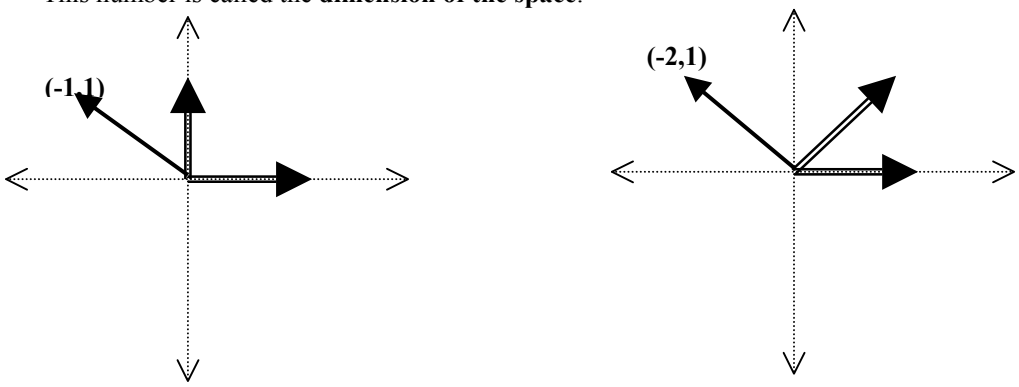
Before moving on to matrices, we need a few more concepts:

The **span** of a set of vectors is the subset of all points that can be reached by linear combiantions of the vectors $(a_1x_1+a_2x_2+\ldots)$

If you add a vector to a set of vectors and it is already in the span of the previous vectors, then the vectors are not **linearly independent**.

A set of linearly independent vectors that span the whole space is called a **basis**. This means that any vector can be written as a linear combination of the basis vectors, and hence we can use coordinate notation to describe a vector: $v=(n_1,n_2,\ldots)_{\mathbf{B}}$. Note that this coordinate notation is dependent on the basis. The same vector has a different coordinate notation under a different basis.

The number of vectors that make up a basis is constant (the vectors may change but the number of them stays constant). This number is called the **dimension of the space**.



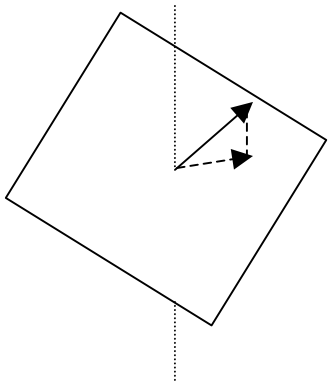
Example: in the above figures, the pair of double lined vectors make up a basis in each case. The entire ∇^2 plane is spanned by the pair of vectors. And conversely, neither vector is a linear combination of the other vector so they are linearly independent. Thus, happily, ∇^2 is 2-dimensional! Note that any third vector can be written as a combination of the basis vectors. Thus the line that goes into the top left quadrant can be written as (-1,1) for the first basis or (-2,1) for the second basis.

Note that our intuition tells us that the first basis is somehow cleaner. The two key facts about the first basis is that the vectors are orthogonal (perpendicular) and of unit length. A basis with these two properties is called **orthonormal**. To see if a basis is orthonormal, we need to know how to define angles and distances. To do this, we need to introduce the concept of the **inner product**. The inner product is a function that takes two vectors as input and returns a single, scalar number. It is often denoted as: (x,y). In ∇^n the obvious inner product is $x_1y_1+x_2y_2+\ldots$ also know as the **dot product**, denoted: $x \bullet y$. In general:

- iff (x,y)=0 then x is perpendicular ($\perp$) to y
- the length of x, denoted ||x||, is defined as $\sqrt{(x,x)}$

for the dot product in ∇^n, we have $x \bullet y = ||x|| \, ||y|| \cos\theta$ which fits into the above general rules. Thus we now have a handle on distances and angles.

The next two concepts we need to introduce pertain to linear transformations when they go from one space to another that have different dimensions. An **embedding** occurs when a space of lower dimension is mapped into a space of higher dimension (e.g. $\nabla^2 \rightarrow \nabla^3$). The transformation can only map into the higher dimensional space a set with dimension of the lower dimensional space. Thus in the example, every transformation would consist of a plane embedded in 3-D space.



A projection is when an object that is outside of a lower dimensional surface, is mapped onto the "best" representation of that object in the lower dimensional surface. In the picture above, the solid vector lies in 3-D space. It is projected onto a 2-D plane. In this case, the "best" projection is the one that leaves only a line $\perp$ to the plane difference between the object and its projection. One general formula for the projection of a line, x, onto another line, y, is:

$$(x,y)\, y / ||y||$$

The final concept which we wish to describe geometrically is the determinant. The **determinant** is a single number that quanitifies how much space is stretched or shrunk by a transformation. For example if it is stretch by a factor of three in one direction and shrunk by a factor of ½ in a $\perp$ direction, then we would want the determinant to be 3/2. Note that this is simply the product of all the eigenvalues. In general, this works. The determinant is a measure of how much space is shrunk/expanded by a transformation and is given simply as the product of the eigenvectors.

How this maps to matrices

Matrices as basis changes

You will recall that in the example of bases, the representation of a vector was different depending on the basis. This is inconvenient – it would be nice to be able to switch rapidly between coordinate notations for different bases. It turns out you can. Simply create a matrix that describes the source basis as column vectors in the destination basis. Then matrix multiplication converts bases!

In our example, lets take the source basis as the basis on the left. Recall that our sample vector was denoted (-1,1) in the source (left) basis (and as (-2,1) in the destination – right – basis). Note that the source basis (on the left) can be written as [1 0] and [-1 1] in the destination basis. Thus if we use the matrix:

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

as a conversion matrix we can switch between the two basis. And indeed if we take the matrix times times the vector notates in the source basis (-1,1) we get (-2,1) which is the vector in the destination basis! To go backwards, we use the same algorithm in reverse (write a matrix of column vectors of the source – right – basis in terms of the destination – left – basis). If we do this, we get:

$$B=\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

We note that if we take AB (matrix multiplication), we get the identity out and hence these two matrices are inverses of each other.

Matrices as representations of tranformations

A more interesting use of matrices is to represent transformations. Recall that we studied linear transformations as abstract, geometrical objects. This is intuitive but imprecise. To make things precise we need to write our linear transformations down as combinations of numbers. It turns out that writing a transformation from $\nabla^n \rightarrow \nabla^m$ as a mXn matrix works well. Then the vector, X, is mapped into AX (where we use standard matrix arithmetic).

The problem is, as we know, that there are many different basis. Which basis to write a transformation relative to?? Well the "obvious" and "natural" basis is to use the eigenvectors as a basis. Then our matrices will take a

particularly simple form – a diagonal matrix with the eigenvalues down the diagonal. There are two exceptions to this. Rotations will not be a 2X2 block of values representing the angle of rotation $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and shears will have the single eigenvalue, λ, repeated on the diagonal and with the value 1 on the super diagonal $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$.

This makes it very easy to convert a transformation into a matrix, but there are two problems. First it requires us to use a different basis for every transformation. Second, it doesn't tell us what basis is used. Thus, we usually choose to write our transformations as matrices relative to the standard basis. This means we must do a change of basis between the standard basis and the eigenvector basis. Recall from above that this simply requires that we write a matrix of column vectors, each column vector containing the destination basis (eigenvector) in terms of the standard basis. Thus, every transformation can be written as:

$V^{-1} J V$

where J is a matrix in Jordan normal form (i.e. it has eigenvalues down the diagonal and in cases of shear or rotation it has a number on the super/sub-diagonal or superdiagonal respectively). V is a matrix consisting of the eigenvectors written in the coordinates of the standard basis, and V^{-1} is the inverse matrix that takes from the basis of the eigenvectors back to standard basis. Thus with matrix multiplication, AX, we take a vector in standard basis, convert it to eigenvector basis, apply a simple, nearly diagonal matrix to it that performs stretches, reflections, rotations, and shears, and then take this new point and convert it back from the eigenvector basis to the standard basis.

This is most obvious in the example of the non-orthogonal stretch where we stretched by a factor of 1.5 and shrunk by 0.5 in the directions of (0,1) and (1,-1) respectively. Lots follow our cookbook and see if this gives us the right matrix. We know how to write the matrix that goes from the eigenvector basis to the standard basis!

$$\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

To get the matrix that goes from the standard to the eigenvector basis, we need to right the standard basis in the eigenvector coordinates. Or we could just take the inverse of the above matrix. Either way we get:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

The diagonal matrix has stretches by 1.5 and 0.5 so we get:

$$\begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Putting this all together, we get:

$$\begin{bmatrix} 0.5 & 0 \\ 1 & 1.5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

And, indeed, the matrix on the left performs exactly the non-orthogonal stretch we expected.

Here the matrices that do all of the transformations described above

Identity	1 0 0 1
Rotation	Cos π/4 -Sin π/4 Sin π/4 Cos π/4
Orthogonal-Stretch	2 0 0 0.5
Reflection	-1 0 0 1

Non-Orthognal stretch	0.5 0 1 1.5
Shear	1 1 0 1
Combinations: e.g. rotate-stretch	√2 -√2 √2 √2
Combination: e.g. orthostretch-reflection	1 2 2 1

How does the above generalize to n-dimensions? Fairly simply. Rotations always take place in a two dimensional plane. Shears can take up 2,3, or n dimensions. Thus to write an n-dimensional matrix, we need to see which directions are stretch/shrunk/refelected which planes are rotated, and which dimensions are sheared. We can then write this as a matrix V^{-1} J V where J is in Jordan normal form (diagonal with eigenvalues/stretch factors on the diagonal and 2X2 blocks for rotations and 1's in the superdiagonal for shears. Thus a potential 4-D transformation would be:

$$[v1\ v2\ v3\ v4]^{-1}\begin{bmatrix}\lambda_1 & 0 & 0 & 0\\ 0 & \lambda_2 & 0 & 0\\ 0 & 0 & \lambda_3 & 1\\ 0 & 0 & 0 & \lambda_3\end{bmatrix}[v1\ v2\ v3\ v4]$$

where [v1 v2 v3 v4] is a matrix with column vectors. This transformation will stretch by factors λ_1 and λ_2 in directions v1 and v2. It will do a shear in direction v3 by amount λ_3 taking along direction v4.

What they usually teach you in linear algebra

Most of linear algebra is taken up with teaching you how to multiply matrices, take their determinants and their inverses. Here we take the position that computers do this numerical computation. All that you need is to understand the concept.

Several things they teach in linear algebra become trivial if you understand the above:

matrix A is invertible $\Leftrightarrow$ $\det(A) \neq 0$ (this means no eigenvalues are 0, so no dimension is collapsed, so it is a 1-1 map, so it is invertible)

an orthogonal matrix is one whre the the eigenvalues are all $\pm 1 \Leftrightarrow A^{-1} = A^T$

a symmetric matrix is one where $A=A^T$ and it is always an orthogonal stretch so it always has orthogonal eigenvectors

an eigenvector has the property that $Ax=\lambda x \Leftrightarrow Ax-\lambda Ix=0 \Leftrightarrow (A-\lambda I)x=0 \Leftrightarrow \det(A-\lambda I)=0 \Leftrightarrow \lambda$ is root of characteristic polynomial

a positive definite matrix is one where $x'Ax \geq 0\ \forall x \Leftrightarrow \lambda_i \geq 0\ \forall i$ and similarly for negative definite matrices

the solution to Ax=b can have one of three outcomes:

1) b is in the span of the columns of A and A is non-singular: there is exactly one solution $x=A^{-1}b$
2) b is in the span of the columns of A and A is singular: there are an infinite number of solutions
3) b is not in the span of the columns of A: no x solves the equation

All cases are covered by using the singular value decomposition:

AA^T is a symmetric matrix and hence has orthogonal eigenvectors (as does A^TA)

Let U be the matrix of the eigenvectors of AA^T, V be the matrix of eigenvectors of A^TA, and $\lambda=\sqrt{}$eigenvalues of A^TA=eigenvalues of A^TA. Then:

$A=U\Lambda V^T$ and let $A^G = V\ 1/\Lambda\ U^T$

if A is invertible, $A^G=A^{-1}$

$A^G b$ gives an optimal solution in all 3 cases

the 15 minute version of 2nd year calculus

f: $\nabla^n \rightarrow \nabla$

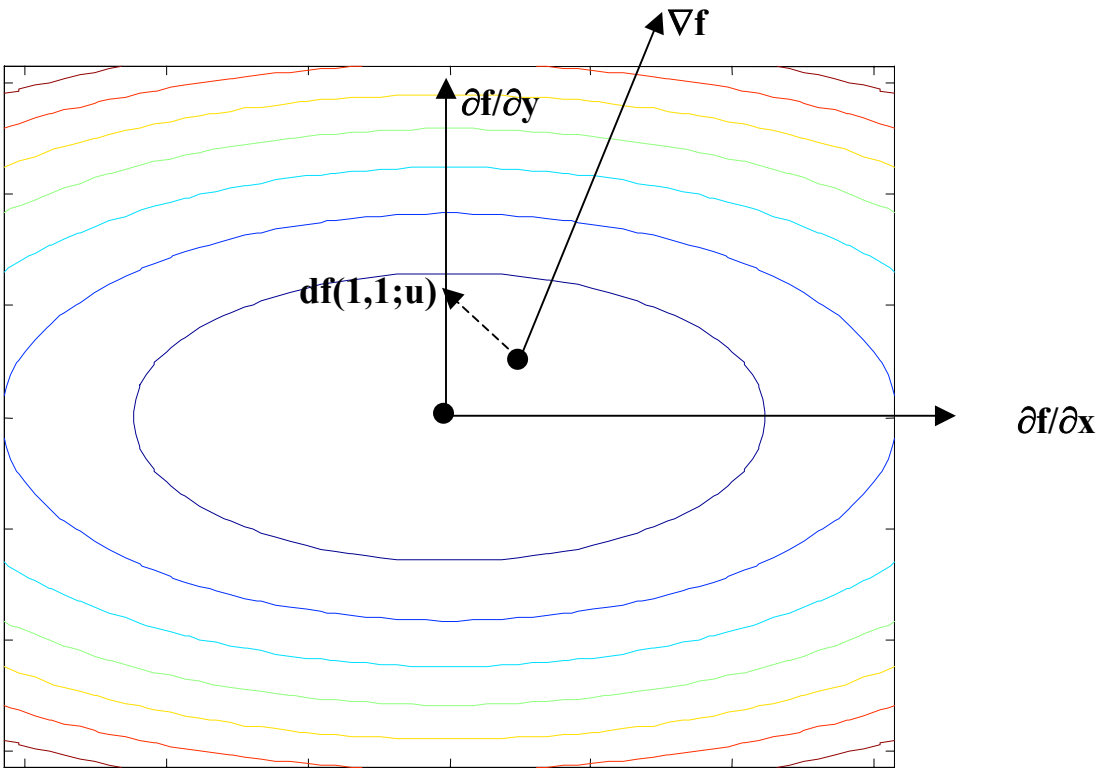
These functions assign a single value to every point in ∇^n (such as temperature to every point on the surface of the earth ∇^2). They can be though of as contour maps (assign height to every point). The key question is what constitutes a derivative?

One valid concept is to take the derivative in the principle directions: x,y, etc. This is called partial derivatives. Hence $\partial f/\partial x$ is the slope of f in the x direction

Another valid concept is when you specify a direction u and then take the slope in that direction. This called a directional derivative

The final valid concept is to take the slope in the direction of maximal ascent (greatest slope). This is called the gradient.

To calculate the gradient we need to use partial derivatives, and in particular $\nabla f=[\partial f/\partial x, \partial f/\partial y, \ldots]$ is the gradient. To calculate a directional derivative, we need to know the gradient, and in particular the derivative in the direction u is given by $\nabla f \bullet u$ (usually u is chosen to be of unit length).



For example take the function: $f(x,y)=x^2+3y^2$. Then:

$\partial f/\partial x=2x$

$\partial f/\partial y=6y$

$\nabla f=[2x,6y]$ so $\nabla f(0,0)=(0,0)$ and $\nabla f(1,1)=(2,6)$

The directional derivative in the direction (-1,1) at (0,0) is (0,0) while at (1,1) it is 8 $(6,2)\bullet(1,1)$

Finally, what works as a second derivative? This is a matrix called the Hessian: $\begin{bmatrix} \frac{\partial f}{\partial x \partial x} & \frac{\partial f}{\partial x \partial y} & \\ \frac{\partial f}{\partial y \partial x} & \frac{\partial f}{\partial y \partial y} & \dots \\ & \dots & \end{bmatrix}$

Note that $\partial f/\partial x \partial x$ is usually written as $\partial^2 f/\partial x^2$ and that since under most conditions $\partial f/\partial x \partial y = \partial f/\partial y \partial x$ the matrix is usually symmetric. The determinant of the hessian can tell us much about the behavior of a function

The generaliation of Taylor's theorem to $f:\nabla^n \rightarrow \nabla$ is:

$f(x+h)=f(x)+\nabla f \bullet h + 1/2\ h'\ H\ h + o(h)$

Note this shows us that for a critical point to occur (maxima, minima, saddle point) the gradient must be zero (i.e all partials must be zero) and the the Hessian must be positive or negative definite to give a minima or a maxima respectively.

