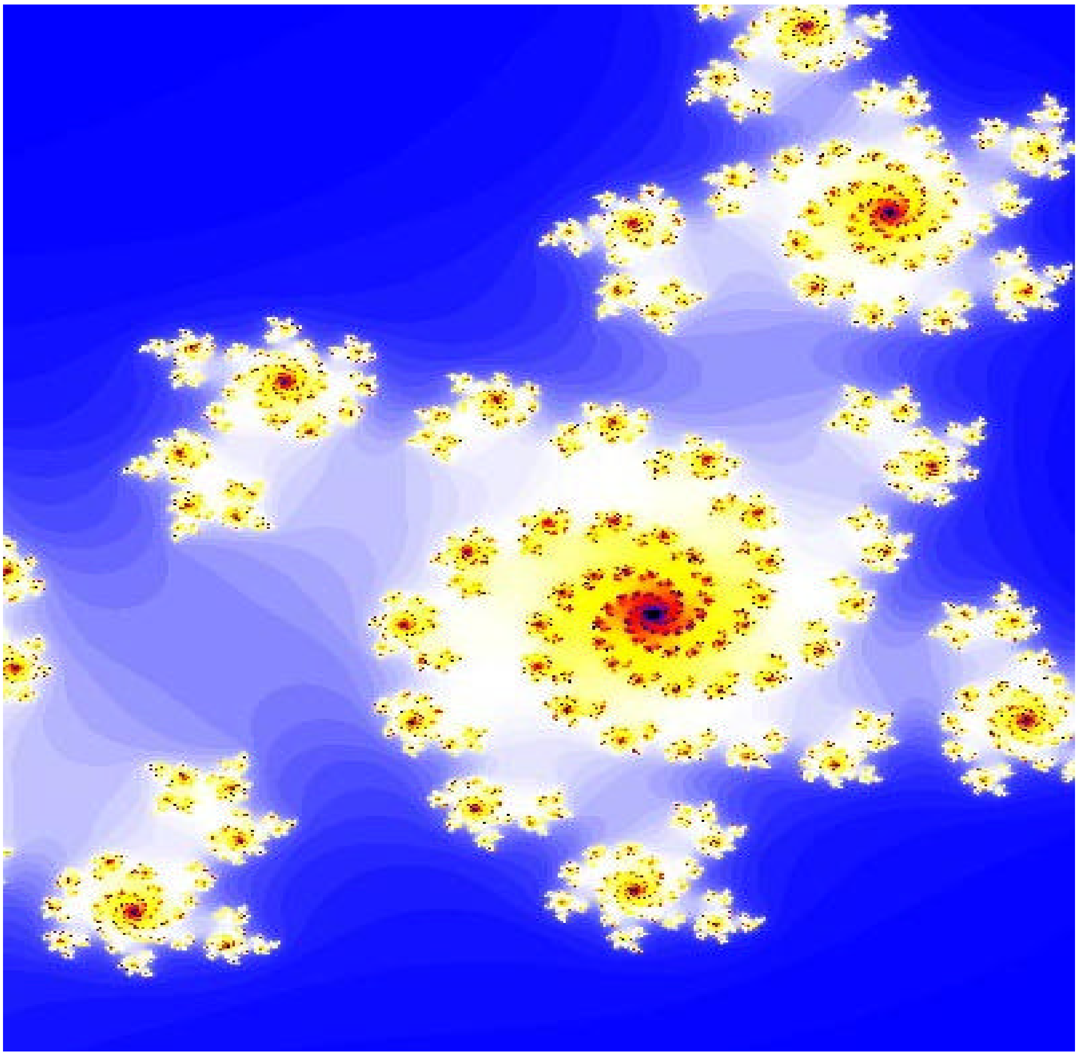
An Introduction to Mathematical Modelling

Michael Alder



HeavenForBooks.com

An Introduction to Mathematical Modelling

by

Michael D Alder

HeavenForBooks.com

An Introduction to Mathematical Modelling

Introduction

This book is based on a course given to first year students doing Calculus in the University of Western Australia's Department of Mathematics and Statistics. The unit was for students mainly from the Life Sciences, with some Economists, Social Scientists, Computer Science students and others, and the aim was to give them some understanding of the uses of Calculus in their areas of work. The book was about half of the complete course, the rest being statistical modelling.

Everything I write in this book from now on is addressed to the reader on the assumption that he or she has a similar background, and similar or broader interests. I assume, in other words, that you are not a mathematician, physicist or engineer or that if you are you have an uncommon and admirable breadth of interest in the rest of the world.

The amount of Mathematics in the soft sciences has been increasing dramatically in the last few decades. You might be puzzled as to why this is. Are mathematicians somehow able to coerce other departments into pushing their own merchandise down everybody's throats? No, actually we are very bad at this. But what is happening is that many areas that used to be done in a literary sort of way, with people arguing in natural language, have suddenly become amenable to *modelling*. The main reason is that computers have come into our lives. This means we can explore much more complex systems than could have been dreamed of twenty years ago. The impact on the research level has been dramatic over the last twenty years, and this is slowly filtering down to the undergraduate courses.

There are other reasons, more fundamental than the computer revolution why this happens: sciences evolve. They start off as collecting butterflies and newts and flowers and in a century they are cloning sheep. The evolution needs an increase in the precision with which you communicate the facts you have discovered, and Mathematics is the language of choice here. The rise of the Physical Sciences and the Engineering that rests upon it has benefited from, and contributed to, the Mathematics that we now have. And practical people wouldn't buy the stuff if it didn't work. Well, exactly the same factors make Mathematics useful to other people too.

Science works by building 'models'. Not little cardboard and plasticine models, but models made out of symbols. We can play with the symbolic models and adjust them until they start to behave in a way which resembles the things we care about. When we have done this, we get an understanding of the things we care about which is much deeper than we could ever get if we stuck to words and pictures. Mathematical models do not replace words and pictures, they sharpen them.

So models deepen our understanding of 'systems', whether we are talking about a mechanism, a robot, a chemical plant, an economy, a virus, an ecology, a cancer or a brain. And it is necessary to understand something about how models are made. This book will try to teach you how to build mathematical models and how to use them.

There is a huge range of useful models invading the Life Sciences: Richard Dawkins' [1, 2, 3] little stick creatures which evolve and mutate can sharpen our ideas, and also dramatise them so you can *see* evolution working. Cellular automata can tell us things about growth and evolution that again sharpen our ideas. The Social Sciences increasingly use models for both numerical predictions and for qualitative behavioural analysis.

Calculus is largely about systems which change in time and the problem of saying something about how this can happen. Since many biological and social systems do evolve, there are plenty of applications of Calculus, and some of them are very illuminating. So in this book I shall restrict myself to Calculus; more specifically to what can be done with Ordinary Differential and Difference Equations. There are lots of types of models we could look at, but it is a good idea to start off with a type of model which has shown itself to be very useful over a colossal range of applications.

I am a friendly, chatty sort of bloke, and this is a friendly, chatty book. I have tried to make it as readable as possible, but it would be a good idea to read bits of other text books as well.

Mathematics is a lot easier if you can see why things are done the way they are, rather than just learning the stuff off by rote. It is also a lot more fun this way. Most text books assume you already see why, but experience suggests that this is in fact where the problem lies. Which is why I am discursive and friendly. Best of luck!

Contents

1 Fundamentals **9**

1.1 Systems and States 9

1.2 Idealisations, Real Numbers and Guns 16

1.3 Bacteria and People 27

1.4 How to Do It Yourself 39

1.5 Summary and Conclusions 44

2 Growth **47**

2.1 Bacteria, People, Money 47

2.1.1 The Logistic Equation Revisited 47

2.1.2 Death and Taxes 53

2.1.3 Money 64

2.2 Of Mice and Men. And Rats and Women 76

2.3 History: Truth, Lies and Radioactivity 83

2.4 Summary and Conclusions 93

3 A Menagerie of Difference Equations **99**

3.1 Some Definitions 100

3.2 Linear (and Affine) Difference Equations 105

3.2.1 First Order Difference Equations 105

3.2.2 Second Order Difference Equations 114

4 Iterates of maps: Stability **125**

4.1 Cobwebs and Chaos 126

4.2 More about Stability 143

4.3 Heartbeats 148

5 Higher Dimensional Systems **151**

5.1 Eating People is Wrong 151

5.2 But Killing them in War is OK 154

5.3 The Dismal Science 155

5.4 A Cheap Trick 160

6 Ordinary Differential Equations **161**

6.1 First and Second Order linear and affine Equations 161

6.1.1 First Order Linear and Affine Equations 161

6.1.2 Second Order Linear and Affine Autonomous Equations 162

6.2 Systems of First Order ODEs 169

6.2.1 A Smart Trick 173
6.3 Some Applications 175
6.3.1 Predator Prey Systems 175
6.3.2 Competition 177
6.3.3 Chemical Kinetics 178
6.3.4 Epidemics 180
6.4 Zeros of Vector Fields: Stability 184

7 Serious Modelling 191

7.1 Why Models Matter 191
7.2 How To Model 193
7.3 One Last Model: Days of Empire 195
7.4 Concluding Thoughts 202

Chapter 1

Fundamentals

I am going to start off by talking about some basic ideas in a chatty sort of way. It is tempting to take for granted that you already know things that you might not. I am going to fight this temptation and am going to give you a sort of scientist's perspective on the world. It is certainly different from the Joe Sixpack or Homer Simpson view of the world, and it may be different from yours. So lean back and relax and enjoy the discussion.

1.1 Systems and States

One of the words which scientists and engineers throw around a lot without ever actually defining it is the term 'system'. I don't know how to define it either, but then I couldn't define an elephant, but I can recognise one when I see it. So I shall give some examples.

Example 1.1.1

1. *a computer*
2. *a gun*

3. *a flush toilet*
4. *an ecology*
5. *a robot*
6. *an economy*
7. *a virus*
8. *a human being*
9. *a tree*
10. *an industrial complex for mining iron ore*

Systems may have sub-systems inside them: a human being is a large number of interacting subsystems, and so is a robot. In fact you *are* a robot, although one made of meat and squishy stuff rather than metal and silicon. This is because the rigid parts of a robot are large, while the building blocks of you are molecules.

This might make it seem as if absolutely anything is a system. So here are some things that usually aren't:

1. a painting
2. a rock-concert
3. a footy match
4. a can of baked beans

The difference is one of attitude as much as anything; if you are a promoter, then putting on a rock-concert does indeed require a system, and so does manufacturing the can and baking the beans or organising the footy match. A gun to James Bond is merely a tool of the trade, not a system, but if you are an engineer in the middle ages and you want to know whether or not you can point a cannon up at the right angle and put just enough gunpowder in, and just the right weight cannon ball to lob the ball over a castle wall without blowing up the gun, then you have a system. And you'd better have a good idea of how it works. Likewise if you are Purdey, or Smith or Wesson.

If I had to define what I mean by a 'system', I think I should start off by saying that it is something that people make measurements on it and get out numbers. That may not seem to be true of all my examples at first sight. What measurements do you make on a flush toilet? Well, you press a button or pull a chain, and some water comes down into the pan and disappears. The casual user just feels glad it happened that way and leaves, but the guy who designed it had some other things to think about; how does he make sure that the water stops coming, how much water does he want to come, how does he fill the tank up afterwards? He really needs to know how *much* water came down, and how long it took, and how fast it came. You only have to think of what might happen if he got the numbers wrong to realise that you have rather relied upon him getting it right. On a good many occasions by now. So there is a qualitative aspect to the system, there is the kind of thing it does, but there is a quantitative aspect too, so some measurements have been made in building it. It may not matter whether it's a pint or only half a litre comes down the pipe, but a thousand gallons or two millilitres would change the usefulness of the machine.

Of course, you can take the view that so long as it works you can ignore the questions of how it does it, but somebody somewhere has to understand these things, or we *would* be in trouble. And you wouldn't be reading this book if you didn't accept that it is worthwhile trying to understand something about some complicated systems. So thinking about a few simple systems is not a bad idea.

A tree takes up water, absorbs carbon dioxide and takes up sunlight. How much of these things? How much of the energy it is exposed to in the form of sunlight does it actually use, and where is it used? Once we start to ask these questions we have begun to understand how a tree works, which is quite a different activity from just admiring it. I am not saying that we should stop admiring trees, I am saying that we shouldn't just stop there. And you get to admire trees a whole lot more when you see how incredibly intricate is the machine that is a tree. You see much more in the tree than some poor philosopher who does not think in these terms. By contrast, a painting is there to be looked at, and unless you are an artist or an art critic, that is about all you can do. And if you are an art critic, the only other thing you can do is to write or talk about seeing it.

Also, paintings are static. While trees grow and change and adapt. And if you want to know how they do it, you need to study what they do quantitatively.

So the second part of my definition of a system is that the measurements you make change in time, and you often want to predict how they will change. When you know, you understand something you didn't understand before. To any scientist, this is a ***GOOD THING***. A scientist wants to understand things.

Exercise 1.1.1 *Make a list of systems that you are likely to have to understand in the course of your professional life. What kinds of measurements are needed to make understanding these systems possible?*

Imagine that you were going to study a quite different area, possibly the one your parents wanted you to work in. What systems would you have to understand and work with there? What sorts of measurements are relevant?

You might like to consider legal systems and financial accounting systems in this light.

What constitutes a measurement can be tricky too. Usually we get numbers out: if I weigh myself to find out if I have been pigging out too much recently, or put a tape measure around my middle, I get hard evidence and some numbers. But if I point a camera at something and get a digital image, then typically I get an array of maybe 1000 by 1000 *pixels.* And each pixel can be described by the amount of red, green and blue light in it. And the amount of light gives the brightness, and is measured by a number. On a computer screen, it is commonly a whole number between 0 (black) and 255. So in a sense, when I take a picture which comes up on a computer, I have made about three million measurements of whatever it was the camera was pointing at. And your eye does something very similar. So in a sense, when you look at a painting, you *are* making measurements of it. Rather a lot of them. But since you don't get to write down the values of the neural excitation, it isn't science. Whereas if a video camera makes the measurements and the values are sampled by a computer, it may be science. It all depends on what we do with the numbers.

How many numbers does it take to measure a tree? This is rather like asking how many beans make lots. You make up your mind which parts of the trees behaviour you want to study and measure those. This still leaves lots of other things for other people to measure if they are studying different aspects of the tree. Trees are mind-boggling in their complexity when you come down to it.

When systems change in time, we say that what changes is the *state* of the system. The state, at any time is measured (in principle) by some set of numbers, and when the numbers change in time, we say the state is changing.

Example 1.1.2 *When we look up at the night sky and see the moon and some planets and stars, we are making measurements. If we are good scientists, we make this a bit more public by saying where the moon is in the sky. We could do this by saying 'just over the roof of the building next door', but saying 'thirty degrees above the horizon and forty-two degrees North of Due East' is a lot more useful to observers at an observatory. Since the moon will be in a different place in an hours time, it would be a good idea to give the time and date. This then becomes a measurement of a small part of the solar system.*

Example 1.1.3 *A bacterial culture develops in a petri-dish. We estimate the area of the culture by counting the number of little squares of a piece of graph paper which are covered by the mould. We measure the amount of sugar in solution in the medium with a glucometer, and the temperature. These three numbers change in time, so we also measure the time, and the results tell us something about the way bacteria reproduce and grow under different conditions.*

Example 1.1.4 *We count the number of Europeans who use the French style of pronouncing the letter 'r' (something between a gargle and a heavy cold) per head of population in sampled towns and villages. We repeat at intervals* *of ten years. We discover that the French style is invading Germany at a rate of metres per day. (It started in Paris in the last century, perhaps by somebody who was wounded in the Napoleonic Wars.) This tells us something about how anxious people are to keep up with fashions.*

Example 1.1.5 *We put two oscilloscopes into an electrical circuit. This tells us something about the voltage and current relationships. Basically, an oscilloscope measures voltages, but it measures them fast.*

Example 1.1.6 *We measure the temperature, pressure and radioactivity at a number of points inside the boiler of a Nuclear Power plant, also the degree of penetration of control rods. This tells us when to head for the hills. Also other useful things.*

Example 1.1.7 *We measure the amount of money created by the government printing presses; we also measure the credit created by banks and other financial institutions. We measure the inflation rate by taking the price of a basket of commodities. We do this for several countries over several years, and learn that inflation* can *be controlled (although there is a price to be paid).*

Example 1.1.8 *We measure the ability of people to remember lists of words and look to see how the recall rate changes in time. So we learn something about how memory works.*

Example 1.1.9 *You ask five hundred people if they would vote for one major party or the other or None of The Above if there was an election tomorrow. You do this every month and make predictions about a general election, or correlate the ups and downs with political events to try to explain the results.*

You probably get the idea by now. The point I want to get to is this: the Scientist decides what to measure, and gets a list of numbers by clapping some kind of instruments onto the system he or she studies. This could mean anything from electronic instruments to sending out people armed with questionnaires. Doing it again later may mean waiting a nanosecond or a decade. But what comes out is a sequence of *vectors*, lists of numbers.

The numbers are not *just* numbers. They tell us things if we listen to them. A physician takes your blood pressure, your temperature, your pulse rate and gets an estimate of other *state variables* by asking you questions. Other variables may be measured in laboratories, or you may have to have X-Ray scans or CAT scans. These describe the state of the system (you) at some time. They don't, of course, tell us *everything* about you, but most of what there is to know about you is of limited interest to the physician.

The Mathematics starts with the sequence of vectors. In order to be any fun (and make no mistake, Science is about having fun) we want to be able to figure out what numbers are coming next. Sometimes we can do things to the system and look to see what it does to the numbers, sometimes (as with stellar systems or brains) this is not feasible. (The stars are too far away and the owner of the brain may object.)

This is how Science is conducted. The basis for figuring out what how the numbers change as you observe or meddle with the system is called a 'model'. If it is a whole family of systems or the Universe or something very big, the model may be promoted to being called a *theory.* A theory may tell you what sort of model to use in a particular case. But the whole point of theories or models is to allow us to make sensible guesses at what will happen if we do *this.* Scientists almost never just poke something to see what happens. They almost always poke something to see if it will do what they thought it would. If it does, the model is validated or the theory is confirmed. This is nice, the scientist feels pretty good. If it doesn't, there is something wrong with the model or the theory, and we have to do some thinking. This is also nice, the scientist has learnt something. It's a win-win situation being a scientist. Aren't you glad you aren't a literary critic?

To summarise, in Science we are quantitative whenever we can be; we measure numbers. We want to know what will happen if we make changes, or even if we just keep on making measurements. A model of what is going on to give rise to the numbers is something we construct in order to understand the system. We then need to test the model to see if it is giving good answers. We never come to absolute truth, but we do get to the point where we can feel that we have a good grasp of how something works. All developed Science looks like this, although the variety of Science is pretty amazing itself. Pseudo-science looks quite different, it uses Mathematics only to impress the illiterate, if at all.

In the section after the next, I shall take the problem of the bacterial mould and think about how to construct a very simple model for just one measurement, a measurement of how many bacteria there are. This is doing the sensible thing and looking at a nice easy case first so as to cut our teeth on it. It will turn out that different systems can have almost identical models- and so it would be a mistake to think that you can stop thinking about the subject right now because you are not interested in mould. Anyway, it will turn out to be quite a lot more complicated than you might think.

1.2 Idealisations, Real Numbers and Guns

I have said that Engineers, Scientists and Mathematicians (and the dividing line is not sharp) like getting the state of a system precisely specified at any time by clapping on a set of measuring instruments and obtaining a list of numbers, and repeating periodically. Numbers do not, of course, tell us everything about the system, but they usually tell us a lot more than words in natural language do.

But what sort of numbers? In practise, you measure your weight and say ‘it’s 78 kilograms’. Or if you are honest, you say ‘I can make it anywhere between 79 and 85 kilograms depending on how I stand on the scales and whether I hold on to the washbasin.’ So there is always some degree of *precision* in the way you specify your number. There is also the question of whether the number is *accurate*, a somewhat different matter. I might announce that my weight is 82.397654522893 kilograms, which is amazingly precise but not at all accurate. Any measuring instrument has some built in limits on how precise it can be, and also some limits on how accurate it can be. And yet Mathematicians build models which produce answers like $\pi/2$, which is a real number with a decimal representation which goes on for ever. You can’t hope to measure to infinite precision, why do the models have it built in, as they mostly do? Well, it is very useful to distinguish between the properties of the world and the properties of the measuring apparatus. We might measure my eight with a domestic weighing scales one day, a medical scales the next, and something found only in expensive chemical laboratories the next. There is some sort of an assumption that we could increase the precision as much as we liked, and if I have actually *got* a weight, then the different systems ought to give answers which have something in common.

So we model systems in what are, seen from some points of view, a very odd way indeed: we have a model which gives infinite precision, but we feel reasonably happy if we get agreement between model and observation to some limited precision which depends on the measuring apparatus.

We may also model the measuring apparatus to some extent: we may replicate a measurement and get a different answer. If we measure what is on the face of it the same thing a hundred times and get a lot of different answers, we talk about observation error or instrument noise. Probability Theory is used for modelling this kind of thing. It is a very important subject, and I shall say absolutely nothing about it because it is not my job to do so, not in this book anyway.

Sometimes our measurement is in fact a count, so we can only have whole numbers, or *integers* to use the two-dollar word. Sometimes we can only make a test as to whether something is there or not, in which case there are only two possibilities. This is sometimes called a *logical* measurement. Filling in questionnaires provides data like this in many cases. But most commonly we use real numbers, sometimes to approximate integers. For instance, if I am interested in the number of bacteria in a culture and I measure an area, I shall get a finite precision approximation to a real number, which is actually roughly proportional to a positive integer. Using real numbers looks rather odd. Similarly, it is possible to have a model of how much money is in my bank account which uses real numbers - despite the fact that it has to be a whole number of cents.

Such things as this are done in order to keep life simple. It isn't at all obvious that it will in fact do so. As you will discover, it actually does, at least some of the time.

Some extremely peculiar things will sometimes happen in models which you might feel are grounds for not using them at all. But there is only one test of whether a model is any good, which is how well does it describe the measurements? The fact that it is clearly loony is something we cheerfully ignore. This often bothers students at first, who feel that since the model is loony we shouldn't have anything to do with it. Looniness may be catching. But we in fact make the most bizarre and clearly false assumptions with a happy smile; so long as we can validate the model by comparing predictions with observations and we get adequate agreement, we are going to do it. So don't go around worrying whether a model is TRUE because it almost never is; ask yourself whether it is useful.

Example 1.2.1 (The Cannon Ball)

Suppose the year is about 1650 and you are the bloke who owns a cannon, a pile of cannon balls and a whole lot of gunpowder. You want to know how the angle at which you aim your cannon affects the distance the cannon ball travels. I shall set up a simple model as an example of how modelling is done, and you should enter into the spirit of the things as a sort of game, and not suppose that the fact that guns and gunpowder do not much interest you is a reason for skipping this.

To people who do it lots, it all looks quite simple, but to beginners it is rather bizarre. So it is worth reading this slowly and carefully.

First I assume that the cannon ball is a point with some mass, or a particle *as it used to be called. This assumption is clearly silly, but I make it anyway.*

Second I assume that my cannon is a line segment of zero length, which is an idea which makes no sense at all. I shall place this line segment of zero length on an infinite horizontal plane, and I shall look only at the line which is the intersection of this infinite plane and the vertical plane through the line segment. Insofar as it is possible to draw anything so unlikely, it is shown in figure 1.1.

The cannon has been made into a short stubby line, the ball into a black blob, the path or trajectory *of the ball into a dotted curve.*

The fact that the cannon is a line segment means that I can describe it as

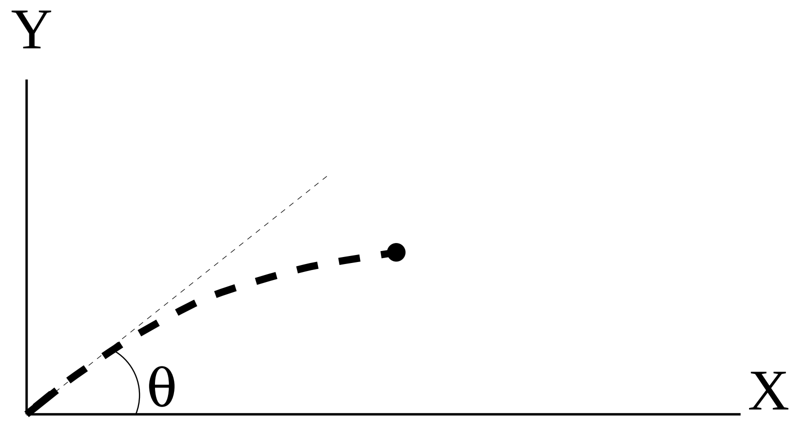


Figure 1.1: An unlikely cannon

having an angle at which the bullet starts off before it leaves the barrel of the cannon, and the fact that it has zero length means that I can give it a precise position. In fact I shall put it at the zero location on the real number line, which is the line underneath the trajectory of the ball. Similarly, the fact that the ball is a point means that I can describe its position at any time by giving its X and Y coordinates and pretend that these can be real numbers. So the third assumption is that I can (at least in principle) measure to infinite precision and get real numbers specifying (a) the distance covered by the shadow along the line over which the cannon ball travels, (b) the height on a real line 'flagpole' erected over the starting position (bent so as not to get in the way of the zero length cannon?) and (c) the time measured from the instant the cannon is fired.

Some of the difficulty students have with model building is the sheer unreasonableness of these asumptions. When Newton modelled the Solar System, he modelled the Earth as a point. This is not reasonable to people who live on it. It is crowded, but not that *crowded.*

I continue with this model even though it is clearly loony, since none of the *assumptions made so far is the least bit reasonable.*

The fourth assumption is that the whole operation takes place in a vacuum. You might ask why the bloke who fires the cannon doesn't die of asphyxiation, but we shall suppose he fired the cannon first. I suppose this is another assumption, but I shall pretend I didn't think of that issue.

The fifth assumption is that the world is flat, and gravity acts downwards over the whole plane, and the sixth assumption is that it is also independent of height. We know that actually the Earth is more or less spherical and gravity obeys inverse square law, so these are all wrong too.

I shall now look at the position of the shadow of the cannon ball along the real line, the position of the height of the cannon ball up the flagpole and the time. In order to keep things clear, I shall give them names, I shall use t *to denote the time in seconds,* $x(t)$ *to be the distance along the horizontal line at time* t*, and* $y(t)$ *to be the height measured at time* t*. So I assume that for every time between starting and stopping there is a decimal (real) number* t *specifying that time, and that* $x(t)$ *and* $y(t)$ *are also two real numbers giving the precise location of the canon ball point at that time.*

For my next assumption, I shall suppose that both x *and* y *are twice differentiable functions. Ask yourself if this assumption is actually true, and you run into some considerable difficulties. It certainly isn't an easy assumption to test, being a sort of philosophical assumption. It is a different kind of assumption from most of the earlier ones, although it has something in common with the assumption that we could use real numbers to say where the cannon ball is and when. This is assumption seven and counting.*

Now if the acceleration due to gravity is constant and downwards (assumptions five and six) and the number $y(t)$ *is a twice differentiable function, and if the derivative* $y'(t)$ *of the function* $y(t)$ *is the velocity of the height (regarded as a sort of shadow on the flagpole at the origin) and the acceleration is* $y''(t)$*, then we can put*

$$y''(t) = -g \tag{1.1}$$

where g *is the constant acceleration due to gravity.*

This translates into English as the observation that everything falls down at a constant acceleration of one gravity if something else isn't pushing it as well, and in this case nothing else is pushing the cannon ball. This doesn't contain a new assumption about the velocity being the first derivative and the acceleration the second, because I define *velocity and acceleration that way. This does raise the question of assumption seven rather pointedly however.*

Now we can integrate 1.1 to get

$$y'(t) = v_0 - gt \tag{1.2}$$

to give the vertical velocity in terms of the time, the constant acceleration down due to gravity and the vertical velocity, v_0*, at the instant the cannon is fired. This is a bit of a problem because until the cannon is fired, the cannon ball is sitting there minding its own business, and after firing it is speeding along the barrel (which has zero length), which means that the function describing it couldn't possibly be twice differentiable. The idea that the cannon ball has two velocities simultaneously, a vertical one and a horizontal one also needs some thought: I suggest you think of a shadow on the ground which is always beneath the cannon ball, and a shadow on the flagpole which is always the same height as the cannon ball, and when I say 'vertical velocity', you think of the velocity of the second shadow, and when I say 'horizontal velocity' you think of the velocity of the first shadow. After all, if I know where both shadows are, I know where the cannon ball is, and if I can do it with position, I can do the same thing with the velocity of the cannon ball. (You could be forgiven if you have some doubts about this and wanted to make it a separate assumption, but I skip over this for the time being.)*

And the explosion that sent the cannon ball on its way happened instantaneously. The eighth assumption is that the function is only twice differentiable from the instant the cannon is fired and the cannon was actually fired in an instant. Whatever an instant is, but it corresponds in this case to the number $0.0000\cdots$ *measuring the time.*

We can integrate equation 1.2 to get:

$$y(t) = v_0 t - g\frac{t^2}{2} \tag{1.3}$$

and the constant of integration is zero because we started at height zero at *time zero.*

If we happened to know the vertical velocity, we could now calculate the time it would take for the cannon ball to return to earth, and if we knew the horizontal velocity we could work out how far it would travel in that time.

For my ninth assumption, I shall suppose that I know the actual speed at which the cannon ball leaves the cannon and the angle θ that the zero length cannon-barrel makes with the horizontal. Then using simple geometry, if s is the initial speed, we have that

$$s \sin\theta = y'(0) = v_0 \tag{1.4}$$

and

$$x'(0) = s\cos\theta \tag{1.5}$$

gives the horizontal velocity at time 0.

For my tenth assumption, I shall assume that because this is in a vacuum, $x'(t)$ is constant. This is a rather well known assumption invoking Newton's first Law of Motion. It is also rather a peculiar one, being neither wholly mathematical nor something which can easily be verified by experiment. It is a theory *that says this is an OK assumption to make.*

This gives me

$$x(t) = (s\cos\theta)\ t \tag{1.6}$$

to give me the horizontal distance travelled by time t.

It is now a simple matter to solve 1.3 for t in terms of s, θ and g, to get the time t_g at which the height of the cannon ball is zero:

$$t_g = 0 \quad \text{or} \quad t_g = \frac{2s}{g}\sin\theta \tag{1.7}$$

The first of these is when the cannon ball starts moving, and the second is when it falls back to earth, and hence the horizontal distance travelled by the cannon ball before it hits the ground is the horizontal speed multiplied by this second time:

$$x(t_g) = \frac{s^2}{g}\ 2\sin\theta\cos\theta = \frac{s^2}{g}\sin 2\theta \tag{1.8}$$

This tell us that if we keep the amount of gunpowder fixed and if we assume that the initial speed depends only on the amount of gunpowder, then the distance is a maximum when $\theta = 45^o$, (because then $\sin(2\theta) = 1$, which is as big as it can get) which seems reasonable. It also makes predictions about what will happen as we change the angle from zero to 360^o. Some of these seem fairly sensible, for example if $90^o < \theta < 180^o$ the cannon ball goes the other way. On the other hand, if θ is negative or greater than 180^o so we fire into the ground, we don't get sensible answers at all. For $\theta = -45^o$, we get the distance travelled is $-s^2/g$, whereas the actual answer is zero.

There are several points to notice about this example.

- Despite the dottiness of most of the assumptions, the model has got some sort of agreement with reality in some regions of the variable θ.
- It is easily tested experimentally to see how well it works. You could also apply it to a bow and arrow and do a fair job of checking it out. A good way of testing the model would be to measure θ and x_g for a range of values of θ, and then to draw the graph of x_g plotted against $\sin 2\theta$. If this is a straight line, you measure the slope to get s^2/g. If the line is a bit wobbly, as is likely due to measurement noise, you can find a reasonably good fit and use that. If it turns out that the graph doesn't look like a straight line, then the model is clearly wrong. If it looks to be a good straight line, then the model is of some use. It might still be wrong because it is possible that the slope is not related to the initial speed the way the model says it is. But we could still use the model for some things.
- There are some values of the variable θ for which the model gives totally daft answers.
- It would not be too hard to modify the model to deal with other cases: for example if you are on top of the only hill for miles around, your cannon ball would obviously go further. And for another example, you could change it a bit to see if you could hit the top of a nearby tower by choosing the angle appropriately.
- There are some numerical constants in the model that you might not know, and you could calculate them by working backwards: choose a

θ which you can measure and look to see how far you get. Now plug into the equation

$$x_g = \frac{s^2}{g} \sin 2\theta$$

to work out what s^2/g is. Or better, take several readings and fit the best straight line and measure the slope. Of course, this assumes the model is correct, and if it isn't you could produce garbage. But you can test at least some aspects of the model. Such constants, often unknown at the time the model was built, are sometimes called *parameters of the model*. The cannon model has only one parameter, but other models may have lots more. Warning: Scientists and Engineers tend to use the word 'model' to mean that we have actually determined what the parameters are; Statisticians tend to use the term 'model' to mean the whole family of models, with all possible values of the parameters. Beware of confusion when in mixed company.

Exercise 1.2.1 *Why do you get daft answers when $\theta = -45^o$? What is happening in the model which shouldn't?*

A defence of some of the assumptions can be given by pointing out that if they are not true, then they are not too far out. The cannon barrel may not actually have zero length, but it is fairly small compared with the distance travelled by the cannon ball. The cannon ball may not be a point mass, but it too is fairly small compared with the distance travelled, and so on. There is, in effect, a natural scale associated with the things you measure, and things that are very small on that scale we simply set to zero. The reason for doing this is that it makes our lives easier. Some people have a lot of trouble with this; they feel that if you throw out that amount of bath water then you must have lost a baby or two. Others feel that it is fun to go around making wild simplifications to see what happens. Since these are only ideas we are playing with, you can always simplify with complete abandon, but you run the risk of getting answers which are badly wrong.

So we seem to be making a *new* assumption, which is a sort of *Stability Assumption* that says that even if all the assumptions are wrong, provided they are not *too* wrong, the answers we get won't be too wrong either. Sometimes this works very well, and sometimes it doesn't. The best policy is to try it out to see what happens.

There is a rather convincing demonstration that the above model is reasonably good.

From equation 1.6 and equation 1.3 we can find an expression for t:

$$\begin{aligned} t &= \frac{x}{s\cos\theta} \\ \text{and since} & \\ y &= (s\ \sin\theta)\ t - \frac{g}{2}\ t^2 \\ \text{we get :} & \\ y &= (\tan\theta)\ x - \frac{g}{2s^2\cos^2\theta}\ x^2 \end{aligned}$$

This tells us that for any angle θ, other than $\theta = \pi/2$, the path of the cannon ball is a parabola. If you want to test this, you can get a garden hose, turn it on and hold it at different angles. The water drops are very small cannon balls, and the path they trace out looks to the eye very much like a parabola. So the scientific modeller gets a good warm feeling about the general properties of the model. I hope you do to.