$f: \nabla^n \rightarrow \nabla^m$

There is one meaningful derivative, defined in terms of partials and it is called the Jacobian:

$J=\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots \\ & \dots & \end{bmatrix}$

why a Taylors series is really a linear algebra problem

So far we have only talked about systems where the dimensionality is finite? So what – why would you want an infinite dimensional system? It turns out that the set of all functions makes a fairly good infinite dimensional linear algebra system. There are several bases for the set of all functions:

x^n is a basis but not orthogonal

$a_n \sin nx + b_n \cos nx$ is an orthogonal system for functions that are periodic (or defined on the finite interval $[0,2\pi]$

e^{nx} is a basis and orthogonal but may be uncountably infinite (the other two are countable infinite, i.e. $n \in \wedge$)

In all above the systems, the inner product function is an integral.

an application of linear algebra to state space models

Suppose your state space is $x_{t+1}=Ax_t$. Then by simple recursion you can show that $x_t=A^t x_0$. Then, assuming A is diagnolizable (i.e. no shearing), we can write $A=V^{-1}DV$. Then $A^t = V^{-1} D^t V$ and D^t can be calculated simply by raising each diagonal element to t. Thus we have a simple means of calculating.

However, suppose what we're really interested in is the long term behavior of the system ($t \rightarrow \infty$). Recall that each eigenvalue/vector pair has $Ax=\lambda x$ (i.e. behavior on an eigenvector axis is simple stretching by the eigenvalue). Now take any start point x_0. Write it in terms of the eigenbasis of A: $x_0=c_1x_1+c_2x_2+ \dots$ where the x_i are eigenvectors and the c_i are the coordinates of x_0 in terms of the eigenbasis. Then $A^t x_0 = c_1\lambda_1^t x_1 + c_2\lambda_2^t x_2+\dots$ This is called the spectral theorem. As t gets large, all terms will be small compared to the one with the largest λ, say λ_1. Thus as $t \rightarrow \infty$, x_t goes to a multiple of x_1 and grows in each time step by a factor of λ_1.

Deterministic dynamical systems

solving differential equations (the 15 minute version and why you don't need more)

So, we have a state space model that specifies our problem using a deterministic rule. One approach would be to solve the state space model so that we know behavior at every point in time. This would take the form $x_t=f(t,x_0,p)$. This means we can calculate x at any time t if we know the initial conditions, x_0, and the parameters p.

A simple example is the logistic equation:

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

we will solve this using a technique known as separation of variables: put all dN and N's on the left and all dt's and t's on the right:

$$\frac{dN}{rN(1-N/K)}=dt$$

in theory, all we need to do to solve this is to take the integral on both sides, but to find the integral on the left, we need to rewrite it as:

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

which gives:

$$\int\frac{1}{rN}dN+\int\frac{1}{K-N}dN=\int dt$$

which gives:

$$\frac{\ln N}{r}+-\ln(K-N)+C=t \Rightarrow \frac{1}{r}\ln\frac{N}{K-N}=t+C \Rightarrow \frac{N}{K-N}=e^{rt+C} \Rightarrow N=Ke^{rt+C}-Ne^{rt+C} \Rightarrow$$

$$N(1+e^{rt+C})=Ke^{rt+C} \Rightarrow N=\frac{Ke^{rt+C}}{1+e^{rt+C}} \Rightarrow N=\frac{K}{1+ce^{-rt}}$$ (multiply top and bottom by $e^{-(rt+C)}$)

note that the C comes from integration (antiderivatives lose a constant, so we have to add it back in when we integrate. Also note that we have been somewhat casual with the exact meaning of C, it keeps changing as we manipulate the equation, all really have kept true is that it is some constant independent of N,t. Now we solve by using the initial condition: at t=0, $N=N_0$:

$$N_0=\frac{K}{1+c} \Rightarrow c=\frac{K-N_0}{N_0}$$

We now can predict N_t at any time t if we know N_0, K and r

Very briefly, we can also solve linear systems such as:

$y'' + 3y' +2y=0$

we just turn this into:

$D^2-3D+2=0$ where $D=d/dt$

we know the solution here:

$D=-1$ or -2

so the solution must be of the form

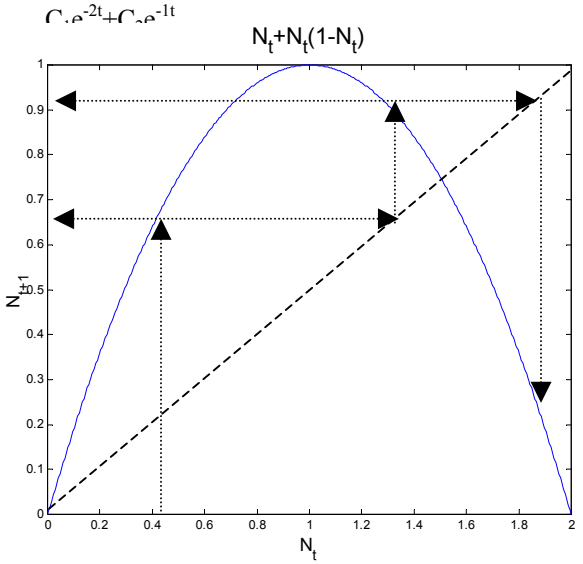
$C_1e^{-2t}+C_2e^{-1t}$

Y nts (C_i) are determined by initial conditions.

(s makes it easier to translate to continuous time (the

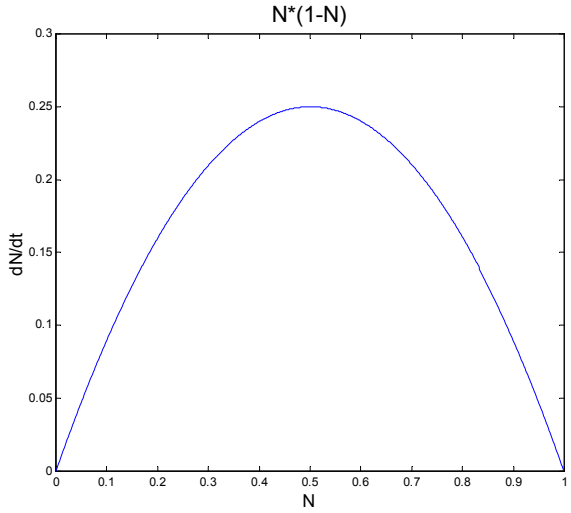
s ı is more tractable.

v vs N_t. We can predict what will happen by cobwebbing



1-D continuous model: $\frac{dN}{dt} = g(N(t))$

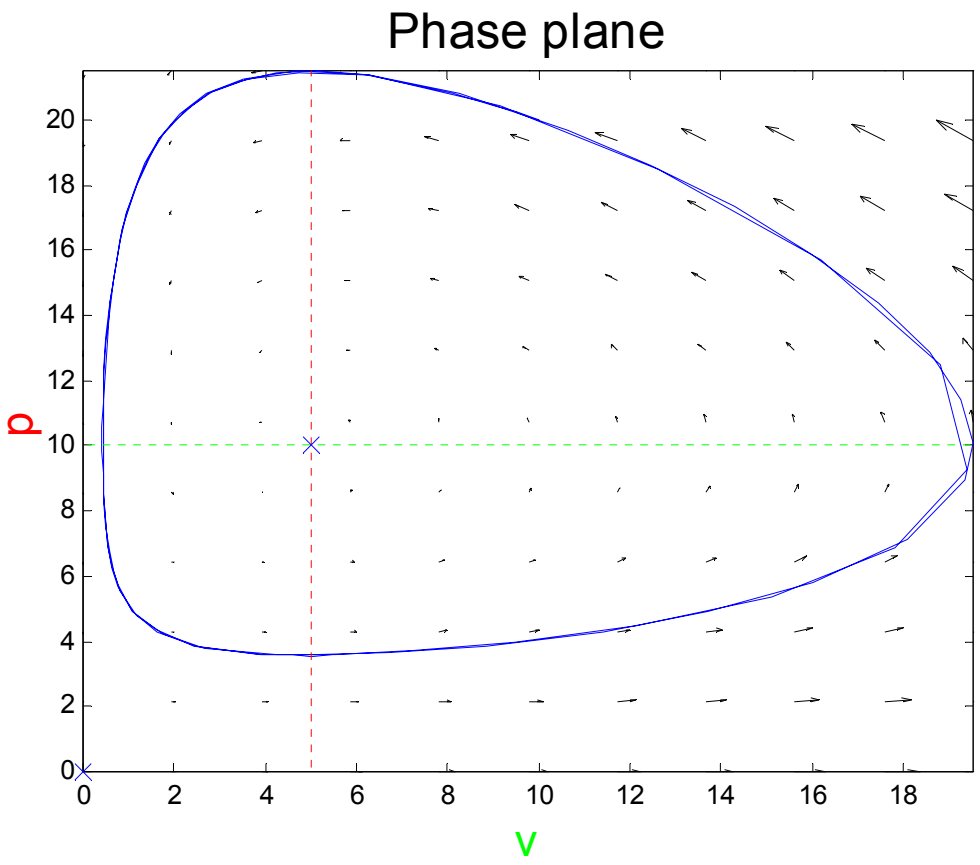
For a 1-D continuous model we plot something similar to the return map but which is technically not the return map which we will call increment map:



2-D continuous: $\frac{dN}{dt} = G(N(t))$

No plot is useful for 2-D discrete or 3-D and above. However, another plot can be useful for a 2-D continuous system. This is called a phase plane. It is a generalization of the increment map. A straight generalization would end up as a 3-D plot the x & y dims would be the two variables and the Z dimension (out of the page) would be the increment (dN/dt). However we typically drop the 3rd dimension. On this we can plot several things:

- isoclines – lines where there is no increment for one variable or the other
- positive and negative increment regions for a variable (separated by the isocline)
- flows – the direction of flow from a given point
- trajectories – an actual path



asymptotic behavior - stability analysis, chaos and everything in between

However, there are several problems with this. First solving this equation was very messy and it is very simple. Solving more complicated equations is often impossible. Second, it may be useful predictively, but it is not useful qualitatively – it is not told us in general what to expect regardless of the exact values of r, K, and N_0 (although simple inspection does show that as $t \rightarrow \infty$, $N \rightarrow K$).

What is usually much easier to do and sometimes more useful is to do **asymptotic analysis**. This involves predicting what happens in the long run (as $t \rightarrow \infty$)

Asymptotic analysis is organized around 4 types of organizing structures and 3 types of approaches.

There are four possible behaviors in the long run:

- Point equilibrium – the system moves to one point and stays there
- Cycle - the system moves to a repeating cycle
- Quasi-periodic – the system moves to an almost repeating cycle
- Chaos – the system moves to a finite region and bounces around in a seemingly random fashion forever
- Goes to infinity – is another option. This is not really of the same status as the above 4, but many systems do this

Another name for these organizing sets is the **α-set** or **ω-set**. The ω-set is the limit set that the system spends its time getting infinitely closer to (for all but an insignificantly small amount of time where transient behavior happens). The α-set is the reverse – it is the ω-limit set that you get if you take a point and run the system backwards around a repellor.

The above 4 organizing structures all have the property that if the system starts there, the system will stay there. However, to be of interest we need to see what happens in the nieghborhood of the organizing structure. A point that starts near the organizing structure can do one of three things:

move (exponentially fast) towards the structure
move (exponentially fast) away from the structure
stay exactly the same distance away (called neutral)

This behavior can happen in each eigenvector direction (one for each dimension of the space – e.g. 3 direction for the 3-D space). Thus we can have a system that moves toward in two directions and moves away in one direction. The directions where it moves towards is called the **stable manifold**, the directions in which it moves away is called the **unstable manifold** and the directions where it doesn't move is called the **center manifold**. Every eigenvector must be a member of one of these manifolds. If all directions are part of the stable manifold, it is called an **attractor**. If even one direction is part of the unstable manifold, it is called a **repeller.** If one direction is stable and one is unstable, it is called a **saddle**.

Point equilibria

This is the most boring but most well understood case. Here the trajectory ends in a point. To find such points we have either:

$N^* = f(N^*)$ or $dN\backslash dt|_{N^*} = g(N^*)=0$

The normal question is whether the point is attracting, repelling, or a saddle. We have two methods for studying this.

Lyupanov's 1st method

We can linearize around the equilibrium point N*. We use the Taylor's series expansion to linearize around the equilibrium point.

$n=N-N^*$ (i.e. n=deviation from equilibrium)

$dn/dt=g(N^*+n)=g(N^*)+ (Jg|_{N^*})\, n + o(n)$ where J is the Jacobian of g or G evaluated at $N=N^*$

There is a theorem which shows that no matter how non-linear the system is, around the equilibrium it behaves the same as the linearized the system.

so $dn/dt=g(N^*)+ (Jg|_{N^*})n$ but $g(N^*)=0$ if it is an equilibrian

so $dn/dt=Jn$

has simple solution: $n=e^{Jt}$ or $N=N^*+e^{Jt}$

Now this system has a simple solution $N(t)=N^*+e^{Jt}$. The behavior of this system is in turn determined by the eigenvalues of J – if they are all negative the system consists of decaying exponentials back towards N*. If they are all positive, the system consists of increasing exponentials away from N*. If they are some negative and some positive then there are corresponding directions where it is moving towards exponentially and directions where it is moving away exponentially. This is a saddle and leads to a phenomenon called a **fly-by**. If they are zero then they don't move towards or away. If the eigenvalues have an imaginary component, then there is rotation as it does it. This leads to the following cases:

stable node: both (all) eigenvalues negative, no imaginary part
unstable node: both (all) eigenvalues positive, no imaginary party
stable/unstable focus: both (all) eigenvalues negative/positive with an imaginary part
saddle: some eigenvalues positive and negative
center: all eigenvalues strictly imaginary

For example: take the Lotka-Volterra predator prey system:

$\frac{dV}{dt} = f_1(V,P) = V(b-aP)$ $\quad \frac{dP}{dt} = f_2(V,P) = P(-d+kaV)$

V^*,P^* occur where $f_1=f_2=0$

$0=V(b-aP) \Rightarrow P^*=b/a$ and $0=P(-d+kaV) \Rightarrow V^*=d/ka$

i.e. equilibrium is (b/a,d/ka)

To see if this is stable or not we need the jacobian

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} b-aP & -aV \\ kaP & -d+akV \end{bmatrix}\Big|_{N^*} = \begin{bmatrix} b-\frac{ab}{a} & -\frac{ad}{ka} \\ \frac{kab}{a} & -d+\frac{akd}{ak} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-d}{k} \\ kb & 0 \end{bmatrix}$$

the eigenvalues solve the equation $\lambda^2 + db = 0$ or $\lambda = 0 \pm i\sqrt{bd}$

the real part of all eigenvalues is zero and there is an imaginary part – hence the system is a center

Lyupanov's 2nd method

sensitivity to initial conditions and parameter values

what can we study if its chaotic – ergodic properties

diffusion equations

Random dynamical systems (stochastic processes)

the 10 things you need to know about probability

the top 10 probability distributions and how to remember them

Markov models

Markov Chains

Counting (Poisson) and Birth Death processes

Stochastic Differential equations

Jump equations

Static systems - easier to solve than dynamic systems

optimization (0 players affect the system)

control theory (1 player affects the system)

game theory (2+ players affect the system)

Descriptions of complex data

Multivariate statistics

Regression

GLIM

Neural nets

Reduction of variables

Principle components

CART/Trees

Classification

Fractals - cutting through the (very thick) hype

Timeseries data

Chaos analysis

Fitting non-linear models

ARIMA

Wavelets

Deterministic dynamical systems

solving differential equations (the 15 minute version and why you don't need more)

So, we have a state space model that specifies our problem using a deterministic rule. One approach would be to solve the state space model so that we know behavior at every point in time. This would take the form $x_t=f(t,x_0,p)$. This means we can calculate x at any time t if we know the initial conditions, x_0, and the parameters p.