If you think about it, this 'agreement with reality' is really quite astonishing: a certain amount of thinking, some scribbling down of symbols on the back of an envelope, a bit of sketching the graph of a function, and bingo! you have predicted the shape of a curve that occurs in the real world. This is amazing. If you haven't noticed that this is really mind bogglingly weird you have a serious lack of imagination. It isn't so much that the rest of us have a general fascination with curves, or with garden hose-pipes, just that we are surprised that you can get so much out with so little going in, that scribbles and doodles which are all symbolic should connect up with reality in such a wild way.

These calculations were done in the seventeenth century, and people got very excited about them. People like Newton calculated things like the distance of the moon and the amount of flattening of the earth at the poles before there were accurate measurements of these things. It looked like black magic to the ordinary folk, and maybe it is. In some places, Mathematics is known as 'Greek Magic', and the more you think about it, the more understandable the name becomes.

Exercise 1.2.2 *Get a garden hose and a friend who doesn't mind getting wet. Make yourself a large protractor out of cardboard. Lie down in the long grass with the hose, choose an angle and switch on the hose. Get your friend to measure the distance the water goes before it hits the ground. Repeat for some different angles.*

Can you confirm the above model? If the acceleration due to gravity is 9.81 *metres per second per second, what is your estimate for the speed with which the water comes out of the hosepipe? How would you confirm this conclusion directly?*

An enormously satisfying afternoon can be had doing this exercise, although it needs to be done in the Summer or you'll catch a nasty cold. If at all possible you should try it. Reading about things and actually doing them are very different, and there are lots of activities which sound incredibly silly before you have tried them but which are actually a lot of fun. Sex, flying kites and riding bicycles are well known examples, and this is another.

I have discussed cannon balls in some detail (although it is obvious that much more could be done). I didn't do this because I am the proud owner of a cannon and a shed full of gunpowder, nor because I think you might be in that position. I did it so that you could see something of the job of building models. You see, there are so many zillions of things to model, and different reader's interests are so diverse, I can't possibly undertake to choose to model something you all care about. It is better to look at something none of you are interested in professionally but all of you can understand when starting. Anyway, it doesn't so much matter which system you think about, what matters is that you learn how to think about systems.

1.3 Bacteria and People

Bacteria matter rather a lot, since some of them live on, or in, us. Of about fifteen hundred known species of bacteria, some two hundred can cause you to be sick. Only a century ago, surviving to age five was a chancy business, and it might become so again. It still is in many parts of the world. So understanding the things is of some pressing importance. I shall look at the job of understanding something about how many of them there are.

Models for understanding the growth of bacteria are also useful in understanding the growth of other things, such as your bank balance or small companies. So please don't switch your brain off just because you are not a biologist. We are working at a high level of abstraction here, and consequently talking about lots of things at once.

To cut down on the complexity, let's measure only one thing, the number of bacteria. If we measure this number at different times, then there is a *function* defined which sends each time t to the number of bacteria, say b(t), at that time. Now we only get a *sample* of this function; presumably we could have measured in between the times we actually did, and we would have got some generally intermediate value. Because we are interested in the bacteria and not just the numbers we actually got, we are mostly concerned with this function $b(t)$, even though we only know some finite set of values of $t, b(t)$. We could name this set

$$\{(t_i, b_i) : 1 \leq i \leq N\}$$

which means that we have N pairs of numbers, the i^{th} pair being the ith time and the measurement b_i of the number of bacteria taken at that time. This collection of observations is our *data*.

It is not generally practical to count bacteria in a colony, the whole colony looks to the naked eye like a greyish spot in many cases. (You will see them on fruit which has been carefully stored behind the furniture for a few months if you live in the same style as I do). What we actually do is to measure the area covered by the mould. This gives us a decimal number (real) to some finite precision. Strictly speaking, the values of b will be integers, whole numbers. But the actual areas will be approximations to real numbers. Should this worry us? If we accept that a model won't be true but ought to be useful, no it shouldn't. The actual number of bacteria will usually be enormous, and we aren't going to get anywhere even close to a perfectly accurate value. So a sloppy, fuzzy, approximation will almost certainly do.

The frequency of sampling might be over periods of the order of minutes, an hour or even a day. We usually try hard to keep the samples evenly spaced in time so as to make predicting and modelling easier.

Now what sort of model can we come up with? It should lead to a function, $b(t)$ which can be fitted to the data. As with the cannon model, we can start off by making simplifying assumptions. The more you know about bacteria, the better placed you are to decide what simplifications you might hope to get away with. Some of mine will be silly. This is because I don't know as much as I would like to about bacteria; you can't know everything and there's lots more things I don't know about.

I shall assume that a bacterium is like an amoeba, it guzzles food and then divides into two amoebae which go and guzzle and grow bigger until they divide in turn. And so on. No doubt it looks like a full and interesting life to an amoeba. I shall assume that this is exactly what bacteria do. I have seen images of them doing it, so I can justify this assumption fairly well. Actual bacteria species vary widely; some bacteria can replicate themselves in about five minutes, others may take up to an hour or more.

I shall also assume that a bacterium needs to have absorbed a certain amount of food before it splits into two. And I shall also assume that they live for ever or until they meet some penicillin, or until they get eaten by something else.

I am now going to start modelling. This is good innocent fun, because I can make any old assumptions and see what happens. Let us start by assuming that a single bacterium in the presence of plenty of sugar or beef soup replicates itself in ten minutes. Suppose at time $t = 0$ I have one of them. At time $t = 1$ measured in units of ten minutes, I have two, after twenty minutes or two time units I have four.

It is easy to write down the rules (I made them easy by (a) choosing a simple life style for my model bacterium and (b) choosing a time unit which saves on arithmetic).

At time $t = n$ I shall have twice as many as I had at time $t = (n-1)$. This works for $n \geq 1$. It is easy to see that at time n I have 2^n of the little dears.

If I wait for a week, how many will this be? A week is 7 days, there are twenty four hours in a day, and six time units in an hour. So the time is now 1008 So the number of bacteria is

$$2^{1008}$$

What is this in ordinary numbers? $2^{10} = 1024 \sim 10^3$ so

$$2^{1008} = (2^8)(2^{10})^{100} > 256 \times (10^3)^{100} > 10^{302}$$

which is even more than Bill Gates has cents. In fact if Bill Gates has ten billion dollars, this is only 10^{12} cents so it is an awful lot more.

Bacteria come in several shapes and sizes. A bacterium may be roughly spherical or rod shaped or spiral, and in size could have a diameter of around 1-10 microns. One micron is 10^{-6} metres, If you had 10^{10} of the bigger ones, the colony would cover an area of about one square metre. If you had 10^{302} of them the colony would cover an area of more than 10^{30} square metres. There are a thousand metres to the kilometre, so there are 10^6 square metres in one square kilometre. So the weeks worth of bacteria would cover 10^{24} square kilometres. Australia is about $5,000$ kilometres wide and about $3,000$ kilometres the other way, so has an area of about 15×10^6 square kilometres. (These are, of course, very rough estimates indeed, but I am having fun, so don't stop me now.) So Australia would be covered by bacteria to a depth of about 10^8 kilometres, or about two hundred times the distance from here to the moon. Alternatively, the area of the planet is around 5×10^8 square kilometres, so the bacteria would cover the entire planet in a blanket of thickness about seven times the distance from here to the moon. Only the area to be covered would go up as you added extra layers, so it wouldn't be as much as seven times the distance.

Exercise 1.3.1 *How far* would *it be?*

And this is after only a week. You'd think that someone would have noticed by now.

What went wrong?

Exercise 1.3.2 *Check my arithmetic to make sure I didn't slip a few million decimal places.*

The model is producing daft answers, so there has to be something seriously wrong with it. You are probably thinking that the supply of sugar or beef soup would have given out at an early point in the proceedings, and of course you are right.

Instead of thinking about bacteria, let's think about people. If we take the best information from the anthropologists, the species Homo Sapiens is between two and three million years old. (We don't know it's birthday so we can't send ourselves birthday cards or sing happy birthday to us on any special day; do it now if you want.)

We know that human beings become capable of reproduction after about fifteen years (twenty one in Queensland) and usually have children before they are thirty. Let us settle on thirty years as one generation.

The number of offspring varies a lot, and is decreasing these days, but let us suppose, to keep the model simple, that they have, on average, three children in that time and never have any more. Let us also suppose that after they have had their children they just drop dead, so we don't have to count them. Remember, I am model building, so I can put whatever assumptions into the model that I feel like. Be brave and bold.

Our unit time period is now thirty years. If we started off with two people, one male and one female (and it could hardly have been less), two million years ago, how many people would be alive today?

The model has been described in words (because it is a very simple one) and the calculations can be done in your head almost. The number h_n of human beings at time n follows the rule

$$h_n = 1.5 h_{n-1}, \; n = 1, 2, 3, 4, \cdots \tag{1.9}$$

where one time unit of thirty years separates the different measurements, and $h_0 = 2$ is the population at time zero.

A little thought says that we can write

$$h_n = 2 \times (1.5)^n$$

for the *solution* to our *difference equation* 1.9.

Then doing some rough and ready approximation, $(1.5)^4 = 5.0625 \approx 5$ and $(1.5)^6 > 10$. So in 180 years the population would be more than ten times its starting population, which was two.

Call it two hundred years, and be pessimistic. Then in a thousand years the population would be multiplied by 10 five times, so would have grown to over 200,000 people. So there is a multiplicative factor of 10^5 per millennium. Which means that in two million years, the population of human beings would be greater than

$$10^{10000}$$

This is not the case. What went wrong?

Exercise 1.3.3 *This is clearly daft if only because a human being takes up a certain amount of space. Making plausible, rough estimates of how much space people take up when crowded together, how many people could you get on the surface of the earth if they all stood up? Look up the Guiness Book of Records to see how many people have been squashed into a telephone box and work out how many telephone boxes you would need for* 10^{10000} *people.*

Make an estimate of the mass of the earth by knowing it is roughly spherical with a radius of about 4000 miles and has an average density of about 5.25 times that of water. Make estimates of the mass of 10^{10000} *people. Would we need to turn the whole planet into human flesh to get those numbers? The whole solar system? Or just Ayers Rock? I strongly recommend playing with these numbers, you will find it much more entertaining than watching television. Strange but true.*

There are only two parameters in our model, the time it takes to have children and the number of kids you get. The other constant obtained from other sources was the two to three million years for the human race to do what comes naturally. So there are four possibilities: (1) The mean period to produce offspring is much greater than thirty years, (2) the mean number of offspring per couple is much closer to two than three (3) The model is just awful for reasons not immediately apparent or (4) The human race is much younger than two million years.

On the first possibility, it is hard to credit that for most of the two million years people waited until they were forty before getting at it, but we could easily modify the model to see what effect it has.

Exercise 1.3.4 *Do some back-of-the envelope calculations to see how late people could leave reproducing in order to get about* 10^{10} *people on earth by the present time. This is about half as much again as the present world population, estimates, but maybe we missed a few.*

War, Famine and Disease are likely to have an effect on the mean number of children per couple. Many couples die childless when the crops fail. Many children get killed by disease. And war kills males in large numbers. Can these factors be enough to explain the discrepancy between prediction and measurement?

Of course you might be a creationist with a firm conviction that the world is only 6000 years old, being created in 4004 BC or thenabouts. Would this give an explanation? I leave you to work out if this is a possible explanation.

There are several ways to look at this problem, the ordinary bloke's and the scientist's being rather different. The ordinary bloke just says to himself that there's bound to be some sort of explanation of this sort and goes back to watching something on Telly. The scientist has a bit more go in him and plays with the numbers to see what is going on.

What we are trying to do is to fiddle the parameters so as to get agreement with the data. If we can do this successfully, we may sit back with a sigh of relief and feel that we have a plausible model. If we can't, we have to think again.

It should be obvious that this is a game of a certain sort, and unlike most games it is a rather creative one. It is also cheap and hygienic. But it can have rather far reaching consequences, Like going back to the people who dig up old bones and asking 'Are you sure these bones are two million years old? And are you sure they are human ancestors?' because it is at least possible that you could come to the conclusion that this is what is wrong with the present calculation.

This isn't likely, because we ran into exactly the same problem with the bacteria: there were far too many of them to be explicable on the model. So on the basis of common sense reasoning, we are probably going to come round to the view that our model was too simplistic. We left something out of the calculation.

Exercise 1.3.5 *Go back and look at the parameter estimates I gave for the human population growth model and get a calculator and a couple of old envelopes. Can you fiddle the parameters so as to get anything like the present population? If the population of the Earth has in fact doubled in the last forty years, what conclusions can you draw about the parameters?*

It is worth pointing out that this problem is essentially the same as the compound interest problem: you put one dollar in the bank two million years ago, the rate of interest is such that you get a fifty percent increase after thirty years, how much money will you have next week?

You should be so lucky.

Let's explore the relationship between the two problems, compound interest and population. In both cases we have

1. A natural time period (the 'compounding period'), one year for most banks, 30 years for human replication and ten minutes for our bacteria.
2. A multiplier which relates the amount of whatever at time n to the amount at time $n-1$.
3. A starting value of the amount of whatever.

In the case of interest, we usually talk about something like 6% per annum, *annum* being Latin for a year. The bank calculates 6% of whatever you had at the end of time period $n-1$ and adds it on to the sum you had. This means you have 1.06 times the amount you had at time $n-1$ in your account at time n. We write the *recurrence relation*:

$$b(n) = 1.06\ b(n-1)$$

We have a starting amount $b(0)$, and we can thus calculate $b(1)$. Once we know that we can calculate $b(2)$, and so on. It is obviously nice to get a formula for $b(n)$ and this is easily seen to be

$$b(n) = (1.06)^n\ b(0)$$

This formula is said to *solve* the recurrence relation. The reason I drop this jargon in is because there are more complicated recurrence relations which we shall have to solve, when the solution is a bit less obvious. And there are some where the solution is absolutely ghastly and the only effective way of handling it is to do it on a computer.

Example 1.3.1 *What is the annual interest rate that leads to a fifty percent increase in thirty years?*

Solution

If the interest rate is written as a number (not as a percentage) and we call it α, then we know

$$(1+\alpha)^{30} = 1.5$$

so

$$1+\alpha = (1.5)^{\frac{1}{30}}$$

or

$$\alpha = (1.5)^{\frac{1}{30}} - 1 = 0.0136073 \approx 1.36073\%$$

At least that is what my calculator does with it to 7 digits. This is all schoolkid stuff, but the idea is at the root of what we shall be doing later. The recurrence relation is said to define a *difference equation.* I thought it would be useful to link what we are going to do with stuff that you are familiar with so as to get you seeing the basic ideas.

There is a serious problem with the compounding period. If you have a choice between 6% per annum, and 1.47% per quarter, which should you choose? If you are a bit of a dill, you figure that 1.47 is less than a quarter of 6, so you pick the latter. But the multiplier of the former over a year is

$$(1.0147)^4 = 1.0601093$$

which means you have just over 6% if you choose the quarterly rate of compounding.

Exercise 1.3.6 *Suppose the bank offered to compound monthly at* 0.04% *per month. Does this beat* 6% *per annum?*

What rate would you accept if they offered to compound daily?

This raises the question of what happens as the compounding period shrinks to zero. Banks aren't likely to offer this option, but mother nature comes close: the human population doesn't have 1.5 babies per couple every thirty years, they have them all the time.

Let us look at the human population problem again, using the same terms that the population experts do. They make it a *continuous* process, like a bank that compounds continuously, and they take into account the birth rate per head of population, the death rate likewise, and then subtract the latter from the former to get the net rate of increase. It is easier to say this in algebra:

Let $B(t)$ be the birth rate at time t, $D(t)$ the death rate at time t. Then the net rate of population increase is

$$B(t) - D(t)$$

at time t.

This is a *continuous* model, where the other models for growth have been *discrete* models. Discrete means 'comes in lumps', and refers to the time periods in this case.

Of course, you can argue that this is another loony idea, since it takes time to actually have a baby, and the population comes in lumps too. But let us approximate cheerfully and pretend that the human population of the earth is some function $P(t)$ which is a continuous function of time. In fact let's assume it is a differentiable function, in defiance of common sense, and write

$$P'(t) = B(t) - D(t) \tag{1.10}$$

where time is measured in years.

Why do this? because I have chosen $P(t)$ to be the name of the thing I care about, which is the number of bodies on the planet at time t. I called it P for population; I could have called it anything,the important thing is that this is what I can actually hope to *measure*. It is a function because it changes in time, and that is what functions were invented for.

Now having decided that $P(t)$ is what I actually care about, I find that the information I am given when they tell me the Birth rate and the Death rate doesn't give me $P(t)$. But it *does* tell me instead how fast it is changing. This is what derivatives were invented for, so I say it with differentials. And this is how I got equation 1.10.

This is a key point about this idea of modelling, and it is something students often screw up on. It is called 'setting up the equations', and getting confidence at doing it is important. The chances are good that one day you will have to do this in earnest in you area of study. Now is the time to get it very clear.

It is true that this is totally unrealistic in one sense, because the value of $P(t)$ is actually an integer. But if the value $P(1997)$ is around[1] 6×10^9, then we can argue that being out by half a human being is no big deal. Anyway, we simply haven't got that accurate a count of the people on the planet. And at this rate of growth, we would have to specify the time fairly precisely to get an answer.

Exercise 1.3.7 *Making crude estimates of the fraction of the population above childbearing age, and including what you know of human nature, about*

[1] This is about the present human population of the Earth.

how many children are born somewhere on earth in a second? An accurate answer is not practical, but is it less than one? More than a thousand? Try to get some rough estimate.

Example 1.3.2 *The excess of Births over Deaths this year is approximately* 1.7% *of the population which is around* 6×10^9*. How long at this rate for the population to double?*

Solution

We have from equation 1.10

$$P'(t) = 0.017P(t)$$

This differential equation has solution

$$P(t) = P_0 \; e^{0.017 \; t}$$

where P_0 *is the population now, which I put at time* $t = 0$.

So the answer is when

$$e^{0.017 \; t} = 2$$

or when

$$t = \frac{\ln 2}{0.017} \approx 40.77$$

In other words, it will take just over forty years.

This is pretty close to what has happened in the last forty years. But take heart, the birth rate is actually declining. On the other hand, so is the death rate. But not so much. Present estimates are that the doubling won't happen for nearly 70 years. You'll be pretty doddery by then and may not notice.

Let's return to the problem of why there aren't lots more people on the earth. Has the rate of population growth been constant for 2 million years? Since we know the birth rate is currently declining, we can speculate that it has been higher in the remote past, but then, so was the death rate. It's a delicate balancing problem. We know that the birth rate over all was bigger than the death rate, or we wouldn't be here.

Suppose it *was* constant; what is it? Or to put it another way, what was the net annual 'interest rate' over the period in order to give us our present population? This is the same problem as asking: if interest was compounded continuously and we started with two dollars in the bank and two million years later we have six billion dollars, what was the interest rate?

Suppose it was α, and we start with two people in year zero. We have

$$P(t) = 2\ e^{\alpha t}$$

as the solution to the continuous compounding problem. So

$$P(2 \times 10^6) = 2e^{\alpha \times 2 \times 10^6} = 6 \times 10^9$$

I make α about 10^{-5} or 0.001%. This makes it look like a very chancy business as to whether there were going to be *any* human beings at all. One thing which emerges immediately is that the rate of births over the last forty years is much higher than it has been for most of the last two million years if that period is indeed about the age of the human race.

Exercise 1.3.8 *A little thought suggests that the estimate for α is a bit shaky. If there had been one hundred people instead of two, all those years ago, then an even slower rate of growth would be required. If it was only one million instead of two million years ago, then we would have a larger value. What would it be in these cases? How about a creationist 6000 years instead?*

It is recommended that you find a quiet corner and do the sums on your calculator in order to get some rough idea of what can happen. You won't find Mathematics fun until you do some on your own that you don't actually *have* to do. Then you will find it a lot more entertaining than most of the other things you can do on your own.

This is not exactly of pressing practical interest in general, although it is rather enjoyable. The trouble is that any extrapolation over time scales of the order of two million years is going to leave us rather out of touch with reality. Our model lacks a certain believability, and we would do well to remember that it is only a model. At the same time, when a model starts giving daft answers, it is a puzzle to find out why, and often rather rewarding.

The model we have just looked at, warts and all, is a very simple example of a *differential equation*. I pulled the solution to 1.10,

$$P(t) = P_0 \; e^{0.017 \; t}$$

out of thin air, but it is easy to check that it is right by differentiating back again. Anyway, you should have done this at school. If not, sue your educators.

The question as to what happened to the bacteria that aren't swamping the planet remains unresolved, and I shall come back to it with a better and more sensible model in the next chapter.

1.4 How to Do It Yourself

I have given you examples of models of two sorts, one sort using difference equations and one sort using differential equations. The cannon ball and the last example were differential equations, and the bacteria and the first human population model were difference equations.

You will probably have seen the last example in an earlier part of your life, or something close to it. I give it again because I want to point out the relation between difference and differential equations. A differential equation is a sort of 'continuous compounding' and they are usually easier to handle than the difference equations. I also want to set the stage for the serious work which will begin in the next chapter.

Of greatest importance is the business of setting up a model to describe something in the real world. It is something students often find very difficult. Students can often tackle problems which have been set up in algebra- such as solve $y'(t) = 3ye^{-t}$ - without knowing where to start when faced with a real phenomenon they actually care about. Looking back over the simple examples I have discussed you will notice the following steps were taken:

1. Work out what thing it is that is being *measured* by some number.

There may be more than one. The things measured may be distances, numbers of people or dollars, temperatures, weights, fraction of a population voting for a particular politician, shoe size, price of fish (or anything else), average number of goals per match, density of a solution, amount of water coming down the pipes when you flush your loo, or a zillion other possibilities. But work out what the measurable quantities are that describe the system.

2. If this is a number that will usually change in time, then there is a function of time which is the value. Give it a name like $b(t)$ or $x(t)$. If there are several things that change in time, give them all names. Choose names that are easy to keep distinct! You don't want to get muddled by some silly coincidence.

3. Occasionally you will have something other than time, but in most cases it will be the time that is important.

4. Make up your mind if the time is continuous or discrete. It may be entirely up to you which you pick.

5. Now you have to take the information that you have about how the system you are looking at changes in time. This is usually in the form of telling you something about the derivative of the function in continuous time, or how it depends on the last time measurement for discrete time. It can depend on several times in the past, perhaps the last one and the one before that, for example. And similarly, the information about the rate of change of the thing you are measuring may involve second derivatives. Don't panic, your job is to translate the information into algebra at this point. How to solve the equation is a separate issue.

6. Check to see if there is anything you know about the system that has not been translated into the algebra. For instance, you often know that the variables are always positive.

   Go backwards and forwards from the algebra to the problem a few times in your head to make sure that everything that is in your common-sense English description is properly stated.

7. Now, and only now, think about solving the equation. First identify the type as one you have met before if possible. It is worth looking at

the difference or differential equations in the text book to see if yours looks a bit like one of theirs. Look up the answer. Follow a text book solution point for point.

8. If your equation is horribly complicated, ask yourself if it could be simplified by putting in some extra assumptions that are unlikely to have a big effect on the answer. You can have not just one model, but a collection of them, some more simplified than others. Simplify down to the bone; you can always put the complications in later.

9. Above all, don't rush things. You get faster with practice anyway; the important thing at this point is to try to make sure that your equations are saying what you want them to say. Anyway, you can always use Mathematica or MATLAB for actually getting a solution. This isn't the interesting bit.

10. If you think you have a solution, *Check it* by differentiating or plugging a few numbers in to see if it is right in at least some cases.

Example 1.4.1 *A mothball evaporates at a rate which is proportional to the surface area. One of them starts off with a diameter of 1 cm and has halved its size after a week. How long until it vanishes?*

Solution

First we have to decide what we care about in respect of mothballs. We could choose the diameter, the radius, the volume or the surface area. We could also choose the weight, the price or the number in the packet. Since the question asks about the thing vanishing, we can probably rule out the last two. It might be better to choose one of the first three. Which one? It doesn't matter a lot, but I think I shall choose the diameter because that's the information that is given. I reserve the right to change to something else if it looks like a good idea.

The diameter changes in time so we have a function which takes in the time and outputs the diameter of a particular mothball at that time. I need a name for this function, and I shall call it $d(t)$ *for diameter at time* t.

The next thing I need to do is to work out what the information given tells *me about $d(t)$ or its derivatives. It says that the rate of evaporation is proportional to the area. What on earth is the rate of evaporation? It is probably the rate at which matter is lost. Which is the rate at which the weight decreases. Oh. Maybe I should have picked weight after all. But the weight and the diameter are related: the weight is the density times the volume, and the volume of a mothball is $4\pi/3\ r^3$ when r is the radius, half the diameter, since mothballs are pretty much spherical.*

So I think I shall invent a name for the volume of the mothball and call it $V(t)$. Then I am told something about $V'(t)$. OK, so maybe $d(t)$ was not the best choice. I shall move to $V(t)$ if necessary. Anyway, when $d(t) = 0, V(t) = 0$, since $V(t)$ is proportional to $(d(t))^3$.

I am told that the value of $V'(t)$ is proportional to the surface area. The formula for the surface area of a ball I would have to work out since I have a terrible memory, but there's not a lot of point in that, since it clearly depends on $(d(t))^2$. Translating this into algebra I get:

$$V'(t) = k(d(t))^2$$

where k is some constant and I don't know what.

I am making progress, but I need to relate V to d by

$$V(t) = \frac{4}{3}\pi(\frac{d(t)}{2})^3$$

If I differentiate this with respect to t (What else?) I get

$$V'(t) = \frac{4}{3}\pi\frac{3}{8}(d(t))^2\ d'(t)$$

Putting these two equations for d' together,

$$d' = k/(\pi/2)$$

or in other words, the derivative $d'(t)$ is some new constant, K.

This I shall write out because it encapsulates what I am told about the behaviour of mothballs, and I have set up the equation:

$$d'(t) = K$$

Now I go into equation solving mode: this is fortunately very easy:

$$d(t) = Kt + C$$

is the general solution.

Now I have the general form for the solution I am nearly finished. I have some parameters in my model, K *and* C*, and I need to nail them down. If my clock starts at* $t = 0$ *I can translate the statement that it starts off with diameter* 1 *cm by putting*

$$d(0) = 1 \Rightarrow C = 1$$

And the statement that it halves its size after a week is translated by

$$d(1) = 0.5 = K + 1 \Rightarrow K = -0.5$$

I have chosen to use weeks as my natural unit of time. I can choose any unit, and this makes the numbers easier.

To write down the final equation which becomes our model for the evaporating mothball:

$$d(t) = 1 - 0.5t$$

It is now blindingly obvious that it halves its original diameter every week, so by week 2 it has vanished into a powerful smell. Note that by week four it has a diameter of -1 *cm according to the model. Common sense tells us that the model stops working when* $d(t) = 0$.

Exercise 1.4.1 *Instead of doing it in terms of* $d(t)$*, do it with* $V(t)$*. Which is easier?*

I have had doubts about choosing $d(t)$ as my function name, particularly when I intend to differentiate it. The expression

$$\frac{dd}{dt}$$

looks truly horrible. You can see why mathematicians use f; a, b and c are constants, d is for differentiating, you know that $e = 2.718\cdots$ and the next available letter is f.

You may or may not care about mothballs; personally I can take them or leave them. But it is missing the point to think that this question is about mothballs. It is about setting up models, a very different thing.

The problems that you will find in your text book are sometimes of the sort cooked up by school-teachers to train you into obedience and proper respect for their wisdom. I have never forgiven the author of a book who set a problem which started off: 'A certain radioactive substance decays at a rate proportional to the square of the amount of radioactive material present..' This is just a silly pretend problem that violates laws of Physics as well as common sense. The trouble with problems of this sort is that they try to get you to switch your brain off, while I am trying hard to get you to keep it switched on. So I shall stick to honest problems, although I shall simplify them a lot.

1.5 Summary and Conclusions

The world is full of a number of things, and many of them can be described as systems, the states of which are specified by making measurements, and the measurements usually change in time.

The question of 'how systems behave', which means how their states change in time, can be tackled in two complementary ways: one is to just keep a list of the results of measurements at different times. Such a thing is a sample of points from the graph of a function of time if we are only measuring one *variable*. If we measure lots of variables, then for each one there will be some function of time saying how it changes, and although we only get a sample of points on each graph, it is hard to deny that there *is* a function and a graph of it, because you could usually have made measurements in between or for ever.

The second thing we do is to try to construct a model of the system, where a model is the function of time which is what you expect to get (subject to measurement noise) at each time you make a measurement, together with an interpretation telling you what the numbers actually mean, i.e how to make the measurements.

Many such functions arrive from being given information about how the state tends to change in time; this can be given in the form of a difference equation or a differential equation. So we need to get from the difference or differential equation to the actual function, which is called *solving* the difference or differential equation. There are a few equations we know how to solve, and an infinite number we don't. If we do, we say we understand the system; otherwise we usually shut up about it but keep on slogging away before going to bed at nights.

Your job will generally be (a) to find in your work some identifiable system specified by measurements of various sorts, (b) to find out enough about how it behaves by making some measurements, i.e. acquire some data on the system, (c) to construct a model for the data based on knowledge of the system. This may or may not involve difference or differential equations (it could involve probabilistic modelling, for example), but the chances are good that it will, and (d) to check the model against the data. If there are discrepancies too small to bother about which can be written off as measurement noise, pat yourself on the back, if not try to figure out what went wrong by playing with the model to see where it goes bung. This is the life of a scientist, it is much of the life of an engineer (until he or she gets turned into a manager) or an applied mathematician (until he or she gets turned into a bureaucrat), and it is an awful lot of fun. It has also changed the world in the last three or four hundred years, mostly for the better by my standards.

Chapter 2

Growth

2.1 Bacteria, People, Money

2.1.1 The Logistic Equation Revisited

In the case of the bacteria growth and the population growth, the models lost touch with reality at some point and gave us daft answers. Let's look harder at the assumptions that went into the models and see if we can get something more sensible.

It is important that you appreciate that this is not done because I have an obsession with bacteria, but because it is worth sorting out the nature of the modelling process on something easy to understand.

If you put some bacteria into a petri dish and feed them sugar solution so they can guzzle and reproduce, the sugar will eventually all get used up and the bacteria will die of starvation. But this isn't a matter of exponential growth ($P(t) = P_0 e^{\alpha t}$) up until the sugar vanishes and then BAM! yer dead. Before that happens, the bacteria will start fighting over the last drops of sugar. Well, they would if bacteria could fight. Anyway, some will get at the sugar and some will not, and those that do will perhaps take longer to reproduce because they aren't so well fed.

This reasoning about how bacteria function in petri dishes leads us to decide our model was too simple minded. Well, of course it was, the numbers proved that, the question is what do we do about it, and the answer is that we build a better model. Let us approach this in a mood of optimism and experiment, like small children playing in a muddy puddle. While I discuss this modelling problem, keep out a careful watch as to how I do it, as well as following what I do, because you are going to have to do it on your own someday.

Suppose the amount of sugar is constant, and a bacterium is effectively a machine for converting sugar into new bacteria. It would seem likely that a bacterium would need something rather more than sugar to do this, since bacteria are not just sugar lumps. In practice they tend to prefer a sort of beef soup in many cases, but the principle is the same, and I shall just refer to it as sugar. It sounds classier than 'Lolly Water' if not quite so classy as Agar. So for our second model, we simply note that if we can estimate the amount of sugar it takes to manufacture a bacterium, we can estimate the population limit by dividing this into the total amount of sugar. Of course, it also takes a certain amount of sugar to keep a bacterium going, so this limit will not be reached, and once all the sugar is gone, the bacteria will kark it shortly afterwards. And what holds for bacteria holds for human beings.

This gives an upper bound on the number of bacteria, but it doesn't tell us how the numbers change in time, and it would be as well to find out. It might and indeed does have a bearing on the human population.

First, suppose I assume that I start out with some amount of sugar, S, and that bacteria eat the sugar. Then I can reasonably guess that there is a sugar function, $s(t)$ giving the amount of sugar at time t. I can also suppose that there is a rate of exchange between sugar and bacteria, the less sugar, the more bacteria. This means I am thinking of a bacterium as a factory for turning sugar into bacteria, and there is some constant amount of sugar it takes to make one bacterium.

In which case, I (cautiously and without conviction that I am arriving at TRUTH, but nevertheless enjoyably) write down the equation:

$$s(t) = S - Ab(t) \tag{2.1}$$

where $b(t)$ is the number of bacteria. S is the starting amount of sugar, and

A is the rate of exchange between sugar and bacteria. Both A and S are positive numbers, constant in this model.

Follow this closely, because I am having fun, and I would like you to as well.

The next assumption I shall make is that the rate of growth of bacteria is proportional now to the product of the number of bacteria and the amount of sugar left. If there is no sugar, then no bacteria will be produced. On the other hand the number of baby bacteria will be proportional to the number of parents if there is enough for all. While if there is some intermediate amount of sugar, this will slow the birth rate.

This assumption is a bit silly, since if I had only one bacterium and a humungous quantity of sugar, the model would spit out new bacteria at a million a New York Minute; it seems reasonable to assume there is an upper limit to this and adding more sugar won't increase the production rate in general. But if we are operating below this maximum, it sounds more reasonable. Halve the amount of sugar and it might mean that it takes a bacterium longer to get enough to manufacture another bacterium.

If you don't like these assumptions, then make up your own. I chose them mainly for simplicity so as to get started. I reserve the right to mess around with them later.

I can write my simple minded assumption in the form:

$$b'(t) = Bb(t)s(t) \tag{2.2}$$

where B is an unknown but positive constant, a third parameter of the model.

Putting these together I get

$$b' = Bb(S - Ab)$$

Now I have that there are really only two parameters, (BA) and S/A. Amalgamating these and changing the names, we get

$$b' = kb(a - b) \tag{2.3}$$

where k and a are constants of the model, or *model parameters* if you prefer the classy name.

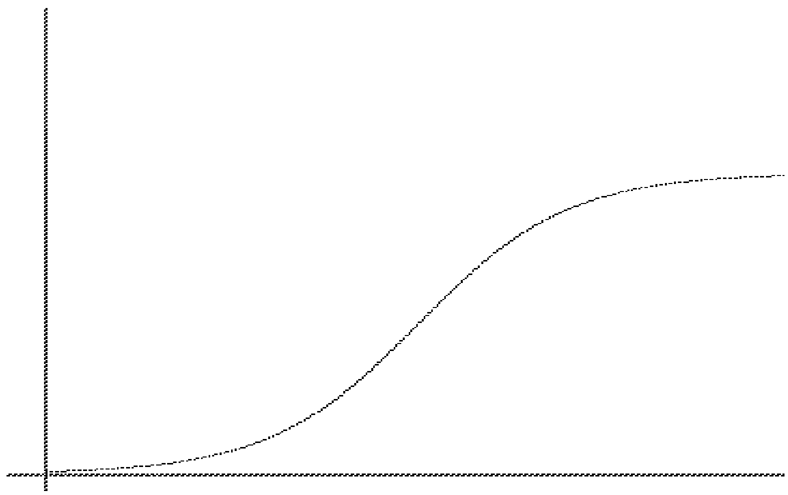


Figure 2.1: A solution to the Logistic Equation

This is a major step on the road to setting up the model. In fact, in many cases, people will say that this *is* the model. All we have to do is to solve the equation 2.3, which comes up in connection with a lot of other models, and is called the *Logistic Equation*. In some exercises, you may be asked to produce an argument to justify using the Logistic Equation for modelling some system.

In figure 2.1, I have used a simple ODE solver using Euler's method for generating a solution.

The actual choice of constants I made was $k = 0.5, a = 1.5$, the starting point is $t = 0, b(0) = 0.01$ and the solution goes up towards a horizontal line of height 1.5 unit above the t axis. I have scrunched up the t-axis so that it is five times as compressed as the b axis.

You can see that it must if it starts off below the axis; think about what happens if you get to $b = 1$ with the equation:

$$b' = b(1 - b)$$

Then $b' = 0$ so you don't go up any higher (or down any lower). In other

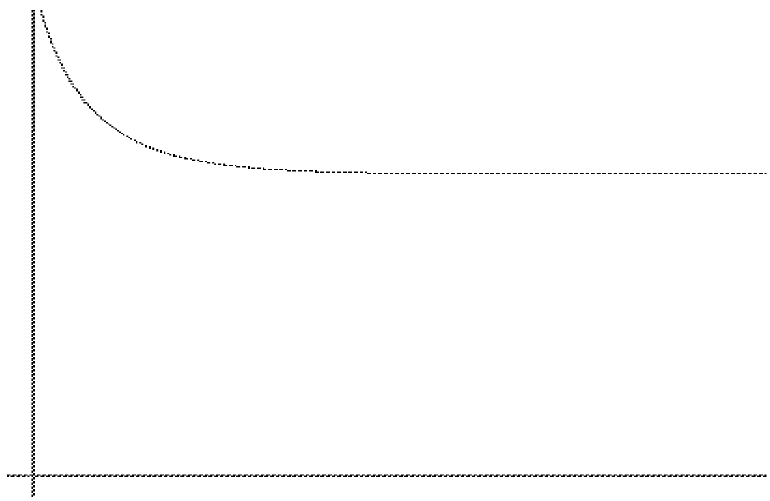


Figure 2.2: Another solution to the Logistic Equation

words, you stay at $b = 1$.

In figure 2.2 we have a solution which starts at $b = 2.5$ at $t = 0$. This decays down towards the horizontal line of height $b = 1.5$.

In general, we can reason our way to seeing what the logistic equation must do. I am now going to argue about the general shape of the solutions; follow this like a hawk, it is a powerful kind of reasoning which has developed into a whole theory.

If we have a and k positive, and if we start off with $0 < b < a$ at time $t = 0$, then we have b' positive, so b increases. If we start off with $b = a$ then we have $b' = 0$ so we stay at the same value. And if $b > a$, then b' is negative, so b decreases, exactly as in the figures.

If a is big and b is small and positive, then $a - b$ is pretty close to a. In this case, we have approximately the equation

$$b'(t) = kab(t)$$

which is just the exponential increase function with solution $b(t) = b_0 e^{kat}$, which will grow like crazy. So the function $b(t)$ starts off looking like exponential growth if it starts off small enough. When b is very close to a and a bit less, the slope, b', will be decreasing (although still positive). If b actually got to a at any time, then b' would be zero, so it would stay at a for ever.

If we rewrite equation 2.3 in the form

$$(a-b)' = -kb(a-b)$$

and argue that this is pretty close to

$$x' = -kax$$

where I have put $x = a - b$ and replaced b by a since they are close, then I get a solution

$$x = Ae^{-kat}$$

or

$$b = a - Ae^{-kat}$$

which is exponential decay towards a from below. This is a pretty sneaky trick and you would be forgiven if you didn't think of it. Still, it gives us some idea of what is happening.

Exercise 2.1.1 *What happens if we start off with b negative? This doesn't make sense for b a number of bacteria or people, but it does for some other situations. What happens when the signs of the constants k, a are changed?*

Exercise 2.1.2 *What happens as k is changed? What happens if it changes sign? What about a?*

You could reasonably ask if you could have the functions shown expressed as some mixture of elementary functions, you can find it on page 360 of Goldstein, Lay and Schneider, [4].

The logistic equation is called a *non-linear, first order, autonomous* equation. It is a bit hard to explain why some of this terminology is used unless you have done some Linear Algebra, which is a bit of a handicap. The 'autonomous' part means that the (time) derivative depends on the value of the function but not on the time. The first order means that there is only one differentiation involved. The linear in this case means that a first order autonomous equation must look like

$$y' = ky + d$$

for some constants k and d. Since it has a power of y other than one, the logistic equation is non-linear. It has proved an excellent fit to real data: In 1840 Verhulst predicted the population of the United States a century ahead. When 1940 turned up, he was within 1% of the actual result. That would be considered a good result in predicting the position of a planet, so models can have their uses.

Mind you, the book doesn't say what other predictions he made; and how well he scored on them. See [4], p360.

2.1.2 Death and Taxes

Let us push the model for bacteria the extra mile. Again, I want to explore the business of setting up equations which describe the world (or a bit of it). I am not totally obsessed with bacteria, although I am a little curious about what went wrong with the last model; mainly though I am interested in the process of model building, so keep an eye on how I go about it. Slow, steady careful thought is the key.

Let's think about one aspect of the model for bacteria which I found rather embarrassing, namely that when all the sugar was used up, the bacteria stopped reproducing, but they didn't die. This is not entirely potty, some bacteria can exist for long periods without food, but it will certainly go wrong for people. For a fixed amount of food, when it is all eaten, the people kark it.

How can we represent this? Well, we need to model a birth rate and a death rate separately, and take the difference to get the rate of growth of population.

I can reasonably claim that the birth rate for bacteria can be regarded as proportional to the product of the number of bacteria with the amount of food, provided that the amount of food is less than some threshold. Past that point, the bacteria are working at maximum capacity and extra food makes no difference. So I can write:

$$B(t) = k_2 b(t) s(t) \tag{2.4}$$

where B(t) is the number of new bacteria being produced, per time unit, per head of population (not that bacteria have heads, but you know what I mean) at time t. k_2 is a constant of proportionality and a parameter of the model.

The death rate, $D(t)$ is the number of bacteria falling off the perch per unit time, per bacterium, at time t. The question we ask ourselves is what will kill off a bacterium? Do they die of old age? If so, the death rate will be some function of t, presumably one that goes up as the bacterium ages, but this is a bit hard to specify. It is certainly true that people work this way. As it happens, bacteria don't; but they need some minimum amount of food to be present per individual in order to keep ticking over. If the amount of food is over the threshold, then there is zero death rate, and if the food supply drops below this threshold, then the bacterium runs out of steam gradually.

The amount of food per bacterium at time t is

$$s(t)/b(t)$$

We now put in our hypothesis about the death rate:

$$D(t) = 0 \ \text{ if } \ s(t)/b(t) \geq k_3, D(t) = 1 \ \text{ if } \ s(t)/b(t) < k_3$$

This is a bit clunky, so I shall make use of the sgn function; $sgn(x) = 1$ if $x \geq 0$, $sgn(x) = -1$ if $x < 0$. Then if I take $sgn(k_3 - s(t)/b(t)$ I get -1 if $s(t)/b(t) \geq k_3$ and 1 if $s(t)/b(t) < k_3$. But I don't want -1 and $+1$, so I add one and divide by 2 to get:

$$D(t) = 0.5(sgn(k_3 - s(t)/b(t)) + 1)$$

This is still a bit clunky, but better.

I shall suppose that there is a constant amount of sugar required per unit time in order to keep a bacterium alive, and call this k_4.

This gives the following equations for the system:

$$\begin{aligned} s'(t) &= -k_1 b'(t) - k_4 b(t) & (2.5)\\ B(t) &= k_2 b(t) s(t) & (2.6)\\ D(t) &= 0.5(sgn(k_3 - s(t)/b(t))) + 1 & (2.7)\\ b'(t) &= B(t) - D(t) & (2.8) \end{aligned}$$

with $s(0) = S$ and all constants (model parameters) positive. We now go through these carefully to see what they are saying, and to make sure that they are saying the right thing.

The first equation says that two things contribute to the decrease in the amount of sugar, the first being the production of new bacteria and the second being what is needed to keep the existing number of bacteria alive.

The second equation says that the birth rate is proportional to the number of bacteria there are ready to make more, multiplied by the amount of sugar available for making them. This is only realistic when the number of bacteria is past some threshold, but it at least allows for the fact that the birth rate will go down when the amount of food available does.

The third equation says that the death rate will be zero provided the amount of food per bacterium is past some threshold, and will be 100% if the amount of food per bacterium falls below the threshold. This is certainly a bit simple minded, and replacing it by a continuous function makes a lot of sense. But I shall worry about this point only if the answers are surprising.

And the last equation says that the rate of growth of the bacterial population is the difference between the birth rate and the death rate. It is hard to argue with this one.

Exercise 2.1.3 *Which of the above assumptions are doubtful for human populations? Can you write down more plausible equations for human beings?*

I now have two linked quantities, the food supply and the population of bacteria which both change in time in a way which is a bit more complicated. I have differential equations for each of them which involve the other. This sort of thing is rather to be expected, and needs some thought.

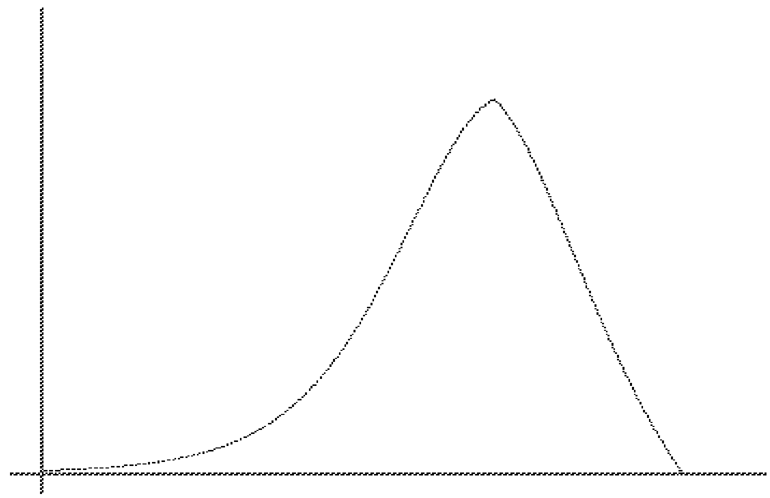


Figure 2.3: A colony with fixed food supply

You might feel that it is all quite hopeless, since you seem to need to know s before you can solve for b, and you also seem to need to know b before you can solve for s. But think of simulating it on a computer. You start off with some initial value of b and s. This makes B(t) and D(t) computable. This makes b' computable, which makes s' computable. So you take an (almost infinitesimal) teeny-weeny little change in b and s, an 'Euler step', and repeat. This will solve for both the variables at once.

In other words, although a closed form solution, i.e. some formula for b, is beyond us at present, it is not particularly hard to do numerically.

I am going to separate the problem of solving the equations from the problem of setting them up in the first place, and will worry about the second problem later, when I have convinced you it is worth finding out how to do it. So don't let the sight of the equations make you nervous.

A numerical solution showing the plot of b against time is shown in figure 2.3.

As one would expect, the colony grows until it starts to saturate, and then gets to the point where the amount of food per bacterium is inadequate, when it starts to die out. Collapse is then quite fast.

In order for you to play with models of this type, you will need to learn to use some of the programs available. You should netsurf until you find some *ode* solvers going free. If you are rich, MATLAB contains enough material to keep you occupied for decades and the student edition is good value for money. Mathematica is to my mind even better value for money, although it costs more.

Some people would regard this as just playing around, and others would regard it as good training in doing Science. Actually, it is both.

In the days before computers, a lot of time was spent on doing shortcuts by hand, and this is often worth doing so as to be able to predict the qualitative features of the output of the programs. Unfortunately, the modern folk who just copy them often fail to see that running it on a computer can be easier and more illuminating than doing a laborious analysis. Some of the old fashioned approaches to modelling still linger on in the books. Fortunately, you now have a modern book, this one, and get told how to do it properly!

It is typical that as the model gets more complicated, the equations get more tricky to work with and computer exploration is a good idea. On the other hand, one tiny bug in your coding can be totally misleading and give you horribly wrong answers. And you may not know the bug exists. For this reason alone, some sort of qualitative analysis is essential. In the present case, the output is pretty much what one would expect, but it is useful to play with the model parameters to see if the changes are what you would expect.

Exercise 2.1.4 *In the illustration, the value of S was set at 4. What would be the result of making it 2 instead?*

The value of k_4 was set at 0.01. What would happen if we increased it?

If instead of a hard threshold we had a continuous one, what effect would you expect it to have on the shape of the curve?

Try to work out the qualitative results of mucking about with the other parameters, i.e. with increasing or decreasing them.

The situation where we have not one relevant variable, the population, but *two*, the population and the food supply in this case, is very common. In our next case study, we look at the situation where the food supply is maintained at a constant rate. This might happen if the food were another species of organism which also reproduced, so it is a model of the case where we have a 'mini-ecology' to study, something we shall look at more closely later.

I turn then to the case of a human population. For most of the two to three million years of human existence, people lived just as the aborigines of Australia did until Captain Cook turned up and started complicating their way of life. They were hunter-gatherers, who wandered around foraging for food. The food reproduced itself, but relatively slowly, so the human beings (and we are talking about your ancestors here) had to keep marching, not returning to their starting point for a year or two.

If instead of the human beings marching around, the food had marched past them, then we would have pretty much the same situation as the bacteria except that we would need to have a constant *rate* of supply of food. It would keep on coming at some level, actually determined in our remote history by the rate at which human beings could travel to where the food was still plentiful. The rate would not have been *really* constant, but would vary from day to day according to whether dad was lucky in spearing a rabbit or catching a fish. Early human beings are thought to have been scavengers, living off the meat left behind by other animals killing for them. You would be, if this is true, a descendant of something with the general habits of a vulture or hyena. Some people still show traces of those habits, but none of my friends and few of my colleagues. Anyway, the food supply would vary a bit, but I shall assume it constant in my first model.

Food of this sort isn't much like sugar, it tries to run away when the hunter-gatherer goes looking for it in some cases, and in the case of the carrion-eater theory, the lion comes back and argues the matter on occasions, or the kids have to fight off the vultures who got there first. So there is some limit on the rate at which the food supply is maintained.

This is all common sense of course, and you will need to think about these things the way I have discussed them.

The next step is to translate these common-sense observations into algebra. This makes it quantitative and allows us to explore the ideas on a computer. This is what modelling is all about: applied common-sense made more precise.

We need to have the food supply, $s(t)$ renewable, which means that instead of a constant starting amount, we have a steady rate. So we write:

$$\begin{aligned} s'(t) &= k_5 - k_1 b'(t) - k_4 b(t) && (2.9) \\ B(t) &= k_2 b(t) s(t) && (2.10) \\ D(t) &= 0.5(sgn(k_3 - s(t)/b(t))) + 1 && (2.11) \\ b'(t) &= B(t) - D(t) && (2.12) \end{aligned}$$

These are the same as the last lot, 2.5 et seq., except for the fact that there is a k_5 term to give the constant rate. I shall suppose that $s(0) = S$, a constant still.

As I have said before, I am going to worry about the job of actually solving the equations later, for the present it is enough for you to see how these equations are obtained. This is the hard bit, because there are programs to do the solving. But programs with common-sense seem to run only on human brains, and not all of those.

The results of solving the equations on a computer may come as a bit of a surprise. The figure 2.4 shows a solution. The population oscillates, zooming up from a pathetically low level initially, hitting a maximum, then diving down. For some parameter values, the population dies out completely after hitting a huge maximum. The oscillations die down and the system hits a steady state after a while, and the steady state is basically in equilibrium with the food supply.

We can see this by making a change to the equations that says that after some time t, the food supply rate triples. This could have happened when our ancestors discovered agriculture, around twelve thousand years ago[1]. I

[1]This happened in several remote parts of the world including the Americas within a few thousand years. Whether this was done independently or whether the word got around is not clear. It wasn't invented in Australia, partly because the raw materials may

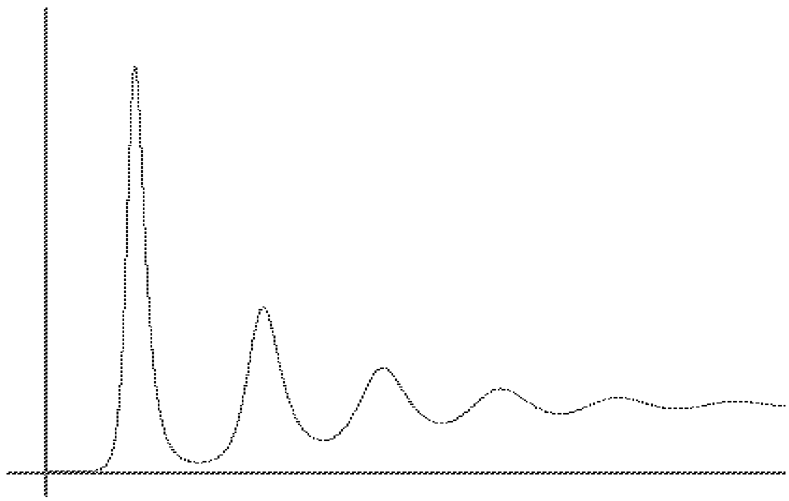


Figure 2.4: A population with a steady state food supply

show this in figure 2.5. The sudden jump in the food supply actually sent the population off the screen.

The model has a number of parameters to play with, and you should try to work out what will happen as you mess around with them. The model is not by any means complete yet, and much could be done to make it more realistic, but it has something to tell us already, namely that it is quite likely that over most of the two million years, the population was stable at some much lower level than presently, with fluctuations as the food rate changed. And that any substantial technological change which affected the food supply rate by increasing it would automatically cause a baby boom, soon followed by a death boom, eventually settling down into something relatively steady,

not have been here, and partly because life was quite comfortable without it. Although it is believed that the aborigines used fire to clear forests. This seems to suggest that the aborigines were just as keen on making life easy for themselves and having a good time as the rest of us. I should like to point out in passing that the belief that any one race is inherently superior to another is extremely silly, and there is no particular reason to suppose that aborigines are either less intelligent or more spiritual than the rest of us, assuming that these words have any precise meaning. Although I expect most of them are more spiritual than me, however you define it. I am notoriously deficient in this respect.

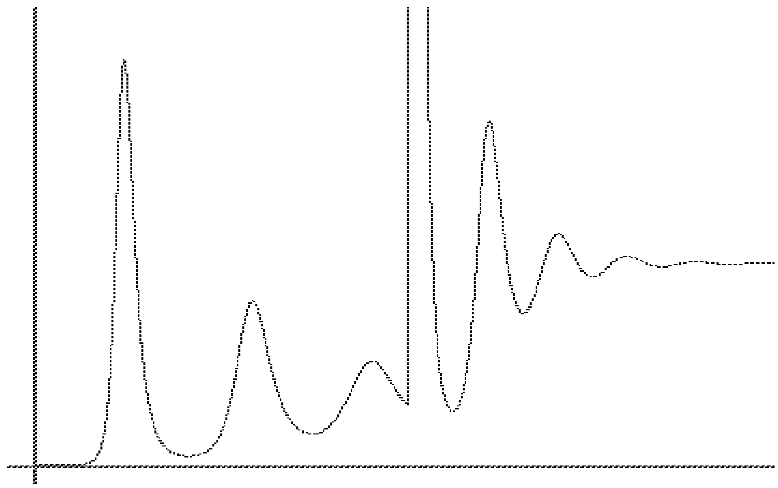


Figure 2.5: A population with a steady state food supply which suddenly invents agriculture

but with more of us than before.

My choice of parameters makes the effect rather striking, which was why I picked those values, but the qualitative effects are generally as described. For some values of the parameters, the wild oscillations do not occur. For example, figure 2.6 is the same model class but with different parameters, and what happens is that the population merely climbs to the new equilibrium value along a logistic curve and stays there (until another change in the food supply happens.)

Of course, the model is still very simple. It makes, for instance, no allowance for the species which eat *us*. The assumption that the food supply was of constant rate(or maybe two constant rates) is not very plausible. It would be quite reasonable to suppose that there is a random element in the food supply, and it would not be too hard to model it in a computer simulation. The resulting equations are called *stochastic differential equations.*

A factor which has been left out of all consideration is the matter of what is delicately called 'waste products'. For human beings this includes what goes

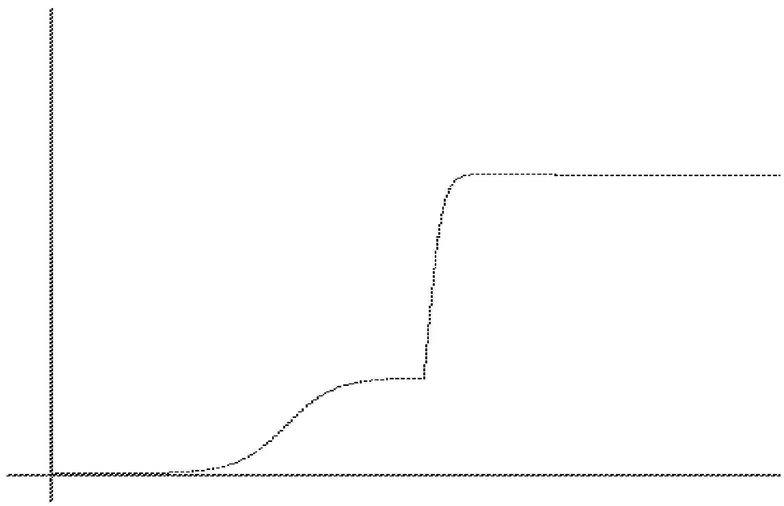


Figure 2.6: A smoother ride on the gravy train

down the loo when you flush the toilet and also empty tinnies and coke-cans in the garbage. For bacteria, it includes alcohol. Now you must realise that alcohol is made principally by fermentation, which is a polite way of saying that you find some bacteria which like eating sugar and which excrete alcohol, put them in a bottle with lots of sugar and wait for them to eat some of the sugar and drown in their own 'waste products'. It would be like putting *you* in a bottle, feeding you stubbies of beer, and waiting for *you* to drown in your own urine. This is how champagne is made, although I prefer bacterial waste products to yours. The bubbles are mostly carbon dioxide, also excreted by the bacteria which 'make' the alcohol. To think of bacteria as 'making' alcohol is exactly like thinking of you as a device for producing excrement. It is a point of view, I suppose, and one which would look perfectly reasonable to some bacteria and plants, were bacteria and plants to bother with points of view.

Now in a closed environment, the waste products can become important. And waste products are toxic to the organism producing them (alcohol is good for killing bacteria). The reasons are obvious: if the organism could use it for nutrient, it would have done so.

Exercise 2.1.5 *Write down some equations for a model of bacterial growth in a champagne bottle; you will need to have a variable $s(t)$ for the amount of sugar, another $b(t)$ for the number of bacteria, and a third $a(t)$ for the amount of alcohol which the bacteria produce as a result of eating the sugar. You need to think about how these quantities affect each other and then say it in algebra.*

Feel free to make simplifying assumptions; it is your *model and you can make it do whatever you like. Be creative but not silly.*

Fortunately, one species pee and poo is another species wine and cheese party. And so there is a reasonably closed loop as far as the material by-products of most species are concerned. On the other hand, there is no bacterium with a taste for PVC (it isn't 'biodegradable') or old car tyres. And there is some sign of population pressure on the environment. Nor have I mentioned the greenhouse effect, global warming and so forth. Some of the waste products of humanity are heat, carbon dioxide and molecules which eat ozone.

So a more comprehensive model for population growth will have to take into account almost all the things that are making modern life complicated.

Sometimes well-meaning people tell us earnestly that we ought to try to live in harmony with nature. They may have some dim intuition about the parameters of models such as those we have been looking at, because too high a consumption level produces the oscillations which can be rather extreme. Or possibly they love trees and want trees to love them back. But they have the wrong idea altogether. The human species is just a part of nature, subject to natural laws the same as trees and comets. Living in accord with nature is not something we have any choice about. The real issue is, are we going to find out what it takes to survive or are we going to kill ourselves off first out of stupidity and ignorance?

We come then to an interesting conclusion about the human population of the planet: it may well have been essentially in equilibrium with the environment for most of the time. But every technological advance that either reduced the death rate (mostly the rate at which other organisms ate us, from sabre-toothed tigers to bacteria) or increased the birth rate (via the food supply rate), would disturb the equilibrium and increase the growth rate, leading in general to oscillations which can be violent or very small or non-existent, but which eventually settle into a new equilibrium with a higher population. It is easy to see that this has happened in history, and must have happened in prehistory too. Are oscillations in population size observed in nature? The answer is yes, indeed they are, and I shall investigate them later in the book. So the model is starting to look a bit closer to reality.

The bad news is that we have been on an explosive exponential growth rate for some time now. The model suggests that this is relatively rare, and can be followed by huge death rates. A more detailed study would try to estimate the model parameters to decide if catastrophic oscillations are going to occur.

We come to this conclusion by making simple theories about the nature of growth of a population of animals, formalising them in algebra, and computing the consequences of the assumptions. This process leads to the evolution of better theories, as we have seen. In some cases it leads to theories which can be used for predicting complex behaviour qualitatively. We are nowhere near the limits of what can be done by these means, and there may be new models with interesting properties waiting to be found by you.

2.1.3 Money

Money is very interesting stuff and behaves in unexpected ways. Your interest in it is probably restricted to making sure you have enough of it, but one way of doing so is to have a clearer idea of how it behaves. So everyone should take an interest in this section.

First of all, it grows. More accurately, if you put some in a bank account and go back later, there may be more of it. Moreover the rule is simple and easily calculated. Whereas I used differential equations for population (despite the fact that the population is actually integer valued) I shall have to use difference equations for money, because banks don't compound continuously.

Just as it grows, it also dies; an individual dollar bill doesn't actually kick the bucket, but it may disappear from your account to buy something. And it may be born; or at least it may appear in your bank account in exchange for work. So as far as the money in an account is concerned, we have a situation similar in some respects to the situation for bacteria or people. The main practical difficulty is one of time scale of the changes. Interest rates vary, and so do expenses. By simplifying the model we can get some sort of answers to questions, but the issue of whether this actually means anything comes up. Some rather impractical folk write books containing questions like the next example:

Example 2.1.1 *You graduate in the year* 2010 *and get a job paying* \$1,000 *per week, the money being paid into your account. You have* \$500 *per week expenses and the bank pays interest at 10% per annum. It also makes a charge for running the account of* \$100 *per quarter. The interest is calculated on the average amount of the balance of the account measured weekly. How much will be in the account at the end of five years if you start off with nothing?*

Solution 1 *We set up the equations: first we look at the money coming in, the income is* \$52,000 *per year. This is the equivalent of a food rate. It follows that at the end of the year there is a total payment in of* \$52,000 *to the account.*

Next we look at the money going out; this is half the income which is \$26,000 *plus the charges of* \$400.

And finally we know that the 'birth rate' is 10% *of the average balance: It seems likely that the bank will make the charges whenever the balance is a maximum, so as to keep down the average and hence reduce the interest it must pay, but I shall be charitable and assume it comes out at the end of the year in a lump. Then at the beginning of the next year, the interest sum is added to your account.*

let $C(n)$ *be the balance at the beginning of year* n*, so that* $C(1) = 0$ *tells us that you start off year 1 stony broke. This is quite traditional for students graduating. Then the balance at the beginning of year* $n+1$*,* $C(n+1)$ *is the sum of the balance* $C(n)$*, the income, less the expenditure, together with the interest.*

Saying it in algebra:

$$C(n+1) = C(n) + 52000 - 26400 + I(n) \tag{2.13}$$

Where $I(n)$ is the interest for year n which is ten percent of the average balance. Estimating this crudely as the average of the sum at the end of the year with that at the beginning, we could say

$$I(n) = (0.1)(1/2)\ (C(n) + 25600 + C(n)) \tag{2.14}$$

This simplifies to

$$I(n) = \frac{1}{10}(C(n) + 12800) \tag{2.15}$$

Putting the two equations together and simplifying we get

$$C(n+1) = 1.1C(n) + 26880 \tag{2.16}$$

This is a first order constant coefficient linear difference equation. *More modern books will use the correct term* affine *instead of* linear.

It is instructive to mess around with it and try to solve it yourself before being shown. (Close the book right now and give it a go!)

The general form or shape of a first order constant coefficient linear difference equation is

$$C(n+1) = mC(n) + c$$

It is easy to write out the first few terms:

$$\begin{aligned} C(1) &= C(1) \\ C(2) &= mC(1) + c \\ C(3) &= m^2C(1) + mc + c \\ C(4) &= m^3C(1) + m^2c + mc + c \\ &\vdots \end{aligned}$$

The general expression looks as though it ought to be:

$$C(n+1) = m^nC(1) + c(1 + m + m^2 + \cdots + m^{n-1})$$

or

$$C(n+1) = m^n C(1) + c(\frac{m^n - 1}{m - 1})$$

by elementary results from geometric series.

Putting n=2 we get that $C(3) = m^2 C(1) + c\frac{m^2-1}{m-1}$ *which is right.*

So after five years and the first day of the sixth we have $n = 5$ $m = 1.1$ *and* $c = 26880$ *so you have in your account the magnificent sum of*

$$\frac{(1.1)^5 - 1}{0.1}\ (26880)$$

dollars.

This works out at $\$164,105.09$ *on my calculator. Nice potatoes. With no interest at all, you would have got* $\$130,000$*. So the kindly bank has paid you over* $\$30,000$ *for the right to use your money for that time.*

Solution 2

The first solution, as anybody who has tried it knows is totally unrealistic. It assumed you saved half your income of $\$52,000$ *per annum for five years.*

But in the first year you met somebody you fancied and you took her/him out to restaurants and concerts or pubs and mosh-pits. In the second year you got married and in the third and fourth years you had children. You also had to buy a car to impress the person you fell for in year one, and had to buy a house in year two and a larger house in year four. Consequently you wound up owing the bank about $\$200,000$ *at the end of year five.*

The kindly bank is doing very nicely out of you, since you *are paying interest on the loan at a lot more than* 10%*.*

As said, the problem is highly artificial, and the second solution faces up to this. People simply do not behave in the way hypothesised. Well, most people don't. It is possible, and I have known one man who became a millionaire by doing essentially this until he found a better rate of return than banks provide by starting a business. He didn't buy a car, he saved bus fares by walking for as many stops as he could, and he avoided women until he could afford to buy them. But you should appreciate that the problem almost certainly isn't describing *you*.

Banks and Insurance Companies can afford to be more dispassionate when contemplating loans to each other, since there is no way a bank is going on a toot with a gorgeous blonde at two of the morning. Or a gorgeous hunk if it is a female bank. And when banks spawn offspring, they send them out to work from day one. So although there is a world where these calculations are done and work, it is all a little removed from the sort of thing you are used to.

A more depressingly realistic problem is the following:

Example 2.1.2 *You owe the bank* $\$250,000$ *and have a 25 year mortgage. The bank lends you the money at* 13% *per annum. The bank compounds fortnightly and your repayments are fortnightly. What are the repayments?*

Solution

Let R *be the repayment rate in dollars per fortnight. Let* $D(n)$ *be the outstanding debt at fortnight* n*, and count from zero so that at* $n = 0$*, the debt is* $\$250,000$*. Then we get that the fortnightly interest rate is* 0.5% *so the recurrence relation or difference equation is*

$$D(n+1) = D(n) + \frac{0.5}{100}D(n) - R = 1.005D(n) - R$$

We have the same equation essentially but with different constants, this time with $D(0) = 250000$*. The expression for* $D(n+1)$ *is 'solved' by putting*

$$D(n) = (1.005)^n D(0) - R(\frac{(1.005)^n - 1}{1.005 - 1})$$

where $D(0) = 250000$*. (This is slightly different from last time because I started at* $n = 1$ *and this time I start at* $n = 0$*.)*

Now the number of fortnights in 25 years is $25 \times 26 = 650$*, and at this time we want* $D(n) = 0$*. So we get*

$$(1.005)^{650} = R(\frac{(1.005)^{650} - 1}{1.005 - 1})$$

So

$$R = \frac{(1.005)^{650} \times 250000 \times 0.005}{(1.005^{650} - 1)}$$

and since $1.005^{650} \approx 25.582333$ *we find*

$$R \approx 1300.8497$$

But don't worry, inflation will mean you can easily afford to pay twice that rate in ten years.

Exercise 2.1.6 *Work out the general formula for the fortnightly repayment rate on a loan over N years of an amount of principal P for interest rate r% per fortnight.*

If the loan were interest free, the formula goes bung. What should it be?

These results are rather simple, and easily calculated with any modern calculator. They are not very exciting unless you are running a bank and multiplying the repayment rate by the number of home loans to calculate your take-home pay.

The Economy, Stupid

Economics must be a matter of concern to any thinking person, because we live in a complex system which we desperately need to understand better if we are to have any sort of control over it. Inflation, unemployment, bank rates affect us all. Leaving these matters to politicians and bureaucrats to implement what may be idealistic but impractical aims just ain't safe. (And leaving them to politicians with a taste for power and no ideals other than self-advancement could be even worse.)

Let us therefore move from this pico-economic scene to a broader look at a macro-economic issue:

We can talk of a *national income* meaning the total of all the separate incomes of people living in one country. It isn't clear that this means very much in reality, since we could double this by just giving ourselves a 100% pay rise, without being in any way better off. You would have twice the money but twice the expenses, since my expenditure is somebody else's income. And one must never forget the small island where everybody 'eked out a precarious living by taking in each other's washing'. It doesn't seem likely that we could all get rich by following this strategy.

When you get your money, there are basically two things you can do with it, spend it or save it. Saving money might mean putting it a bank, having superannuation, buying stocks and shares, or keeping it in an old sock under the mattress. It isn't at all clear that the two categories, saving and expenditure, are really distinct. If you have children with the intention of relying on them to look after you in your old age (still done in many parts of the world), is the amount spent on their education an expense or an investment? If you start a company to give investment advice, it will cost you plenty to do such things as advertising and printing your letterhead and renting office space. Does this come under investment or expenditure? Or if your investment strategy is to nick down to the casino or the horse races and put something on the red or the two o'clock, is this an investment or an expenditure?