A simple example is the logistic equation:

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

we will solve this using a technique known as separation of variables: put all dN and N's on the left and all dt's and t's on the right:

$$\frac{dN}{rN(1-N/K)}=dt$$

in theory, all we need to do to solve this is to take the integral on both sides, but to find the integral on the left, we need to rewrite it as:

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

which gives:

$$\int\frac{1}{rN}dN+\int\frac{1}{K-N}dN=\int dt$$

which gives:

$$\frac{\ln N}{r}+-\ln(K-N)+C=t \Rightarrow \frac{1}{r}\ln\frac{N}{K-N}=t+C \Rightarrow \frac{N}{K-N}=e^{rt+C} \Rightarrow N=Ke^{rt+C}-Ne^{rt+C} \Rightarrow$$

$$N(1+e^{rt+C})=Ke^{rt+C} \Rightarrow N=\frac{Ke^{rt+C}}{1+e^{rt+C}} \Rightarrow N=\frac{K}{1+ce^{-rt}}$$ (multiply top and bottom by $e^{-(rt+C)}$)

note that the C comes from integration (antiderivatives lose a constant, so we have to add it back in when we integrate. Also note that we have been somewhat casual with the exact meaning of C, it keeps changing as we manipulate the equation, all really have kept true is that it is some constant independent of N,t. Now we solve by using the initial condition: at t=0, $N=N_0$:

$$N_0=\frac{K}{1+c} \Rightarrow c=\frac{K-N_0}{N_0}$$

We now can predict N_t at any time t if we know N_0, K and r

Very briefly, we can also solve linear systems such as:

$y'' + 3y' +2y=0$

we just turn this into (we can do this if we transform into laplace space where differentiation by x = multiplication by x):

$D^2-3D+2=0$ where $D=d/dt$

we know the solution here:

$D=-1$ or -2

so the solution must be of the form

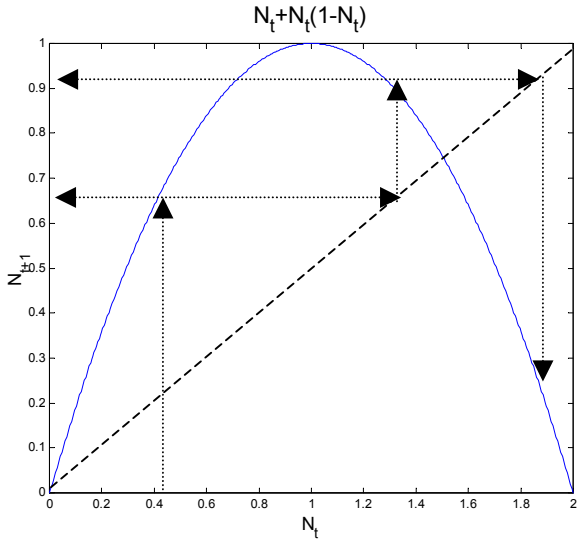
$$C_1e^{-2t}+C_2e^{-1t}$$

You can check that this solution works. The values of the constants (C_i) are determined by initial conditions.

Graphing deterministic systems:

1-D discrete model: $N_{t+1}=f(N_t)$

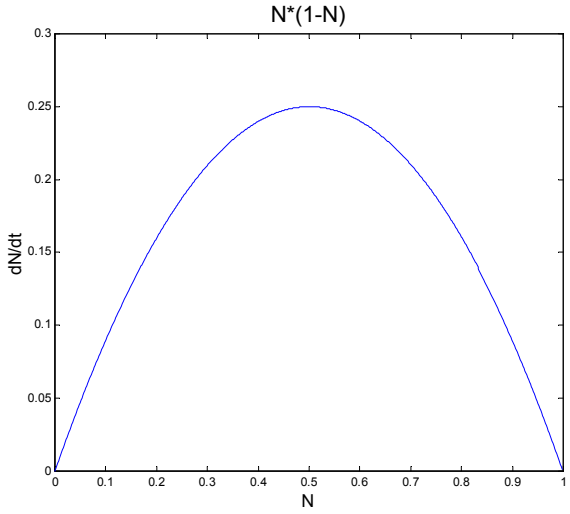
These are also described in the form $N_{t+1}=N_t+f(N_t)$. This makes it easier to translate to continuous time (the same function, f, can be used, but mathematically, the other form is more tractable.



For a 1-D discrete model, we plot the return map – N_{t+1} vs N_t. We can predict what will happen by cobwebbing which is just really a graphical shortcut for the math.

1-D continuous model: $\frac{dN}{dt} = g(N(t))$

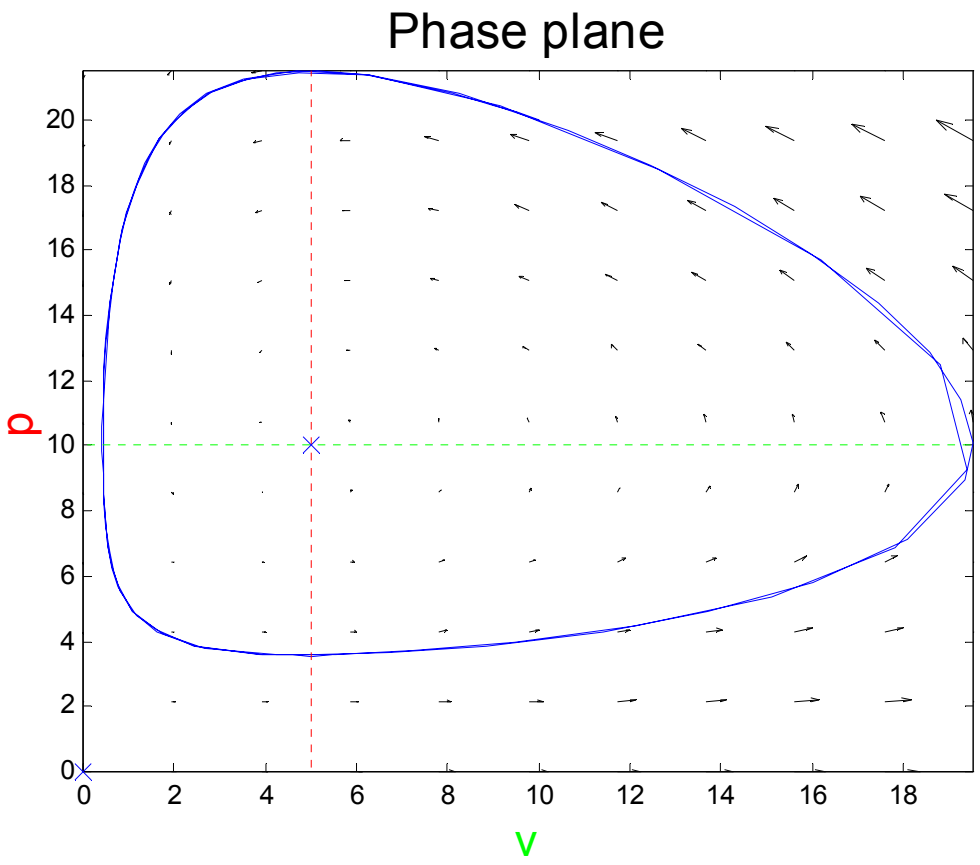
For a 1-D continuous model we plot something similar to the return map but which is technically not the return map which we will call increment map:



2-D continuous: $\frac{dN}{dt} = G(N(t))$

No plot is useful for 2-D discrete or 3-D and above. However, another plot can be useful for a 2-D continuous system. This is called a phase plane. It is a generalization of the increment map. A straight generalization would end up as a 3-D plot the x & y dims would be the two variables and the Z dimension (out of the page) would be the increment (dN/dt). However we typically drop the 3rd dimension. On this we can plot several things:

- isoclines – lines where there is no increment for one variable or the other
- positive and negative increment regions for a variable (separated by the isocline)
- flows – the direction of flow from a given point
- trajectories – an actual path



asymptotic behavior - stability analysis, chaos and everything in between

However, there are several problems with this. First solving this equation was very messy and it is very simple. Solving more complicated equations is often impossible. Second, it may be useful predictively, but it is not useful qualitatively – it is not told us in general what to expect regardless of the exact values of r, K, and N_0 (although simple inspection does show that as $t \rightarrow \infty$, $N \rightarrow K$).

What is usually much easier to do and sometimes more useful is to do **asymptotic analysis**. This involves predicting what happens in the long run (as $t \rightarrow \infty$)

Asymptotic analysis is organized around 4 types of organizing structures and 3 types of approaches to the organizing sets.

There are four possible behaviors in the long run:

- **Point equilibrium** – the system moves to one point and stays there
- **Cycle** - the system moves to a repeating cycle
- **Quasiperiodic** – the system moves to an almost repeating cycle
- **Chaos** – the system moves to a finite region and bounces around in a seemingly random fashion forever
- **Goes to infinity** – is another option. This is not really of the same status as the above 4, but many systems do this

Another name for these organizing structures is an **invariant set** — any point that starts on an organizing structure stays on this organizing structure for ever. Another name for these organizing sets is the **α-set** or **ω-set**. The ω-set is the limit set that the system spends its time getting infinitely closer to (for all but an insignificantly small amount of time where transient behavior happens). The α-set is the reverse – it is the ω-limit set that you get if you take a point and run the system backwards around a repellor.

These organizing sets can also be distinguished by their dimension:

Organizing set	Dimension
Point	0
Cycle	1
Quasiperiodic	1,2,3,4, …
Chaos	1.x, 2.x, ….

The above 4 organizing structures all have the property that if the system starts there, the system will stay there. However, to be of interest we need to see what happens in the neighborhood of the organizing structure. A point that starts near the organizing structure can do one of three things:

move (exponentially fast) towards the structure
move (exponentially fast) away from the structure
stay exactly the same distance away (called neutral)

This behavior can happen in each eigenvector direction (one for each dimension of the space – e.g. 3 direction for the 3-D space). Thus we can have a system that moves toward in two directions and moves away in one direction. The directions where it moves towards is called the **stable manifold**, the directions in which it moves away is called the **unstable manifold** and the directions where it doesn't move is called the **center manifold**. Every eigenvector must be a member of one of these manifolds. If all directions are part of the stable manifold, it is called an **attractor**. If even one direction is part of the unstable manifold, it is called a **repeller.** If one direction is stable and one is unstable, it is called a **saddle**.

Another way to look at it is to look at whether the map is shrinking or stretching space around the organizing structure. This can be determined by looking at the determinant of the Jacobian. If it is <1 then space is shrinking and it is an attractor.

Attractor #1 - Point equilibria

This is the most boring but most well understood case. Here the trajectory ends in a point. To find such points we have either:

$N^* = f(N^*)$ or $dN\backslash dt|_{N^*} = g(N^*)=0$

The normal question is whether the point is attracting, repelling, or a saddle. We have two methods for studying this. There are two types of stability:

neutral stability – if the it is perturbed off the equilibrium point, it doesn't move any further away
asymptotic stability – if perturbed off the equilibrium point, it will move asymptotically back to the point

There are two major methods for testing the stability of a point, both due to Lyupanov.

Lyupanov's 1st method

We can linearize around the equilibrium point N*. We use the Taylor's series expansion to linearize around the equilibrium point.

$n=N-N^*$ (i.e. n=deviation from equilibrium)

$dn/dt=g(N^*+n)=g(N^*)+ (Jg|_{N^*})\, n + o(n)$ where J is the Jacobian of g or G evaluated at $N=N^*$

There is a theorem which shows that no matter how non-linear the system is, around the equilibrium it behaves the same as the linearized the system.

so $dn/dt=g(N^*)+ (Jg|_{N^*})n$ but $g(N^*)=0$ if it is an equilibrian

so $dn/dt=Jn$

has simple solution: $n=e^{Jt}$ or $N=N^*+e^{Jt}$

Now this system has a simple solution $N(t)=N^*+e^{Jt}$. The behavior of this system is in turn determined by the eigenvalues of J – if they are all negative the system consists of decaying exponentials back towards N*. If they are all positive, the system consists of increasing exponentials away from N*. If they are some negative and some positive then there are corresponding directions where it is moving towards exponentially and directions where it is moving away exponentially. This is a saddle and leads to a phenomenon called a **fly-by**. If they are zero then they don't move towards or away. If the eigenvalues have an imaginary component, then there is rotation as it does it. This leads to the following cases:

stable node: both (all) eigenvalues negative, no imaginary part
unstable node: both (all) eigenvalues positive, no imaginary party
stable/unstable focus: both (all) eigenvalues negative/positive with an imaginary part
saddle: some eigenvalues positive and negative
center: all eigenvalues strictly imaginary

For example: take the Lotka-Volterra predator prey system:

$\frac{dV}{dt} = f_1(V,P) = V(b-aP)$ $\quad$ $\frac{dP}{dt} = f_2(V,P) = P(-d+kaV)$

V^*,P^* occur where $f_1=f_2=0$

$0=V(b-aP) \Rightarrow P^*=b/a$ and $0=P(-d+kaV) \Rightarrow V^*=d/ka$

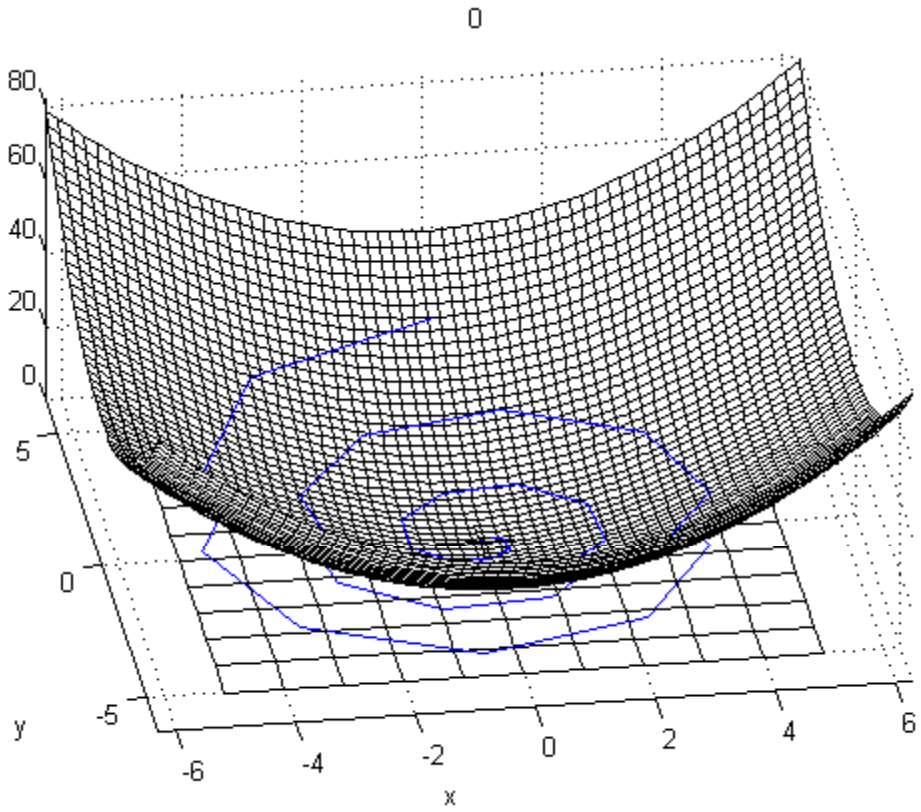
i.e. equilibrium is (b/a,d/ka)
To see if this is stable or not we need the jacobian

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} b - aP & -aV \\ kaP & -d + akV \end{bmatrix}\Big|_{N^*} = \begin{bmatrix} b - \frac{ab}{a} & -\frac{ad}{ka} \\ \frac{kab}{a} & -d + \frac{akd}{ak} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-d}{k} \\ kb & 0 \end{bmatrix}$$

the eigenvalues solve the equation $\lambda^2 + db = 0$ or $\lambda = 0 \pm i\sqrt{bd}$
the real part of all eigenvalues is zero and there is an imaginary part – hence the system is a center

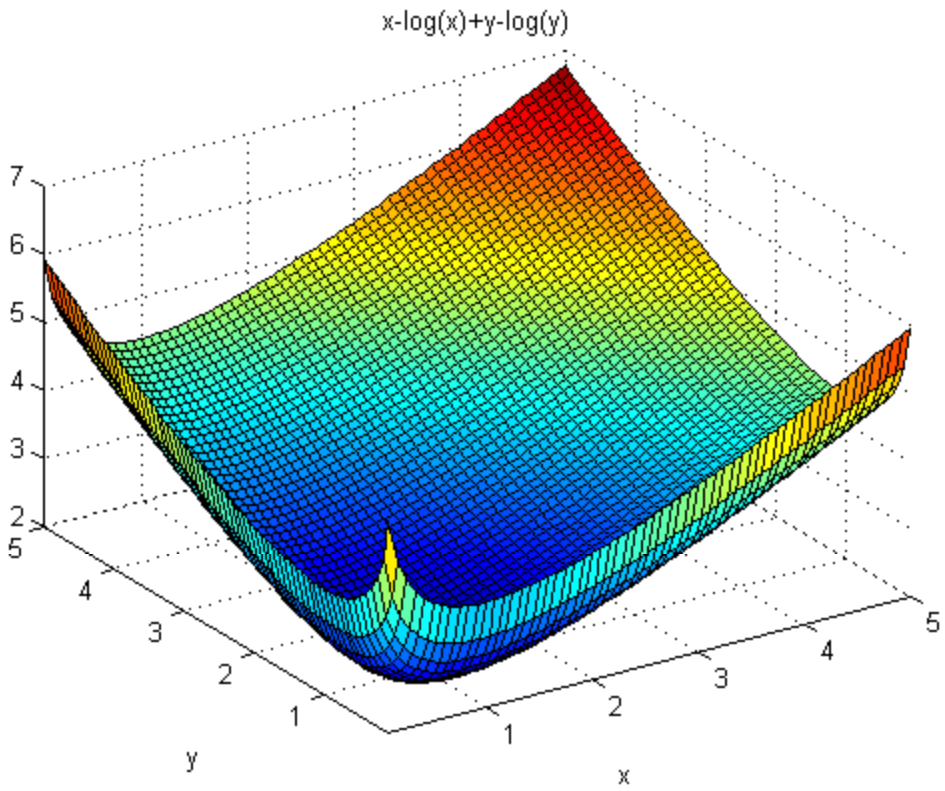
Lyupanov's 2nd method

Lyapunov provided a 2nd method for testing for stability. It is actually much more robust because it tells us over what domain the point attracts. His second is based on finding a Lyapunov function. A Lyapunov function is basically an energy function – it maps every point in phase space to a scalar value representing energy. Lyapunovs insight is that if we can show that at every point the state variables go down the Lyapunov function (lower energy), then we must end up at a minimum of the Lyupanov function.