The tax office might see it one way, you another. And is the tax you pay

- a payment for services provided by government,
- an investment in a well ordered society or
- just your contribution to the biggest extortion racket in town?

It isn't particularly easy to answer these questions, which casts some doubt on whether the categories really make sense.

Let us proceed with the assumption that it is possible to classify each dollar you lose as either investment/ savings or expenditure. I have the gravest doubts about the feasibility of this, but I shall follow the economists in pretending that it can be done.

Then we can add up all the investment money and all the expenditure money for everybody in the country and get numbers for national investment and national expenditure. This also neglects the fact that quite a lot of money goes overseas, and quite a lot of overseas money comes here, but we are used to simplifying assumptions. The question of whether the stability assumption, that the simplifications won't change the conclusions to any great extent, is justified, needs a lot more thought than it usually gets. I shall again follow economists and assume that this is also OK, and I really *do* think this is extremely implausible. Anything that leaves common-sense behind is suspect: it may be great stuff which goes beyond the obvious, or it may be the ramblings of a bunch of nutters off their collective heads. In order to decide which, we have to look at the extent to which the models are useful, and we should approach them without prejudice.

Having warned you that trusting models without a hard look at the assumptions, and an even harder look at the extent to which the model agrees with reality, is a sign of a sort of lunacy common among academics, I shall carry on. Kindly remember however that a model may be utterly lunatic and still be useful.

Let us suppose that there is some level of unemployment and that we can also measure this. Again, it is one thing to define the unemployment level as the number of people out of work, but how do you count a bloke on welfare who is a part-time student? Or another who paints pictures and sells them when he can find a buyer? Is he 'self-employed'? And if he is, aren't we all selling our labour to someone? Can't you regard a bloke on the dole as selling his time to society so that society will get a good feeling about looking after the poor? So again, there is some doubt as to whether the numbers are measuring anything that is 'really there' or whether they are just a convenience for economists to make a living by an elaborate and possibly unconscious confidence trick.

If we go along with the economists in treating all these things as measurable quantities describing an economy, the first thing we claim is that *the national income at any time is the sum of the expenditure and the investment.* Now this is more complicated than it looks, since if you get the sack you might have to live off your savings for a while, but if we allow that this is a negative investment we might get away with it.

We take the statement in italics at face value and say it in algebra for the usual reasons:

I define $Y(n)$ as the national income for a country over year n, $C(n)$ as the national consumption or expenditure for year n, and $I(n)$ as the investment for year n. Then we have:

$$Y(n) = C(n) + I(n) \tag{2.17}$$

We now assume that as the national income goes up, the consumption goes up proportionately, with some adjustment because some consumption is going to be needed to keep the people alive. So we write

$$C(n) = mY(n) + c \tag{2.18}$$

This assumption would mean that if nobody in Australia had any money next year, they would still spend c dollars on food and drink and housing, which makes the thoughtful person wonder how they do this. If they use up last years savings, then equation 2.17 is wrong. So equation 2.18 and equation 2.17 cannot both be right. Also, the assumption that the equation 2.18 is affine (linear plus a constant) is clearly not true for any individual. We would expect consumption to saturate at some point, since there must be some limit to the amount of food, drink, shelter, video-recorders, yachts and overseas holidays a human being can get through.

Since this is a classical economic model, (I didn't make it up) and since we judge models by their consequences and not by whether they make sense, I shall persist.

The constant m is called the *marginal propensity to consume*. It is a number between 0 and 1, on the hypotheses we have made so far. This is a translation of the observation that if we get more money, we spend some fraction of the increase on extra goodies. The claim that it is always the same fraction is one we regard as a simplification, but approximately true over some range.

Now we turn to investment. If a company is making widgets and it is selling lots of them at a price which leaves some money over, it makes sense for it to make more widgets. This will require it to expand its workforce and possibly to buy extra widget-making machinery and a factory to put them in. So an increase in investment leads to more jobs. If you stick your hard earned savings into a bank, the widget making company can borrow your money off the bank in order to pay the extra workers until the widget sales produce extra income. So your putting money in banks does something to reduce unemployment. Keeping it in a sock under the mattress encourages burglars to come looking for it, and if they find it, this encourages the burglarious profession which expands. All doing something to keep down unemployment, although not necessarily as measured by government statisticians.

Economic growth means a larger national income next year, so we write

$$Y(n+1) = Y(n) + r\ I(n) \tag{2.19}$$

where r is assumed to be a constant called the *growth factor*. We hope that $r > 0$.

Putting together the above equations we find

$$\begin{aligned} Y(n+1) &= Y(n) + r\ (Y(n) - C(n)) && (2.20) \\ &= Y(n) + r(Y(n) - (mY(n) + c)) && (2.21) \\ &= [1 + r(1-m)]Y(n) - rc && (2.22) \end{aligned}$$

This is yet another first order constant coefficient affine difference equation, which we can easily solve either by memorising the formula or by working it out from first principles. I leave you to write down the formula for $Y(n)$ in terms of $Y(0)$.

Exercise 2.1.7 *Get an explicit formula for $Y(n)$ in terms of $Y(0)$ and the model parameters.*

Since all the constants are positive, what we get is exponential growth of national income, so we will all be filthy rich one day. Well, collectively we will be, or our descendants will.

In western societies there has indeed been a large growth of national income, and we are hugely wealthy by comparison with, for example, mediaeval peasants. Whether this is due to steady investment or whether it is due to particular forms of investment, e.g. research into science and technology, I leave you to think about. Whether it can be expected to grow indefinitely like the bacteria in our first population model, or whether it is going to saturate at some level is to be decided by thinking more heavily about the model.

The investment level $I(n)$ also grows exponentially on this model. We can get an expression for a difference equation in $I(n)$ as follows:

Since $Y(n) - I(n) = mY(n) + c$ we have

$$\begin{aligned}
Y(n+1) - I(n+1) &= mY(n+1) + c \\
&= m(Y(n) + rI(n)) + c \\
\text{so} & \\
Y(n+1) - Y(n) &= I(n+1) + (m-1)Y(n) + mrI(n) + c \\
\text{so} & \\
rI(n) &= I(n+1) + (m-1)Y(n) + mrI(n) + c \\
\text{and} & \\
I(n) &= Y(n) - C(n) \\
&= Y(n) - (m(Y(n) + c) \\
\text{so} & \\
Y(n) = \frac{I(n) + c}{1 - m} &
\end{aligned}$$

eliminating $Y(n)$ and rearranging we get:

$$I(n+1) = I(n)(1 + r(1-m)) \tag{2.23}$$

This has solution

$$I(n) = (1 + r(1-m))^n \; I(0)$$

Now all the terms are positive and so we get exponential growth of investment. At some time n the investment levels will go through the roof, no matter how high the roof is. Of course, a really high roof will take a little longer.

There is no explicit connection with unemployment, but it is rather hoped that if the rate of growth of investment is big enough, unemployment can be held constant or even reduced. Consumption therefore has to reduced so as to bring unemployment down. If you are an employer of thousands, you are doing your bit for unemployment if you pay your staff less, thus forcing them to reduce their consumption. A pay rise, on the other hand is forbidden under these circumstances, since it can only push unemployment higher. One person's pay rise is another person's job.

A better model would be one in which unemployment appeared as a term in the model, with some explicit dependence which could be translated into algebra. It would then be possible to try to determine the constants and see if the model was actually any good. There is, as I hope is clear to all of you, a lot of room for improvement in this model (which dates from about the fifties); some improvements have been made to it since then. Also, I have given a slightly simplified version.

Exercise 2.1.8 *Make some assumption about how unemployment might depend on the variables discussed above. Write it down in algebra. What happens to it in time on this model?*

What kinds of things might be done to make this model closer to reality?

That economics is a pretty important subject is fairly obvious. That economists don't have a very good grasp of it is also fairly obvious. Should we reduce tariffs or should we not? What are the probable consequences of doing so? What ought the bank rate to be? (This last question translates into 'how much interest should I pay on my mortgage?' but also into 'how much interest should mother and father get on their life savings when they retire?') How do we control inflation? How do we reduce unemployment? These are clearly important things to understand. And we cannot altogether rule out the possibility that economists do understand some of these matters but are having a bit of trouble getting the answers into our thick heads.

You need to understand that economists construct models (more complicated than the above) in order to find out what is going on just as physicists do. They have a lot of trouble checking their predictions however, and they don't usually get the chance to do much experimenting. They do it in the same way you would. It would be foolish to suppose that the models are an infallible guide therefore. It would also be foolish to ignore them.

For many years, economic arguments were conducted in words. It is clear that some good models of what goes on would be very, very useful.

2.2 Of Mice and Men. And Rats and Women

There is a famous paper on *learning* by Bush and Mosteller which also dates from the fifties. It was a hideous disappointment to me when I first read it, because it struck me as totally uninformative about what is actually happening in brains, rat brains or human brains. Like many people interested in Psychology, I expected something which would illuminate the nature of man[2]. Most students of Psychology, it is said, study the subject to find out what kinds of nuts they are. Others seek insights into the human condition. It is a disappointment to both groups to discover that it is mostly Rats and Stats. The great men of the past like Freud and Jung and Adler (no relation) seemed to have made their disciples dizzy with insights, but none of the insights actually work. It seems, in fact, that the best way to understand human beings is to spend some time talking to them and watching what they do.

One of the things they do is to learn things. They learn to read and write and do simple sums, for instance. They learn to play basketball and to drive cars in many cases, which involves what is called motor-coordination, as well as a grasp of the rules. They learn people's names, and get to be better at predicting the reactions of their wives or husbands. So the word 'learning' seems to apply to rather a large range of activities, from getting better at hitting balls with a stick to memorising telephone numbers.

It is good practice in Science to study the simplest phenomena first, and if you want to study the functioning of the brain, you could start on seeing how good people are at memorising telephone numbers and work your way up. Put like this, it sounds rather disappointing to those of us who hoped for

[2]I was young at the time, and optimistic, and I also hoped to understand something of the nature of woman. I was rather afraid of the female of the species and hoped to learn more about her by reading books, which seemed safer than the, ah, hands-on approach. This seems very naive in retrospect. In the intervening years I have learnt a lot about the female of the species, and she frightens me a lot more.

great insights into the human mind, but a long journey begins with a single step, and maybe telephone numbers will lead onto other things[3]. It is in this spirit of optimism that we look at a simple model for learning.

It is quite easy to remember the telephone number of your sweetheart whom you ring three times a day, much harder to remember your own since you seldom telephone yourself. And hardest of all is to remember the number of the plumber when the septic tank overflows and you are up to your ears in sewage. On the other hand if you called out a plumber twice a day, it seems likely that you would eventually learn his number although not until after you were bankrupt. We conclude that repetition helps. If you break up with your sweetheart and need to call two years later when the current number has thrown you over, you will probably have to check out your little black book. So forgetting occurs, and the longer the time, the more forgetting occurs.

Let's make it look a bit more like serious science by putting it in a more general setting of Stimulus-Response Theory. Why this is called a 'theory' baffles me, it is more a choice of language and not a very intelligent one, although it does go back to Pavlov, who was some sort of scientist, unlike Freud, Jung or Adler. The idea is that you regard an animal, whether a dog, a rat or a human being as a sort of 'black-box', you put some sort of input in such as offering it a pizza or ringing a bell or both, and you see what it does next by way of output. The input is called the 'stimulus' and the output is called the 'response'.

If you imagine trying to understand a computer program by this means you can see that it would work only for the very simplest. Still, it's a start.

Now we imagine taking a human subject and going through his address book and calling out his friends names and seeing if he can respond with their telephone numbers. This would tell us rather less about his memory than about the present state of his social life, so we change it so that we give him a list of pairs of made up non-words, so that any significance he attaches to real words will be irrelevant. When we say 'Greej' he is supposed to say 'Wubnop', when we say 'Groose' he is supposed to reply 'Suglit' and so on. We give him

[3]And then again, maybe they won't. Maybe we could find better simple phenomena to study.

a score of how many he gets right, or maybe we penalise a Wubnop when it should have been a Wopnub. And we tell him what he should have said. Then we look to see if he gets better, and how fast. Students are sometimes paid real money for shouting 'Wubnop' at people. Golly, ain't Science grand?

You can do something similar with rats. Since rats can't talk and don't seem anxious to distinguish Greejim from Grooses, we need a different class of stimulus. For example we can put them in little mazes where they get rewarded by getting bits of cheese when they learn the right way through the maze, and we can see how long they take. Then we can put the cheese somewhere else and see how long they take to abandon the old ways of doing things and learn the new. We can give them electric shocks for doing things wrong, and cheese for doing things right. In the case of human students there is an implied convention that the money is for trying to do thing right, and most students are obliging enough to comply with this. A student who decided to get things wrong consistently in order to teach the psychologist to scrutinise his or her assumptions would be reversing the natural order[4].

In general, take an animal and give it a stimulus and reward it if it gives what the experimenter has decided is the right response, and punish it if it gives the wrong one. This includes giving cheese and electric shocks to rats, and ticks and crosses on their assignments to students (although some of us think cheese and electric shocks would work better with the students too). Measure the score it gets of successes (cheese, ticks) over the failures (shock, crosses) and see how this changes as the data is repeated. The work of Skinner, [5, 6, 7] is based on the assumption that all, or almost all, of human behaviour is made up from conditioned responses in this way. He invented the famous 'Skinner Box', of which a lecture theatre or classroom is a pale and malfunctioning copy.

So we can certainly get some data of this type, for a range of animals and a range of stimuli/response sets. It is far from clear that the results will give any information about anything much, but at least the data is, sort of, real. I say sort of, because there is an implication that two rats would behave in a fairly similar manner, which means that you can infer from a sample of rats which have learnt a given maze, roughly how long it will take another rat to learn the same maze. This has obvious problems, like striking a really

[4] Always worth considering.

obtuse rat. It is far from clear that you can use traditional methods to teach difference and differential equations to a student and infer how long it would take to teach them to another student. Some pick it up in seconds, others years, and some never. Still less is it clear why you can't teach differential and difference equations to rats if you are prepared to code it appropriately. Still, there are bound to be questions outside the scope of any model. In the case of students learning random pseudo-word pairings it would seem that the only questions that get answered are those nobody actually cares about. But this may be harsh.

Let us suppose then that there is a natural frequency with which a subject produces the correct response. I shall turn this into a probability and suppose that $p(n)$ is the probability that at presentation n the subject gives the correct response. Think of the psych. student having to search his memory to come up with 'Wubnop' on the n^{th} time of the experimenter enquiring 'Greej?'. We can estimate this probability rather poorly by assuming it is the same as the probability of getting any other pseudo-word pairing correct at the n^{th} repetition. This is a big assumption, and might cause you to raise your eyebrows and have doubts about the whole business, but you should be used to that by now.

Now it is not wholly daft to suppose that $p(n)$ increases with n. This is the reflection of our experience with telephone numbers of people we ring a lot. And Bush and Mosteller wrote down the equation

$$p(n+1) = m\ p(n) + c \tag{2.24}$$

Whether they chose this equation because it has some basis in neurophysiology or whether it was because they knew how to solve it is an interesting question. My money is on the latter.

Now we know exactly how to solve this equation, it is dead easy (which is how Bush and Mosteller came to know it) and the result is

$$p(n) = m^n p(0) + c\frac{1-m^n}{1-m} \tag{2.25}$$

where p(0) is the initial probability of a correct guess.

Now the question is, what values of m and c give us reasonable sort of behaviour? We can say something about this without actually going near a

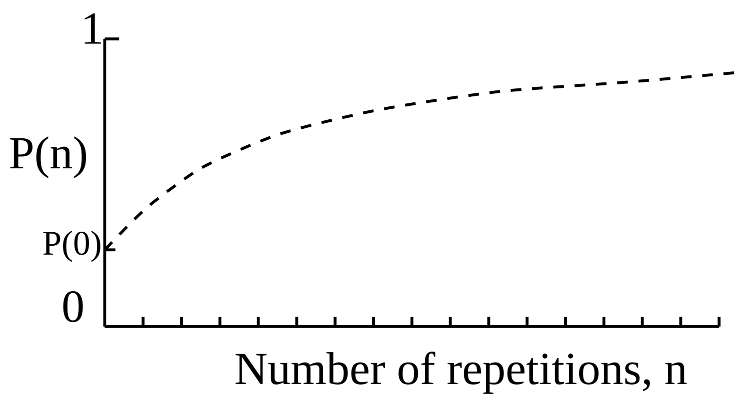


Figure 2.7: A possible shape for the learning curve

rat or a student. We do know that the probability has to be between 0 and 1, and we expect that it will be increasing. We might also conjecture that it gets closer and closer to 1 as n gets bigger.

If we make this assumption, then the function $P(n)$ has to look something like figure 2.7.

Nearly all of our first order affine difference equations thus far have gone zapping off to infinity at a rate of knots, leaving you, perhaps, with the impression that they all must. Not so.

It should be clear that if there *is* a point y such that the equation

$$y(n+1) = m\, y(n) + c$$

has this y unchanged or *fixed* or *an equilibrium point* then we must have

$$y = my + c \tag{2.26}$$

which means that

$$y = \frac{c}{1-m}$$

Now if we assume that repetition eventually makes you perfect, then $p(n) \to 1$ as $n \to \infty$. In which case we conclude that

$$c = 1 - m$$

must hold.

Of course, this assumption that you get better and approach the perfect score may be false; maybe you approach some less than perfect score. I am sure I would. (In fact I am sure that I would get worse after a very few repetitions indeed, possibly none.) In which case we deduce that

$$c < 1 - m$$

But let us stay simple minded for a while.

If this is true, then it doesn't take many measurements to determine the whole curve. This is called *parameter estimation* and is usually a bit harder than this.

Example 2.2.1 *Suppose we take two possible stimuli and paired responses, Greej and Wubnop being the first pair and Groos and Suglit the second.*

We take a whole class of Psych students, who know the two response words but not which one is right for any stimulus word, and say one of the stimulus words at them, and if they get it right we give them a dollar, while if they get it wrong we give them an electric shock. We keep a record of the fraction of the class which get the right answer after n repetitions. We find that after only three repetitions 75% of the class get the right responses. How many will get it right after five repetitions if Bush and Mosteller are not talking through their hats?

Solution

If we suppose that $p(0) = 1/2$ since the students can't be expected to do better than chance, and if we trust the model, and also assume that the students are independent estimates of a probability in each students head (and moreover the same one), then we put $p(3) = 0.75$. This gives

$$p(3) = m^3(1/2) + 1 - m^3 = 0.75$$

and leads (check me carefully) to

$$m = (0.5)^{1/3} \approx 0.79370053$$

Then $p(5) = m^5(1/2) + 1 - m^5 \approx 0.84250987$ *which means that more than* 15% *of the class will still get it wrong. Well, that's Psych students for you some would say, but it might be a trifle more charitable to suspect the model. Or the data in the question.*

It is necessary to point out that this is an extremely simple minded model (a physicist would probably say 'moronic') but at least it has the basics of science in it: we can check up to see if it works reasonably well and ditch it when it doesn't. This is a considerable advance on Sigmund Freud's work, which has had a great influence on literature but very little by way of a success rate at curing neuroses, which is what it was intended to do. In fact it is well known that all the psychoanalytic methods of curing neuroses have about the same success rate, and this includes not doing anything at all. Which is at least cheap. This hasn't stopped the practice of psychotherapy or noticeably slowed it down. By contrast, it is fairly easy to cure say, bed-wetting, on a Skinner model, by making the bed-wetter sleep on a gizmo which detects the presence of urine and sets off a buzzer. This works very well. Freudians mutter darkly about repressed neuroses coming out in other forms, but there is no particular evidence for this.

There is nothing very difficult in the ideas discussed so far, and were it not for the propensity of some psychologists to take essentially banal ideas and dress them up in ill-defined jargon to make the material look more important, anybody could read the literature. There are also some very well written papers and books in the area. I recommend the books of Professor Hans Eysenck [8, 9, 10] because they are well written and rather fun. I heard Eysenck speak to a MENSA meeting about this stuff a long time ago, when I was young and relatively charming, and the effect it had on the MENSA intellectuals was very educational.

Since I work in the area of trying to understand something of how human beings make sense of the world, and write computer programs to test my ideas (or get my students to do it), it is only fair to point out that the Bush-Mosteller model would be regarded as moronic by most psychologists these days, and that far more sophisticated modelling processes are available today. Read David Marr's beautiful book *Vision*, [11], to see how much progress has been made in constructing serious models. One of my colleagues and a former student currently in the department of Electrical and Electronic Engineering is working to test some of Marr's ideas at the present time.

Exercise 2.2.1 *Investigate a model for not just two words but lots more. Where does it make a difference?*

The model makes no allowance for forgetting. *One possibility is that people simply lose a response to a stimulus over time at some fixed rate. Write down some equations describing this possibility. Another is that the thing that causes you to forget is learning other things. You could imagine that you just have something fall off the end of a list of what you can remember, the theory that your memory is a bit like a bucket, and it has some definite capacity. Another possibility is that you don't lose items, you just lose a bit of skill at each recall, it hits you in the probabilities.*

Investigate to see whether you can write down equations for these possibilities to turn them into respectable mathematical models.

2.3 History: Truth, Lies and Radioactivity

There is, in a cathedral in Turin in Italy, a piece of cloth called the *Shroud of Turin* which has got on it marks which look like the face of Christ. Well, the face of somebody. You can see marks which look a lot like a man's face with blood stains where the thorns would have been.

It has been around for quite a while, and the faithful (or credulous, depending on your perspective) have taken it for the shroud in which the body of Christ was wrapped after he was removed from the Cross. Some mysterious process is supposed to have made it act like a photographic negative and absorb the appearance of Christ himself.

The face (figure 2.8) is certainly impressive to me: it is possible to read whatever you want into a poor quality image, and if I'd been a poor peasant

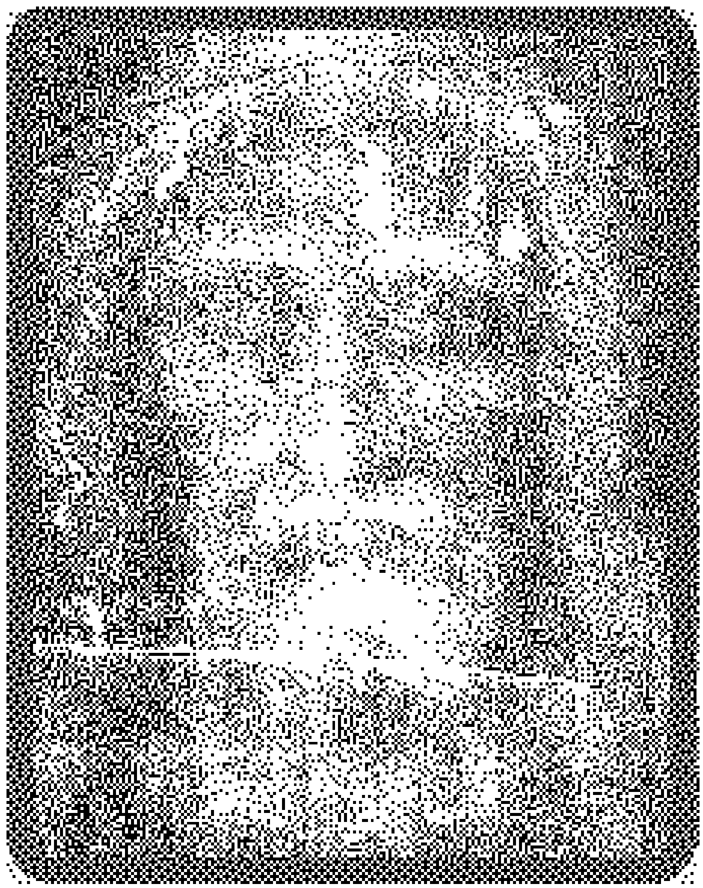


Figure 2.8: The Face of Christ?

and had just heard the sermon on the mount, I would probably have followed Christ if he'd looked like Danny DeVito, but there is something about the face looking out at you that has moved a lot of people.

If it were the face of Christ, it would of course be very exciting. After all, Christ was (probably) a historical figure of some importance. There are those who believe he was the Son of God, the bloke who created the entire universe. There are others who think the claim sounds a lot too much like claims made by the bunch of California whackos who planned to go up to the UFO escorting comet Hale-Bopp. And a lot of other people would take the view that whether JC was the Son of God is rather doubtful, but he certainly had an ethical view which we can only admire and respect. It should be obvious which group I come into. I rate him a human being but a good guy. Opinions differ, but that's mine. I tell you this so that you can allow for my prejudices in what follows, and not so as to give a hint of what the right answer is for an exam. Like most academics, I should admire you more for sticking to your guns and producing arguments against my point of view, than for accepting a belief out of some form of political correctness or because you want me to like you. Anyway, I like people who argue with me.

I tread on dangerous ground here, because I might give deep offence to some people who hold their religious beliefs very strongly. I have wondered whether to choose a less contentious subject. But after some thought, I concluded that if there is a God and he cares about us then he probably wants us to tell the truth as we see it and to be as honest as we can manage. And telling lies on his behalf, or even just suppressing the truth, is probably not a good move. It is hard to see God wanting us to deceive ourselves or others. So in the spirit of honest pursuit of truth, essentially a rather spiritual ideal and one which marks the honest scientist, let us proceed.

So the question is, is the Shroud of Turin a genuine picture of Christ himself, or is it a forgery or some other face altogether?

It would seem to be impossible to decide, and the Roman Catholic Church for centuries took no definitive position on it. Maybe it is, maybe it isn't. It is certainly quite old.

But is it old enough? If it were the shroud in which the body of Christ were wrapped, it would have to be pretty close to two thousand years old. If it *were* two thousand years old, this wouldn't prove it was the face of Christ, but if it were a lot less old it would prove it wasn't. So can we date linen fabrics that well? The answer is yes, and the Shroud of Turin has been exposed as a well constructed fake. It is still kept on in Turin.

The faithful may have doubts about this process, so I shall explain the technique so you can scrutinise the assumptions and decide which ones you want to doubt. Any fundamentalist christians who believe the world was created around six thousand years ago will want to be full bottle on this because Jericho dates back to more than 9000 years ago when it had a population of about 2,500 people, and of course many human remains predate Jericho; the Lascaux cave paintings are more than 13,000 years old and the Willendorf Venus, a sort of palaeolithic pin-up girl, is over thirty thousand years old. There is Australian aboriginal art dating from around fifty thousand years ago, the oldest known human art, and there are good reasons for dating human origins to several million years. The question is, why does anyone at all believe these estimates? How reliable are they? Is it an elaborate con put about by atheists to lead the faithful astray? Or is it a conclusion that any reasonably honest and well informed person would be led to? These are interesting questions, and of some importance if you are worried that scientists are trying to swindle you. They are also generally fascinating to anybody whose brain is not permanently damaged by too much television.

The best known technique is that of radio-carbon dating, so I shall explain that one. The others are similar in general form.

The first thing to note is that radioactive substances are made of atoms which spontaneously change from being one element to another. It came as something of a shock to find that this could happen, and Marie Curie was the first person to establish that it happened with Thorium. I am going to skip the Physics, fascinating though it is, and pass on to Carbon.

The Carbon atom usually has an atomic weight of twelve. There is a form of Carbon called Carbon-14 which has an extra two neutrons in the nucleus, and which spontaneously goes SPUNG every so often, spitting out an electron and turning into Nitrogen-14. We can work out the amount of Carbon-14 in a sample of ordinary Carbon by weighing it; this requires something more exotic than a bathroom scales however. Anyway, by heating the stuff up, enough to ionise it so as to get rid of the electrons in the shell, squirting the stuff through a slit and passing it through a space over which there is a magnetic or electric field, we can measure the extent to which the path is curved by the field: this measures the mass of the atoms. Carbon-12 goes one way and Carbon-14 another. Nitrogen-14 goes the same way as Carbon-14 pretty nearly, since the masses are almost the same. So this doesn't allow you to distinguish Carbon-14 from Nitrogen-14, but just leaving the stuff around at room temperature will get rid of the Nitrogen. If you have doubts about the feasibility of anything so unlikely as estimating the amount of C-14 in a mixture of ordinary carbon by these means, you should realise that your television set works on pretty much the same principle, except that there the mass of the electrons is constant and the electrostatic field varies.

Carbon-14 decays spontaneously then. It does so pretty much at random. One day it is Carbon-14, and after waiting a while it goes SPUNG and it's Nitrogen. There is no way that you can predict when it will happen for any atom of Carbon-14. With a whole stack of the stuff however, you can say what fraction will go SPUNG in any given time.

The reason is precisely that the process is random. There is a fixed and definite probability for any atom of C-14 that it will SPUNG in the next 24 hours. It is very small. If, however, you have a kilogram of the stuff, you have an awful lot of atoms, and you can confidently bet the farm that a certain number of them will go SPUNG in any minute. And you can count them by putting on a *beta* particle counter near the stuff. You can hear the separate clicks as the electrons hit the detector. By counting over a lot of minutes, you can get a good estimate of the average number of SPUNGS per minute. It turns out to be proportional to the amount of radioactive material present, which should not surprise you.

We can write down therefore the equation:

$$N'(t) = -kN(t)$$

where $N(t)$ is the number of atoms of C-14 present and $N'(t)$ is the SPUNG rate of the Carbon atoms, alternatively the rate of shrinkage of N. It follows that $-k$ is some negative number.

Writing
$$\frac{dN}{dt} = -kN$$
and doing some fiddling we get
$$\frac{dN}{N} = -k\ dt$$
and integrating we get
$$\ln N = -kt + A$$
or

$$N(t) = Ae^{-kt}$$

where A is the amount of stuff present at time zero.

Different atoms will have different values for k. We could actually give a k value for every element. It would be zero for elements that are not radioactive at all, and relatively big for atoms which decay (the technical term for 'go SPUNG') relatively fast. what we actually do is to give the time it would take for any amount to have half the atoms decay. This is called the *half life* of the substance. For example, C-14 has a half life of 5,568 years. Thus we can put
$$N(5568) = Ae^{-k5568} = A/2$$
so
$$e^{-5568k} = 1/2$$
or
$$k = \frac{\ln 2}{5568} \approx 0.0001244876$$
This is the average number of SPUNGS per atom per year.

Decay rates for different elements vary enormously. Half lives range from microseconds to gigayears. Naturally, the longer the half life, the fewer atoms go SPUNG, sorry, *decay*, in any given time.

All this is well known and easily checked up on, and if you were brought up by some neurotic to believe that radioactivity is evil, you should find out a little of how many medical applications there are.

The next question is, if Carbon-14 has a half life of just over five thousand years, how come there's any of it around now if the earth is nearly 5×10^9 years old? Either there was an awful lot of it that long ago, or something is producing it. The answer is that something *is* producing it, and the something is cosmic rays. Cosmic rays are superhigh energy particles which come from outer space, mainly in the form of gamma rays, very high frequency radiation. They hit the atmosphere and transmute what they hit into something else, not always but every so often. They cause a number of things to become radioactive, and do all the damage that radioactivity does. Cosmic ray bombardment would have to be considered 'natural', but it isn't very healthy, at least not to individuals. On the other hand, it powers evolution, which has produced you and me, so it isn't all bad.

But how does this allow us to date the Shroud of Turin? Well, it is a matter of living things versus dead things. When flax or wood or other organics are alive and well, they absorb carbon in the form of carbon-dioxide from the atmosphere. If we assume that the rate of cosmic ray impact hasn't changed dramatically over the last few thousand years, then the fraction of atoms of carbon which are C-14 is about the same as it is today in living wood or in the atmosphere. When somebody cuts down a tree or some flax and makes it into a table or a shroud, then there is no more balance between the atmosphere and the substance, because it doesn't take in any more carbon. The radioactive carbon decays. The amount of the stuff goes down in time. So the SPUNG rate for a gram of the mixture of ordinary carbon and C-14 decreases over time. We measure the number of SPUNGS from a gram of the mixture.

For living wood, or flax, this is about 6.68 per minute per gram. By counting for a lot of minutes, we can get quite good estimates of the decay rate. Suppose it was precisely 0.0000070 per minute for a microgram of the mixture. By timing over a period of about two years we can get an estimate for just a microgram. Ten micrograms (a VERY small amount) is sufficient to get fair estimates to the required precision. Of course, we have to make sure that we are shielded from other things that the detector could respond to.

If we make the measurements, we get about 5.6 per minute per gram from the Turin Shroud. How do we calculate the age? More accurately, how do we calculate when the flax from which it was made was harvested?

Well, if we burn the carbon we lose the nitrogen, and we have to do a correction because a gram of the mixture of the two carbons of time t ago has lost the Carbon-14 which got turned into Nitrogen. So it doesn't weigh a gram today. And if we take a gram of the mixture of the stuff that we are trying to date, we get more of the C-14 than we ought to have. Actually, the correction is small because only a small fraction of the carbon found in nature is C-14. So we usually ignore this consideration in giving an explanation, although not, of course, in doing the calculation. So I just take a gram of the mixture today and assume it started off as a gram of the mixture t years ago. Then the decay rate (which is the proper term, I made SPUNG up) is proportional to the amount of C-14 in the sample which is e^{-kt} times the mass of C-14. We therefore have

$$5.6 = e^{-kt}6.08$$

or

$$t = \frac{1}{k}\ln(\frac{6.08}{5.6}) \approx 661$$

This makes it living material around the time of the crusades, which makes the idea that it was woven before Christ died rather hard to credit.

Exercise 2.3.1 *The half life of C-14 is not known to better than plus or minus 40 years. What is the probable error in the age of the Shroud of Turin arising from this uncertainty?*

There are also methods of dating other things: it is common to be told in geology books that the age of a rock is something like 2,700 million years. What, you might ask, does it mean to give the age of a rock? Are rocks born at some reasonably well defined time? Wouldn't you feel silly singing 'happy birthday' to a rock? Think about this one while looking at the next exercise:

Exercise 2.3.2 *Potassium 40 has a half life of* 1.28×10^9 *years and decays into Argon 40, a gas. Let us suppose that we can find the amount of argon in a rock sample by crushing and heating, and that we can also weigh the amount of Potassium 40 in the rock. Suppose that such a sample provides* 3.5 *grams of Argon-40 and* 0.5 *gram of Potassium-40. When was the rock formed?*

Explain clearly your assumptions.

How do we know that the relative abundance of C-12 and C-14 hasn't changed over the time? Well, we can calibrate the method by checking with objects where the date is known by other means. There are rivers which carry pollen and seeds down stream, the pollen sinks and is covered by silt. Since pollen is produced annually, we can dig up a section of river bed, look at the sections under a microscope and count years. These are called *varves*. If we find some carbon containing object embedded in the mud at some depth, we can estimate its age two ways and see how well they agree. If the agreement is very good over a reasonable range, we can feel a bit more confidence in both methods. Another method is *dendochronology* which means counting tree rings. There are scads of ways of dating things, and they all have margins of error and elaborate checking procedures exist. You may recall the furore recently at the possibility that Aborigines have been in Australia for well over $100,000$ years. This used a thermoluminescence dating procedure and the smart money says the guys who came up with this very early date stuffed up. About half that is the usual estimate.

So there are lots of dating techniques, all with various degrees of reliability and accuracy, and they have all been cross calibrated with each other when this is possible. It should be clear that radiocarbon dating will not work well for very recent or very, very old objects, but other methods can be used in these cases. You must understand that these techniques represent a lot of man (and woman) hours of hard work and careful measurement, with a good deal of care and checking required. I have given you the basic ideas,but there is a lot more to the details than that. There usually is.

The following quotes are taken from the web sites relating to the Shroud of Turin:

'Over the centuries, dozens of shrouds–some with images and some without–have surfaced claiming to be the burial cloth of Jesus. In the case of the Shroud of Turin, it has been publicly declared a forgery by both Roman Catholic Church officials and prominent scientists. In 1389 the local bishop of Troyes denounced the Shroud claiming an artist had confessed to forging it. More recently, in 1988, after three different laboratories Carbon-14 dated the Shroud and found it to be some 1200 years younger than it should have been, the Roman Catholic Church announced to the world the results of the test.'

'In 1988, the Vatican allowed the shroud to be dated by three independent sources–Oxford University, the University of Arizona, and the Swiss Federal Institute of Technology–and each of them dated the cloth as originating in medieval times.'

The fabric has also been dated and the pigments analysed (alas, no blood so we don't know Christ's blood type) by microscopic analysis. The website

`http://www.mcri.org/`

says:

'The carbon-dating results from three different internationally known laboratories agreed well with this date: 1355 by microscopy and 1325 by C-14 dating. The suggestion that the 1532 Chambery fire changed the date of the cloth is ludicrous. Samples for C-dating are routinely and completely burned to CO2 as part of a well-tested purification procedure. The suggestions that modern biological contaminants were sufficient to modernize the date are also ridiculous. A weight of 20th century carbon equaling nearly two times the weight of the Shroud carbon itself would be required to change a 1st century date to the fourteenth century.'

Three papers, two by McCrone [12, 13] and the other by Damon, [14] give the scientific details.

There are still a number of dedicated 'shrouders' who are arguing the matter, but the Roman Catholic Church has accepted the report of the committee (which included theologians) and accepted that it is a fake. Pity really, but that's the way the cookie crumbles.

The following websites can be consulted by the curious for 'balance', although they look to me to be mainly rather amateurish and arguably dishonest counter arguments to me.

http://205.177.10.11/npacheco/shroud/turin.html
http://www.shroud.com/papers.htm
http://nen.sedona.net//nhne/shroud.html
http://www.shroud.com/.

2.4 Summary and Conclusions

In this chapter I have used some relatively simple mathematics to model a variety of systems of interest in the social and life sciences, ranging from population models to anthropology. I have skipped over the actual solving of the equations because (a) you probably did most of the easy cases in school and (b) I wanted to convince you that modelling is good clean fun and takes you interesting places before I got stuck into serious sums.

In a number of cases, I took some quantity which varied in time, and tried to work out the way it actually does change. This was done by finding a *function* which took input the time and gave output the value of the quantity at that time. There were a number of different quantities considered, namely:

1. the number of bacteria in a petri dish
2. the number of people on the planet
3. the amount of money in a bank account
4. the national income and national investment
5. the probability of correctly remembering a nonsense word
6. the amount of radioactivity in the Turin Shroud.

There are a number of things to be thought about here. One is the degree of credibility you might attach to the value of the modelling. Is modelling worth the time and trouble involved in setting up the equations and solving them? Is it *fun*? It certainly is a new and strange sort of fun if so, but I hope that you will come round to thinking that it is.

Another is the strange way in which rather a lot of different measurable quantities seem to be described by the same sort of equations. Is this a coincidence, a sign that the examples have been carefully selected, or is it that the universe runs along fundamentally similar lines in quite different areas? If the last, this would seem to be much stranger than the X-Files.

Looking at the first issue, it has to be admitted that the models discussed varied rather a lot in terms of their value. You could, of course, take the view that none of this stuff is very interesting and that you'd rather be watching television. On the other hand, you might actually wonder whether the human race will overpopulate the planet in the near future. Or you might wonder if the Shroud of Turin is genuinely the face of Christ. If you are interested in these matters, you have that curiosity which has led to our present civilisation. If not, you are basically a house-trained monkey.

The bacteria and human population models struck me as having some interest because they looked as though they were just starting to get somewhere. The explanation that the human population has mostly been in equilibrium with its food supply and its predators, and that the modern spurt in growth is due to technology looks at least plausible, and probably wouldn't have occurred to us if we hadn't tried to get a better model than exponential growth. The logistic model did an impressive job of predicting population growth over a span of a century. It was very testable, and it passed a fairly demanding test. The more general model we developed with its changes of food supply needs a lot of work, but it is plausible that we could get better models by doing some more thinking, and the qualitative features of it were interesting. You can also see, I hope, that this could be a fun way to spend an evening with a computer.

The financial models looked as though they could be useful to some people, although they were pretty simple. I wish I had enough money so that they could be more useful to me. Again, testing them is possible, and making them more complicated to allow, for instance, for estimates of forward values of bonds under uncertainty about interest rates would be fairly straightforward. This looks honest stuff, although perhaps not as exciting as I would like. The reason I feel that way is probably because I am poor and had a deprived childhood.

The economic model struck me as rather dubious, so I should explain now that it was made up by an economist and not a mathematician. I felt that it was very difficult to see how to test it, and also very difficult to see how to measure the quantities in a way we could all agree was the right way. The result that investment would increase without limit worried me. I rather doubt if this sort of thing fits real data if real data can be obtained, And if it can't, we really are wasting our time. As it happens, much more down to earth economic models now exist and are both more plausible and more useful. They still do not do a very good job of predicting the economy, despite using a lot of computer power. Qualitative analysis of economic behaviour can give useful insights, and I shall say more about this later.

The psychological model was probably the silliest we looked at. Again, I didn't invent it, it really was produced by a couple of psychologists, and may be telling us that psychologists around the middle of the century were really bad at maths. I don't mean that they weren't good at solving equations (although I expect they were fairly awful at that in general), I mean they didn't do a very good job of selecting the right thing to model. All the same, the model can be tested experimentally, which makes it a huge improvement on the waffly theories which cannot. Again, much more complicated models of much more interesting things are being developed in psychology at the present time. I chose this one because (a) it is in all the books and (b) it is very simple and we can easily solve the equations. This isn't fair to the people who work in the area today. On the other hand, there is still a lot of waffly stuff in psychology which is never made precise and never made testable. In serious sciences, theories and models get rejected by appeal to the data. In less serious subjects, theories or models simply become unfashionable. This is a sure sign that nobody takes *any* of the theories or models seriously. They are a means to advance one's career, not a means to advance human knowledge or understanding. Third rate minds consider this much more valuable. First rate minds want hard data and clear theories, and they dream dreams of what might be.

The radiocarbon dating strikes me as rather nice, and I find it more useful and more believable than the economic or memory models. It is simple enough to understand fairly well in principle. There are a number of points which I think I would need careful persuading on, and if I believed deeply that the Shroud of Turin was authentic, I could pick some holes in the case as presented here. But I would definitely be worried. I would want a hard look at the numbers as actually produced by the committees which did the analysis (three groups independently which agreed pretty well). From what I know of how careful and honest most scientists are, I should expect to be beaten down by the huge mass of detail supporting the arguments. Similarly, there are other applications of radiocarbon dating going back tens of thousands of years, others (like thermoluminescence) going back comparable times, and others like Potassium Argon dating going back millions of years.

A very nice feature of all the models which may strike you is how the same equations kept coming up all the time. It is only a little bit faked: I selected models that could be analysed with only simple mathematics when I could do so. This had the effect of making the economic and psychology models look rather sillier than was altogether fair. The population model for bacteria looked rather silly too when it was started, but I developed that one a bit and I didn't the others. But is it something mysterious that makes the same ideas come up again and again in different areas? Many people have been puzzled by this. Some have claimed that 'God is a Mathematician'. This was said by a theoretical physicist, and is rather an extreme view, but if you knew how some of the theories in the physical sciences seem almost uncannily good, enough to send a shiver of awe up the spine, you'd see what he meant. Anyway, I do not think this is a very satisfactory explanation myself. I leave you to form your own opinions as you get more information.

So is modelling a good idea? Well, it has made a big difference in the hard sciences. In some of the softer sciences it does give interesting results in some cases, and in others it looks pretty bogus. I shall examine some other models later in these notes which may reassure you that the implausibility of some of the models can be fixed up in the same way as we improved the population model and came to learn something in the process.

You should however be sceptical about modelling in general, and suspicious that mathematicians are selling you something which is basically high-tech garbage. I don't believe this is true, but I prefer to make a case to hard-headed doubters rather than the far more cynical people who don't give a damn either way as long as they pass the exam.

In this chapter I have been mainly concerned with showing the power of only two abstract ideas, the difference equation and the differential equation as machinery for studying systems which change in time. I have limited myself to simple cases in the main, in some cases too simple to be of any genuine value. You ought to be impressed, all the same, with the unreasonable effectiveness of the modelling process in some of the cases.

In the next chapter I shall forget about reality briefly and concentrate on the mathematics. I want to study more complicated equations and figure out how to solve them, so we can make some headway with more complicated models. If you have followed me so far, you will see that separating the process of modelling from the tools for actually solving the equations is a good idea, because totally different models can have the same underlying equations.

Chapter 3

A Menagerie of Difference Equations

In this chapter I am going to give you the lowdown on the standard types of difference equation which are easy enough to have closed form solutions. I shall also say a bit about how to use computer programs to get some pictures showing what is going on.

After this, we can really start the course proper, which is going to be on the art of developing models for real applications. At this point I should warn you not to get your hopes too high: being able to computer model a hurricane or even the day to day weather is not on the syllabus. Neither is developing a full economic model to figure out what to do about unemployment. It would take many years to get you able to understand the basics of how to do the weather modelling, and we do know how to cut unemployment but you wouldn't like it. We are rather limited by the need to make sure you can understand enough of the background to see how to set up the equations; otherwise you just learn to use the jargon without understanding it, which is a silly and pointless exercise.

3.1 Some Definitions

A *Difference Equation* or *discrete one-dimensional dynamical system* is a process for generating an infinite sequence of numbers by giving a rule for calculating each number, using in the calculation only where it comes in the sequence and those numbers already known.

Example 3.1.1 $y(n) = y(n-1) + y(n-2); y(0) = 0, y(1) = 1$

This is a Fibonnaci sequence*,*

$$0, 1, 1, 2, 3, 5, 8, 13, \cdots$$

We write $y(n)$ for the n^{th} number. It can be any real number, not just an integer as in the above example. This means we can think of y as a function from the natural numbers, $\mathbb{N} = \{0, 1, 2, 3, \cdots\}$ to the real numbers, sending n to $y(n)$. Sometimes instead of $\mathbb{N}$ we use $\mathbb{Z}^+ = \{1, 2, 3, 4, \cdots\}$. It doesn't really make much difference, the Americans start with floor one and the British start with the ground floor, and in Australia, the first floor is the first one you get to after you climb the stairs. In the examples, I used both, but from now on I shall be consistent and use the Aussie way; 0 is the starting number.

A difference equation is *of first order* if the value $y(n)$ depends only on $y(n-1), n$ and constants. It is of *second order* if it depends on the last two values, n and constants. And so on.

Example 3.1.2

$$y(n) = 3\ y(n-1) - n^2 + 2; y(0) = 0.1$$

is a first order difference equation.

$$0.1, 1.1, 1.3, -3.1, \cdots$$

The Fibonnaci sequence given above is a second order difference equation.

Exercise 3.1.1

$$ny(n)y(n-1)^2 = (k-1)y(n-2);\ \ y(0) = y(1) = 1$$

is a second order difference equation. Write down the first four terms.

A difference equation is said to be *autonomous* if its calculation does not use n. Of course, the term $y(n)$ will certainly 'depend on n' in one sense; the question is does it use n in the computation.

The Fibonacci equation was autonomous but the last example was not.

This is a property of the equation and not a property of the actual sequence, because

$$y(n) = n$$

and

$$y(n) = y(n-1) + 1; y(0) = 0$$

are the same sequence, but the second version is autonomous while the first is not. Because the first equation does not contain any reference to earlier values of y, it is called a *solution* to the second equation.

Exercise 3.1.2 *Is $y(n) = 2^n$ replaceable by a non-autonomous equation? Alternatively, find a difference equation for which this is a solution.*

A first order difference equation of order k+1 is said to be *linear* if it has the form

$$y(n) = f_n(n)y(n-1) + f_{n-1}(n)y(n-2) + \cdots + f_{n-k}(n)y(n-k)$$

where the $k+1$ functions are any functions of n.

It is said to be *affine* if it is 'linear plus a shift', i.e. if it has the form:

$$y(n) = f_n(n)y(n-1) + f_{n-1}(n)y(n-2) + \cdots + f_{n-k}(n)y(n-k) + g(n)$$

Example 3.1.3 *An autonomous first order affine difference equation must therefore look like*

$$y(n) = m\ y(n-1) + c$$

for some constants m and c. Many of those discussed in the first two chapters were of this form. That was because it is easy to solve them.

For a really concrete example,

$$y(n) = \frac{1}{2}y(n-1) - \frac{1}{2};\ y(0) = 2$$

which is easily seen to go:

$$2, \frac{1}{2}, -\frac{1}{4}, -\frac{5}{8}, \cdots$$

There is a close relation between differential equations and difference equations, the matter being all tied up with the question of how often you compound the interest.

Suppose I have the differential equation

$$y' = ky; \quad y(0) = 1 \tag{3.1}$$

We all know this has solution

$$y = e^{kt}$$

And it is easily checked that this is correct by differentiating.

If differentiation had never been invented but computers had, we might have written 3.1 as

$$\delta(y) = ky\ \delta(t); \quad y(0) = 1 \tag{3.2}$$

in which case we could have taken steps of size $\delta(t)$ and started off with $y(0) = 1$ to get a sequence of values of y, $\delta(t)$ apart:

$$1, 1 + k\delta(t), 1 + k\delta(t) + k(1 + k\delta(t))\delta(t) + \cdots$$

This simplifies a bit:

$$1, 1 + k\delta(t), 1 + 2k\delta(t) + k^2(\delta(t)^2) + \cdots$$

and you can probably see how to go further to get:

$$1, 1 + k\delta(t), (1 + k\delta(t))^2, \cdots$$

This is simply because we have

$$\delta(y) = y(n+1) - y(n) = ky(n)\delta(t)$$

or

$$y(n+1) = (1 + k\delta(t))y(n)$$

with the solution

$$y(n) = (1 + k\delta(t))^n y(0)$$

Note that if we make $\delta(t)$ half the size, then to find out what is happening at some time T after starting, we would have to calculate twice the number of terms. So we could write

$$y(T) = (1 + k\delta(t))^{\frac{T}{\delta(t)}} y(0)$$

to give an estimate of the value of y at time T. In the limit as $\delta(t)$ gets smaller and smaller, we approach the result of going along a smooth path. The fact that $\lim_{x\to\infty}(1+m/x)^x = e^m$ explains where the exponential function comes in. It is all about the limit of 'continuous compounding', or taking the limit of smaller and smaller 'Euler Steps' as we approximate an ordinary differential equation better and better by difference equations. I shall come back to this point later when we look at some more differential equations.

Exercise 3.1.3 *Check through this carefully to make sure you follow the argument, and to confirm that the solution is right.*

This is exactly how computers obtain numerical approximations to differential equations. Naturally, the approximations get better as the size of $\delta(t)$ (the 'compounding period') is reduced, and naturally there are a lot of smart tricks which have been dreamed up by mathematicians over the years to get out the best possible answer with the least amount of work. and to work out when the answers they get are unreliable. This is the subject of Numerical Analysis. Whenever anyone talks of *solving a differential equation numerically*, they mean that they have converted it into a difference equation. And usually that they have 'solved' the difference equation by bunging numbers, and the rule for getting more, into a computer.

Example 3.1.4 *The logistic differential equation*

$$y' = ky(a - y)$$

might be solved numerically by replacing it with the difference equation:

$$y(n) = y(n-1) + k\ y(n-1)(a - y(n-1))\Delta$$

where Δ *is some scaling factor to make sure that the approximation is not too bad. The figures 2.1 were drawn by precisely this means.*

The above difference equation is first order but NOT a linear or affine equation because it contains a square term for y*.*

To *solve in closed form* a difference equation with $y(n)$ as the name of the function value at n is to find some expression in terms of known functions of n for y. For example:

Example 3.1.5 *Solve*

$$y(n) = \frac{1}{2}y(n-1) - \frac{1}{2};\ y(0) = 2$$

Solution:

$$y(n) = \frac{2}{2^n} - (1 - \frac{1}{2^n})$$

I shall show you exactly how I got this expression soon, but you should see if you can figure it our for yourself first. Try writing down the first three or four terms. This is good clean fun and will keep you busy for a while.

Usually we just say that we *solve* the equation when we mean 'solve in closed form'.

A *numerical solution* means that we start off applying the rule from the starting number or numbers and just keep going. Given a computer we can do a lot of these. There might be more efficient ways of doing it, but this will work.

Sandefur, in [15], shows how to use a spread sheet to calculate some of the terms in a difference equation. If you are used to spread sheets, you might like to try it yourself.

3.2 Linear (and Affine) Difference Equations

3.2.1 First Order Difference Equations

Awful Warning

The older books lump together affine and linear equations and call them both 'linear'. Watch out for this. The distinction is important, and the linear equations are easier, which is why we study them first.

End of Awful Warning

The simplest possible equation is the first order linear autonomous equation. Autonomous equations are often called 'constant coefficient' equations, for obvious reasons.

Every first order linear autonomous equation looks like

$$y(n) = my(n-1)$$

where m is any real number, and if we start off with $y(0)$ we get

$$y(0), my(0), m^2y(0), m^3y(0), \cdots$$

which has solution

$$y(n) = m^n \, y(0)$$

To be absolutely cast iron certain about this, we can see that if $y(n) = m^n \, y(0)$ then $y(n-1) = m^{n-1}y(0)$ and dividing the first equation by the second, we get

$$\frac{y(n)}{y(n-1)} = m$$

which is the difference equation we started with. So this is definitely a solution. Could there be another one? Obviously not if $y(0)$ is given, because any two solutions would have to agree on $y(0)$ and couldn't ever disagree, since the difference equation always tells us how to construct the next term, and there is only one possible answer. Of course, if we start from a different $y(0)$ then we will always get a different solution, if by the term 'solution' we mean the actual sequence.

The solution doesn't allow many possibilities. If m were zero, then every term after $y(0)$ is zero. If m is bigger than one, the sequence goes belting off, getting bigger and bigger. Try putting $m = 2$ in the sequence $y = m^n\ y(0)$. The lingo is: y grows without bound. Or sometimes 'y tends to infinity', which is misleading since it sounds as if infinity is some number which y gets closer to.

If m is 1, then every term is the same, namely $y(0)$. If m lies between 0 and 1 then the sequence gets closer to zero quite fast. Try putting $m = 1/2$ and watch things shrink.

If m is negative, then the sequence oscillates, getting closer to zero for m between 0 and -1, and getting wilder for $m < -1$. Try putting in the numbers to see what happens.

That says a fair amount about first order liner autonomous equations. They are pretty simple things, although they describe growth in interest rates fairly well. The fact that they are rather simple is a ***GOOD THING*** therefore. We can understand them and move on to more complicated things.

Exercise 3.2.1 *Draw graphs of all the cases for different values for m with $y(0) = 1$.*

The next most complicated is the first order autonomous affine equation:

$$y(n) = m\ y(n-1) + c$$

This came up in the Bush and Mosteller model for memory, and elsewhere. Let us look at the solution. I have done this three times already, but I shall do it again for completeness.

We write down the first few terms to see what is happening:

$$\begin{aligned} & y(0), my(0) + c, m(my(0) + c) + c + \cdots \\ = \quad & y(0), my(0) + c, m^2y(0) + mc + c, m(m^2y(0) + mc + c) + c + \cdots \\ = \quad & y(0), my(0) + c, m^2y(0) + mc + c, m^3y(0) + m^2c + mc + c, + \cdots \end{aligned}$$

It is fairly easy to guess:

$$y(n) = m^n y(0) + c(1 + m + m^2 + m^3 + \cdots m^{n-1})$$

You might wonder whether guessing is allowed. Yes. Anything is allowed at this point. You just have to nail it down carefully and show that your guess is right.

Writing the expression for $y(n)$ and the expression for $y(n-1)$ and dividing the first by the second we get

$$\frac{y(n)}{y(n-1)} = \frac{m^n y(0) + mc(1 + m + m^2 + \cdots + m^{n-1}) + c}{m^{n-1})y(0) + c(1 + m + m^2 + \cdots + m^{n-1})}$$

$$= m + \frac{c}{y(n-1)}$$

and multiplying both sides by $y(n-1)$ gives us the original difference equation. So the formula gives a solution, and by the same argument as before, the only one.

The string

$$(1 + m + m^2 + m^3 + \cdots m^{n-1})$$

can be written more compactly: if

$$S = (1 + m + m^2 + m^3 + \cdots m^{n-1})$$

then

$$mS = m + m^2 + m^3 + \cdots + m^n$$

and subtracting the first equation from the second:

$$(m-1)S = m^n - 1$$

and if $m \neq 1$ we can divide to get

$$S = \frac{m^n - 1}{m - 1}$$

If $m = 1$ then we just have to sum n numbers, all of them 1, which is n. So we can finish by writing down the equation

$$y(n) = m^n y(0) + c\,\frac{m^n - 1}{m - 1} \quad \text{if} m \neq 1$$

and

$$y(n) = y(0) + cn$$

if $m = 1$.

The next step, having done the algebra, is to get a clear idea of what it means. This entails drawing pictures. Some students think this is not really part of mathematics which is x's and y's and n's, but this is a serious error. The x's and y's and n's are just language, precise language it is true, but just language. What is important is what they mean. Anything we can do to understand what the stuff means is a ***GOOD THING***. Translating into a pictorial language is therefore a smart move.

The best way of working out what is going on here is to use a computer to plot some values and draw the graph. This means that we have to choose some values for m and c. Since the sorts of graph we get will depend on m and c, we can develop some understanding of what is going on in all the future models we may investigate.

A good graph drawing program can save you lots of time, but it is probably simpler to get MATLAB or Mathematica. Hours of fun guaranteed.

This is the simplest sort of linear equation. Play with it until you are happy that you can make it jump through hoops.

If you change the value of m to be $-1/2$ and c to be $1/2$ you will get a graph for $y(t)$ that settles down and gets closer to a certain fixed value.

Can you work out what the value is that it is settling down to?

Experiment suggests it is settling down to some number under 0.4. This value is clearly rather special. It is called an 'equilibrium point' or 'fixed point' of the difference equation.

To see what it has to be, write

$$y(n) = -\frac{1}{2}y(n-1) + \frac{1}{2}$$

This is the difference equation we have been plotting.

Now if it really settled down, we would have to have that $y(n) = y(n-1) = y$ so

$$y = -\frac{1}{2}y + \frac{1}{2}$$

or

$$\frac{3y}{2} = \frac{1}{2}$$

or

$$y = \frac{1}{3}$$

Change the starting point to different values to see what happens to different initial conditions. Put some negative starting points in to see what happens.

You will discover that there are three things that can happen from a given starting point: the graph can increase, decrease, or stay the same. For a really easy example, if we write

$$y(n) = 2y(n-1)$$

then it is easy to see that the solution is $y(n) = 2^n y(0)$. Now if $y(0)$ is positive, the solution goes zapping up, if it is negative it goes zapping down, and if it is zero, it stays zero for ever.

We say, again, that 0 is an *equilibrium point* or *fixed point* of this difference equation.

If we take the equation:

$$y(n) = \frac{1}{2}y(n-1) + 1;\ y(0) = 0$$

we get the sequence:

$$0, 1, 1.5, 1.75, \cdots$$

which looks as if it converges to 2. Plotting it on a computer certainly gives the impression that this is correct.

If we start off from $y(0) = -2$, we get

$$-2, 0, 1, 1.5, \cdots$$

and if we start from $y(0) = 4$ we get

$$4, 3, 2.5, 2.25, \cdots$$

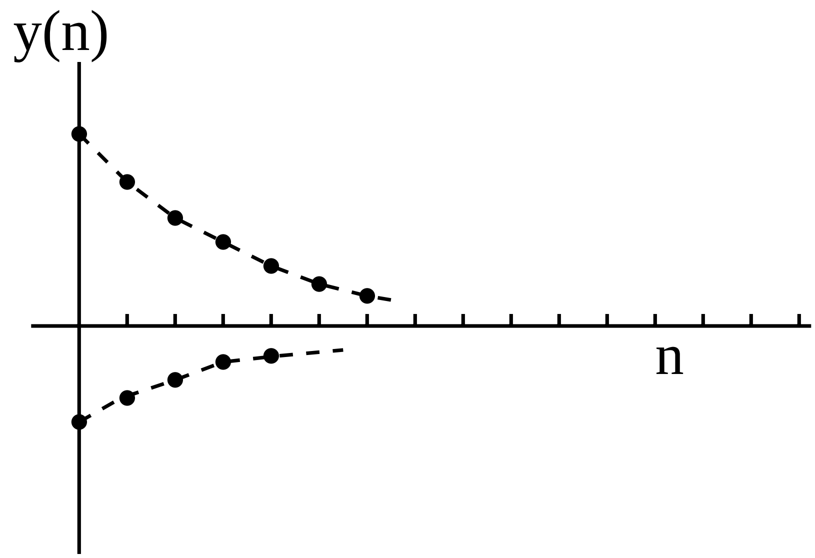


Figure 3.1: A stable equilibrium at the origin

which also looks as if it converges to 2. If we start off from $y(0) = 2$ then we stay there for ever. So 2 is an equilibrium point.

It is rather more than that, it is an equilibrium point having the property that if you start near it you get even closer. This is called a *stable* equilibrium. The example of the equilibrium point 0 for the equation

$$y(n) = 2y(n-1)$$

is clearly not stable, since if you start off close to it but not on it, you move further away from it. This is called an *unstable* equilibrium. The terms come from thinking of a ball rolling on a table: if the table is curved like a breakfast cereal bowl, the ball will stay at the bottom of the bowl if it is placed there, and if it is put nearby, it will roll down to the bottom of the bowl. If you put a ball on your head you might find it will stay there for a while, but if you move a little bit, it will fall off. This shows that there is an equilibrium point on top of your head but it is unstable.

Exercise 3.2.2 *Find a first order affine autonomous difference equation having zero as a stable fixed point, and another having 1 as an unstable*

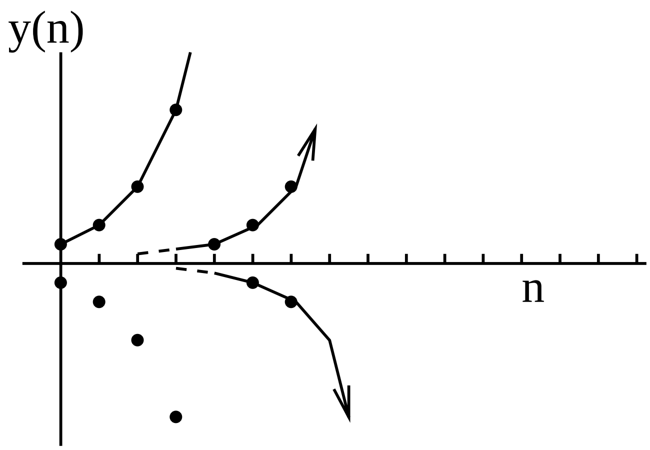


Figure 3.2: An unstable equilibrium at the origin

fixed point. Is it possible for a first order affine difference equation to have precisely two fixed points? More than two? None at all?

You should investigate the last exercise by writing down a general affine constant coefficient first order difference equation and asking what it would mean for $y(n) = y$ to be an equilibrium. Then you should find some cases where you can make 0 and 1 equilibrium points. Finally you should experiment to see what the effect is of changing the parameters in such a way as to keep the equilibrium point unchanged. I shall come back to this, but some mucking about with numbers will be very useful for you, and much more useful than just taking my word for things.

To see why the existence of fixed points is important, think of population growth: a stable fixed point means that the population tends to some stable limit if it gets near to the fixed point. It will not matter if your state is disturbed a little bit by some outside factor, it will recover its equilibrium. An unstable fixed point means that you rapidly get into one regime or the other, depending on which side you start. Things could get better or worse, but they won't stay the same!

Or think of a state of an economy; a stable equilibrium will survive some random wobbles, but an unstable one will leave the state of the system going somewhere else.

You should discover, if you investigate the possibility of more than one equilibrium, that for first order constant coefficient affine equations, either every point is fixed or none is or precisely one is. But other equations can have any number of fixed points, and in higher dimensions, more exotic sorts of things can happen, as we shall see.

Exercise 3.2.3 *You have investigated the stability of a fixed point of an affine first order autonomous difference equation. Now we ask the following question: what happens if we wobble the parameters a little bit? If y is a stable equilibrium for the equation $y(n) = my(n-1) + c$ and we change m a little bit, do we still have a stable equilibrium point close to y? Or can it disappear or become unstable? How about if we wobble c a little bit?*

A reason for worrying about this is that we could discover that if we wobble the parameters a bit, the qualitative features do not change. In this case, the model parameters may be measured by an experiment and it won't matter too much if we get them a little bit wrong. On the other hand, we might find that a small measurement error translates into something totally different.

Example 3.2.1

$$y(n) = 1.01y(n-1) - 0.01$$

has y as a fixed point if $y(n) = y$ for every n, in which case

$$y = 1.01y - 0.01 \Rightarrow y = 1$$

so 1 is a fixed point. If we start with $y(0) = 1.01$ we get

$$1.01, 1.0101, 1.010201, 1.010303, 1.010406,$$

and if we start with $y(0) = 0.99$ we get:

$$0.99, 0.9899, 0.989799, 0.98969699, \cdots$$

This gives the impression that 1 *is an unstable fixed point, which is indeed the case. Of course, we might worry a bit that if we had started off much closer to the fixed point we would have got closer again, so* 1 *is really stable after all. We shall therefore have to be a bit more careful than to just bung some numbers in as I just did, but bunging numbers in is often a good first move.*

Now what happens if we wobble the parameters 1.01 *and* -0.01 *a little bit? Say to* 0.99 *and* $+0.01$*. Well,*

$$y = 0.99y + 0.01 \Rightarrow y = 1$$

still, so we haven't moved the fixed point. But if we start with $y(0) = 1.01$ *we get:*

$$1.01, 1.0099, 1.009801, 1.009703, 1.009606, \cdots$$

which looks as though it is getting smaller and closer to the fixed point. And I leave it to you to see what happens if you start a bit below the fixed point.

So a 'perturbation' (the posh word for wobble) of the model parameters might give totally different behaviour. This could blow up any reliance whatever on long term behaviour of a system. If the earth is spiralling in towards the sun, buy shares in refrigerators and air-conditioning systems. If it is spiralling out, buy firewood and shares in the gas and electricity companies. But if it could be either, just gnaw on your knuckles and guess.

I shall come back to these ideas of equilibrium points and stability again later in a more general setting. It is obvious to the thinking man or woman that this is important and we need to know about these properties of a model.

There are some folk who imagine that computers have made mathematicians redundant; you don't need a human being to do the sums, you can get the computer to do them. Mathematicians wince a bit when they hear this sort of thing. The awful truth of what happens to people who put their innocent faith in computers ought to be clear enough by now. The belief that computers have made thinking about things unnecessary should strike you as rather unlikely. Computers can stuff things up to a degree unprecedented in human history if we do not use them sensibly. So mathematicians are needed far more, not less. And as far as solving equations is concerned, if you take a crude approximation to estimating something, and if the something is sensitive to the initial conditions, or the parameters, then you will probably get garbage out. People can run a huge program for months to get a 'yes' or 'no' answer to a question and discovered that their program was, in effect, a huge and inefficient random number generator. Tossing a coin would have been about as reliable and much cheaper. It would also not delude anybody into believing the answer means something just because a very expensive computer was used to obtain it. I have seen people do exactly this, although it would be tactless to say where.

3.2.2 Second Order Difference Equations

The only second order equations we have looked at so far are those that give rise to the Fibonacci sequences. This was first studied in connection with rabbits. Yes, that's right, rabbits. It is a variation on the population model for bacteria or people.

We assume that we are going to measure the number of rabbits, $R(n)$, at time n measured in months. We divide the rabbit population into two classes, the baby rabbits and the adult rabbits, the difference being that adult rabbits can go off into a quiet corner and produce more rabbits, while baby rabbits just play with their rattles and look cute. Let $B(n)$ be the number of baby rabbit pairs in month n, and $A(n)$ the total number of *pairs* of adult rabbits. Then

$$R(n) = 2(A(n) + B(n))$$

We assume that a pair of rabbits produces another pair of baby rabbits after a month. that after another month a baby rabbit turns into a grown up rabbit with a grown up rabbit's idea of fun. We also assume rabbits are immortal, which is the simplest first assumption.

Then the first assumption, that a pair of adult rabbits produces another pair after a month comes out as:

$$B(n + 1) = A(n)$$

And the assumption that the baby rabbits grow up and become adults in another month translates into algebra as:

$$A(n+1) = A(n) + B(n)$$

Now we can eliminate $B(n) = A(n-1)$ and get:

$$A(n+1) = A(n) + A(n-1)$$

or

$$A(n) = A(n-1) + A(n-2)$$

Now if $A(0) = 0$ and $A(1) = B(0) = 1$, which means we start off with just one pair of cute little bunny-woppets, we get

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \cdots$$

As the rabbit, sorry, *Fibonacci Sequence* with the two numbers given as initial conditions.

It is easy to draw some plots, and it is easy to see that within a short time we have a plague of rabbits.

Exercise 3.2.4 *How many rabbits at the end of one year? Two years? Five years? Suppose a rabbit lives four years and then expires with a smile on its face. How many offspring will have offsprung? Modify your equations to take into account the fact that the rabbits are not immortal.*

It would be satisfying to solve the equation and get an explicit formula for the Fibonacci Function. Can it be done? Yes, and it is rather a surprising one. And quite famous. You will see why when we get to it.

Instead of focussing on Fibonacci, we shall look more generally at the equations of the form:

$$ay(n+2) + by(n+1) + cy(n) = 0 \tag{3.3}$$

which certainly includes rabbits with $a = 1, b = c = -1$.

The differential equation

$$a\ y'' + b\ y' + cy = 0$$

is also second order. These two equations have a good deal to do with each other. You recall how we can approximate a differential equation with a difference equation, If we think of a difference operator applied to the sequence $y(n)$ as giving a new sequence

$$(\Delta y)(n) = y(n+1) - y(n)$$

and if we then apply this operator again, we can see a close parallel.