The exact criteria for a Lyupanov function is:
let N* be an equilibrium of dN/dt=G(N) and E:state space →∇
Then if over some region W:
1) E(N*)=0 and E(N)>0 ∀ N ≠ N*
and 2) d/dt E(N) </≤ 0 ∀ N ∈ V
⇒ N* is stable/asymptotically stable
Note that by the chain rule d/dt E(N) = ∇E(N) • f(N)

Example: look at the Lotka-Volterra predator prey system:
we claim E(P,V)=kaV-d ln V + aP- r ln P - Cis a Lyapunov function:
the minimum of E occurs at a location where ∂E/∂P=a-r/P=0 and ∂E/∂V=ka-d/V=0 ⇒ (V*,P*)=(r/a,d/ka)



which is identical to the equilibrium we found above – thus the minimum occurs at the right place, but it is too high $E(V^*,P^*)=d-d\ln(d/ka)+r-r\ln(r/a)$ so we set $C=-E(V^*,P^*)$ and we've met criterion #1.

we also note that $\nabla E \bullet f=(ka-d/V,a-r/P)\bullet(rV-aPV,-dP+kaPV)$ which reduces to 0 by simple algebra

this tells us that the state variable system stays on level surfaces of the Lyupanov function (if it had been negative we would have gone to the equilibrium).

Notice that in general, **fitness serves as a Lyupanov function for population genetics** – at each step the allele proportions change to increase fitness and they stop changing when fitness reaches a local maximum.

Graphical tests for stability:

- for 1-D discrete – equilibrium points are where it crosses y=x line and it is stable if the slope is between –1 and 1
- for 1-D continuous – equilibrium points are when it crosses the x-axis andit is stable if it crosses going down ($f'(x^*)<0$)
- for 2-D continuous – equilibrium points are where the isoclines cross and you can often tell if it is stable by looking at the flow

These **tests match the algebraic methods**

Attractor #2 - Cycles

The man result on cycles is the **Poincare-Bendixson theorem** – in 2-dimensions if a trajectory is bounded (never goes to infinity) and it doesn't go to a fixed point, then it must cycle (this applies only to continuous systems). This has two results: first it bans chaos in continuous systems until there are at least 3 dimensions (and chaotic 3-d systems are knonw). Secondly it gives us a way of proving a limit cycle exists even if we can't write an equation for it.; For example with the Lotka-Volterra predator-prey system all we really had to do was show that the system was bounded and that the equilibrium point was unstable to know that there was a limit cycle.

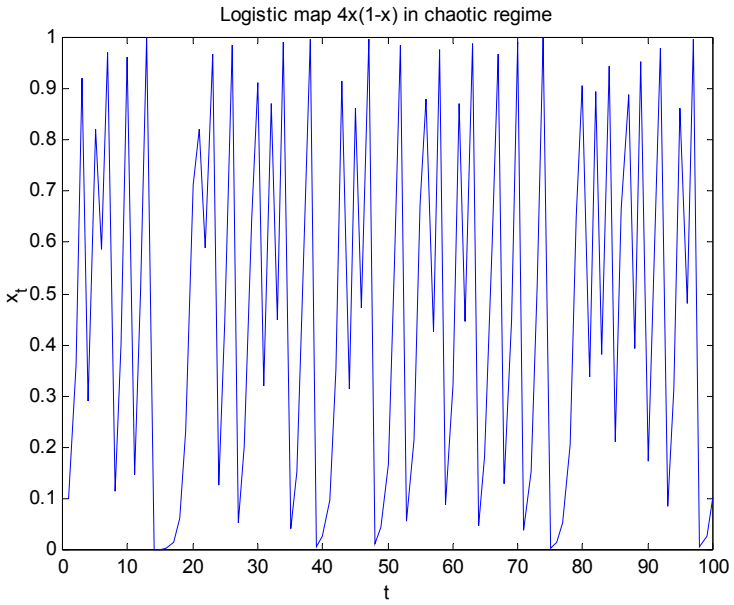
A system has a cycle (is periodic) when **there exists a T such that $N_{t+T}=N_t$ for all t**

In discrete maps we can test whether a cycle is attracting or repelling by looking at a compound map. In particular, if we are looking at a **k-cycle, we can look at $f^k(x)=f(f(f(\ldots(x))))$**. This compound map will have a single fixed point representing the whole cycle. If this fixed point is stable, then the cycle is attracting. For a continuous map, we look at the discrete map given by $f(x)=g_T(x)$. This is called a **Poincare map.**

Attractor #3 – Quasiperiodicity

This is rarely seen in biological systems. It usually occurs in conservative (area preserving systems). A simple example is $x_{t+1}= x_t+\pi \bmod 3$. Because $\pi/3$ is irrational, it will never come back to the same, exact spot but it will come very close. It is commonly found when the system is conservative (no area contraction). Such systems are common in physics but the only well-known example in EEB is the simple Lotka-Volterra model with neutral cycles.

Attractor #4 - Chaos



Chaos is the most notorious, but it is amazingly difficult to pin down – mathematicians still can't agree on the exact definition. One thing we know is that its not a limit point, limit cycle, quasi-periodic, or going to infinity. In particular we **know it is bounded and a-periodic**.(never repeats itself). It turns out the last key condition is **sensitivity to initial conditions**. Intuitively this means that two points that start very near to each other end up a long ways away – this differentiates chaos from quasi-periodicity (where nearby points stay nearby even though it is bounded and aperiodic). To make this rigorous takes some work. The key measure is the Lyapunov number or exponent.

To understand Lyupanov exponents, we once again need to bring back our friend, the Taylor expansion. Assume we have a 1-D discrete map: $N_{t+1}=f(N_t)$. Then lets take two points N and N+h which are separated by a distance h. How far will the next points be: $|f(N+h)-f(N)|=|f(N)+f'(N)h+ 1/2\ f''(N)h^2+o(h)-f(N)|\approx|f'(N)h|$. Thus an initial distance of h has been magnified by a factor of f'(N). If $|f'(N)|<1$ then the next pair of points are closer, and further if $|f'(N)|>1$. Thus lets imagine we have a long series of points $\{N_t\}$ that represents a trajectory of the map f. Now at some points we might be moving further apart and at some points we might be moving closer together, but the overall magnification would be: $|f'(N_1)f'(N_2) \ldots f'(N_t)|$ and the average magnification with each step would be: $|f'(N_1)f'(N_2) \ldots f'(N_t)|^{1/t}$ (i.e. the geometric mean separation factor). This is the **Lyapunov number** and there is an overall spreading of nearby points (sensitivity to initial conditions) when it is greater than 1. As a calculational convenience we sometimes look athe log of the Lyupanov number, ln : $|f'(N_1)f'(N_2) \ldots f'(N_t)|^{1/t} = 1/t\ \Sigma \ln f'(n_i)$.and this is called the **Lyapunov number**. Spreading occurs when the Lyapunov number is >0. In practice to calculate a Lyapunov number we start at any point and proceed to **calculate a very long series of points** (take $t\rightarrow\infty$) and assume this gets us close to the Lyapunov number since it is hard to determine all the points in the orbit. Note that at a limit point, the Lyapunov number is just a **test for stability.** To generalize Lyapunov numbers to an **n-D discrete map**, we simply keep the geometric mean of all the 1st eigenvalues, 2nd eigenvalues, etc. of the Jacobians and end up with n separate Lyapunov numbers. chaos occurs when none of the Lyapunov numbers are =1, and one of them is >1.

This sensitivity to initial conditions has great **implications1 on the use of deterministic models for prediction**.

A key factor that seems to make chaos is a **stretch and fold** dynamic, similar to that by which salt-water taffy is pulled. If you look at the logistic map, it is first stretched since the region [0, ½] and the region [½ ,1] are both mapped to [0,1]. This is then folded since the region [0,1] is then mapped back to two effectively independent regions [0, ½] and [½ ,1].

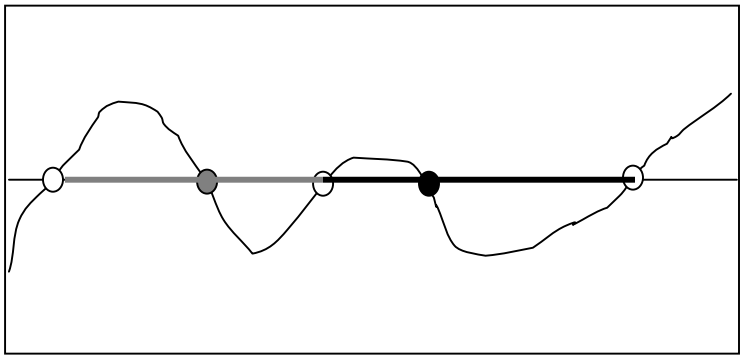
An alternative to the sensitivity to initial conditions criteria is that the dimension of the attractor be fractal (non-integer). Thus an attractor that is bounded, aperiodic and fractal is called a **strange attractor**. While there are chaotic attractors that aren't strange and vice a versa, this set is largely overlapping.

Sensitivity to initial conditions – multiple attractors

All of the above analysis assumes that there is only one attracting point. In reality, many non-linear systems have multiple attractors and repellors. Thus there is **not one asymptotic ($t\rightarrow\infty$) behavior** but many (one for each attractor). The long term behavior depends on the initial condition (starting point).

This situation is analyzed by looking at basins of attraction. The set of all points that go to a given attractor is called its basin of attraction. Often times the basins of attraction are neatly divided by a line or curve. Sometimes however, the basins of attraction are divided by extremely complicated shapes with a fractal dimension >1. These are called "fractal" basins of attraction. The most famous of these is the "Mandelbrot set", although this takes place in the complex plane and has little relevance to biology. Fractal basins of attraction are sometimes found in biology too.

As an example consider the following increment map:

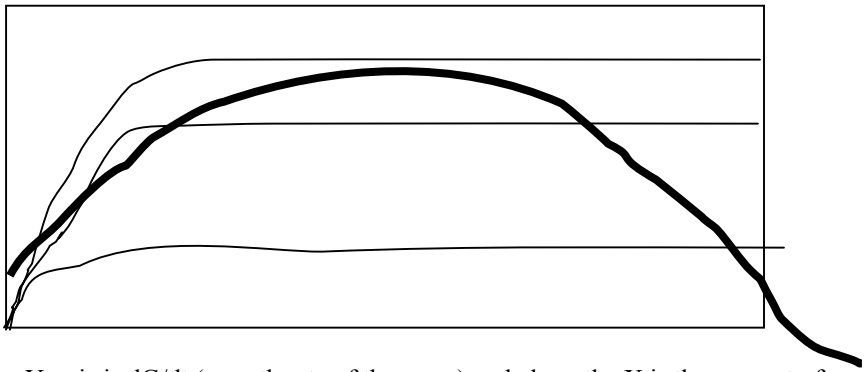


There are two stable points shown in gray and black and two unstable points shown in empty circles. The basins of attraction are shown by lines that match the color of the stable point.

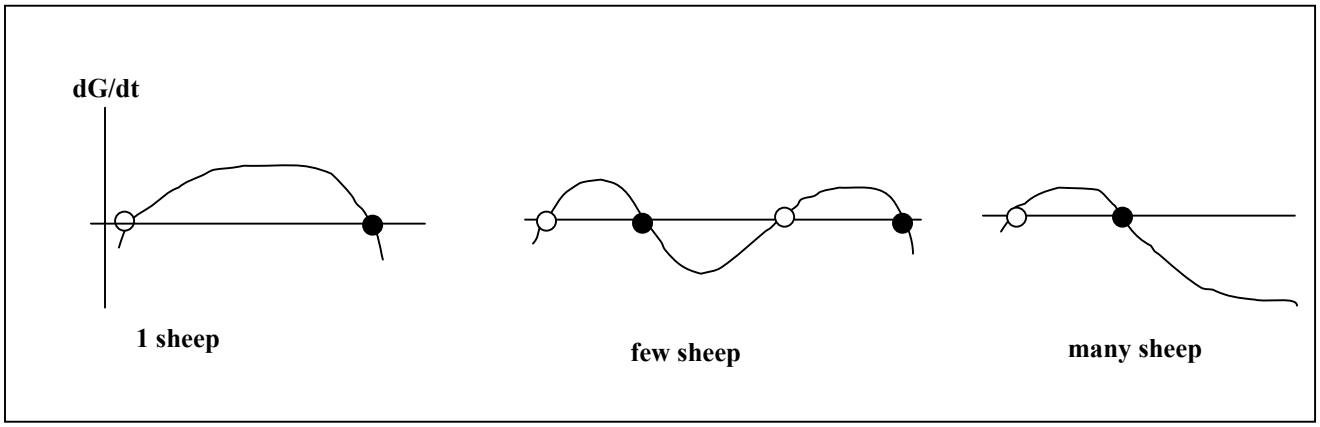
sensitivity to parameter values

One big assumption of asymptotic analysis is that all initial conditions go the same place. We addressed this above. The other big assumption is that the asymptotic behavior is the same regardless of parameter values. Often this is true – at least qualitatively. In the logistic model $dN/dt=rN(1-N/K)$. The system has the same asymptotic behavior regardless of the value of r. If we change K, the exact location of the equilibrium point changes, but qualitatively the behavior is unchanged.

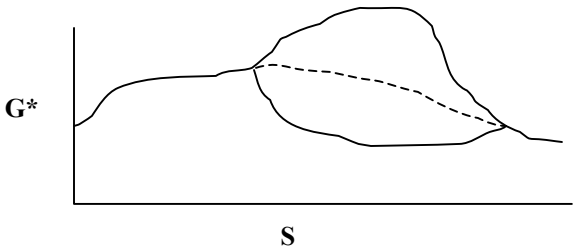
However, many times this is not true — as we change a parameter the asymptotic dynamics of a system change drastically with equilibrium points appearing and disappearing. As an example – lets look at the famous sheep grazing example. The grass grows logistically, and the sheep graze with a type I functional response:



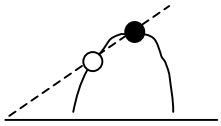
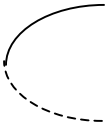
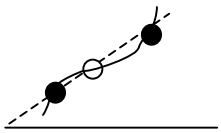
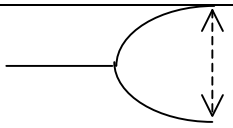
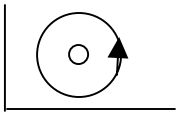
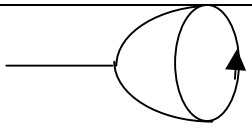
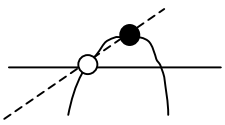
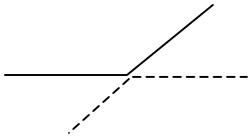
Along the Y-axis is dG/dt (growth rate of the grass) and along the X is the amount of grass. The increment map function g, is given by subtracting the logistic growth (dark line) from the appropriate saturation consumption line. We will look at the behavior of the system as we increase the number of sheep (leading to higher saturation levels of consumption). The return maps go as follows:



we can see that as we vary the parameter S, the number of sheep, we go from 1 equilibrium point to 2 to 1. This information, might usefully be contained in what is called a **bifurcation diagram** where we plot the equilibrium value(s) vs. the parameter value:

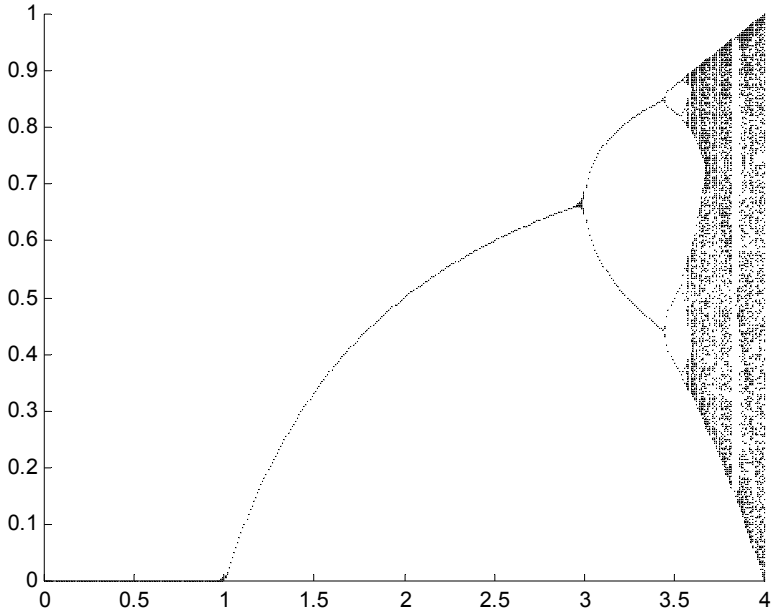


The above bifurcation diagrams shows two bifurcations (changes in attractor structure). In each case, this occurred because a hump of the increment crossed the magical y=0 line – causing us to go from no equilibrium points associated with the hump to two (one stable, one unstable). Similarly in discrete systems, bifurcations happen when the hump goes over the magical y=x line (f' leaves the region –1 to 1). Thus if we want to come up with a classification of possible bifurcations we just need to find a complete list of the ways we can cross the dividing line between stability and instability:

Type	Discrete event	Continuous event	Return map just after	Bifurcation map
Saddle-node	f'>1	g>0		
Period doubling	f'<-1	#N/A – discrete only		
Hopf Neiman-Sach in discrete 2+ Dimensions only	conjugate pair a±bi go out unit circle	conjugate pair cross y-axis (a>0)	2+D – draw phase map 	
Transcritical (non-generic, but very common in biological systems)	f>1, but one side of hump anchored were crosses critical line	g>0 as in discrete		

These bifurcations can also occur in reverse direction – called a **subcritical** bifurcation. These same maps occur in **continuous time** (except we lose the period doubling). When we go to **two dimensions**, we also pick up Hopf bifurcations (creation of a stable limit cycle from a stable limit point). We also pick up the **1-D bifurcations of limit**

cycles. Thus a limit cycle can undergo a saddle node bifurcation or a period doubling bifurcation. We use Poincare maps/sections to study these. There are several **bifurcation routes to chaos**. One is called the period doubling cascade, where we see 1,2,4,8,16, …, 3 cycles then chaos. Another is through quasiperiodicity (this is usually seen in



conservative=area-preserving maps which are not that common in biology).

Here we see the bifurcation map for the simplified logistic function (rx(1-x)) where r is the bifurcation parameter. This is a numerically calculated bifurcation diagram. We do this by taking a random starting point (here 0.1) and running it for parameter values r=0, 0.01, 0.02 … For each a parameter value we run the logistic for 200 steps. We take the last 100 steps and plot it. Here we can see several bifurcations:

- At 1.0 we see a transcritical bifurcation where the point x=0 transfers its stability to the other equilibrium point.
- At 3.0 we see a period double bifurcation
- At ~3.445 we see another period doubling (giving us a period of 4)
- The period doublings continues
- After ~3.570 we see chaotic behavior with occasional brief returns to periodic behavior.

You may have noticed that in all these bifurcations there seemed to be a preservation of some number of equilibrium points. In fact there is. If you count an index for the phase space where:

stable points are +1
unstable points (& saddles) are –1
cycles (2-cycles or limit cycles) are +2

Then the sum of the indices gives an index which remains constant as you vary any parameter. Thus a Hopf bifurcation takes a stable point (+1) and turns it into a stable cycle and an unstable point (+2 and –1 = +1) and the index remains constant.

What can we study if its chaotic – ergodic properties

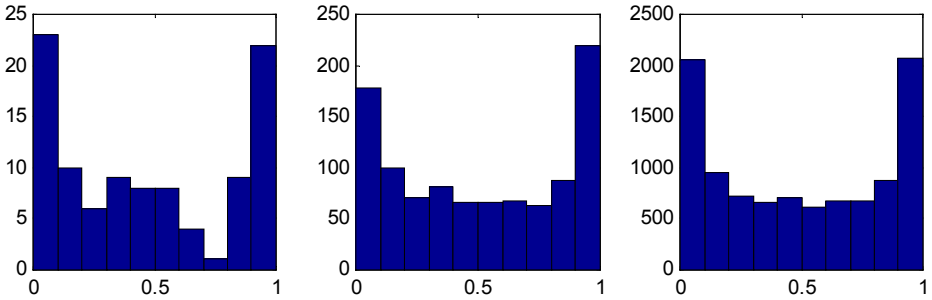
Although the fact that chaos has sensitivity to initial conditions might seem to doom us to being unable to make predictions in chaotic systems, there is some hope. This is due to a field that treats deterministic systems as random systems. And the prediction we can make is the probability distribution of where we will find the point at any given time.

A system is **ergodic** if when you **start at any point x, you will eventually reach the point y for any points x and y**. Technically this statement needs the qualifier "**almost any**" x and "almost any" y. Another way to look at it is small **intervals** (ε-balls to a mathematician) – any ε-ball carried forward will intersect any other ε-ball if you wait long enough, no matter how small the ε-balls. Another definition is that any invariant set is has measure 0 or 1. Notice that **ergodicity is relative to the set you're using**. A system that is ergodic over its attractor is not ergodic if you include a transient trajectory leading into the attractor. A system that is ergodic has two key properties:

- **measure preserving** - there is a measure (probability distribution) that is untouched by the transformation (all ergodic systems are measure preserving but not vice a versa)
- a **time average=ensemble average** –this result means that if you take one starting point and carry it forward over time, waiting long enough for the transients to disappear, then follow the behavior of one trajectory over time it will match the natural measure (that is preserved) and it will be **independent of initial conditions**

$$\lim_{n\to\infty}\frac{1}{n}\sum_{0}^{n-1} f(M(x_0)) = \int_0^1 f(x)\rho(x)dx$$

In other words, **if a system is ergodic, there is some probability distribution that describes the time average of the system.** The time average is necessary because a cycle is ergodic and a bolus of initial conditions can all be moving through the cyle together, leaving all probability concentrated at a single point in the cycle at any one point in time.



The above diagram shows the natural measure for the discrete logistic map over 100, 1,000 and 10,000 iterations.

An even stronger condition than ergodic is **mixing**. A mixing system prevents a bolus from moving through as a group. It sprays any interval out. It can be thought of as "ergodic without cycles". This has the stronger consequence that at any point in time far in the future, the behavior of a single initial condition will look like the ensemble average

$$\lim_{n\to\infty} f(M^n(x_0)) = \int_0^1 f(x)\rho(x)dx$$

In other words, **if a system is mixing, there is some probability distribution that describes the behavior of an ensemble at a single point in time (far into the future).**

Notice that in this equation and the related one for the ergodic system, not just does the time average of x equal the ensemble average of x, but the time average of any function, f(x) = the ensemble average. Hence for example the variances and higher moments will also be equal.

The strongest condition is called **exact** and it is very similar to mixing, but subtler stronger – it assumes you also have a dense number of periodic points. This gives rise to the stretching and folding of chaos, and in fact exactness implies chaos (sensitivity to initial conditions occurs.

The famous Perron-Froebenius theorem is really just a special case of this. The Perron-Froebenius theorem addresses the eigenvalues of positive matrices and is applied to linear systems: $x_{t+1}=Ax_t$. Two results tell us that the dominant eigenvalue (one with largest magnitude) is bounded by the lowest and highest row sums (or column sums) and that there is a real (non-complex) eigenvalue with this magnitude. We can add to this that if the matrix is ergodic (irreducible in matrix terms), then the system is independent of initial conditions and all starting points will converge s probability distribution given by the dominant eigenvector – there may be cycling within this state (and hence more than one eigenvalue with this magnitude). If the system is mixing (primitive in matrix terms), then there is no cycle and we converge to a single probability distribution found in the dominant eigenvector, which has a single dominant eigenvalue.

diffusion equations

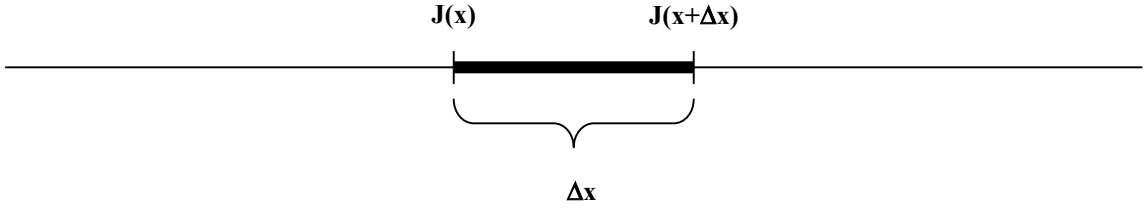
So far we have only looked at ordinary differential equations (ODE's). Now we will look at partial differential equations (**PDE**'s). These are differential equations that use partial derivatives. They are much harder to solve than ODE's. There are three general classes of ODE's, but it turns out that in EEB, really only one type is useful. This type is called the reaction/advection/diffusion equation. It is primarily useful for describing patterns in space. It takes the form:

$$\frac{\partial}{\partial t}n(x,t) = r(n(x,t)) - \frac{\partial}{\partial x}a(n(x,t),x,t)n(x,t) + \frac{\partial^2}{\partial x^2}d(n(x,t),x,t)n(x,t)$$

Here, n(x,t) is the state equation. It tells us how much of something (individuals, genes, etc) there is at the point x (along the real line) at any time t. It has straight-forward generalizations if x is a vector (2-D plane, 3-D space).

- Reaction (r) – describes creation and destruction of the objects being counted (e.g. birth, death) and depends only on the count at the given location (no neighborhood effects)
- Advection (a) – describes directed movement of individuals and gives the average distance moved by an individual
- Diffusion (d) – describes random, aimless, random-walk-like movement and describes the variance in the distance moved.

To build some intuition about this, let us first start by looking at a single interval and the current or flow, denoted by J, into or out of this interval:

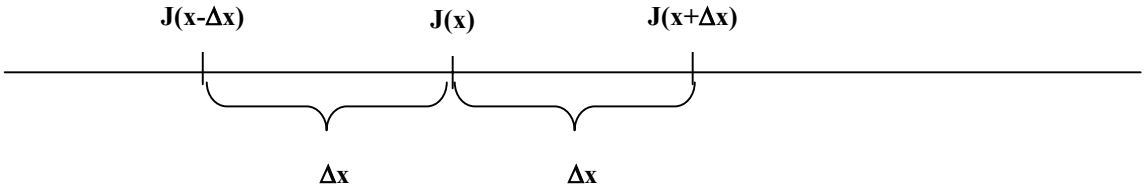


Then n(x,t)Δx is the number of individuals found in the interval

The change of number of individuals is $\frac{\partial}{\partial t}\big(n(x,t)\Delta x\big) = r(n(x,t),x,t) + -[J(x+\Delta x,t) - J(x,t)]$

If we divide both sides by Δx and take the limit as Δx→0, we get: $\frac{\partial}{\partial t}n(x,t) = r(n(x,t),x,t) + -\frac{\partial}{\partial x}J(x,t)$

Now, lets take a closer look at J(x,t), the current past the point x assuming the particles behave as if in a random walk:



Take Δx= mean distance traveled in Δt time and let p, q be odds of a step right or left respectively. Then (p-q)Δx/Δt = velocity to the right. If we look at variance for a binomial system, we can see that the variance of the distance covered in unit time is: (p Δx)(q Δx)/Δt.

$$J(x,t) = \frac{volume\ through\ x}{time} = \frac{p\,n(x-\Delta x,t)\,\Delta x - q\,n(x+\Delta x,t)\,\Delta x}{\Delta t}$$

$$= \frac{\Delta x}{\Delta t}\left[p\left(n(x,t) - \Delta x\frac{\partial}{\partial x}n(x,t)\right) - q\left(n(x,t) + \frac{\Delta x}{2}\frac{\partial}{\partial x}n(x,t)\right)\right] = \frac{\Delta x}{\Delta t}\left[(p-q)n(x,t) - \frac{\Delta x}{2}\frac{\partial}{\partial x}n(x,t)\right]$$

$$= (p-q)\frac{\Delta x}{\Delta t}n(x,t) - \frac{\Delta x^2}{2\Delta t}\frac{\partial}{\partial x}n(x,t) = \bar{v}\,n(x,t) - \frac{1}{2}\mathrm{var}(\bar{v})\frac{\partial}{\partial x}n(x,t)$$

The above assumed constant velocity and diffusion, moregenerally, we can write:

$J(x,t) = a(n(x,t),x,t)n(x,t) - \frac{\partial}{\partial x}d(n(x,t),x,t)n(x,t)$ Which completes the equation. Note that flow being negatively proportionate to change in concentration makes sense. Very often, we will see scenarios where a(n(x,t),x,t)=0 and d(n(x,t),x,t)=D, a constant.

The simplest case is where r=0 also. Let us assume we start with all individuals concentrated at the point x=0. In this case the solution can be quickly checked to be a **gaussian distribution** whose mean remains 0, and whose **standard deviation, σ, grows as √t**. If we instead put boundaries (say at plus and minus C), then the system spreads out until it is uniformly distributed. If we instead put a different kind of boundary where items that hit the boundary

disappear (r(-C,t)=0, r(+C,t)=0) then the population eventually disappears to zero. These two types of boundaries are called **reflecting and absorbing** respectively.

If we add in a reaction term we can represent population growth as well. If we let population growth vary as a function of x – and say there is a central patch where it grows at a rate>1 and the two edges it grows at a rate <1, then we have modeled a source sink population. We find that depending on the size of the patch and the proportion of the two growth rates relative to 1, we may either get a self-sustaining population or a population that goes extinct.

It is also possible to get more complicated situations. If we model growth with a carrying capacity with the population starting all at x=0, we get a **standing wave** whose front moves at a constant spead. Under other scenarios (e.g. predator prey) we can get an **oscillating wave**, where at one point over time or at one point in time over space, we see oscillations.

Numerical simulation

Very often since differential equations are so hard to solve, we end up calculating numerical solutions. Although this in the end is just a form of simulation it is fairly powerful for several reasons. One is that we may be able to get analytical results in certain, degenerate cases that allow us to test our simulations. Another is that we inherit a rich set of machinery (computer code) written by others to simulate differential equations. Finally, it gives us a very concise notation to describe our model for others even if the solution is messy.

If time is discrete – all is well, we can simulate each time step forward and we will be accurate to within the rounding limits of the computer.

If time is **continuous it is more problematic**. Continuous time does not exist in the computer. Some sort of discretization of time is needed. The simplest and most intuitive approach is called **Euler's method**.

Say we have a 1-D continuous differential equation: $dx/dt=f(x)$. Then we pick a very small time interval, Δt (maybe =0.001). We then approximate by $x_{t+\Delta t}=x_t+f(x)\,\Delta t$. We repeat this over and over and we get a reasonable approximation to the solution. However, now in addition to the rounding error in x, we have discretized time which introduces error as well. More sophisticated methods (such as Runge-Kutta exist). In practice, Euler's method should only be used as a pedagogical device and more sophisticated methods should always be used.

We can also numerically solve partial differential equations (**PDE's**). Again the simplest approach is a simple discretization of space and time. For the diffusion equation we take small intervals Δx in space Δt in time. The simplest discretization of the reaction-diffusion equation is:

$$N_{x,t+\Delta t}=r(x,t)+D[(n(x+\Delta x,t)-n(x,t))-(n(x,t)-n(x-\Delta x,t))]$$

If we have initial conditions for all points x for t=0, then we can just roll this equation forward, solving for all points x for $t=t+\Delta t$ and then so on. Again this naïve method has problems and more sophisticated methods should be used.

Homework #2

1) Find the dominant eigenvector and eigenvalue of the matrix $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Use repeated iteration (take any random initial vector x_0 and calculate $x_t=A^t x_0=A(A(Ax))$). Repeat until it converges.

2)Analyze the simplified discrete logistic $x_{t+1}=\mu x_t(1-x_t)$ by numerical methods.

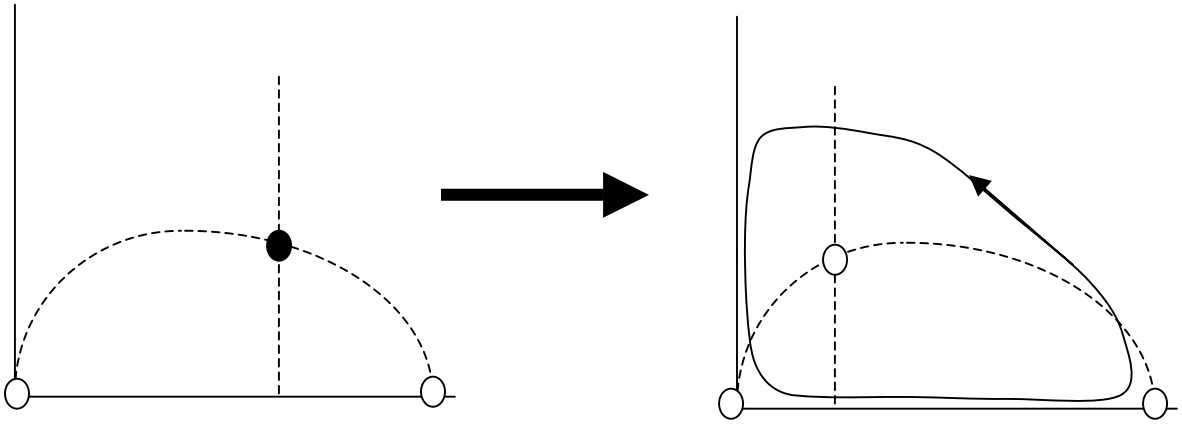
2a) analyze sensitivity to initial conditions for $\mu=2$. Start at x_0= each of -0.1, 0, 0.1, 0.5, 1, 2, 5. Iterate the system forward and see what attractor you end up at. Use this information to guess at basins of attraction.