You may like to explore this idea and try to see why it is that the following works: be warned that it took some very smart guys to sort this out the first time! I shall come back to this when I discuss second order differential equations,

In order to solve the differential equation, we put $y = e^{kt}$ as a trial solution and tried to find out what k can be. Since there are generally two values, we get two possible solutions, and we take a mixture of them as our *General Solution.*

In order to solve the difference equation 3.3 we try a solution

$$y(n) = r^n$$

to see what happens. The question 'How on earth did you come to trying *that* and not something totally different?' is a good one which is not altogether easy to explain, I pass over it in mysterious silence.

If we do this with

$$ay(n+2) + by(n+1) + cy(n) = 0$$

we get

$$ar^{n+2} + br^{n+1} + cr^n = 0$$

and factoring out the r^n term we deduce

$$r^n(ar^2 + br + c) = 0$$

Now this holds for any n, so unless $r^n = 0$ we must have

$$ar^2 + br + c = 0$$

which has solution

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

This gives us, usually, two possible solutions, and it is not too hard to show that any mixture of these two solutions is also a solution. So the *general solution* is going to be

$$y(n) = C_1 \left(\frac{-b + \sqrt{b^2 - 4ac}}{2a} \right)^n + C_2 \left(\frac{-b + \sqrt{b^2 - 4ac}}{2a} \right)^n \tag{3.4}$$

The 'arbitrary constants', C_1, C_2 are fixed up to make sure that the difference equation agrees with its initial values.

Example 3.2.2 *For the case of the randy rabbits, we had* $a = 1, b = c = -1; y(0) = 0, y(1) = 1$.

So the solution is:

$$y(n) = C_1 (\frac{1 + \sqrt{5}}{2})^n + C_2 (\frac{1 - \sqrt{5}}{2})^n$$

And if $y(0) = 0, y(1) = 1$ *we have* $C_1 + C_2 = 0$, *and* $\frac{C_1 + C_2}{2} + \frac{(C_1 - C_2)\sqrt{5}}{2} = 1$. *From which we deduce that* $C_1 = 1/\sqrt{5}, C_2 = -1/\sqrt{5}$.

Or to put the final formula explicitly:

$$y(n) = \frac{1}{\sqrt{5}} (\frac{1 + \sqrt{5}}{2})^n - \frac{1}{\sqrt{5}} (\frac{1 - \sqrt{5}}{2})^n \tag{3.5}$$

This has to come as something of a surprise given that we know that the value of $y(n)$ *is always an integer. It doesn't seem likely that this is true of the formula. But it is!*

Exercise 3.2.5 *Try calculating the first four or five terms of the sequence from the above formula and directly from the difference equation.*

You can probably see that the interesting thing for the mathematician is not the cute little bunny-woppets, but the cute little formula 3.5. It is rather unexpected and we see it as rather cool. Of more practical value is the reasoning which led to it, because we can use that for solving many more problems.

In the example, we found two roots of a quadratic equation and there are two things that can go bung. The first is when the roots coincide. The second is when they are complex numbers, i.e. when the *discriminant*, $b^2 - 4ac$ is negative. Both of these lead us into territory which can be unexpected, and consequently interesting.

If two roots are the same, then it can be shown that there are still two independent solutions, one is r^n for some number r, and the other is nr^n for the same number r. An example will make this clear:

Example 3.2.3

$$y(n+2) = 2y(n+1) - y(n); y(0) = 0, y(1) = 1$$

Putting $y(n) = r^n$ *we get* $r^2 - 2r + 1 = 0$ *hence* $r = 1$. *This seems like the lot, but there is a solution* $y(n) = (1)(n) = n$. *The general solution therefore is* $C_1 n + C_2$ *since* $1^n = 1$. *Since* $y(0) = 0$ *we have* $C_2 = 0$ *and since* $y(1) = 1$ *we have* $C_1 = 1$. *The solution is therefore* $y(n) = n$.

Example 3.2.4 *For a more complicated case,*

$$y(n+2) = 4y(n+1) - 4y(n)$$

Putting $y(n) = r^n$ *and moving things around we get the* auxiliary equation

$$r^2 - 4r + 4 = 0$$

which is just $(r-2)^2 = 0$ *so* $r = 2$. *The general solution is therefore*

$$y(n) = C_1 n 2^n + C_2 2^n$$

Exercise 3.2.6 *Find the explicit solution for the above case when* $y(0) = y(1) = 1$.

And finally, what happens when the discriminant, $b^2 - 4ac$ is negative and we have an imaginary part of the solution? Well, the mysterious complex numbers have a great deal to do with running round in circles, and hence with the cos and sin functions. This is a frustrating remark for both of us, because if you are the lively minded person I hope you are, you will want to know *why* the complex numbers are all about these things. There is no obvious connection between a strange number the square of which is -1 and the circle, but there is. And it is maddening for both of us that I do not have time to explain it. Likewise, to understand why we suddenly stick an n in when the two roots of the auxiliary equation coincide is, for you, a weird fix. The explanation requires some linear algebra, and there is no time for it in this book. So you will have to live with a certain amount of mystery in your life, unless you do more Mathematics, when you might have it all made clear. If you are a genius, buy the complex function theory book also supplied by *HeavenForBooks* for an explanation of why $e^{i\pi} + 1 = 0$.

To give the rule, deMoivre's theorem states:

$$(r(\cos\theta + i\sin\theta))^n = r^n(\cos n\theta + i\sin n\theta)$$

Now we get solutions

$$\frac{-b + i\sqrt{4ac - b^2}}{2a}, \frac{-b - i\sqrt{4ac - b^2}}{2a}$$

which we write in polars (as conjugate solutions) to give a general solution:

$$y(n) = C_1 r^n(\cos n\theta + i\sin n\theta) + C_2 r^n(\cos n\theta - i\sin n\theta)$$

where r and θ are easily calculated from a, b, c. This simplifies to

$$y(n) = Ar^n \cos n\theta \ + Br^n \sin n\theta$$

as the general solution.

Exercise 3.2.7 *Check the argument through for yourself. If your algebra gives you trouble, try Dr. Math.*

Example 3.2.5

$$y(n+2) = 2y(n+1) - 2y(n);\ y(0) = 0, y(1) = 1$$

Transforms to

$$y(n+2) - 2y(n+1) + 2y(n) = 0$$

Putting $y(n) = r^n$ we get the auxiliary equation

$$r^2 - 2r + 2 = 0$$

The solutions are

$$\frac{2 + i\sqrt{4}}{2}, \frac{2 - i\sqrt{4}}{2}$$

These come out as $1+i$ and $1-i$ which translate into polars as $\sqrt{2}\cos(\pi/4) \pm \sin(\pi/4)$ which means the general solution is

$$y(n) = A(\sqrt{2})^n \cos(n\pi/4) + B(\sqrt{2})^n \sin(n\pi/4)$$

If $y(0) = 0$ then $A = 0$ and if $y(1) = 1$ we have $B\sqrt{2}/\sqrt{2} = 1$ so $B = 1$. Thus the solution is

$$y(n) = (\sqrt{2})^n \sin(n\pi/4)$$

Exercise 3.2.8 *Check by calculating the first 8 values by hand. A sequence like this would represent a fairly catastrophic sort of system. Can you imagine any physical or biological system which might behave in a similar way?*

Exercise 3.2.9 *Can you find a more soothing sort of difference equation where the oscillations die down in size (amplitude) instead of getting bigger?*

The equations of second order I have looked at have been linear, constant coefficient equations,

$$y(n+2) = \alpha y(n-1) + \beta y(n) \tag{3.6}$$

although I have moved things around a bit so as to get a recognisable quadratic equation out. According to the general philosophy of making things more complicated but only by a little bit at a time, we shall now go on to *affine* equations:

$$y(n+2) = \alpha y(n+1) + \beta y(n) + \gamma \tag{3.7}$$

We tackle this by first solving the linear part. We have already dealt with the three cases, discriminant positive, zero and negative. I shall write this out as if the discriminant were positive, the method is the same as before in the other two cases.

Suppose that we have

$$y(n) = C_1 r^n + C_2 s^n$$

is the solution to the case where $\gamma = 0$, which is called the *homogeneous equation*. I shall not expect you to remember the jargon, but I tell you this in case you want to check this out in other books.

Now suppose there is some constant C_3 such that

$$y(n) = C_1 r^n + C_2 s^n + C_3$$

is the solution to 3.7. Then $y(n+1) = C_1 r^{n+1} + C_2 s^{n+1} + C_3$, and $y(n+2) = C_1 r^{n+2} + C_2 s^{n+2} + C_3$. Substituting in 3.7 we get

$$C_1 r^{n+2} + C_2 s^{n+2} + C_3 = \alpha(C_1 r^{n+1} + C_2 s^{n+1} + C_3) + \beta(C_1 r^n + C_2 s^n + C_3) + \gamma$$

Now we can rewrite this to

$$(C_1 r^{n+2} + C_1 s^{n+2}) - \alpha(C_1 r^{n+1} + C_2 s^{n+1}) - \beta(C_1 r^n + C_2 s^n) = \alpha C_3 + \beta C_3 - C_3 + \gamma$$

But the left hand side of this equation is zero, since we fixed up the solution so that it would be. This gives:

$$C_3(\alpha + \beta - 1) + \gamma = 0$$

or

$$C_3 = \frac{\gamma}{1 - \alpha - \beta}$$

So the general solution to equation 3.7 is

$$y(n) = C_1 r^n + C_2 s^n + \frac{\gamma}{1 - \alpha - \beta}$$

An example will make this easy to follow:

Example 3.2.6 *Solve:*

$$y(n+2) - 2y(n+1) + 4y(n) = 5; \quad y(0) = 1,\ y(1) = 0 \tag{3.8}$$

Trying a solution of the form $y(n) = r^n$ for the homogeneous equation, we have the auxiliary equation

$$r^2 - 2r + 4 = 0$$

which has roots $1 + 1\sqrt{3}$, $1 - i\sqrt{3}$. Putting this into polars we get

$$2(\cos \pi/3 + i \sin \pi/3), \quad 2(\cos \pi/3 - i \sin \pi/3)$$

The general solution (to the homogeneous equation) is therefore

$$y(n) = 2^n(C_1 \cos(n\pi/3) + C_2 \sin(n\pi/3))$$

Trying now a solution to the original equation 3.8 of the form

$$y(n) = 2^n(C_1 \cos(n\pi/3) + C_2 \sin(n\pi/3)) + A$$

we get

$$\begin{aligned}
& 2^{n+2}(C_1 \cos((n+2)\pi/3) + C_2 \sin((n+2)\pi/3)) \\
+ \ & A - 2\left(2^{n+1}(C_1 \cos((n+1)\pi/3) + C_2 \sin((n+1)\pi/3)) + A\right) \\
+ \ & 4\left(2^n(C_1 \cos((n)\pi/3) + C_2 \sin((n)\pi/3)) + A\right) \\
= \ & 5
\end{aligned}$$

And we can separate this into

$$\begin{aligned}
& 2^{n+2}(C_1 \cos((n+2)\pi/3) + C_2 \sin((n+2)\pi/3) \\
- \ & 2\left(2^{n+1}(C_1 \cos((n+1)\pi/3) + C_2 \sin((n+1)\pi/3)\right) \\
+ \ & 4\left(2^n(C_1 \cos((n)\pi/3) + C_2 \sin((n)\pi/3)\right) + A - 2A + 4A \\
= \ & 5
\end{aligned}$$

Since all the complicated bits go to zero, we get

$$A - 2A + 4A = 5$$

or

$$A = 5/3$$

We therefore have the general solution to equation 3.8 is

$$y(n) = 2^n(C_1 \cos(n\pi/3) + C_2 \sin(n\pi/3)) + 5/3$$

Putting $y(0) = C_1 + 5/3 = 1$ *we get* $C_1 = -2/3$ *and putting* $y(1) = 2C_1 \cos(\pi/3) + 2C_2 \sin(\pi/3) = 0$ *we get* $C_2 = -1/\sqrt{3}$. *So*

$$y(n) = (-2/3)2^n \cos((n+2)\pi/3) - (1/\sqrt{3})2^n \sin((n+2)\pi/3) + 5/3$$

The others are all essentially the same, but some practice will get you able to do these problems quickly and efficiently.

The reason for doing these general problems of second order linear and affine constant coefficient difference equations is that once you can rattle them off fast, you can focus on the interesting part, which is setting up the equations for some real problems.

It has to be said that these are only the tip of the iceberg. The ways in which one can complicate equations is considerable. And there are scads of mathematical tricks for solving some of them in closed form. Some old fashioned courses would assume that students were prepared to knuckle down and learn these techniques in the hope that someday they would come in useful. I doubt if we can expect that degree of dedication and trust in these days, at least in the English speaking world. In Japan, Korea and Singapore, they still do it this way. So they are technically way ahead of the West, having been doing this sort of thing since school; but a deeper feel for why you need the techniques has its advantages too.

Chapter 4

Iterates of maps: Stability

In the last chapter we discussed the first and second order linear and affine autonomous difference equations. As remarked, this is only a small part of what mathematicians have learnt over the years, but this is a course on modelling, not on difference or differential equations as such. And it represents a healthy start. Quite a lot of interesting systems can be modelled to a useful extent by such equations.

One of the features which starts to emerge is the question of equilibrium points, and the question of whether they are stable or unstable. It must be clear to you that this is a matter of very considerable practical importance; if we are studying a system, perhaps a biological one involving the functioning of an organ in the body, or the concentrations of reagents in a cell, perhaps an economic one involving inflation, the difference between a system which runs wild and one which settles down is usually more important than the details of just how long it takes to do either of these things. This is the *qualitative theory* of difference or differential equations, and it complements the numerical solution of equations in an important way. Again, there is a huge amount known about these matters, and we shall only skip over the surface. The subject was started seriously in the last years of the last century, so it goes back about one hundred years, and has evolved into Topology sometimes known as Rubber-Sheet Geometry, (which turns up everywhere these days). I tell you this sort of thing because there are more important things in life than exams, and getting some sort of perspective on which way things are going is one of them.

You can think of stability as studying the question of what happens if you wobble things a little bit. It sounds classier if we talk of the effects of perturbations on systems, but it means the same thing. In this chapter we shall look at stability in the simplest cases and get a clearer picture of what to look for and what kinds of things can happen.

4.1 Cobwebs and Chaos

Going back to first order equations, when I ask you to calculate the first ten terms of the sequence which solves

$$y(n+1) = my(n) + c; \quad y(0) = d$$

what you did was to start off by putting $n = 0$ to get $y(1) = md + c$. Then you took the answer to this calculation and fed it back into the same process to get out the next value.

A useful way to look at this is to think of the function

$$y = mx + c$$

and to put in a value for x, $y(0)$ and see it as outputting $y = y(1)$. Then we get the next value by feeding the output back to the input. The stream of numbers we get out is the solution sequence. This is, of course, a very easy thing to put on a computer.

This works with other first order difference equations too; if we had the non-linear logistic approximation equation

$$y(n+1) = y(n) + my(n)(a - y(n))\Delta$$

we could just take the function

$$f(x) = x + m(x)(a - x)\Delta$$

and again feed the output back into the input. Thinking of a function as an input-output device is very useful, and thinking of this feedback process as generating a sequence of numbers is also very useful.

There is a graphical way of doing this called a *cobweb diagram* which has been much used by economists, and we shall see why shortly. Let me explain the idea of a cobweb diagram for the simplest sort of equation first.

The difference equation I shall look at first is

$$y(n+1) = 3y(n)(1-y(n)); \quad y(0) = 1/2 \tag{4.1}$$

If we start off by writing down the *iteration map*

$$y = 3(x)(1-x)$$

and initialise our sequence generator with $x = 1/2$, we get

$$1/2, 3/4, 9/16, 189/256, \cdots$$

and it is hard to see any particular structure to what is going on.

If we run this on a suitable computer program, we see that the system oscillates and seems to settle down to about 2/3 which looks as though it could be an equilibrium point. If so it is a *fixed point* of the iteration map; we simply say that

$$x = 3x(1-x)$$

when x is a fixed point and solve to get

$$x = 3x - 3x^2$$

or $3x^2 = 2x$ giving $x = 0$ and $x = 2/3$. So there are exactly two fixed points of this function.

We can get a better picture of what is going on by drawing the graph of $y = 3x(1-x)$ and looking to see where it cuts the graph of $y = x$.

We get the cobweb by starting at input 0.5. We go up to the curve which is the graph of the iteration function. This gives the output on the vertical

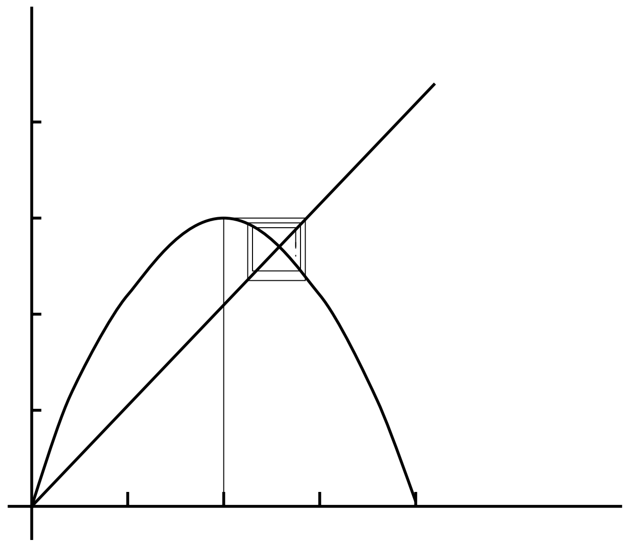


Figure 4.1: A cobweb plot

axis. To turn it back into an input we move horizontally until we hit the line $y = x$. Now we move vertically until we hit the curve, and so on. It is now obvious from the diagram that 2/3 is a stable fixed point of the iteration. You can see the 'moving point' homing in by a sort of square spiral towards the fixed point.

If we do a numerical plot of the solution to the differential equation, we get something like figure 4.2

With a bit of thought you can see the relation between the two plots: the oscillations in figure 4.2 as it settles down to 2/3 are the projections on the Y-axis of the points on the cobweb.

Many students want to know the one true way of thinking about things, and if Mathematics has any generally important thing to say, it is that there IS no one right way. On the contrary, the more ways of looking at something the better. Be faithful to your friends, be faithful to your principles, but never insist on staying faithful to your point of view:

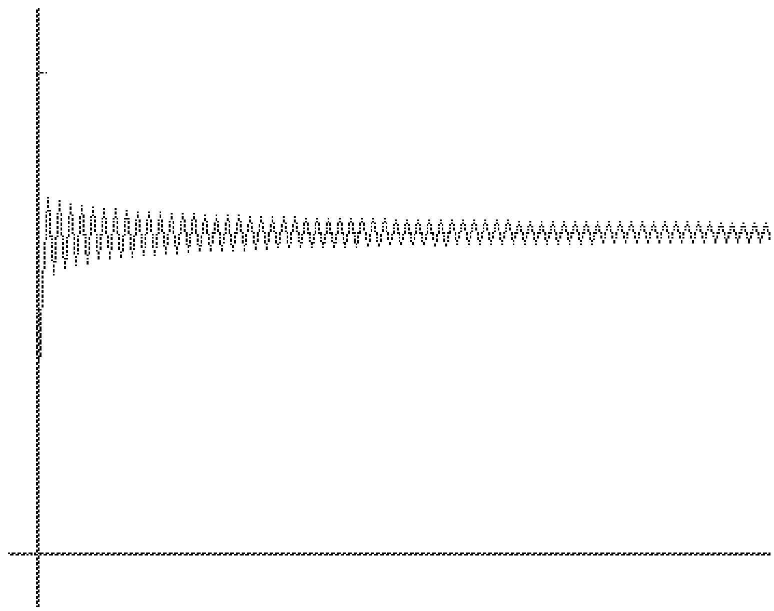


Figure 4.2: A *diffeqn* plot

> "There are nine and sixty ways of constructing tribal lays,
> "And every single one of them is right!"

So get to see the cobweb plot *and* the graph of the time series as both good ways of looking at the process.

By contrast, if we use the iteration function

$$y = 4x(1-x); \quad y(0) = 0.5$$

the cobweb plot go to zero and stays there if we start at 0.5, in two moves;

$$0.5, 1.0, 0.0, 0.0, 0.0, \cdots$$

If we start anywhere else, it seems to go all over the place. The cobweb plot for starting at about 0.45 is shown in figure 4.3.

If we look at the *diffeqn* plot, we get something totally unexpected if you compare it with the earlier difference equation (which had a three instead of a four), but which makes sense from the cobweb diagram. It looks rather

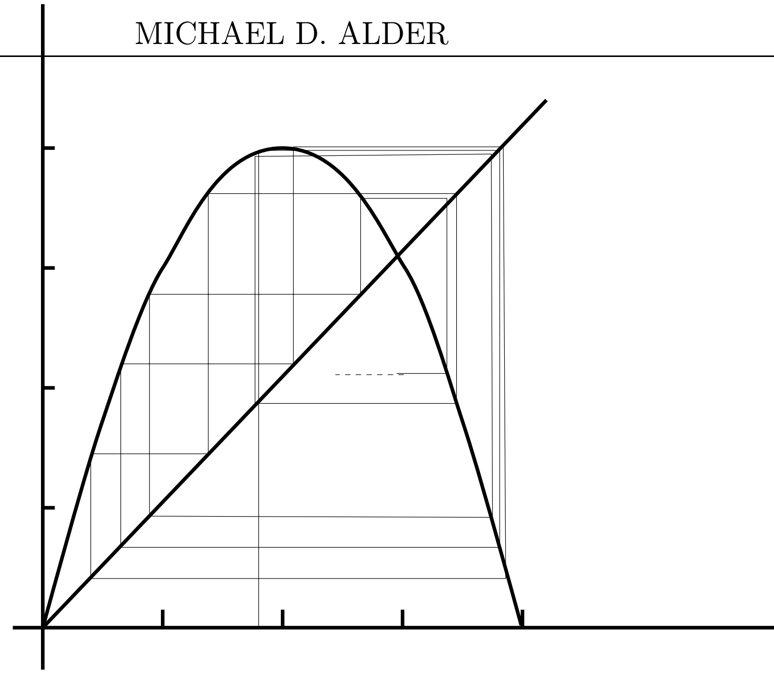


Figure 4.3: A cobweb plot for chaos

like a speech waveform, or a random 'noise' wave. If we start off at 0.25, it jumps quickly to 0.75 and stays there. But if we start at 0.24, it does a random sort of output.

This is the phenomenon of *Chaos* of which you may have heard.

In figure 4.4 we start off at about 1/3 and immediately generate what looks like a random signal. In figure 4.5 we start off at 0.249995 which is very close to 0.25. On the next iteration we move very close to the fixed point at 0.75. It stays there for a while but gradually slips back into a random pattern. So it seems clear that 0.75 is not a stable fixed point.

Chaos is very much about pseudo-randomness. It looks awfully random, but in fact is generated by a very simple process. This means it can easily happen in the world we live in. It is virtually certain that the turbulence you get in water in rapids and waterfalls is a chaotic phenomenon, as is much of our weather at certain scales.

Exercise 4.1.1 *Draw the graph of $f(x) = 4x(1-x)$ on a BIG piece of graph*

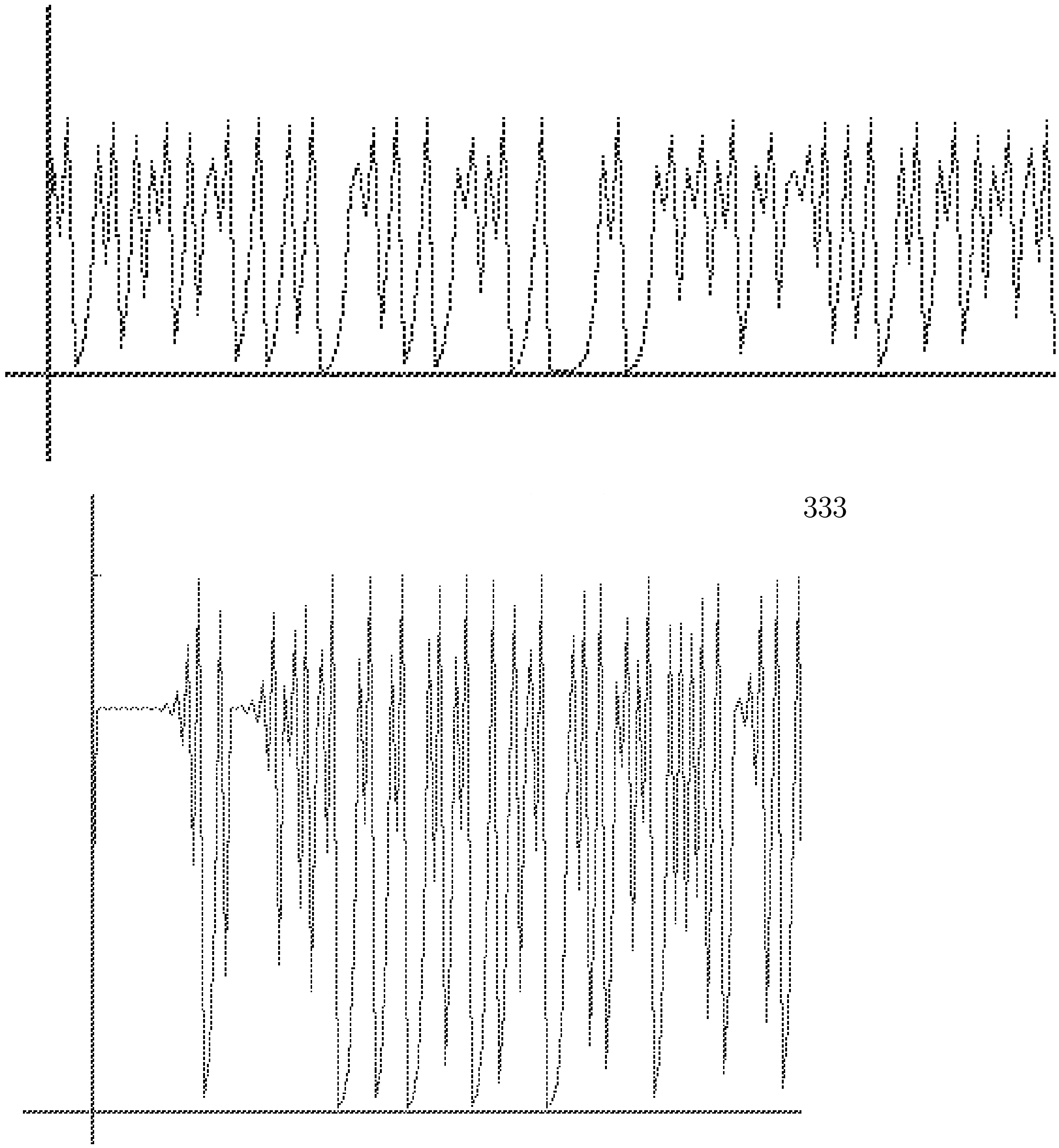


Figure 4.5: Starting at 0.249995

paper. Draw it only for the unit interval, where $f(x) \geq 0$*. Draw also the line* $y = x$*. Note that it cuts the curve at* $x = y = 3/4$*.*

Take a random start on the interval at the bottom, go up to the curve, vertically, then across to the straight line horizontally. Keep doing this, drawing the cobweb diagram, until you get totally fed up. Do you hit either of the two fixed points ever?

This should give you reason to stop and think rather hard. Suppose you came across a model like this in studying economics, or population dynamics, or any of another million things. You seem to be getting random numbers out. And such a simple equation, as well! Could this happen in real life? The answer is 'yes'. It can and probably does, rather a lot.

There seems to be a big difference between 3 and 4 too. Presumably at some magic number between three and four the process turns into a random number generator. Incidentally, this is more or less how random number generators are made.

Example 4.1.1 (Cobwebs) *This is an important economic model which everyone needs to understand so you don't let the economists snow you. All educated people need to get an idea of what is going on here, and fortunately, when stripped of jargon, it is quite easy to understand.*

Suppose we have what the economists call a commodity *such as wheat, something some blokes produce and other blokes consume. (The term 'bloke' is supposed to be sexually inclusive; there are lady farmers and quite a lot of lady shoppers, including some who buy wheat for companies that make bread. But saying 'bloke or sheila' all the time is a bit of a drag. This is an Australian book, so it is Bloke/Sheila or just bloke.)*

The farmers, in the case of wheat, grow a quantity of the stuff for sale one year on. They don't want to grow too much in case they can't sell it, and they don't want to grow too little in case they don't make as much money as they could. If they got a good price in year n*, then the average farmer (according to the average economist) is dumb enough to just decide to grow lots in year* n *so as to make a load of money in year* $n+1$*. (A little thought by the farmer* *would lead him to conclude that this is what everyone else will think, so the smart farmer won't go this way, but let us pick a thick farmer. There may or may not be any, but we can imagine one.)*

So loads of wheat is produced in year $n+1$*. Since the demand for wheat has not gone up as much as the supply of wheat, the price goes down. After all, there is scads of wheat and all those farmers want you to buy theirs; how else except by offering you a bargain?*

The discouraged farmers, left with wheat on their hands or a lousy price conclude that wheat is not such a good idea after all and they all plant cabbages or something. Result, next year hardly any wheat and the price goes through the roof, and the cycle starts again. And the cabbages are the same as the wheat but one year out of phase.

Only mathematicians of a certain academic fatheadedness, and economists of a similar stripe would expect farmers to be so dumb. Real farmers have always been several IQ points above this, and the model is hardly to be taken seriously. Still, let us turn it into algebra to see if anything can be done with it. There may be situations in which this rather stupid behaviour happens, maybe when these economists and academic mathematicians take up farming. It should also be pointed out that most economists and mathematicians are not this dumb. But it makes a good story.

Let $p(n)$ *be the price of wheat in year* n*. Let* $S(n)$ *be the quantity of wheat produced by farmers in year* n*, and* $D(n)$ *be the amount demanded by the wheat eating public in year* n*. The terms 'Supply' and 'Demand' are traditional economist's jargon.*

We make a number of assumptions:

1. *The demand,* $D(n)$ *is a continuous function of* $p(n)$*. Moreover, if* p *increases,* D *decreases.*

2. *The Supply* $S(n)$ *is a continuous function of* $p(n-1)$*. Moreover if* p *increases,* S *increases.*

3. *The price is determined by the Supply and the Demand and nothing else. More particularly, the price is adapted until the supply and demand are*

equal.

To examine these assumptions one at a time, it is clear that for most things people buy, as the price increases people will look for substitutes or abandon the commodity as a hopeless luxury. If beer is taxed enough, we turn to wine. Or, in an extreme case, clean living. It is also clear that nobody would bother to spend $10,000 for a watch just to tell the time, and if Rolex dropped the price to $100 a throw, hardly anyone would bother to buy one. So there are cases when things work backwards, but they don't happen very often. With wheat, it seems safe to say that the higher the fewer. The higher the price, the fewer bushels will be bought.

The situation with the second assumption is rather more complicated. The idea that I shall look at this years high price and say 'Wow, lotsa moolah, do it again next year only BIGTIME!', credits me with rather a low intelligence. It could work for some commodities, as when I am the only supplier, but a thinking man would reason that everyone else would think this, so the thing to do is to go for cabbages not despite but because *cabbages did badly this year. Then of course you have to bear in mind that all the thinking men will think what you just thunk. So maybe it's wheat after all. But they can think of that too! Only most will have given up by now and taken a cold shower to stop their heads spinning.*

Anyway, it is decidedly iffy, and will need attention should we decide to jack up the realism of the model.

The final assumption is also a bit suspect. What it is saying is that if the last farmer is left with a hundred bags of wheat and everybody is carting huge quantities off in the car, he will offer it in exchange for whatever he can get rather than have to pay for the trouble of burning it. He, or conceivably she, will give it away for a smile and a handful of loose change. And knowing this will keep up competition to cut the price until each farmer gets shot of the lot, but not so fast that he leaves someone else to make money by selling at a higher price. You can see the price of Vegies at the markets doing this sort of thing on a Sunday afternoon. Well, taking your wheat home for fuel might not be altogether a bad thing in all circumstances, so there is a strong element of 'averaged out behaviour' and of simplifications. But we are used to that in Physical models, so it should not worry us unduly.

In order to turn this into algebra, I need to make the assumptions even stronger. I shall model the supply and demand functions as affine *functions.*

So suppose:

$$D(n) = -mp(n) + c \tag{4.2}$$

and

$$S(n) = qp(n-1) + b \tag{4.3}$$

where m, q are positive constants. We may also assume that $c > 0$ since otherwise customers would be fighting to get rid of the stuff in the case when the price is zero. There are commodities like this (e.g. a punch in the nose), but we can reasonably suppose that wheat is not one of them. We may be unsure whether $b \geq 0$. A linear model is not flagrantly stupider than an affine one, and the situation where the supply might be zero even for a small positive price cannot be ruled out. Some books do rule it out, but they don't trouble to try to convince you that they aren't pulling numbers out of their ears.

The third condition is:

$$S(n) = D(n) \tag{4.4}$$

The above equations give, rather easily,

$$-mp(n) + c = qp(n-1) + b$$

or

$$p(n+1) = -\frac{q}{m}p(n) + \frac{c-b}{m}$$

We can explore the issue of a fixed point for the price, otherwise an equilibrium price *by the usual means: put $p(n+1) = p(n) = p$ to get*

$$p = -\frac{q}{m}p(n) + \frac{c-b}{m} \Rightarrow p = \frac{c-b}{m+q}$$

This is well defined, and positive so long as $c > b$.

We can investigate the stability of this price. This is answering the question: 'Is there some price so that if we get to it we stay at it, and if we are near to it, do we get closer or further away?'

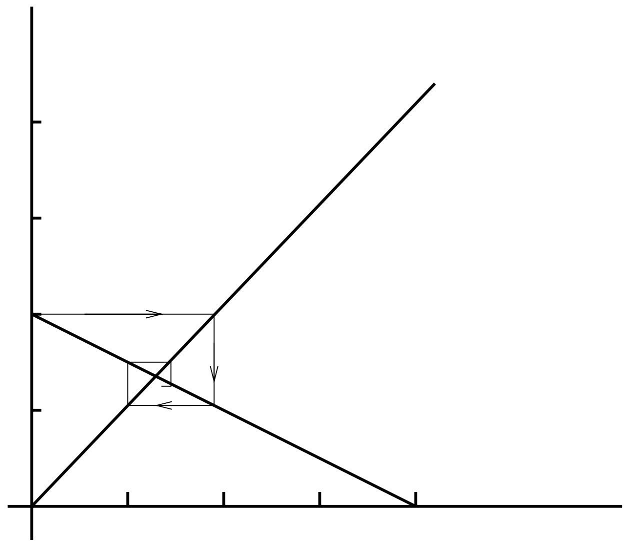


Figure 4.6: The Price of Wheat

We do this by the mathematicians cobweb first: we draw the graph of

$$y = -\frac{q}{m}x + \frac{c-b}{m}$$

and the graph of $y = x$ *and start off a cobweb somewhere and see what happens.*

In figure 4.6 I have drawn $y = -1/2x + 2$ *and started the cobweb from zero to get the sequence:*

$$0, 2, 1, 1.5, 1.25, 1.375, \cdots$$

It certainly would seem that the unique fixed point of the iteration map gives a stable equilibrium in this case. It is easy to see that the fixed point is at 4/3.