2b) analyze sensitivity to the parameter, μ. Starting always at $x_0=0.1$, iterate the system forward for μ=each of 0.5, 1, 2, 3, 3.3, 3.5, 3.9. See what happens in each of these cases. Use this to draw a crude bifurcation diagram.

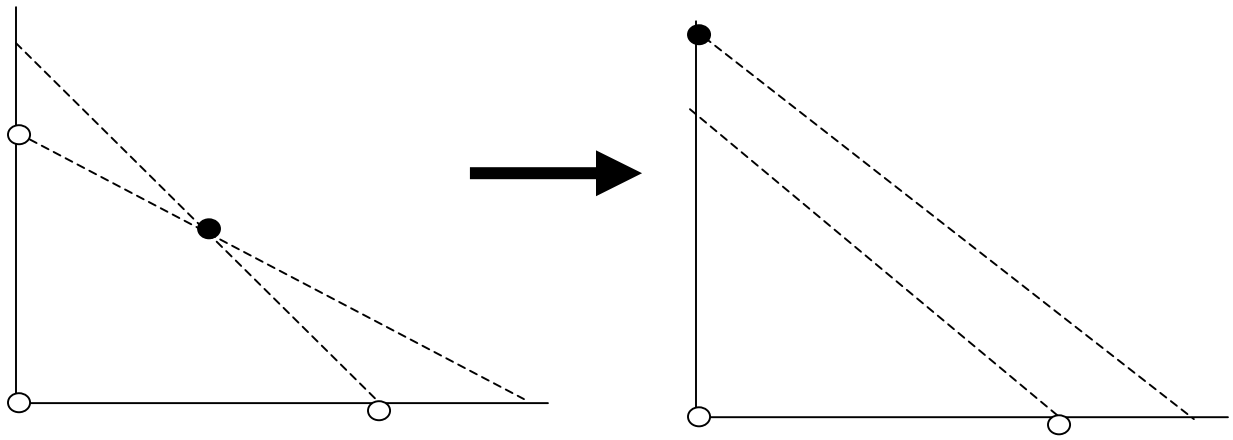
2c) analyze the ergodic properties in the chaotic regime ($\mu=3.9$) again starting at x=0.1, iterate forward 1000 steps, throw away the 1st 100, and then plot a histogram.

Hint: in Matlab, the following piece of code will perform a simple iteration and save all steps in a vector variable (x): x=0.1; for i=1:100, x=[x 4*x(end)*(1-x(end))];end

3) What kind of bifurcation occurred in the below diagram showing the Rosenzweig-MacArthur predator prey equations as we vary the prey carrying capacity, K?



4) Below are two diagrams showing phase plane analysis for the classic Lotka-Volterra competition equations. As we vary the parameter K_1 we move from the diagram on the left to the right. What kind of bifurcation occurred? This is tricky! Hint – is the equilibrium conservation rule followed? What happened? Just because biologists are only interested in the positive quadrant doesn't mean no math goes on elsewhere.



Random dynamical systems

by Brian McGill

September, 2000

Top 10 (okay, 14) things you need to know about probability

Simple probability

Let Ω be a set of possible outcomes. Let $A,B \subset \Omega$ be subsets of possible outcomes. Then:

1) Law of simultaneous events:

$P(\text{both A and B occur})=P(A \cup B)= P(A)+P(B)-P(A \cap B)$

1a) Law of addition

a special case of above: if $A \cap B=\varnothing \Rightarrow P(A \cup B)= P(A)+P(B)$

2) Conditional probability

$P(A|B)=P(A \cap B)/P(B)$

3) Independence

If $P(A|B)=P(A)$ then A & B are independent

23a) Law of multiplication

combining 2 & 3, if A & B are independent then $P(A \cap B)=P(A)P(B)$

4) Law of total probability

Let A_i be a disjoint set that spans Ω (i.e. $A_i \cap A_j = \varnothing$ and $\cup_i A_i = \Omega$). Then for any B, $P(B)=\Sigma P(B \cap A_i)=\Sigma P(B|A_i)P(A_i)$

Example: roll a single, six-sided die. Let A={1,2,3} and B={2 ,4, 6} then:

$P(A \cup B)=P\{1,2,3,4,6\}=P(A)+P(B)-P(A \cap B)=3/6+3/6-1/6=5/6$

$P(A|B)=P\{1,2,3|2,4,6\}=P(A \cap B)/P(B) = 1/6/1/2=2/6=1/3$

A and B are *not* independent

Probability distributions

5) Definition of a random variable

A random variable is a variable that takes on values by chance according to a distribution function, $F(x)=P(X \le x)$. Random variables are usually denoted by a capital letter. Note $F(-\infty)=0$ and $F(\infty)=1$ and F is monotonically increasing. This is sometimes called the cumulative distribution function or CDF.

6) Probability functions

For any random variable and distribution function, there is a more obvious function. For a discrete distribution (X takes on only discrete values), it is called **the probability mass function** $p(x_i)=P(X=x_i)$ and $F(x)=\sum_{i \ni x_i \le x} p(x_i)$. For continuous distributions, it is called the **probability density function (PDF)** and is the derivative of the CDF (and hence the CDF is the integral of the PDF: $F(x)=\int_{-\infty}^{x} p(x)dx$. In either case the probability function can be though of as describing how a probability mass of 1 (think of sand weight 1 gram) is spread out over the probability space.

7) Expectation of a random variable

Expectation is a fancy word for weighted average or mean (the weighting is by the probability of a particular value). It is often denoted simply by the "E" operator. So if X is a random variable, then $E(X)=\sum_i x_i\, p(x_i)=\int_{-\infty}^{\infty} xf(x)dx$.

8) Expectation of functions of random variables

If X is a random variable, then it is reasonable to take the expectation of the function of X — i.e. E(f(X)). **DO NOT EVER ASSUME that E(f(X))=f(E(X))!** This is only true in the very special case where f is a linear operator (f(ax+b)=af(x)+b). In fact Jensens inequality tells us that if f is convex and not linear, then E(f(X))<f(E(X)

8) Higher moments

The most important function of a random variable is the power function. $E(X^n)$ is called the n-th moment. Thus the 1st moment is the average, often denoted μ . Often we prefer to look at a central moment: $E((X-E(X))^n)$ which is called the n^{th} central moment. The 2nd central moment is called the variance, often denoted σ^2. Note, the 1st central moment is always equal to zero. The 2nd central moment has a shorthand calculation formula: $E((X-\mu^2) = E(X^2-2\mu E(X)+\mu^2) = E(X^2)-2\mu E(X)+\mu^2=E(X^2)-E(X)^2$.

9) Conditional expectation

Another useful expectation of a function of a random variable is conditional expectation: E(X|Y)

10) Sums and convolution

Suppose we have two independent random variables X, Y and we want to know what the random variable X+Y behaves like. This is determined by an operation called **convolution:** $P(X+Y=c)=\Sigma_a P(X=a)P(Y=c-a)$.

11) Generating Functions

Generating functions are a nifty trick used a lot. For any given discrete probability distribution p(x=i) for a random variable X, there is one and only one generating function associated with it given by $G(s)=E(s^X)$. For example, for the Bernoulli variable P(X=0)=p and P(X=1)=1-p, the generating function is $G(s)=E(s^X)=s^0P(X=0)+s^1P(X=1)=p+(1-p)s$.

Generating functions are your friend because:

1) They often **take infinite series and reduce them to a finite function** via the geometric series rule and thus can be manipulated.
2) A specific example of #1 is that they **easily perform convolution – i.e. sums of random variables**: if X & Y are independent random variables, then $G_{X+Y}(s)=G_X(s)G_Y(s)$ – i.e. convolution of infinite series is reduced to multiplication
   Example: if X, Y are bernoulli variables, then $G_X(s)=G_Y(s)=(p+qs)$ so $G_{X+Y}(s)=(p+qs)^2$ and more generally the sum of n Bernoulli variables has generating function $G(s)=(p+qs)^n$ which we can readily look up in a table of generating functions to be the Binomial distribution B(n,p). Perhaps even more useful, we can quickly see that the sum of two generating functions B(n,p)+B(m,p)=B(n+m,p).
   Even more generally if we take a sum of N random variables $S=X_1+ \ldots + X_N$ where N is also a random variable, the generating function is $G_N(G_X(s))$. So compounding of random variables is given by compounding of generating functions.
3) **Moments are readily recovered**: E(X)=G'(1) and $E[X(X-1)\ldots(X-k+1)]=G^k(1)$ and thus $var(X)=G''(1)+G'(1)-G'(1)^2$. For example: $E[B(n,p)]=G'(1)=n(p+qs)^{n-1}q\,|_{s=1} = nq = n(1-p)$
4) Continuous functions work nearly identically except we use the characteristic function: $\varphi(t)=E(e^{itX})$ which has identical properties except that $E(X^k)=\varphi^k(0)/i^k$ instead of the formula in #3

13) Law of Large Numbers (sample average → ensemble average)

Let X_i be a sequence of independent, identically distributed random variables. Then $\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mu$. In other words, the average of the X_i approaches the average of the distribution if you take a big enough sample.

14) Central Limit Theorem (sample distribution → normal distribution)

Let X_i be a sequence of independent, identically distributed random variables. Then $\sum_{i=1}^{n} X_i \rightarrow N(n\mu,\sqrt{n}\sigma^2)$. In other words, no matter what distribution, when you sum a bunch together it starts to look normally distributed!

Top 13 probability distributions

Probability distributions play an important role in probability theory and in stochastic processes. It is easiest to understand and remember distributions if we group them into key processes that lead to these distributions.

The counting process distributions

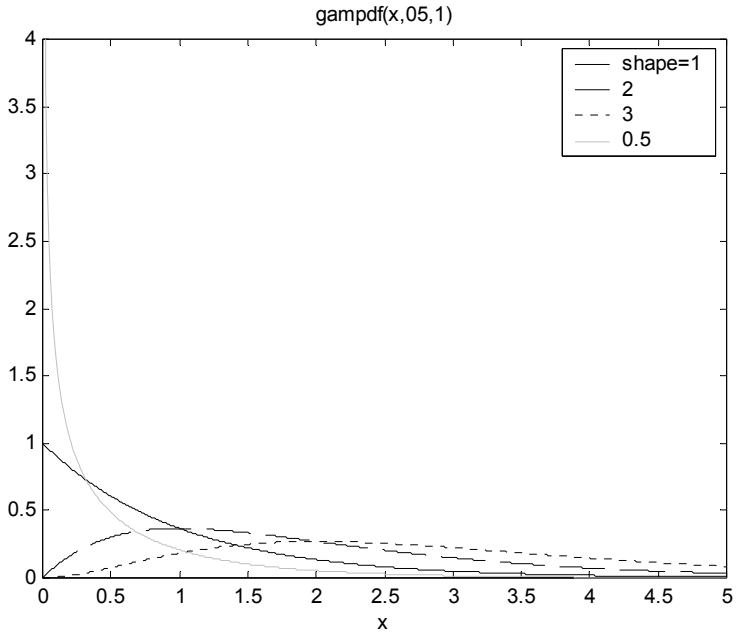
Imagine a process in which we are waiting over time for an event to occur. When this event occurs we count it and start waiting again. This general process occurs in everything from Geiger counters to meteorite strikes on earth to roll a natural 7 in the dice game craps. There are three things we could be interested in:

- the time to the next event
- the time to the nth event
- the number of events that happen in a given amount of time

Another key factor is whether time is discrete or continuous. Do the events happen at any time (continuous) or is there a periodic sampling regime in which we check (dice rolls, annual samples). Combining these factors we have six distributions:

	Discrete Time			**Continuous Time**		
	Time=n, rate at single interval=p			Time=t, instantaneous rate=λ		
	Distribution	**E(X)**	**Var(X)**	**Distribution**	**E(X)**	**Var(X)**
time to next event	Geometric	1/p	$(1-p)/p^2$	Exponential	1/λ	$1/\lambda^2$
time to the nth event	Negative binomial	n/p	$n(1-p)/p^2$	Gamma	n/λ	n/λ^2
expected (mean) # events in given time	Binomial	np	np(1-p)	Poisson	λt	λt

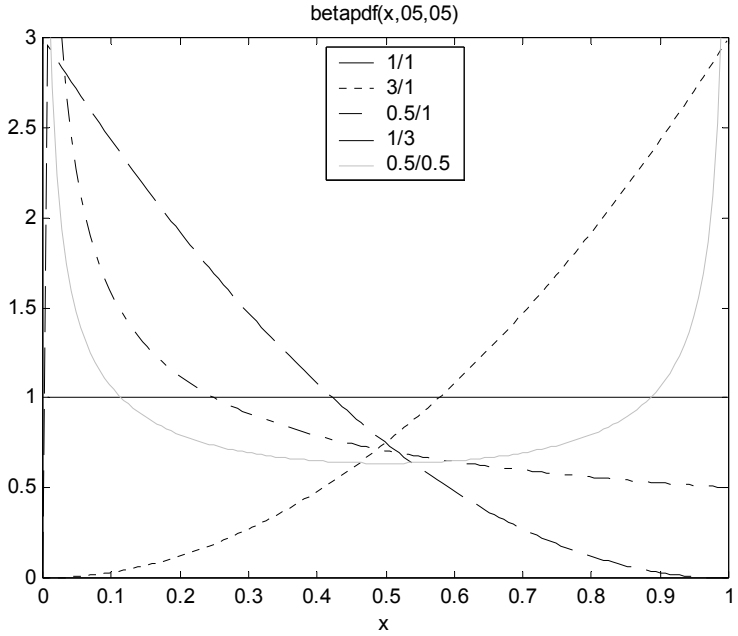
Notice that the time to the **nth event is the same as the sum of n random variables** of the time to the next event (hence the gamma is the sum of n exponential variables). Not surprisingly, the **binomial distribution approaches the continuous distribution** if you take the limit as $n \to \infty$ (smaller time intervals), $p \to 0$ such that $\lambda = np$ is a constant. It is also possible to use these distributions to examine what happens as you **move across space instead of through time**. Hence the Poisson is used frequently as a model of random spatial distribution. The exponential distribution has the special property of **having no memory**. If you have waited time T for an event to happen, and it hasn't happened, then the odds of it happening are the same as if you hadn't waited time T already. This means it can be shown that all Markov processes in continuous time must have exponential distribution between occurrences. The **Gamma distribution is a highly-plastic function on the interval (0,∞)** and hence is used as a generalized probability function to fit errors that must be positive. The **poisson distribution is sometimes called the discrete normal**, because it is so commonly found in nature when you look at discrete counts of events.



The finite interval distributions

Most distributions occur over an infinite interval (either $-\infty$ to ∞ or else 0 to ∞). However, sometimes the state variable of interest occurs on a finite range, which without loss of generality we can consider to be [0,1]. The simplest such distribution is the **uniform distribution** which assigns equal probability to every point over the range. The uniform distribution can be discrete or continuous.

A generalization of the continuous uniform distribution is the **Beta distribution**. The Beta distribution on the inteval zero one has the form $p(x)=Cx^{p-1}(1-x)^{q-1}$, where C is a constant that makes the total probability =1. Thus p and q determine the shape of the distribution. p describes behavior near zero. If $p>1$ then it behaves like x^2 and increases, if $p<1$ then it behaves like p^{-1} and asymptotes down and if $p=1$, it behaves like $p^0=c$ a line. Similarly q describes the behavior near 1 only in reverse (since it is 1-x). Thus we can make basically any shaped function over the interval 0-1.



The normal generalization distributions

The central limit theorem tells us to expect that many measured distributions should look normal. This is the basis of normal distriubtions in statistics (measurement errors should be nearly normal). Thus it is important to know the behavior of various combinations of normal variables. Recall that the normal is defined by two parameters, μ, σ and is denoted $N(\mu,\sigma)$. N(0,1) is called the standard normal distribution.

Combination	Distribution	Range
Sample n normals $N(\mu,\sigma)$	$N(\mu,\sigma/\sqrt{n})$	$(-\infty,\infty)$
Sum n normals squared: N(0,1)	Chi-squared χ_n	$[0,\infty)$
$N(0,1)/\sqrt{\chi_n}$ (found in sample distribution of variance)	t-distribution	$(-\infty,\infty)$
χ_n/χ_m (found in ANOVA tests)	Fisher	$[0,\infty)$
log N(0,1) (found as central limit of *product* of many variables)	Log-normal	$(0,\infty)$

The binomial generalization distributions

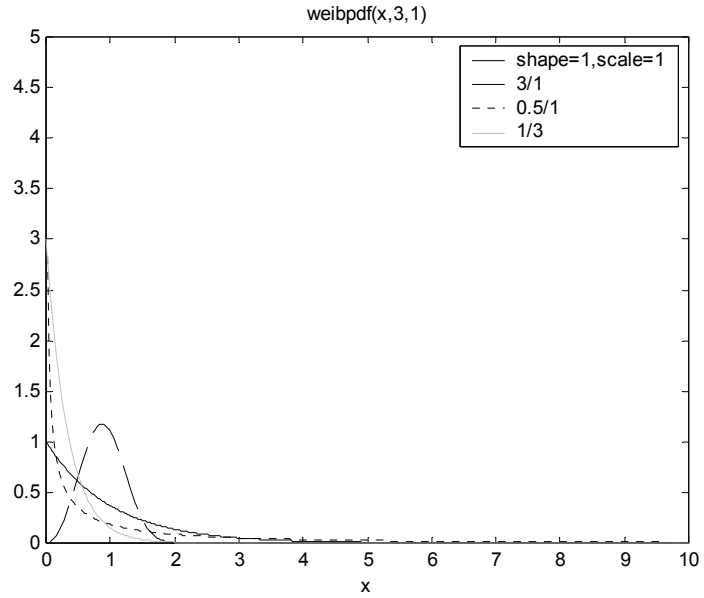
The hypergeometric distribution is a non-Markovian generalization of the binomial. The binomial assumes that whether success or failure occurs in this interval is independent of how many have already occurred. The hypergeoemetric assumes there is a finite number of total successes, N.