Now suppose we change the parameters to get the equation:

$$y = -2x + 4$$

If we start off with zero again, we get

$$0, 4, -4, 12, -20, \cdots.$$

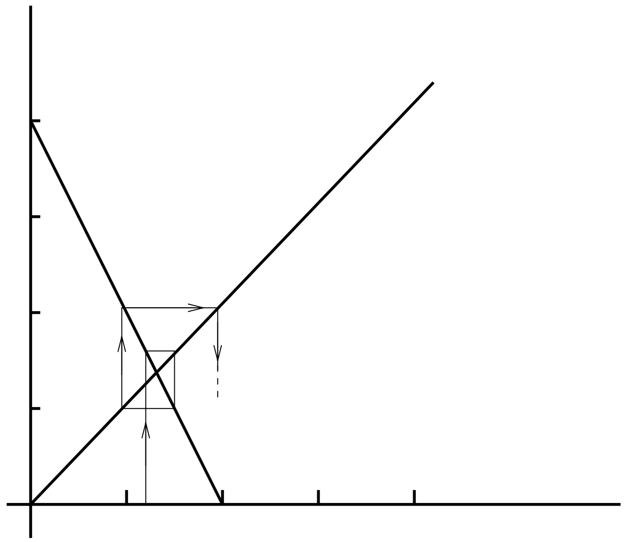


Figure 4.7: The unstable Price of Wheat

There is still a unique fixed point at $4/3$*, but it looks to be unstable. If we start at* 1 *we get*

$$1, 2, 0, 4, \cdots$$

and if we start at $5/4$ *we get*

$$5/4, 6/4, 1, 2, \cdots$$

If we draw a cobweb diagram we get figure 4.7.

It is clear enough from the diagrams that we have an unstable situation, and the price goes wild.

The economists draw slightly different diagrams, but that is not important. They draw the Demand graph and the Supply graph on the same scale, and go up from an initial price to the supply line. Then they go horizontally until they hit the demand line. Then vertically to the supply, horizontally to the demand, and so on until they either converge or fall off the graph paper.

It is obviously of some interest to know when the fixed point of an iteration map is stable. The idea here is very simple and usually made rather opaque, so I shall explain it as a Topologist sees it, and hope that you are an intuitive, geometrically minded reader.

You are used to thinking of functions in terms of their graphs. This is so useful that you have never really learnt any other way of thinking. Which is a pity, because there are some cool alternatives.

I shall offer you another way. Imagine a very thin linear ruler made of chewing gume or something stretchy. Imagine a second ruler made of wood or something else which is rigid. Suppose both rulers have all the numbers marked on them. Now pick up the first, stretchy, ruler, stretch it or twist it, tie knots in it, but do not tear it. You can do pretty much what you like with it, since it is only an imaginary ruler anyway. Now finally, put it down on top of the wooden ruler, so that every number on the chewing gum ruler is on top of some number on the wooden ruler. Then the function which takes chewing gum numbers to wooden numbers has been turned into a process of twisting and stretching.

Example 4.1.2 *The function $y = 2x$ just stretches the line uniformly by a factor of two. The diagram 4.8, shows a picture of the process.*

Exercise 4.1.2 *Draw a picture similar to figure 4.8 for the functions (a) $y = |x|$, (b) $y = -x$, (c) $y = 2x - 1$, (d) $y = x^2$*

You will note that the inverse operation is the stretching operation run backwards, and that if you do the function and follow it by its inverse, each point goes back to where it started.

Now what we are doing with iterating a function, is doing it, and then doing it again, and taking some initial point and looking to see what happens to it.

The functions of the form $y = mx + c$ just do a stretching by m and a shift by c. If $0 < m < 1$, then the compression is a shrinking. (Look at $y = x/2$ which is the inverse to $y = 2x$.) If m is negative, then there is also a reflection in the origin; the line gets twisted around. So it is clear that if $-1 < m < 1$

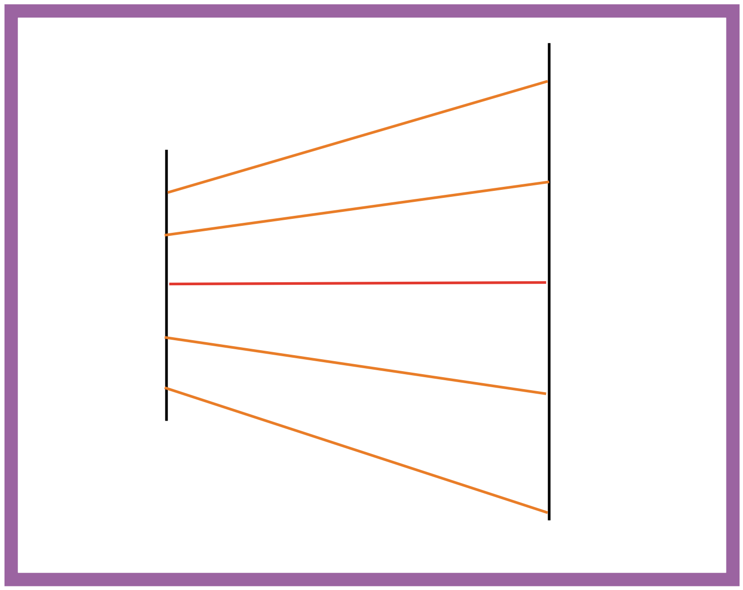


Figure 4.8: Functions as operations: y = 2x

then the map is squashing things in, while if $|m| > 1$ then it is stretching points further apart.

For this reason we have the following result, which is a bit too obvious to really be called a theorem:

Proposition 4.1.1 *A fixed point of the affine map $y = mx + c$ is* stable *if $|m| < 1$, and unstable if $|m| > 1$.*

Proof

The distance between $y_1 = mx_1 + c$ and $y_2 = mx_2 + c$ is $|m(x_1 - x_2)|$. If $|m| < 1$ this is less than the distance between x_1 and x_2, so the function moves points together. So any point must be moved closer to the fixed point. Conversely if $|m| > 1$ any point must be moved away from a fixed point, so the fixed point must be unstable.

If the function is not affine but is differentiable, we have another trick up our sleeves. The idea here is to define the amount of stretching that the function,

f does at x. I shall call this $S(f,x)$. The idea is simple enough: I take a little piece of the chewing gum ruler starting at x and going up to $x+\Delta$ for some little distance Δ. Then I look to see where f takes x and where it takes $x+\Delta$. The first guess at the stretch factor of f at x is the ratio

$$\frac{f(x+\Delta)-f(x)}{\Delta}$$

The only thing wrong with this is that most of the stretch may have occurred somewhere between x and $x+\Delta$. So I make Δ smaller and get another estimate. In the limit, if there is one, I get $S(f,x)$, the amount of stretch that f does to the chewing gum at the point x. The formula for it, which may ring a bell, is:

$$S(f,x) = \lim_{\Delta\to 0}\frac{f(x+\Delta)-f(x)}{\Delta}$$

To see whether f is stable at x if x should be a fixed point of f, all I have to do is to calculate this stretching factor. If the absolute value of the stretching factor is less than one, it is stable. This is exactly the same idea as for the affine case, but I might have to look inside very small regions around x. In the affine case, $y = mx+c$, the stretching factor of this function at x is the same no matter where you are, and is just m.

So all we need to do is to calculate $S(f,x)$ and test its absolute value. But you must surely have realised that $S(f,x)$ is just dy/dx, the derivative. So we get the following:

Proposition 4.1.2 *If f is a differentiable function with a fixed point at x then it is a stable fixed point if $|f'(x)| < 1$ and an unstable fixed point if $|f'(x)| > 1$.*

No Proof

In the case where the 'stretch factor' is precisely 1, we have no information, and it may be that it is stable. This can happen when the stretch factor is a maximum at the fixed point, and less in some neighbourhood. Or it can be unstable because the stretch factor is greater than one in a neighbourhood.

Exercise 4.1.3 *The demand equation for a commodity is given by*

$$D(n) = -3p(n) + 2$$

and the supply equation is

$$S(n+1) = 4p(n) + 1$$

Find the equilibrium price and determine if it is stable or not.

If an initial price of $1/10$ *is negotiated in year* 0*, calculate the price in years 1,2,3,4 and 5.*

Take a big piece of graph-paper and draw the graphs of the demand function $D(n)$ *against the price, and the supply function* $S(n+1)$ *against the price, both on the same graph. Starting at* $p = 1/10$*, go up to the supply graph, then go horizontally to the demand graph, then vertically to the supply, and repeat to get the classical economist's cobweb.*

Rewrite the two equations into a difference equation in p and starting from the same point, generate the mathematician's cobweb. Explain what you find.

Make a list of every single assumption that you use.

Example 4.1.3 *For the system*

$$y = 3x(1-x)$$

determine all fixed points and determine if they are stable

Solution

For a fixed point,

$$x = 3x - 3x^2 \Rightarrow 2x = 3x^2 \Rightarrow x = 0 \text{ or } x = 2/3$$

The derivative of y *at* x *is* $dy/dx = 3 - 6x$ *so*

$$\frac{dy}{dx}|_0 = 3 > 1$$

so 0 *is an unstable fixed point, and*

$$\frac{dy}{dx}|_{2/3} = -1$$

There is no information provided by the test in this case, although in fact as we know it is stable.

Exercise 4.1.4 *See if you can reason your way to showing that the point* 2/3 *in the last example is stable. Hint: if the stretching factor on one side is less than one, and if the stretching factor on the other side folds the region over, then you can hope to do it by looking at the result of doing the function twice. Call the result of doing it twice* g. *Then*

$$g(x) = 3f(x)(1 - f(x)) = 3(f(x) - (f(x))^2)$$

and substituting for $f(x) = 3x(1 - x)$ *we get an explicit form for* g. *Now testing* $g'(2/3)$ *shows that we still get* 1. *However it is easy to confirm that this is a local maximum value of* g', *and that in a neighbourhood of* 2/3 *it is always less.*

Exercise 4.1.5 *Examine the fixed points of*

$$f(x) = 4x(1 - x)$$

to determine if they are stable or unstable.

You should easily establish that there are two fixed points, both unstable. If you think about what kinds of things can happen in the case of precisely two unstable fixed points, you will begin to see that chaos is almost inevitable.

Exercise 4.1.6 *Do the same with*

$$f(x) = 2x(1 - x)$$

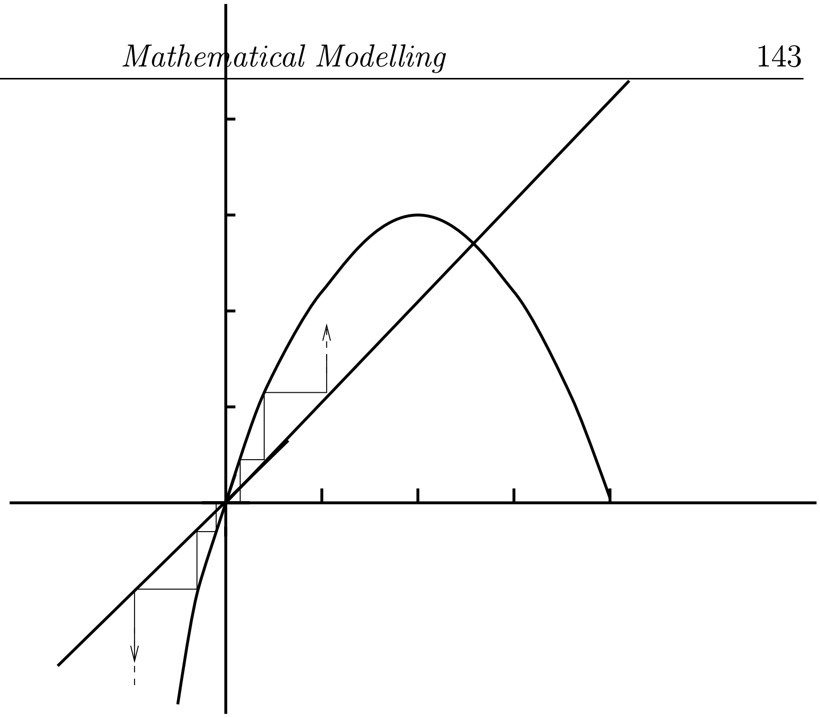


Figure 4.9: An unstable equilibrium

4.2 More about Stability

We can get a convincing picture of an unstable equilibrium by doing cobweb plots.

Example 4.2.1 *For the difference equation*

$$y(n+1) = 3y(n)(1 - y(n))$$

Do a cobweb plot near the origin and show that zero is an unstable fixed point of the iteration map.

Solution *We get results that look like figure 4.9; it is easy to see that the two starting points must indeed lead to the points moving further from the origin.*

In the above case, we had $f'(0) > 1$ so of course we expect an unstable fixed point. Conversely, we get a stable fixed point at x when $0 < f'(x) < 1$:

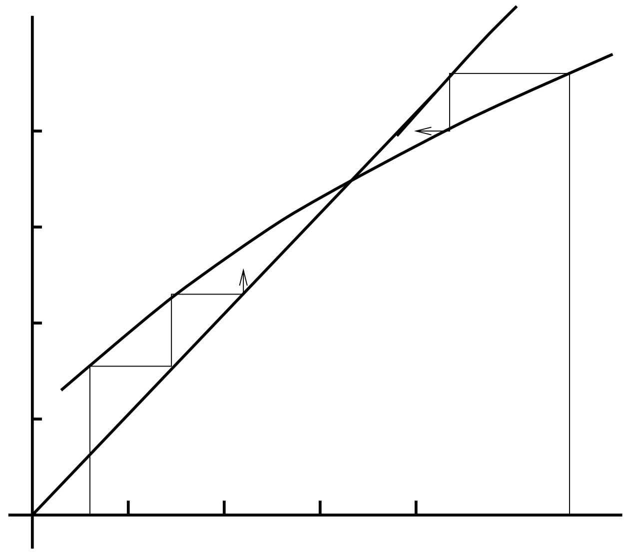


Figure 4.10: A stable equilibrium

Exercise 4.2.1 *Draw the corresponding plots for the case where* $-1 < f'(x) < 0$ *and* $f'(x) < -1$ *for any iteration function* f *with these properties.*

If you think about the situation with two unstable fixed points, you must wonder about the kinds of things that can happen to a starting point in between them that stays in between. It is obvious that if you start out between 0 and 1 for the iteration map

$$y = 4x(1-x)$$

then you must stay inside the region. But you can only get to 0 by either (a) starting there or (b) coming from 1, or (c) starting from somewhere that gets you to 1 after a finite number of moves. In order to get to 1 we must have

$$4x(1-x) = 1 \Rightarrow x = 1/2$$

And in order to get to 1/2 we have to come from

$$\frac{1}{2} \pm \frac{\sqrt{2}}{4}$$

by solving the quadratic. It is clear that we can run the iteration map backwards, but there will be two solutions in general. All the solutions must be roots of a quadratic equation

$$x^2 - x + Q = 0$$

where Q is made up from taking fractions with powers of 2 in them and square roots of numbers obtained by taking square roots, and so on.

So the set of starting points which can wind up at 0 is, although infinite, a rather special set. If we choose a starting point at random, we shall miss this set. Similarly the only way to wind up at 3/4 is to start there or to get there from 1/4, or to get to 1/4 in some finite number of jumps. If you try solving the quadratic equation

$$x^2 - x + 1/16 = 0$$

you can get the possible precursors to 1/4, and working backwards we again get a relatively small set of points which can get us to the fixed point 3/4. Well, it is infinite, but not **very** infinite[1].

What can happen if we start anywhere else? We obviously have to hop around for ever, never arriving at a fixed point. We can't even get closer and closer to any other point. For suppose there is a point $p \in [0, 1]$ and a sequence which converges to p under the iteration map. We would have

$$\lim_{n\to\infty} x(n) = p$$

where $x(n) = f(x(n-1))$. Taking f on both sides and using the fact that it is continuous, we get

$$\lim_{n\to\infty} f(x(n)) = f(p)$$

But the sequence $f(x(n))$ is just the sequence $x(n+1)$ which converges to p. So $f(p) = p$ and p is a fixed point. But there aren't any other such points. So if we start anywhere other than in one of the two 'special sets', we don't converge to any point at all. We hop for ever, never getting home, like the Flying Dutchman.

[1] This sounds crazy but is actually quite sound. It would be fun to explain this but it would take me too long.

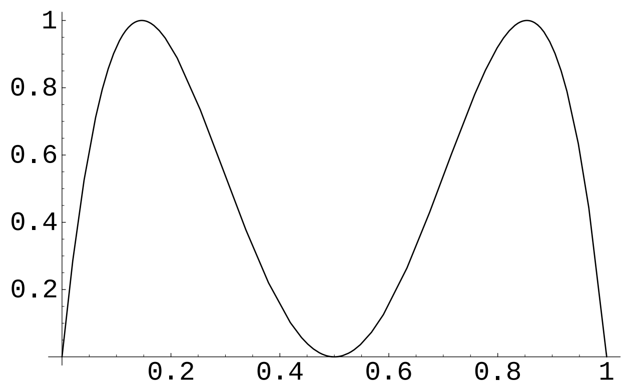


Figure 4.11: The doubling up map

What is there except chaos, darting about apparently at random? The only other possibility is a periodic 'orbit'. We might settle into approaching some pair of points. Can this happen?

To investigate this possibility, we look at the map which takes two steps instead of one.

$$\begin{aligned} y &= 4x(1-x) \\ \& \ z &= 4y(1-y) \\ \Rightarrow z &= 4(4x-4x^2)(1-4x+4x^2) \\ \Rightarrow z &= 16x - 80x^2 + 128x^3 - 64x^4 \\ &= 8(2x - 10x^2 + 16x^3 - 8x^4) \end{aligned}$$

The graph of this over the interval $[0,1]$ is shown in figure 4.11

I have also drawn the line $y = x$ and the fact that it cuts the curve in four places tells you that there are four fixed points of this iteration map. Any such fixed point has to be either a fixed point of the original map, or a point of period two. Think about it! Any fixed point of the original map is still fixed, which means there are two new fixed points of the doubled map, Can you see where they are?

Exercise 4.2.2 *Draw the two graphs on the same sheet of graph paper; find the new fixed points which have to correspond to points which are fixed if you do the original iteration map twice but not if you do it once.*

This must therefore give points which oscillate between two values under the original iteration map. Are these stable or unstable fixed points of the doubled map? What does this mean for the behaviour of nearby points under the original iteration map?

Exercise 4.2.3 *Do the same exercise for the case where the iteration map is*

$$y = 3x(1 - x)$$

and investigate graphically the stability of the period two orbits. Repeat for

$$y = 2x(1 - x)$$

We can clearly look at period three, four, or however many we feel like points by the same method.

Exercise 4.2.4 *Using some graph drawing program, find period four points of the non-linear iteration maps considered so far.*

Exercise 4.2.5 *Investigate the behaviour of the dynamical system with iteration equation*

$$y = 3.2x(1 - x)$$

with initial values in $[0, 1]$.

4.3 Heartbeats

Many biological systems have feedback control to keep some variable within bounds. For example, your heart starts beating before birth, and keeps going for about seventy years. The heart is a pump, and its regular beat is influenced by a number of factors which speed it up or slow it down according to the necessity of the circumstances. If the heart is given a sufficient shock, it goes into *fibrillation*, a far faster, random looking mode which looks like a chaotic state. This mode is not sufficient to keep your brain functioning, and if you are not snapped out of it, you will die in short order. Snapping your heart out of fibrillation can be done by a massive electric shock. Sometimes.

Under other conditions the heart may stop altogether but be restarted by electro-shock. As is well known to all who watch the medical soapies.

The heart, regarded as a dynamical system is rather complicated, and I shall not attempt to build a serious model. What I shall do is to show you a difference equation which has two modes, a chaotic one and a periodic one.

Let

$$y(n+1) = 2.2y(n)(1-(y(n))^2) + 0.3(y(n))^2$$

give the state of the heart at time n, measured by some sort of potential obtained from Electrocardiograms, (ECGs).

If we start the heart at $y(0) = -0.4$, it converges rapidly to a stable oscillation. This is shown in figure 4.12.

If we start at $y(0) = 0.4$ it becomes chaotic, as in figure 4.13.

If as a result of some massive trauma the operating point of the heart is greatly displaced, it can go into the chaotic mode. If it is kicked back, it returns to its normal state.

Exercise 4.3.1 *Use a differential equation plotting program to investigate the equation*

$$F(k, y, z) = 2.2y * (1-y^2) + 0.3y^2$$

for various initial points $y(0)$.

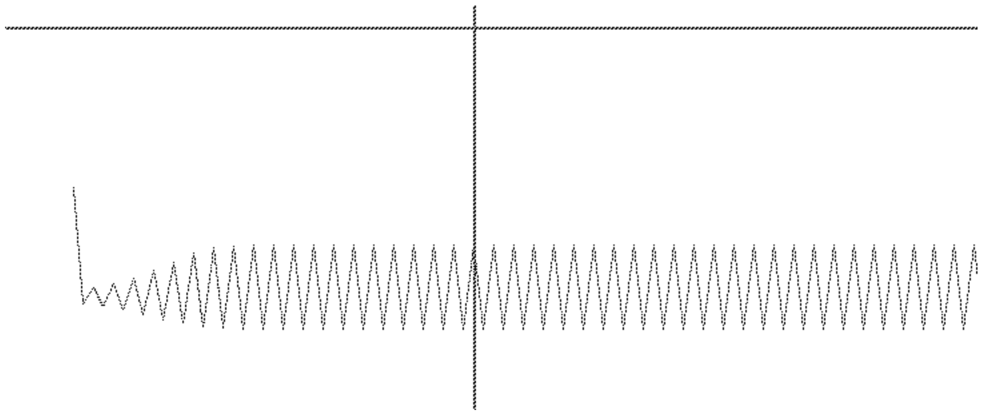


Figure 4.12: The normal heart rhythm

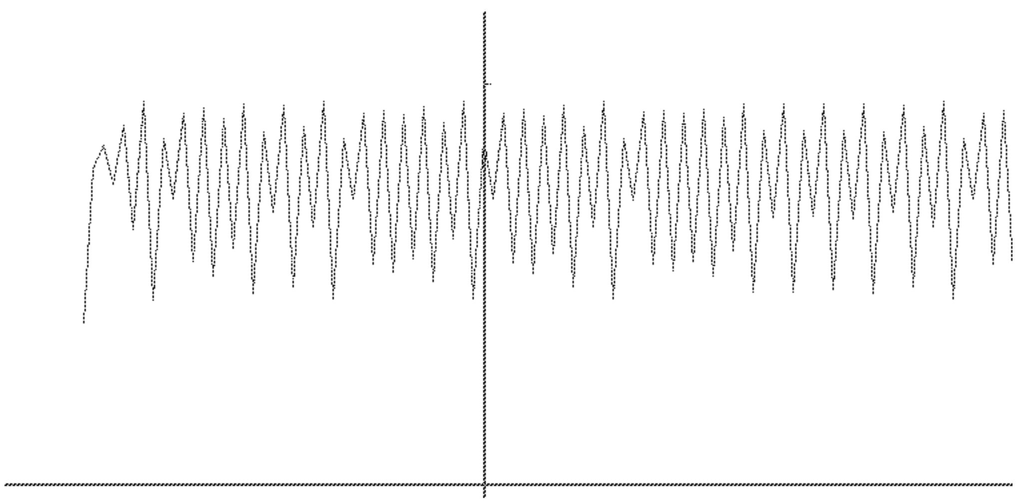


Figure 4.13: The chaotic heart

Using a graph-drawing program, find all the fixed points of the dynamical system, and decide if they are stable or not. What about the periodic points?

Exercise 4.3.2 (For the brighter students) *Try to figure out how I got the above dynamical system. Can you construct a dynamical system with two chaotic modes? With two stable oscillation modes separated by an unstable fixed point?*

When mathematicians investigate these systems, they are preparing themselves for being able to understand some of the things than can happen in the real world.

If you are superbright, you should enjoy exploring these matters, because you will be able to see the creative side of mathematics coming out, perhaps for the first time.

Chapter 5

Higher Dimensional Systems

One of the simplifications which has been made in almost all the systems we have modelled so far, is that there has been only one variable. This is a colossal simplification which I have chafed under, because it stopped me looking at heaps of interesting cases. The closest we got to dealing with more than one variable was the discussion on sugar when we were looking at bacteria. But we didn't worry too much about the sugar, we cared only about the bacteria. This isn't fair to sugar. Nor for some of the other things bacteria can eat. And there are lots of other species besides the bacteria. The case of two species, one of which eats the other, is called a *Predator-Prey* system. Even a tenuous contact with reality would suggest that this is still a huge oversimplification and of limited value, but there are cases when it holds. Do not be too narrow minded about what constitutes a species here. We might be talking about marauding financiers and company take-overs, so choose an interpretation which interests you.

5.1 Eating People is Wrong

The belief that 'eating people is wrong' is widely shared by people, but not, for instance, by tigers or the 'flu virus. It seems likely that sheep and wolves would differ on the morality of eating sheep, were such issues likely to trouble either species, but it seems useless to debate the issue. So in order to avoid personal involvement we shall look at species other than our own.

Suppose we have two species, one of which eats the other, say wolves and sheep. If there are hardly any wolves and lots of sheep, then the wolves find the sheep, feed and produce more wolves, and eat even more sheep. Maybe all the sheep get eaten, whereupon the wolves starve to death shortly afterwards. Or maybe a few sheep survive where the wolves don't find them, and the wolves mostly starve to death. With hardly any wolves, the sheep soon breed again and we are back to lots of sheep and few wolves; we have completed a cycle.

This is observed to happen with some species. The behaviour is usually modelled by differential equations, and we shall look at this case soon, but it can also be done in discrete time (years).

Let us try to turn the description of the sheep and the wolves into a precise model. Please note that this can be done in many ways, like painting a picture; please note that there is no one right way to do it, but that, as with pictures, some come out looking better than others, and again, as with pictures, practice improves performance. So follow my reasoning and if you don't like my assumptions, put in some of your own and see what happens. This is like your kindergarten class where they gave you some mud and water and told you to make a mess. You are supposed to let yourself go and enjoy the experience.

Example 5.1.1 (Predator and Prey) *First we think about what numbers we are going to measure, and this is an easy one: the number of wolves at year n and the number of sheep. So we let $W(n)$ be the number of wolves in year n, and $S(n)$ the number of sheep.*

The wolves eat the sheep, and, in effect, turn them into baby wolves after a time delay.

I simply turn this statement into algebra:

$$W(n) = W(n-1) + A\ S(n-1)W(n-1) \tag{5.1}$$

Here, A is some positive constant saying how many sheep it takes to produce a new wolf. It may also take into account that some of the sheep may be hard to find, but the more sheep, the easier it is to find some of them. The rate of supply of new wolves goes up as the amount of food goes up, and also as the number of wolves ready to go off into a corner with each other and bay at the moon afterwards. I have just multiplied these numbers together, with a fine disregard of detail. If you have a better idea, try it out. In this model, old wolves never die but keep on eating and baying at the moon for ever. This is hardly realistic, but again if you want to make it more sensible, go ahead.

The sheep in turn reproduce by turning grass into new sheep; I shall suppose an indefinitely large supply of grass. The number of sheep in the next generation is, I suppose, increased in proportion to the number of sheep already around. It is decreased by the effect of the wolves; the more wolves, the more sheep get eaten, and the more sheep there are to eat the more sheep get eaten. Turning this into algebra:

$$S(n) - S(n-1) = b\ S(n-1) - C\ S(n-1)W(n-1)$$

and rewriting we get

$$S(n) = B\ S(n-1) - C\ S(n-1)W(n-1) \tag{5.2}$$

where $B = b + 1$ is a positive constant and C is another.

Now go to your trusty computer and program it to compute solutions to these equations for various starting points.

Note that we now move into the plane. A state of the system is given by two numbers, $(W(n), S(n))$ at time n.

I shall let $x(k+1)$ denote the number of sheep, and $y(k+1)$ the number of wolves at year $k+1$.

You should explore possible settings for the initial state and see what happens, then try to explain it in terms of the verbal description of the dynamics given above.

It should be remarked that the above equations are seriously defective as a predator-prey model; I shall explain why later when I discuss systems of Ordinary Differential Equations.

5.2 But Killing them in War is OK

Another example of a system modelling several measurable quantities goes back to Lewis Fry Richardson in the Nineteen Thirties and Forties. You will find a paper by him in 'The World of Mathematics' edited by James Newman, [16]. He discussed arms races. There is a discussion on a simplified version of some of his ideas in [15]. There are also some notes on the topic, treated via differential equations rather than difference equations, in the Adelaide Second year Course Notes to be found at

`http://www.maths.adelaide.edu.au/jdenier/dyn_mod/index.html`

Exercise 5.2.1 *Let A and B be two countries and let $A(n)$ and $B(n)$ denote the amount of money each country spends on armaments in year n. Then we can suppose that if there was a large expenditure this year, all other things being equal there will be a smaller expenditure next year. On the other hand, if the enemy spent more money last year, there will be a tendency to spend more this year. This gives:*

$$A(n) - A(n-1) = -rA(n-1) + sB(n-1) + a$$

for some positive constants r, s. Similarly,

$$B(n) - B(n-1) = -r'B(n-1) + s'A(n-1) + a'$$

Use a suitable computer program to explore various choices of parameter values. You might simplify to make $s = s'$ and $r = r'$ in a first pass. If the expenditure of either country heads off to infinity, we predict war. Try to find conditions under which war can be avoided. Investigate various initial values for $A(0), B(0)$.

If one country is richer than the other, what should their policy be as to handling provocation?

Richardson showed that World War One was inevitable, given estimates of the 'mutual distrust', s, s', and the initial values. Since World War One had already been lost and won, this had limited value. All scientists love a successful prediction; best are those which seem least likely to come true but do. Predictions of the form 'Either it will rain tomorrow or it won't' do not cut much ice. And retrodictions fail to impress too, unless they bring together an awful lot of things that had seemed to be unrelated.

5.3 The Dismal Science

A similar system can model two economies.

Example 5.3.1 (Unemployment) *Let A and B be two countries, A initially very rich and B very poor. Each country has the chance to sell to the other, and its own national income is derived from exported goods to the other country, and the result of investment. The sum earned by exports is proportional to the other country's wealth. Set up the equations to describe this situation, and deduce implications for unemployment in your country, and the extent to which politicians are our friends.*

Solution

Let $A(n)$ denote the national income of country A and likewise $B(n)$ for country B in year n. We put $A(0) = 1, \quad B(0) = 0.1$ to represent the initial conditions.

The growth of income in country A, $A(n) - A(n-1)$, is the sum of two terms, one proportional to $A(n-1)$ and representing investment. The constant here is unlikely to be more than a few percent and may be negative when investment levels are below natural depreciation. The second term is derived from exports to B and will be proportional to $B(n-1)$ since the amount of money B has will determine how much it can buy. The value of the constant will be a small positive number. This gives:

$$A(n) - A(n-1) = a\ A(n-1) + b\ B(n-1)$$

Similarly

$$B(n) - B(n-1) = a' \; B(n-1) + b' \; A(n-1)$$

I shall choose $a = a' = 0.0001, b = b' = 0.01$ *to see what happens.*

In this case, we need

$$\begin{aligned} F1(k,x,y) &= 1.0001x + 0.01y \\ F2(k,x,y) &= 1.0001y + 0.01x \end{aligned}$$

We see in this case a development which increases the income of the poor country more than of the rich country, i.e. the incomes tend to equalise. Both countries profit however.

If the countries, for whatever reason, have a negative term for a, a' *(as could occur when countries run a deficit and their income is spent on other things than investment) then we might have:*

$$\begin{aligned} F1(k,x,y) &= 0.999x + 0.01y \\ F2(k,x,y) &= 0.999y + 0.01x \end{aligned}$$

The behaviour now is different; the incomes tend to equalise, but there is a drop in the income of the richer country. Plot it on diffeqn2. *In time, country A is overtaken by country B, but the process now goes into reverse, and country A catches up. Both countries eventually get richer, but in the short term, B gets rich while A gets poor.*

If the situation is even more extreme,

$$\begin{aligned} F1(k,x,y) &= 0.99x + 0.01y \\ F2(k,x,y) &= 0.99y + 0.01x \end{aligned}$$

then A gets poorer, B gets richer, and there is a stable fixed point of the iteration map which sends (x,y) *to* (u,v) *by* $u = 0.99x + 0.01y$, $v = 0.01x + 0.99y$. *B benefits, A loses, and it is a zero sum game. On the other hand, if* either *country can increase its investment marginally,* both *countries eventually do better than this case, and eventually both countries get richer than they ever were.*

The really unpleasant case is when the investment is even lower. If we have

$$\begin{aligned} F1(k, x, y) &= 0.97x + 0.01y \\ F2(k, x, y) &= 0.97y + 0.01x \end{aligned}$$

In this case, A loses badly, B gains a little bit initially, but both wind up bankrupt eventually.

If country A loses income, then the average income drops, which may be accomplished in various ways. Everybody might take a pay cut, or, more likely, unemployment will rise. What is likely to happen is that the jobs are being done by the people in the poorer country because their costs of production are lower. This wasn't included in the model, but can only make matters worse from the point of view of country A.

If it were, then it seems rather likely that one would conclude that the reason unemployment is high in Australia is because wages are high and it is cheaper to hire a Bangladeshi or a Vietnamese to manufacture shoes. So Bangladesh gets richer and Australia gets poorer. And the prospect of a politician being able to change this state of affairs does not look good. So the promises to cut unemployment would seem to be either naive or hypocritical.

The conclusion that we draw from this model is that income equalisation between countries appears to be forced, and that initially wealth is going to be transferred from the rich country to the poor country unless the investment levels in A are high. If net investment is too low, in either country, then it becomes a straight transfer of income from rich to poor.

Since investment is done by people who already have lots of money ('There's no point giving extra money to the poor, they only waste it on food, drink, clothing and shelter'), the rather unpleasant conclusions for the people in country A would seem to be that the national income will go down, and incomes will be equalised between countries, but incomes must become less equal inside a country in order to maintain investment. The alternative is eventual poverty for everybody. This is hardly a conclusion I like (since I am poor), but it would seem to be forced by the model.