The **multinomial distribution allows more than two outcomes** (just success and failure).

Exponential generalizations

While the exponential model is a nice model for waiting times since it is Markovian, many waiting times are not Markovian (how long you've waited does affect the odds of how long you still have to wait – e.g. a fire occurring is more likely the longer you've waited). The Markovian nature of the exponential is quite a strong statement and is often referred to by saying that the **exponential is memoryless.**

The **Weibull distribution** is a generalization of the exponential that engineers often use to model waiting times. It does allow for memory. The Weibull has two parameters, a shape and a scale parameter. When the shape parameter is =1, then the Weibull reduces to exponential. If the shape parameter is less than one, you get a long-tailed distribution (likely to happen very quickly but a long wait if it doesn't). If the shape parameter is greater than one, you get a hump (likely to happen pretty close to some mean time).



The **Pareto or Power-Law** distribution is sometime used for waiting times but more often is used for counts or sizes of objects (e.g. the species abundance distribution, the size of asteroids, the distribution of incomes in humans, etc). Over finite intervals it behaves much like the pareto with shapes ≤1, but the tail of the Weibull and exponential is longer over an infinite range.

Stochastic processes

Introduction

Recall that in deterministic state space systems, we had either:

$N_{t+1}=f(N_t)$ or $dN/dt=G(N)$

How can we extend this to a random transition rule? The first thing we need to do is to c**hange the state variable form a deterministic value to a probability distribution**. IE N_t doesn't give the exact value of N at time t, but the probability distribution of values at time t. The **probability distribution may be continuous or discrete, unlike deterministic systems**. We typically denote this as π_t the probability distribution of the state variable at time t and thus $\pi_t(x)$ gives the pdf or pmf that the state variable is equal to x at time t. What about the transition rule? The notation remains the same, but it is different – we can't just take a probability distribution and plug it into an equation. Thus we think of probability distributions evolving over time. I sometimes find this helpful to think of as an **amount of sand weighing 1 gram spread over the possible state values, and the sand shifts over time**.

Most of the time we will make the same assumptions that we did for deterministic systems, namely that N_{t+1} depends only on N_t and not on prior values (no lags) — this is called the **Markov assumption** and also that the transition rule (f,G) doesn't change with time — this is called **a homogenous system**, unlike the deterministic world where this is called an autonomous system and homogenous means something different.

Technically, a **stochastic process is a series of random variables indexed by time** (X_s, X_t, X_u, X_v, …) with a known joint probability distribution ($p(X_s=x_1, X_t=x_2, X_u=x_3, X_v=x_4, \ldots)=f(x_1, x_2, x_3, x_4, \ldots)$.

A full solution to a random (stochastic) process is a formula $\pi_t(x)=p_t(x|y)\pi_0(y)$ very similar to a full solution to a differential equation. Note that **$p_t(x|y)$ denotes the probability that you will be at state x given that you were at state y an amount time=t ago**. The Markov assumption allows us to specify only y, the initial condition, and the homogeneity assumption means we only need the Δt, not the exact times. Just like in deterministic systems, we often can't solve for $p_t(x|y)$ or it is very difficult to do so. In this case we **may do asymptotic analysis and look for a stable distribution $\pi^*(x)$** that is approached in the long run. We might also ask other questions such as what is the **odds that I end up at this attractor instead of that attractor** (which makes sense in the deterministic world too) or what is the **average time until I pass this point** (which has no exact analogue in the deterministic world).

Classification of Stochastic Processes

Stochastic processes is currently in the "zoo" stage of development — there are lots of individual processes described with their solutions, but there is very little in the way of general machinery.

There are two key types of classification of stochastic processes.

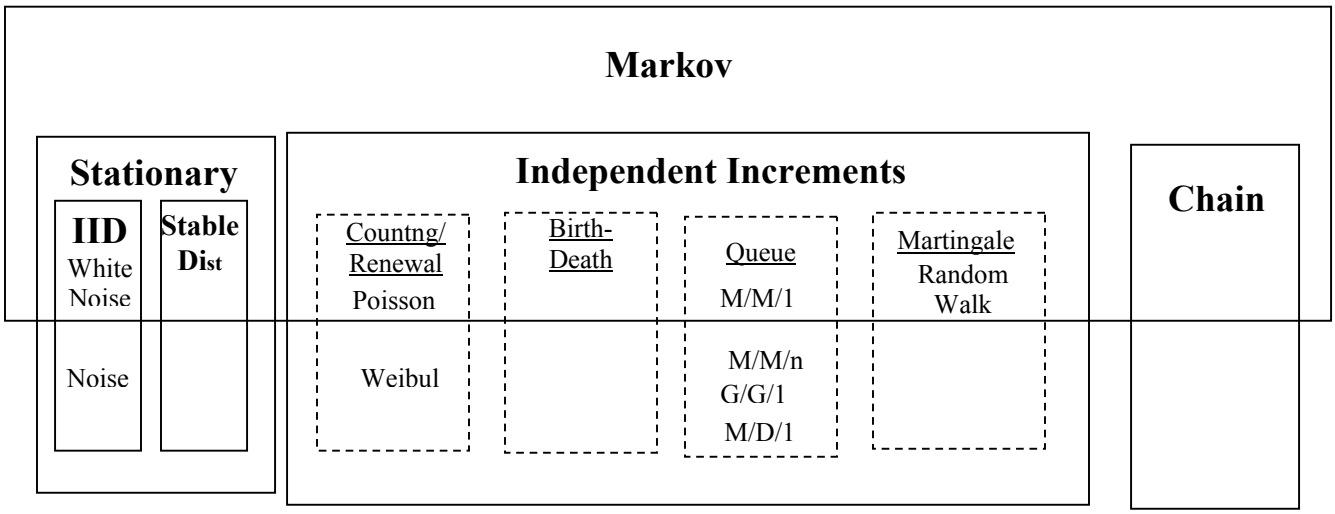
- **Markovian vs non-Markovian:** as we shall see shortly, the Markov assumption allows us to prove a number of results which we can use simply by knowing the process is Markovian.
- **Mathematical model:** there are a number of real-world physical processes that we have in mind. These physical processes can be described mathematically as falling in one of three categories:
    - **Stationary:** This models processes that don't go anywhere – in fact, by definition a stationary process is one whose mean and variance don't change over time ($E\{X_t\}=c$, $E\{X_t2\}=d$ where the constants are independent of time). Stationary processes typically have two physical models in mind: a sequence of independent, identically distributed variables and a Markov process that has achieved a stable distribution that doesn't change over time. This may model processes with state variables on □, $\wedge$, ∇.
    - **Independent increments:** This models processes where the state value in the next period is equal to the current state value plus a random variable. In other words, we study $S_{t+1}=S_t+X_t$ where X_t is a random variable with some known distribution, usually taking on integer values and S_t is the state variable whose distribution we study. Typically, we are modeling on $\wedge$ (although Brownian motion is on ∇). There are several physical models that fit the general description of independent increments.
    - **Chains:** These models are on □ and model situations where you might realistically jump from any state to almost any other state.

In general, once we work down to a specific physical process (such as counting/renewal), we find that the canonical example that is most thoroughly analyzed is Markovian but there is a more general field that gives general results that apply to both Markovian and non-Markovian examples. Thus the organization of the remaining material is as follows:

1. Briefly examine general truths about Markov models
2. Examine each mathematical model and general truths about it
3. Look at a specific physical model nested within the mathematical model

3a. Look at the Markovian example and its analysis

3b. Present more general results that can be derived for the physical model.



Interestingly, stochastic models can handle a richer set of combinations of types of time and state space:

	Time Discrete	**Time Continuous**
State □	chain	chain
State ∧	independent increments	independent increments
State ∇	noise	noise

Recall that deterministic systems can only handle the shaded area.

General terminology and approach

We will use the following notation:

$p(y,t|x,s; w,r, \ldots)$ means probability that the system is in state y at time t, given that it was at state x at time s (<t), at state w at time r (<t) …

A stochastic process is Markovian if it depends only on the previous time period ($p(y,t \mid x,s; w, r; v, q; \ldots) = p(y, t \mid x, s)$)

A stochastic process is time homogenous if the transition probabilities do not change over time ($p(y,t+\delta|x,s+\delta)=p(y,t \mid x,s)$).

A process that is stochastic and time homogenous thus depends only on the amount of time and the state at that the earlier time. Thus we can use a simplified notation:

$\pi_t(i)$ is the probability that the system is found in state i at time t

$p_t(i|j)$ is the probability that the system is at state i when it was at state j an amount t of time ago

P_t is the matrix of transition probabilities $P(t)=[p_{ij}(t)]$

Solely being a homogeneous Markovian system has several implications:

Implications of Markov, homogeneous systems

(1) Markov, homogeneous

(2) $\Leftrightarrow$ Chapman-Kolmogorov: $p_{t+s}(y|x)=\int p_t(y|z)\, p_s(z|x)\, dz = \Sigma_z\, p_t(y|z)\, p_s(z|x)$

i.e. can sum over all intermediate points for some intermediate time (have to get there somehow)

(3) $\Leftrightarrow \pi_t(y)=\int_x p_t(y|x)\, \pi_0(x)\, dx = \Sigma_x\, p_t(y|x)\, \pi_0(x)$

i.e. can solve for distribution π at any time if know $p_t(y|x)$ generally and know initial condition π_0

(4) $\Leftrightarrow$ Chapman-Kolmogorov forward differential equation:

$$\frac{\partial}{\partial t}p_t(y|x)=\int W(y|z)p_t(z|x)-W(z|y)p_t(y|x)\, dz \; - \frac{\partial}{\partial y}[m(y)p_t(y|x)] + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}[v(y)p_t(y|x)] \quad \text{w/ i.c. } p_t(y|x)=\delta(y-x)$$

(5) $\Leftrightarrow$ Chapman-Kolmogorov backward differential equation:

$$\frac{\partial}{\partial t}p_t(y|x)=\int W(z|x)[p_t(y|x)-p_t(y|z)]\, dz \; - m(y)\frac{\partial}{\partial y}\, p_t(y|x) - \tfrac{1}{2}\, v(y)\frac{\partial^2}{\partial y^2}\, p_t(y|x) \quad \text{w/ i.c. } p_t(y|x)=\delta(y-x)$$

(2) is just a tool. (3) is useful if we can find p(). Equations (4) & (5) are useful tools for finding p() by giving us a differential equation for p in terms of something we can get our arms around: the instantaneous rates W, m, v.

(6) $W(y|x) = \lim_{\Delta t \to 0} \frac{p_{\Delta t}(y|x)}{\Delta t}$ (for $|y-x| \geq e$) represents discontinuous jumps from x to y

(7) $m(x)=\frac{1}{\Delta t} \int_{|y-x|<e} (y-x)p_{\Delta t}(y|x)\, dy$ represents small non-random movements (drift)

(8) $v(x)=\frac{1}{\Delta t} \int_{|y-x|<e} (y-x)^2 p_{\Delta t}(y|x)\, dy$ represents small random movements (diffusion)

higher moments (where (7) and (8) are the 1st and 2nd moments) are all zero

thus we have covered all possible movements

Questions

What is $p_t(y|x)$ for all time t (the full solution analogous to solving a differential equation)
Does $\pi^*(x)=\pi_\infty(x)$ exist, if so what is it? (analogous to asymptotic analysis of differential equations)

Chains

Definition and example

A chain is a markov model with the state space being □, an unordered (usually) set of discrete states.
One classic example is behavior, say the behavior of a pollinator where the states are:

1. examine flower
2. eat at flower
3. move to next flower on same plant
4. move to new plant

We will start with a discrete time chain. In this case, the system is quite easy to deal with. If we can specify the one-step transition probabilities the rest can easily be solved. For example:

$$\begin{bmatrix} .1 & 0 & .5 & .8 \\ .5 & .3 & 0 & 0 \\ .3 & .4 & .3 & .2 \\ .1 & 0 & .2 & 0 \end{bmatrix}$$

Thus for example, if we have just moved to the next flower on the same plant, there is a 50% probability we will next examine the flower, no chance that we will immediately eat at the flower, and a 30% and 20% chance that we move on to another flower, plant respectively.

Full solution

The Chapman-Kolmogorov equation tells us that if the one-step transition matrix is P, then the n-step transition matrix P_n is simply the n-th power of $P=P^n$. Moreover, if we know the initial condition π_0, then $\pi_t=P^t\pi_0$. This should look similar to the deterministic systems of linear difference equations we have already seen. Thus we have an exact solution. We can also solve for the equilibrium or steady state distribution, *if it exists*. We can do this in two ways:

$\pi^*=P^\infty\pi_0$ which in turn is solved by finding the dominant eigenvalue of A and π^* is the corresponding dominant eigenvector

$\pi^*=P\ \pi^*$

Following in analogy with the deterministic system, lets think about this from an asymptotic point of view. What type of end scenarios can we end up with?

Example:

Take a simple two state chain. For example, Schaffer's good year/bad year model of stochastic environmental variation. — let the first state represent a good year and the second state represent a bad year.

$P=\begin{bmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{bmatrix}$, say $\alpha=0.2$ and $\beta=0.3$ so $P=\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ So there is a 80% chance a good year is followed by a good year.

Say the year 2000 is a good year. What is the probability distribution of good vs bad years in 2001?

$\pi_{2000}=\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ so $\pi_{2001}=P\ \pi_{2000}=\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}$

More challenging: what is the distribution of good vs bad years in 2003?

$\pi_{2003}=P^3\ \pi_{2000}=\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}^3\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} .65 & .525 \\ .35 & .475 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}=\begin{bmatrix} 0.65 \\ 0.35 \end{bmatrix}$

Note that if 2000 had been a bad year, we would have a different distribution [.525 .475].
Is there a stable distribution?

solve $\pi^*=P\pi^* \Rightarrow \pi^*=[3/5\ 2/5]$

alternatively, look at P^∞. Well, numerically $P^7=\begin{bmatrix}0.6 & 0.6\\ 0.4 & 0.4\end{bmatrix}$ No matter, what π_{2000} is, π_{2007} is going to be [3/5 2/5] as are all ensuing years. Alternatively, analytically, the dominant eigenvector of P (of course associated with an eigenvalue of 1) is [.6 .4]. This is a general rule you may recall from the linear algebra section, if we take a matrix to higher and higher powers, it eventually converges to have columns of the dominant eigenvalue.

Asymptotic analysis

There is a theorem that says that the set of states in a chain can always be broken up as follows:

$S=T \cup E_1 \cup E_2 \cup \ldots$

T represents the transient set (the system eventually leaves T and goes to one of the E_i, never to return). States in T communicate with each other and with states in the E_i's

E_i represents a set with two characteristics:

- you can eventually get from any state in E_i to any other state in E_i (i.e. irreducible)
- once in E_i you can't leave E_i (i.e. closed or absorbing)

within E_i one of two things can happen:

- the system can either spread out to a steady state π^* within E_i (mixing case)
- the system can move periodically around all states within E_i (periodic case)

Thus we have asymptotic analysis similar to the deterministic case. There are attractors (closed irreducible sets) and points that are not attractors (transient). The attractors can be a single point, a cyclic set, or a mixing set (chaos). Look at the following transition matrix to make this clear:

$$\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0.1 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0.3 & 0.2 & 0 & 0.2 & 0\\
0 & 0 & 0 & 0.7 & 0.8 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.3\\
0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0.4\\
0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.3
\end{bmatrix}$$

This might seem like a lot to parse, but look at it now:

$$\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0.1 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0.3 & 0.2 & 0 & 0.2 & 0\\
0 & 0 & 0 & 0.7 & 0.8 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.3\\
0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0.4\\
0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.3
\end{bmatrix}$$

We have three closed, irreducible sets (the E_i's) found in the solid boxes. One is periodic with a 3-cycle. One is mixing with two elements, and one has a single element. There are also two transient states (in the dotted box). These transient states communicate with each other, and with the closed, ergodic sets. We can tell that we eventually end up in one of the closed, ergodic sets, and that the behavior is easy to determine with each of them. The only place where the behavior

might be unclear is in the 2nd ergodic set. We just need to solve the equation:

$$\begin{bmatrix}\pi_1\\ \pi_2\end{bmatrix}=\begin{bmatrix}.3 & .2\\ .7 & .8\end{bmatrix}\begin{bmatrix}\pi_1\\ \pi_2\end{bmatrix}\Rightarrow\begin{bmatrix}\pi_1\\ \pi_2\end{bmatrix}=\begin{bmatrix}2/9\\ 7/9\end{bmatrix}$$

and we see that in the long run, if we hit the 2nd ergodic state, we will spend 2/9 of our time in state 4, and 7/9 of our time in state 5.

Thus we can do a kind of decomposition: determine what our attracting sets are, and then independently analyze the behavior of each attracting (aborbing) set.

Just as in the deterministic world we can make some general statements about behavior within a closed, irreducible set:

since it is closed and irreducible, it is ergodic (time average = π^* = ensemble average)

if it is mixing (aperiodic or primitive) then lim as $t\to\infty$ $\pi_t=\pi^*$= ensemble average *at a single point in time*

also recall that in a finite chain, at least one closed, irreducible set exists

First step analysis

How do we tell which ergodic set we will end up in and how long it takes to get there? We do this using one-step analysis. For example what is the probability that we will end up in the periodic ergodic set (states 1-3). Define:

u_i = P{end up in cyclic set | start at state i}

Clearly:

$u_1, u_2, u_3 = 1$

$u_4 - u_6 = 0$

The tricky part is figuring out u_7 and u_8. Here we need one step analysis:

$u_7=0.1\ u_1 + 0.3\ u_7 + 0.4\ u_8$

$u_8=0.4\ u_7 +0.3\ u_8$

Because if we're at 7, there is a .1 chance we goto the set, a 0.3 chance of 7 and a 0.4 chance of 8. Then our probabilities start over again.

Recalling $u_1=1$ and solving using simple algebra gives: $u_7=0.2121$ and $u_8=0.1212$

Similar analysis can give the odds of ending up in the other two ergodic states. We can use a similar technique to calculate time to leave the transient set T={7,8}.

Define v_i=time first leave T, given start at i. Then

$v_7=1+0.3\ v_7+0.4\ v_8$

$v_8=1+0.4\ v_7+0.3\ v_8$

giving $v_7=v_8=3.333$ time steps

We can generalize the 1st step analysis for any chain, A as follows.

Rearrange P to P=$\begin{bmatrix}Q & R\\ 0 & I\end{bmatrix}$ by replacing ergodic sets with a single absorbing state, and reordering. Thus Q gives a matrix mapping transients to transients and R gives a matrix mapping transients to ergodic sets. Then

$$P^n=\begin{bmatrix}Q^n & (I+Q+\ldots+Q^{n-1})R\\ 0 & I\end{bmatrix}\text{ and }P^\infty=\begin{bmatrix}Q^\infty & WR\\ 0 & I\end{bmatrix}\text{ where } W=(I+Q+Q^2+\ldots)=(I-Q)^{-1}$$

Then WR is a matrix that gives the odds of transition to any absorbing set (rows) from any transient state (cols). Similarly, W **1** gives the mean time to absorption given starting from transient state (row).

Continuous time

Note that a continuous time chain can easily be decomposed into two separate random processes:

- A discrete time chain conditioned on the fact that a change has occurred
- A continuous time process giving the time between changes

Lets briefly examine the second process. In order to be Markov, the odds of jumping after to be the same at time t or t+τ. This is called memoryless. We will show that the exponential distribution is Markov in this sense (it can be shown that the exponential is the only memoryless distribution).

Let T=time of jump. Then, we need P{T>t+τ|T>t}=P{T>τ}, but P{T>t+τ|T>t}=P{T>t+τ and T>t}/P{T>t}=P{T>t+τ}/P{T>t} since τ>0 = exp(-λ(t+τ))/exp(-λt)=exp(-λτ) QED.

Thus in any continuous Markov chain, the time between jumps must be exponential!

It is fairly easy to figure out how to specify transition probabilities for a discrete time chain – just use the one step transition probabilities. This then easily generalizes to $P_n=P^n$.

It is not nearly so obvious what the appropriate transition matrix for continuous time should be. But think about the deterministic case – difference equations are replaced by differential equations with infinitesimal time steps. This is what we need to do for a chain as well.

Let P_t be the transition matrix for time t. Note that as $t\to 0$, $P_t\to I$ which is not very useful. Lets instead use:

$G=\lim_{t\to 0}\frac{P_t-I}{t}$ which conveniently gives us: $P=I+Gt+o(t)$

G is a matrix whose every row sums to zero. The diagonal element gives the rate at which a state is entered (non-diagonal) or left (negative of diagonal). In particular, these rates are the rate parameter (λ) in an exponential distribution. The negative diagonal gives the rate parameter for an exponential distribution giving the time to jump out of a state. The ratio of a non-diagonal to the negative diagonal gives the probability that if a jump occurs, it will be to that state.

Given the generating matrix, G, we can quickly deduce several equations:

$P'_t=GP_t$ (since $P'_t=(P_{h+t}-P_t+o(h))/h=GP_t$) This is called the Forward Chapman-Kolmogorov equation

$P_t=e^{tG}$ (solution to above)

$\pi=P\pi \Leftrightarrow G\pi=0$

Example:

Consider the discrete two-state Markov chain given by:

$$P=\begin{bmatrix}1-\alpha & \beta\\ \alpha & 1-\beta\end{bmatrix}$$

I.E. it leaves state 1 with probability α and leaves state 2 with probability β. Recall that the stable distribution is given by: $\pi=P\pi$ or $\pi=[\alpha/(\alpha+\beta)\ \ \beta/(\alpha+\beta)]$

The continuous time equivalent would be:

$$G=\begin{bmatrix}-\alpha & \beta\\ \alpha & -\beta\end{bmatrix}$$

Note that columns sum up to zero instead of one.

So the general solution is $P_t=e^{tG}$ and the stable distriubtion is given by $\pi=G\pi$.

To solve for P_t we need to diagonalize G. Let $B=\begin{bmatrix}-1 & \beta\\ 1 & \alpha\end{bmatrix}$ $\Lambda=\begin{bmatrix}-(\alpha+\beta) & 0\\ 0 & 0\end{bmatrix}$ and $G=B\Lambda B^{-1}$

so $P_t=e^{tG}=B\,e^{t\Lambda}\,B^{-1}=B\begin{bmatrix}e^{-(\alpha+\beta)t} & 0\\ 0 & e^0\end{bmatrix}B^{-1}=\frac{1}{\alpha+\beta}\begin{bmatrix}\alpha h(t)+\beta & \beta(1-h(t))\\ \alpha(1-h(t)) & \alpha+\beta h(t)\end{bmatrix}$ where $h(t)=\exp(-(t(\alpha+\beta))$

note that as t ranges from 0 to ∞, h(t) varies from 1 to 0. So at $t\to 0$, $P_t\to I$, and as $t\to\infty$, $P_t\to\frac{1}{\alpha+\beta}\begin{bmatrix}\beta & \beta\\ \alpha & \alpha\end{bmatrix}$, the stable distribution which turns out to be the same as the discrete case. However the time to a jump from state 1 is given by an exponential distribution with parameter $\lambda=\alpha$ instead of a geometric distribution with parameter $p=1-\alpha$.

Independent Increments (Add/Remove)

When we study chains we study a finite state space where we can conceivably jump from any state to any other state.

Another very useful type of probabilistic model is what I call add/remove models. It is intended to represent the count of a number of objects. Thus the state space is infinite and goes from 0,1,2, … ∞. Moreover, each stochastic event represents at most one object being added or removed. Thus we only ever jump to neighboring states.

We will give some more concrete examples in a minute, but first we will present two general results

General results for add/remove models

Setup

let λ_i represent the probability (or rate in a continuous model) of increasing – i.e. the birth rate. Let μ_i represent the probability of decreasing – i.e. the death rate. Note that the birth and death rates are functions of i — i.e. the birth and death rates can be different if the state is 1 vs if the state is 100.

Thus: $G = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \\ \mu_1 & -(\lambda_1\mu_1) & \lambda_1 & 0 & \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \mu_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \\ & & \dots & & \end{bmatrix}$ where G is an infinite matrix

Stable distribution

A stable distribution, π^* is given by $0=\pi^* G$. This gives us an infinite set of equations:

$-\lambda_0\pi_0 + \mu_1 \pi_1 = 0$ i.e. net of flows out of 0 and into 0 =0

$\lambda_{n-1}\pi_{n-1} -(\lambda_n+\mu_n)\pi_n + \mu_{n+1}\pi_{n+1} =0$ i.e. net of flows out of n and into n =0

solving we get:

$\pi_1=\lambda_0/\mu_1 \pi_0$

and more generally: $\pi_n = \frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n}\pi_{n-1}$ let $\theta_n=\frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n}$ ($\theta_0=1$)

Now, since $\Sigma\pi_i=1$, then $\pi_0=1/\Sigma\theta_i$. Note that if $\Sigma\theta_i=\infty$, then $\pi_0=0$ and $\pi_i=0$ — i.e. there is no stable distribution.

First step analyses

A first step analysis and solution of recursive equations gives us u_i=P{hit 0 | start at i}

define $\rho_i \equiv \frac{\lambda_1\lambda_2\dots\lambda_n}{\mu_1\mu_2\dots\mu_n}$ ($\rho_0=1$)

Then $u_m=\frac{\sum_{i=m}^{\infty}\rho_i}{\sum_{i=0}^{\infty}\rho_i}$ Note that if m is finite and $\sum_{i=0}^{\infty}\rho_i = \infty$, then the top term is also ∞ and u_m=1

Define $T_{0|m}$=E(T where first hit 0 | start at m)

Then $T_{0|m}=\sum_{i=1}^{\infty}\frac{1}{\lambda_i\rho_i}+\sum_{k=1}^{m-1}\rho_k$

Existence of a stable distribution

Recall that in the finite case, we could divide up the state space as:

$S=T \cup EC_1 \cup EC_2 \cup \dots \cup EP_1 \cup EP_2 \cup \dots$

where the E_i's are each closed and irreducible (and the EC's are mixing and the EP's are periodic, and the T is transient.

In the finite case, at least one E (either EC or EP) must exist.

In the infinite case we get several complications:

- EP's go away

- some E does not have to exist (can have S=T)
- The EC's split into two: E* and E^0

So we have:
$S=T\ [\cup E^*_1 \cup E^*_2 \cup \ldots \cup E^0_1 \cup E^0_2 \cup \ldots]$

E*'s have a stable distribution and may be finite or infinite
E^0 are always infinite and are called null recurrent – the stable distribution is $\pi^*_i=0$ for all i
All states could be transient with no E's

Count/Renewal

Poisson

The simplest example of an add/remove model is the Poisson process. The Poisson process counts occurences of rare random events (clicks of a geiger counter, speciations, long distance dispersals, fires, etc).

We can imagine three variables of interest:
N(t) = # of events occurred in amount of time t
M=W(1) = amount of time between events
W(N) = amount of time for N events to occur (note that this is a sum of N independent N random variables)
Each of these is a random variable.

The distribution functions for these are well known:
N(t) is distributed as the Poisson with rate λ
M is distributed as the Exponential with rate λ
W(N) is distributed as the Gamma with rate λ and shape N

Let's look briefly at how this fits into our add/remove model
$\lambda_i=\lambda$ for all i (i.e is constant)
$\mu_i=0$ for all i

$$\text{So } G=\begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \\ 0 & -\lambda_1 & \lambda_1 & 0 & \\ 0 & 0 & -\lambda_2 & \lambda_2 & \ldots \\ 0 & 0 & 0 & -\lambda_3 & \\ & & \ldots & & \end{bmatrix}$$

giving
$p'_0(t)=-\lambda p_0(t)$
$p'_n(t)=\lambda p_{n-1}(t)-\lambda p_n(t)$

so
$p_0(t)=e^{-\lambda t}$
and substituting into the 2nd equation gives
$p'_1(t)=\lambda e^{-\lambda t}-\lambda p_1(t)$ gives $p_1(t)=\lambda t\, e^{-\lambda t}$ assuming $p_1(0)=0$ (check that it satisifies the dif eq)
and generally $p_n(t)=(\lambda t)^n/n!e^{-\lambda t}$
which is the Poisson distribution!

We can substitute into the general equation for π^* and we see $\theta_n=\lambda^n/0=\infty$ so clearly $\Sigma\theta_i=\infty$ and there is no stable distribution. Hence all states are transient (which makes sense as the processes is ever increasing to infinity)

Non-Markovian=Renewal
The non-Markovian generalization i

Queue

M/M/1

Non-markovian

Birth/Death

Random Walk

Bounded

Infinite

Continuous time

Martingale

Stationary

IID
derivative of RW

Stationary

Stochastic Differential Equations

SDE's as approximations to deterministic
$dX=f(X(t),t)dt + g(X(t),t)dB(t)$ $(dB(t)=W(t))$

Technical condition to meet	Practical requirements
Additive noise	• substitution of a parameter with a parameter plus noise should lead back to an additive model • absorbing states should be absent (or one should have assurance that absorbing states are unimportant for the system - e.g. population sizes are large) • log transforms or other transforms may make the noise additive
White noise	• system under study should be memoryless (deterministic recovery times should be fast relative to the frequency and magnitude of the noise) • alternatively, the noise should be white
Continuity of noise	• deterministic recovery times should be fast relative to the frequency and magnitude of the jumps if any
Multiple attractors	• apply deterministic systems to systems with one global attractor • alternatively, ensure that the basins of attraction are large relative to the disturbances
Chaos	• if the deterministic system is chaotic, do not use the model to do anything more than make one step predictions (i.e. one most have an actual noise realization that can be combined with the deterministic system) • more work needs to be done on exploring this mathematically
Mean path questions only	• although not discussed above, it is obvious that deterministic systems are only good at predicting average behavior. If one wants to look at variation, stochastic systems are needed.

Movement	**Jump**							**Continuous**
Time	**Discrete**			**Continuous**				**Continuous**
State space	{}	∧	**∧ multiplicative**	{}, ∧	∧	**∧ mult**	∇	∇
Examples	Markov chain	Counting Random walk	Branching process	Markov jump chain	Rand walk Poisson	Birth-Death	Logistic w/ catastrophe	Brownian motion SDE
atom	P	P or (λ, μ)	$p_k=P\{\#\text{ offspring}=k\}$	G			???	m(x), v(x)
differn* eq for P	$P_n=P^n$		$G(s)=\Sigma p_k s^k$	$P_t'=PG=GP$ w/ $P_0=I$			???	(4) & (5) above w/ W=0
$\pi_t(x)$	$P^n\pi_0$		$G_n(s)=G^n(s)$ $E(\pi_n)=G_n'(1)=E(\pi_1)^n$ $Var(\pi_n)= G_n''(1)+G_n'(1)-G_n'(1)^2$ $=n\sigma^2$ or $\sigma^2(\mu^n-1)\mu^{n-1}/(\mu-1)$	$e^{Gt}\pi_0$		$\Sigma s^j P'=\Sigma s^j G$ $G(s,t)=\Sigma s^j p_j(t)=E(s^{X(t)})$ solve difeq in G use G', G"	???	$f(x,x_0)=G(x,x_0)/E(T_{a\|x0})$ pseudostationary distr (i.e | not extinct)
$\pi^*=\pi_\infty$	$\pi^*=P\,\pi^*$		N/A Goes to 0 if $\mu\le 1$ Goes to 0 or ∞ if $\mu>1$	$G\,\pi^*=0$		see branching	???	$C\frac{1}{v(x)}*\exp\{2\frac{m(x)}{v(x)}dx\}$ w/ C a constant to give 1
first passage, T_x starting at state x	first step method let $P=\begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}$ then $E(T_i)=((I-Q)^{-1}\bullet\mathbf{1})_i$		????	calc $E(T_i)$ for embedded $P=W/(W\bullet\mathbf{1})$ then $\Gamma(E(T_i))$		G(0,t)	???	$E(T_{a\|x0})=\int_a^\infty G(x,x_0)dx$* a=lower limit (e.g. 1)
U_{yx} absorption into y given start at x	$U_{ij}=((I-Q)^{-1}R)_{ji}$		if $\mu>1$ then P{extinction} found smallest >0 s ə s=G(s)	same as embedded chain P		???	???	$p(b\|x_0,a)=[S(x_0)-S(a)]/[S(b)-S(a)]$

* = formula for Green's function (Karlin & Taylor 1981)

$s(x)=\exp(-2\int_x m(u)/v(u)\,du)$

$S(x)=\int_x s(u)\,du$

$G(x_0,x)=2[S(x)-S(a)]/[s(x)v(x)]\ \forall x<x_0$

$=2[s(x_0)-s(a)]/[s(x)v(x)]\ \forall x>x_0$

for simple exponential birth-death

$E(X(t))=I\,e^{(\lambda-\mu)t}$ $\quad Var(X(t))=2I\lambda t$ if $\lambda=\mu$ $\quad$ else $I\,(\lambda+\mu)/(\lambda-\mu)e^{(\lambda-\mu)t}(e^{(\lambda-\mu)t}-1)$