Of course, the model may be totally and gloriously wrong. The main effect discussed is known as a trade imbalance, and they seldom become as extreme as I have suggested. But it may in fact *underestimate* the tendency to equalisation of national income instead of overestimating it, since I did not take into account the effects of companies preferring to use third world countries to do their manufacturing because of lower labour costs. I have left out of account a large number of niceties, such as population size, cost and standard of living (so as to allow for an estimate of disposable income) and many more elements.

Exercise 5.3.1 *Extend the above model by taking into account the costs of production in the two countries and hence the profits and hence the amount to invest. Investigate experimentally the dependence on initial ratio of wealth and also the dependence on the investment levels. Model the investment levels in terms of investment per capita as a function of disposable income.*

How would you advise politicians in the richer country to handle the drop in national wealth in the case when investment was low? Would it be better to share the money out equally or to head towards greater inequality of income?

Explain why economics is known as the Dismal Science.

You ought to be able to see that when the populace is largely ignorant of science, of modelling and of economics, demagogues who promise to repeal the laws of nature can gain power. And when they do, they are able to buy popularity, if only for a time, by spending precious reserves of money desperately needed to build a future. The voice of reason is always a still, quiet voice, and when it advises unpopular courses of action it is likely to be stifled. So an educated populace is of immense importance. Learn as much as you can, and help others to learn too.

There are a lot of important issues here which I can only indicate briefly. My main aim here is one which may seem very strange: it is to encourage you to admit you don't know much, and to take steps to reduce your ignorance. So ask yourself whether the above model, with its depressing implications for social inequality (it makes you poorer), is oversimplifying. It is not enough to merely say you feel deeply that it is. You have to offer an alternative which fits the data better. The time is past when a thinker like Marx could tell us all what he thought would happen, and sell it to us by sheer eloquence. If he came along today, he'd be told to build a convincing computer model of his system. As one Russian woman is reported to have said, 'If Marxist-Leninism was really scientific, they'd have tried it out on rats first.' These days, they'd have had to try it on a computer. It might have saved a lot of people a lot of pain if they'd had computers a century ago.

You do not need to be an economist to see that there are some rather important reasons for understanding the nature of modelling. In particular, the analysis of stability for two and higher dimensions (because there is nothing that says that only two variables will determine the state of a system) is of considerable importance. I shall have a bit more to say about this when I get onto Systems of Differential Equations, but you simply don't know enough other mathematics, in particular linear algebra, to understand much of the issues yet. So think of this as pointing out the way ahead rather than leading you to a high place.

Another element to be investigated is 'sensitivity analysis'. If you change the parameters in your model, you get out different behaviour. But is it hugely different or only a little bit different? Well, this will depend on the model, and also on the parameters. Some models may be stable under minor perturbations of the parameters for some parameter values, while going wildly different for other values. This again can be explored numerically.

This is important: the Club of Rome frightened people in the sixties by predicting doom and gloom about the future of the industrialised world. We were going to run out of oil about last week. Similarly there have been prophecies of global warming, global freezing, population explosion and various other disasters including economic catastrophes. All good fun for the journalists, and capable of selling almost as many books as Nostradamus. But are these prophecies any better than those of Nostradamus? Do Difference Equations solved by computer necessarily do better than a crystal ball and a nasty headache? Well, they are more explicit, and more easily proved wrong (which is a great virtue; predictions which cannot be shown to be wrong are pretty silly things to waste time on). But the Club of Rome discovered that when it wobbled its parameters a bit, it got totally different outcomes. So it would be silly to sell the farm and buy a nuclear fallout shelter on the strength of something that could depend crucially on the last decimal place in a guesstimate. So there is plenty to think about here.

5.4 A Cheap Trick

Finally, a well known trick which allows us to turn a second order equation in one variable into a first order system in two: given

$$x(n) = a\ x(n-1) + b\ x(n-2)$$

introduce the variable $y(n) = x(n-1)$ and rewrite the above equation as:

$$(x(n), y(n)) = (a\ x(n-1) + b\ y(n-1), x(n-1))$$

It can be done for more complicated non-linear equations.

Exercise 5.4.1 *Turn the second order equation in variable $x(n)$ given by*

$$x(n) = n^2 x(n-1) - (x(n-2))^2 + 3$$

into a system of two first order equations.

This ends the modelling by difference equations theory.

The next chapter deals with some differential equation theory. The rest is on modelling.

Chapter 6

Ordinary Differential Equations

6.1 First and Second Order linear and affine Equations

I am going to recapitulate briefly the material which you will have covered in first semester and then move onto some minor extensions of it.

Remember that ordinary differential equations are mostly about doing in continuous time what difference equations do in discrete time. That is a simplification, but a useful one. We shall keep running into the exponential function for reasons already discussed: it comes up naturally when we go to 'continuous compounding'.

6.1.1 First Order Linear and Affine Equations

The first order linear equations all look like

$$f'(t) = g(t)f(t)$$

for some given function $g(t)$ and the *autonomous* or constant coefficient linear equations all look like

$$f'(t) = mf(t)$$

for some constant m and have solution

$$f(t) = Ae^{mt}$$

for some constant A. If we have (as we always will) an *initial value problem*, then we are told the value of $f(0)$ and can work out what A must be.

The linear equations are called *homogeneous* for reasons too silly to mention, and the affine equations are of the form:

$$f'(t) = g(t)f(t) + h(t)$$

for some particular functions g and h. The special case of an equation

$$f'(t) = mf(t) + c$$

where m, c are numbers is a rather simple extension of the linear case, and I shan't insult your intelligence by discussing it further. This is code for 'I shall assume you are full bottle on this so watch out if you aren't.'

Equations of the form

$$f'(t) = mf(t) + h(t)$$

are of some interest, and can be solved for many functions h. You will have spent some time with equations of this form in first semester, and indeed of more complicated form. Again, I shall assume that you are able to dispose of these when they turn up in a model.

6.1.2 Second Order Linear and Affine Autonomous Equations

Suppose we have an equation involving higher derivatives; the order of the highest derivative is called the order of the equation.

Example 6.1.1

$$f''(t) - 3f'(t) + 2f(t) = \sin(t)$$

is a second order equation. It is also an affine equation.

I shall show you how to solve equations of this type; fortunately it is dead easy. It is also done in all the text books, so if you can't understand my treatment, you can find lots of other sources.

First we look at the linear constant coefficient equations of second order:

$$af''(t) + bf'(t) + cf(t) = 0 \tag{6.1}$$

Please note that the '0' on the right hand side is not the number zero, it is the function '0' which takes the value zero everywhere. I apologise for the lousy notation which confuses these things, but it is traditional. It is also traditional not to tell you and leave you to be muddled about this for ever, but I am not that fond of traditions.

We guess that this has a solution of the form

$$f(t) = e^{mt}$$

for some unknown m. How anybody ever dreamt *that* one up will puzzle some of you; the simplest answer is that mathematicians did it backwards when they were messing around with the exponential function to see what it did when they tried differentiating it a few times.

If we do this, we get

$$(am^2 + bm + c)(e^{mt}) = 0$$

Since the function e^{mt} is not zero anywhere, then we must have

$$am^2 + bm + c = 0 \tag{6.2}$$

The '0' on the right hand side *is* the number '0' now.

Since this is a simple quadratic equation we can solve it to get values for m. There are three possibilities.

1. the two roots of 6.2 are real and distinct
2. the two roots of 6.2 are complex and distinct
3. the two roots of 6.2 are the same.

The procedure is almost exactly as for second order difference equations; if we have distinct roots which are real, m_1, m_2, then we write

$$f(t) = Ae^{m_1 t} + Be^{m_2 t}$$

as the general solution to 6.1. If we are given some initial values of f and f' at $t = 0$ then we get a solution which is an explicit function of time.

Example 6.1.2 *Solve $f'' - 3f' + 2f = 0$ when $f(0) = 0,\ f'(0) = 1.$*

Solution

Putting $f(t) = e^{mt}$ we get the auxiliary equation

$$m^2 - 3m + 2 = 0$$

which has roots $m = 1, m = 2$. So the general solution is

$$f(t) = A\ e^t + B\ e^{2t}$$

Putting $t = 0$ we get $A + B = 0$. Differentiating we get

$$f'(t) = A\ e^t + 2B\ e^{2t}$$

and putting $t = 0$ again we deduce that $A + 2B = 1$ Hence $B = 1, A = -1$ and the complete solution is

$$f(t) = e^{2t} - e^t$$

As with the case of difference equations, the complex roots case reduces to a mixture of sin and cos functions. If a, b, c are all real, the roots must be complex conjugates.

Example 6.1.3 *Solve $f'' - 2f + 2 = 0,\ f(0) = 1,\ f(\pi/2) = 1$*

Solution

The auxiliary equation is $n^2 - 2m + 2 = 0$, the roots are $1 + i,\ 1 - i$. The solution is therefore

$$f(t) = A\ e^{(1+i)t} + B\ e^{(1-i)t}$$

which we write (using polars) as

$$f(t) = e^t(A(\cos t + i\sin t) + B(\cos(-t) + i\sin(-t)))$$

which reduces to:

$$f(t) = e^t((A + B)\cos t + (A - B)\sin t)$$

and changing some names this is:

$$f(t) = e^t(C\cos t + D\sin t)$$

Putting $t = 0$ *we get* $f(0) = C = 1$ *and putting* $t = \pi/2$ *we get* $f(\pi/2) = e^{\pi/2}D = 1$ *so* $D = 2/\pi$. *The complete solution is therefore*

$$f(t) = e^t(\cos t + \frac{2}{\pi}\sin t)$$

Finally, if the two roots coincide (and this works for both real and complex cases) we have a trick to make sure that we get two independent solutions. We have that the general solution is

$$f(t) = A\ e^{mt} + Bt\ e^{mt}$$

Example 6.1.4 *Solve* $f'' - 2f' + f = 0$ *when* $f(0) = 1,\ f(1) = 0$

Solution

We get the auxiliary equation

$$m^2 - 2m + 1 = 0$$

which we factorise to

$$(m - 1)(m - 1) = 0$$

which has $m = 1$ *(twice). So the general solution is*

$$f(t) = A\ e^t + Bt\ e^t$$

Putting $t = 0$ *we deduce that* $f(0) = A = 1$ *and putting* $t = 1$ *we deduce that* $f(1) = Ae + Be = 0$ *so* $B = -1$ *and the complete solution is*

$$f(t) = e^t - te^t$$

You should check that in all cases the given functions satisfy the differential equations and the initial conditions.

Exercise 6.1.1 *Practice by making some equations up and solving them. Check your answers!*

I am not much interested, in this course, in the reason why this method works, although some aspects of it may puzzle you. This shows your brain is working correctly; there is clearly something going on which needs explanation, but the explanation lies in Linear Algebra and you may not have done any. (But see the book provided by *HeavenForBooks.com*!) So your attitude should be that you will buy this stuff on the basis that it is OK to just use it when dealing with a real problem, and explanations of what is going on will come later if you do more Mathematics. Which I hope you will. If this book is doing its stuff, then you will be impressed with what can be done with the mathematics and should be coming around to the view that it would be a good idea to know more of it. It ain't easy, but it sure is useful. Also, very, very cool.

The affine equations we shall look at are a very restricted case; the first is:

$$a\ f''(t) + b\ f'(t) + c\ f(t) = d$$

where d is some constant.

Example 6.1.5 *Solve*

$$y'' - 3y' + 2y = 4 \tag{6.3}$$

Solution *First we solve the 'homogeneous equation'* $y'' - 3y' + 2y = 0$

This is $y = A\ e^t + B\ e^{2t}$.

Now we try a solution to the original equation of the form

$$y = A\ e^t + B\ e^{2t} + c$$

Differentiating the above equation we get

$$y' = A\ e^t + 2B\ e^{2t}$$

and again:

$$y'' = A\ e^t + 4B\ e^{2t}$$

Substituting in 6.3 we find

$$A\ e^t + 4B\ e^{2t} - 3A\ e^t - 6B\ e^{2t} + 2A\ e^t + 2B\ e^{2t} + 2c = 4$$

And with some cancelling we find $c = 2$. *The solution is therefore:*

$$y = A\ e^t + B\ e^{2t} + 2$$

We can handle the more complicated case of a function on the right hand side provided it is simple. Suppose we had some polynomial in t such as

Example 6.1.6 *Solve*

$$y'' - 3y' + 2y = 2t + 3 \tag{6.4}$$

We reason that if we can solve the homogeneous case, to get

$$y = A\ e^t + B\ e^{2t}$$

we can try to find a solution to 6.4 by adding on a polynomial to the above equation. After all, the rest of the process is one of differentiating a few times, adding the results of differentiating once multiplied by a scalar, and adding some scalar times the polynomial. The result has to be a polynomial, and maybe we can fiddle the numbers to get the polynomial we want.

So we try a solution to 6.4 of the form

$$y = A\ e^t + B\ e^{2t} + ct + d$$

to obtain:

$$A\ e^t + 4B\ e^{2t} - 3A\ e^t - 6B\ e^{2t} - 3c + 2A\ e^t + 2B\ e^{2t} + 2ct + 2d = 2t + 3$$

Cancelling massively we find

$$2ct + 2d - 3c = 2t + 3$$

Since this has to hold for any number t *we must have* $c = 1, d = 3$ *so the complete general solution is*

$$y = A\ e^t + B\ e^{2t} + t + 3$$

Exercise 6.1.2 *Check the above carefully. make up your own problem of the same kind and work it through. Check the answer.*

The same trick will work for other functions on the right hand side; we need to find some possible 'space' of things which can be taken to the right hand side.

Exercise 6.1.3 *Solve* $y''-5y'+6y = 2\sin(t)$ *by trying a solution of the form*

$$y = A\ e^{m_1 t} + B\ e^{m_2 t} + c\sin t + d\cos t$$

You can run into trouble if the function on the right hand side is one which is a solution to the homogeneous equation. The fix is to multiply by t:

Example 6.1.7 *Solve*

$$y'' - 3y' + 2y = 2e^t \tag{6.5}$$

We know that a solution of the form

$$y = A\ e^t + B\ e^{2t} + c\ e^t$$

can't give us the $2e^t$ *on the right hand side, since we just add* A *and* c *to get a new constant. So we try a solution of the form:*

$$y = A\ e^t + B\ e^{2t} + ct\ e^t$$

Then differentiating gives us

$$y' = A\ e^t + 2B\ e^{2t} + ct\ e^t + ce^t$$

and differentiating again gives us:

$$y'' = A\ e^t + 4B\ e^{2t} + ct\ e^t + 2ce^t$$

Substituting in 6.5 we get:

$$A\ e^t + 4B\ e^{2t} + ct\ e^t + 2ce^t - 3(A\ e^t + 2B\ e^{2t} + ct\ e^t + ce^t) + 2(A\ e^t + B\ e^{2t} + ct\ e^t) = 2e^t$$

Doing the cancelling we find:

$$-ce^t = 2e^t$$

and hence $c = -2$.

This ends the recipe material; I know that some folk find learning recipes very comforting. Contrary to any impressions you may have formed, this isn't what mathematics is about.

6.2 Systems of First Order ODEs

We next do something which is a lot more fun and also a lot more useful. First a definition so that we are clear about the subject under investigation:

Definition 6.2.1 *An autonomous two dimensional system of first order Ordinary differential equations involves two functions, f(t), g(t) and an expression for* f'*, the derivative of* f *in terms of* f *and* g*, and another expression for the derivative of* g *also in terms of* f *and* g*.*

If the expression is some constant multiplied by f *plus a constant times* g*, we say we have a* linear, autonomous first order system of equations.

Example 6.2.1

$$\frac{dx}{dt} = -y(t) \tag{6.6}$$

$$\frac{dy}{dt} = x(t) \tag{6.7}$$

is a linear, autonomous first order system of equations. Of course f *has turned into* x *and* g *into* y*.*

You might at first have a sinking feeling at the thought of having to solve this system or even to make a guess at what it all means. A little cool thought however will cheer you immensely.

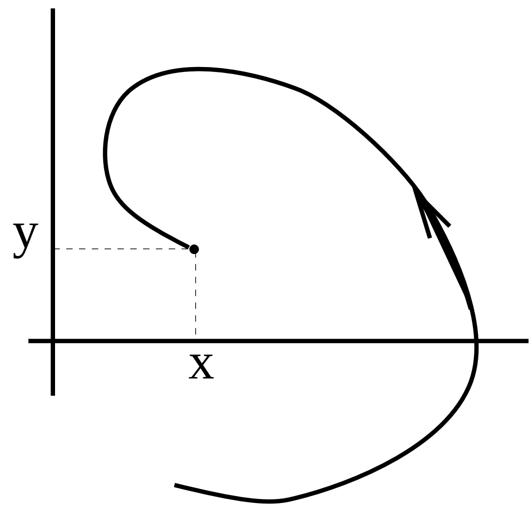


Figure 6.1: The Moving Bug

First of all, what would a solution mean? Well, presumably it would be some pair of functions, $(x(t), y(t))$. And what would *that* mean?

Well, imagine a little bug crawling around on a sheet of graph paper in two dimensions, as in figure 6.1.

At any time t it has an X-coordinate and a Y-coordinate, just like the cannon ball. I again imagine it is a very *small* bug, so we can take it as a point. It is easy enough to see that its shadow on the X-axis is given by the function $x(t)$; at time t, $x(t)$ gives his shadow's position. Similarly, $y(t)$ tells us the position of a shadow on the Y-axis. And if we know both shadows, we know where the bug is.

So the two functions can be thought of as one function from $\mathbb{R}$ to $\mathbb{R}^2$ saying how a bug moves about in two dimensions.

Now what does the system of differential equations mean? If $(x(t), y(t))$ is the solution trajectory of a bug, then dx/dt is the speed of the shadow of the bug on the X-axis at time t, and dy/dt is the speed of the shadow on the Y-axis. And these two numbers are the components of the velocity vector at

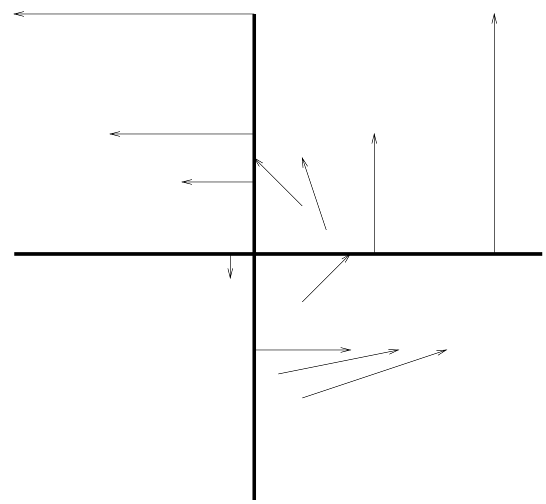


Figure 6.2: A Vector Field

time t. So the pair $(dx/dt, dy/dt)$ at any time t is a little arrow stuck on the point $(x(t), y(t))$ which is where the bug actually *is*, showing which direction it is travelling in and how fast it is moving. And the system of ODE's tells us how to calculate all these little arrows.

We draw some pictures.

Example 6.2.2 *We draw the velocity* vector field *for the system of equations*

$$\frac{dx}{dt} = -y(t) \tag{6.8}$$

$$\frac{dy}{dt} = x(t) \tag{6.9}$$

I have taken a few points (x, y) and drawn the vectors $(-y, x)$ moved to have their tails at (x, y). For example, at the point $(1, 1)$ I draw the arrow parallel to $(-1, 1)$.

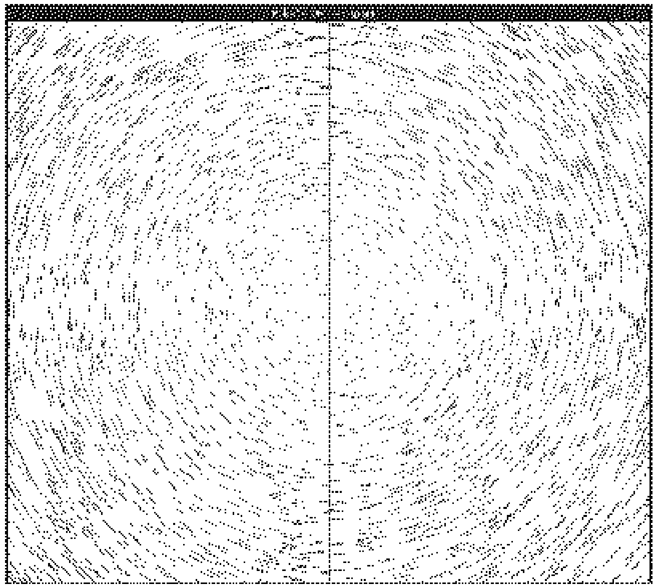


Figure 6.3: Lots more vectors, scaled

This is a bit hard to read, so in the next figure 6.3 I have used a computer to generate a lot more points, and I have scaled the lengths of the vectors and made them shorter.

I have also left the arrows off, because they make even more mess. You must remember that the line segments are really little tangent lines to a solution curve. It ought to be obvious that the solutions are in fact circles centred on the origin.

In fact, if we put

$$x = A\cos(t) - B\sin(t)$$
$$y = A\sin(t) + B\cos(t)$$

we rapidly see by differentiating separately, that the original system has the curve given by these equations as solution.

It is possible to find software which will draw vector fields and also produce solution curves for them. Typically, you type in the equations for $x'(t)$ and $y'(t)$ and the program draws the vector field. Then you choose an initial position usually by positioning the cursor with the mouse and clicking.

The program generates a solution curve by starting at the mouse point, calculating the tangent vector, and then taking a little step along it. It then recalculates the new tangent, and repeats until you click the mouse button or the solution goes off the screen.

You will notice that the path is *not* a circle, but an outward spiral. If you go back and change to a time step of 0.01 the movement is slower, but the solution is much closer to a genuine circle. This is an **Awful Warning**; the numerical approximation to a solution can be qualitatively different from the actual solution. We are turning the differential equation into a difference equation, and there is a price to be paid.

Exercise 6.2.1 *Now make up some more complicated systems and find out what happens when you try to solve them. Some judicious net-surfing might allow you to get some free software for your machine.*

6.2.1 A Smart Trick

The trick we used for turning a second order difference equation into a system of two first order equations also works with differential equations. A simple example will suffice:

Example 6.2.3 *Turn the second order ODE*

$$\ddot{x} + x = 0$$

into a system of first order ODEs. Hence or otherwise, solve the system

$$\dot{x} = -y$$

$$\dot{y} = x$$

Solution

Put $\dot{x} = y$*. Then differentiating again we get* $\ddot{x} = \dot{y}$*. So the equation* $\ddot{x}+x = 0$ *becomes*

$$\dot{x} = y$$

$$\dot{y} = -x$$

As an alternative, we might have tried putting $\dot{x} = -y$, in which case we should have wound up with the system

$$\dot{x} = -y$$

$$\dot{y} = x$$

But we know how to solve

$$\ddot{x} + x = 0$$

The solution is

$$x(t) = A\ \cos t + B\ \sin t$$

If we put $\dot{x} = -y$ we get the solution to the system given as

$$x(t) = A\ \cos t + B\ \sin t$$

$$y(t) = A\ \sin t - B \cos t$$

Exercise 6.2.2 *Turn the second order ODE*

$$x'' - 3x' + 2x = 0$$

into a system of first order ODEs.

Sketch the vector field (by using a program if you wish) and a few solutions. Do these agree with the solution to the original second order equation?

In the case when the vector field arises from turning a second order ODE into a two dimensional system, we call the diagram the *phase-plane* representation, for reasons too tiresome to explain.

6.3 Some Applications

In these applications, I shall rely on you sketching the vector fields using a program and plotting a few solution curves using the same program. All the problems about producing garbage this way recur as they did for difference equations, and so some degree of thought about the nature of what can happen is desirable. First, however, I shall just look at some systems and the associated vector fields and solutions.

6.3.1 Predator Prey Systems

Let me return to the predator prey model for two species, this time using differential equations. I shall let $x(t)$ be the number of sheep at time t, now a continuous variable (despite the fact that sheep come in lumps)! Indeed I shall suppose it is a differentiable function.

Similarly, let $y(t)$ be the number of wolves at time t.

Now the number of sheep at time t could be expected to grow exponentially from some small positive number of sheep if there were no predators and unlimited grass. But the number of sheep will be decreased by the number of wolves. There are many ways we might try to model this, but it is common to argue that the rate of loss of sheep due to the wolves is proportional to the product of the number of sheep and the number of wolves. This might make some sort of sense for some values of the variables, although not much for other values. Bearing this in mind and being properly sceptical about any conclusions we might be led to, we write

$$\dot{x} = ax - bxy$$

for positive numbers a, b as the first part of the system.

Similarly, we imagine that the wolves will die out in the absence of sheep, and will increase in their presence by an amount proportional to the product of the numbers of each. Again, other more complicated models can be thought of. This one is the *Lotka-Volterra* model.

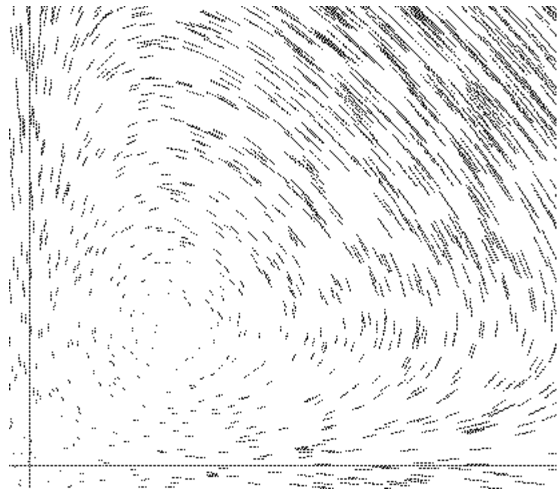


Figure 6.4: The Vector Field for the Predator-Prey model

This gives the equation

$$\dot{y} = -cy + dxy$$

for positive constants c, d.

The system can now be put into a suitable program. If you choose all the constants equal to one, you get a vector field that looks like figure 6.4.

If you run this with a sufficiently good program, you should get a closed orbit. This is telling us that the population of both species will oscillate and that neither species will become extinct.

Exercise 6.3.1 *How do we know that the system really does oscillate and not spiral out by a very small amount? Or spiral in? Do we know this in fact? If you could prove that the given model had closed orbits, would you feel justified in believing that some actual species pairs would always oscillate regularly?*

If you start at different places with this vector-field, you can see an easy explanation for the so called *DDT paradox* [17].

The Citrus parasite known as the Scale insect is eaten by the common ladybird beetle. It was found that when the trees on which they hung out were dusted with DDT, which is known to kill both, the number of Scale insects actually increased. In fact both populations increased.

If you hit them when they are at the maximum of their population by a moderate amount, you kill them off and move to an internal orbit. But if you hit them by a large amount, or if you catch them when both populations are small, you actually push them into a wider orbit!. So go easy on the DDT and hit them when they are in large numbers, not when they are weak.

The difference equation we considered earlier which studied the predator-prey system may be considered a rather crude approximation to the continuous dynamics. The phenomenon whereby a crude approximation to a circular flow spirals outwards is operative, and ensures that there is little possibility of a stable cycle.

6.3.2 Competition

Suppose we have two species which are competing with one another for food, or for anything else which helps the species reproduce. We need to have a limited rate of supply of food, so we postulate for each species individually that it follows a logistic curve in the absence of the other, that is, if $x(t)$ is the population of the first species at time t we have

$$\frac{dx}{dt} = mx(a - x)$$

when $y = 0$, and

$$\frac{dy}{dt} = m'y(a' - y)$$

where $x = 0$, where $y(t)$ is the size of the population of the second species at time t.

We assume that the 'interaction term' saying how each species affects the other is in both cases proportional to the product of the populations and has a negative effect on the growth rate. I don't claim that this is inherently very plausible. If you think you can do better, go ahead and try a different model.

We wind up then with the system:

$$\begin{aligned} \dot{x} &= max - mx^2 - cxy && (6.10)\\ \dot{y} &= m'a'y - m'y^2 - c'xy && (6.11) \end{aligned}$$

If you run this on a computer, you will find that there are, for most parameter values, two equilibrium states when one species is wiped out or the other is. And it turns on a hair as to which of them it will be; two initial states can be almost identical, and one will lead to the extinction of x, the other to the extinction of y. Play with the program, and as you watch the curves turning to one state or the other, reflect that your ancestors have gone this way, or you wouldn't be here. Understand that we can be talking about restaurants competing for customers as well as early hominids competing with our ancestors.

6.3.3 Chemical Kinetics

Chemical reactions occur in accordance with the *Law of Mass Action* which says that the rate of a reaction is proportional to the active concentrations of the reactants, [18]. Chemists write

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$$

to indicate that reactants A and B are combining to produce C, and that the process is reversible. The k_1 tells us how fast the reaction proceeds from left to right, the k_2 says how fast it goes in the opposite direction. If $x(t), y(t), z(t)$ are the concentrations of A, B, C respectively, we have

$$\begin{aligned} \frac{dx}{dt} &= k_2 z - k_1 xy && (6.12)\\ \frac{dy}{dt} &= k_2 z - k_1 xy && (6.13) \end{aligned}$$

$$\frac{dz}{dt} = k_1 xy - k_2 z \tag{6.14}$$

Then equation 6.12 tells us that the rate of increase of x is the rate k_2 times the concentration of z less the rate k_1 times the product of the concentrations of x and y. Similarly, equation 6.13 tells us the same thing about y, while equation 6.14 tells us that the rate of increase of z is k_2 times the product concentration while there is a decrease in the opposite direction as C turns back into A and B.

Since we are investigating matters graphically, we shall stick to two reactants, so we make the two reactants on the left hand side the same to get

$$A + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$$

Using $x(t)$ to denote the concentration of A and $y(t)$ that of B we find

$$\frac{dx}{dt} = 2k_2 y - 2k_1 x^2 \tag{6.15}$$

$$\frac{dy}{dt} = k_1 x^2 - k_2 y \tag{6.16}$$

If you run this on a vector field drawing program with $k_1 = k_2 = 1$, you will find that from any starting point in the first quadrant (concentrations have to be positive!), the trajectory is a straight line which slows down as it moves towards the line $y = x^2$.

If you put $y = x^2$ into the above system, you find that any point on this curve is an equilibrium point; the vector at any such point has zero length. It is an *equilibrium line.*

Exercise 6.3.2 *Set up the system of equations*

$$\frac{dx}{dt} = 2y - 3x^2$$

$$\frac{dy}{dt} = 1.5x^2 - y$$

on a vector field drawing program. Start at a lot of points along the positive Y-axis. Note the terminating positions (although it won't stop until you tell it!). Now start at a lot of points along the positive X-axis. What is the equation of the line of equilibria?

Exercise 6.3.3 *Prove that the trajectories are always straight lines.*

6.3.4 Epidemics

This section also follows [18]. The book is most readable, and has worn very well. You could do worse than buy a copy. Any difficulties with my version may be cleared up by a careful reading of the reference.

Suppose we have a small community such as a lecture group who meet each other regularly. One person comes down with the 'flu; instead of having the good sense to stay in bed and get drunk on Scotch and Disprin he turns up as usual. Before long all those who haven't had 'flu shots are also coughing over their nearest and dearest, turning the lecture into a sort of 'flu virus holiday camp.

We wish to study the spread of infection throughout the group. After studying this case, we can go on and think about more urgent issues of the same type from AIDS to religion.

The first step is to consider the measurables. We have three classes of people: those who have it and can spread it, those who haven't had it yet, and those who have had it and recovered (or died). I shall call these $I(t), S(t), R(t)$ as in [18], for *Infectives*, *Susceptibles* and *Removals*. There is also a fourth class, those who have natural immunity to the disease; we can lump them in with the last group. We ask how the numbers in these three categories change in time.

If N is the number of people in the lecture class, then the three numbers sum to N at all times:

$$I(t) + S(t) + R(t) = N$$

Now we ask what happens to the Infected class $I(t)$.

We reason that the rate of change of $I(t)$ will increase in proportion to the product of the number of infected people in the class with the number of susceptibles, but also be subject to a steady decrease proportional to the number of people infected as the infected recover or die off. There is, in other words, recruitment from the susceptibles and loss to the removables. We write:

$$\frac{dI}{dt} = rSI - \gamma I \tag{6.17}$$

We note that both r and γ are positive constants.

Next we consider the Susceptibles. This group loses bodies at a rate proportional to SI again, with the same constant, because we have already considered this group, it is the people moving into the I category. So we have:

$$\frac{dS}{dt} = -rSI \tag{6.18}$$

Finally, we have the Removals, who get recruits from the Infectives at the rate given above:

$$\frac{dR}{dt} = \gamma I \tag{6.19}$$

This sets up the equations. r is called the *infection rate*, γ the *removal rate*. It might be the same as the recovery rate, or it might be the same as the death rate, in extreme cases.

Note the similarity to the chemical kinetics equations as people change from one state to another instead of atoms.

Since we have $S + I + R = N$ we need worry about only two of the variables,

Using x for the Susceptibles and y for the Infectives, draw the vector field. Let the range in X and Y be from 0 to 3 and let both the constants be 1. You will find that you get results like figure 6.5.

Example 6.3.1 *Suppose there is a new category, those who are infected but not yet able to transmit the disease. Call these the inCubators. There is a*

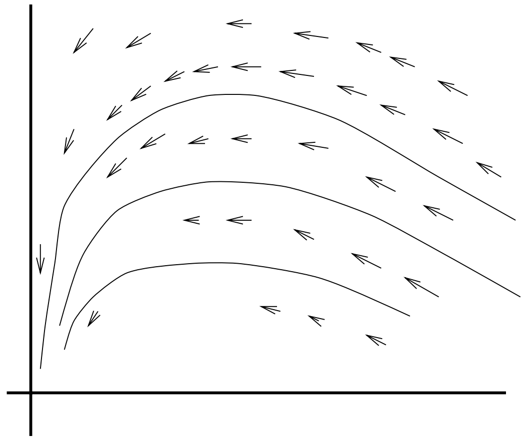


Figure 6.5: The Field and some solutions for the Epidemic Equations

progression from S to C to I to R, where I stands for the infective people, not just the infected ones.

We write down equations for the new system on the assumption that there is a fixed time, T, for a person who is infected to pass through the incubation phase before being able to infect others:

$$\frac{dS}{dt} = -rI(t)S(t) \tag{6.20}$$

$$\frac{dC}{dt} = rI(t)S(t) - C(t-T) \tag{6.21}$$

$$\frac{dI}{dt} = C(t-T) - \gamma I(t) \tag{6.22}$$

$$\frac{dR}{dt} = \gamma I(t) \tag{6.23}$$

This is a system of delay-differential equations. It is harder to find software which can handle this. A pity, because along with stochastic differential equations where there is a random element, they crop up rather a lot. In particular in epidemiology; it is rather important that there are drugs which can delay the onset of AIDS. There are obvious humanitarian reasons for increasing this time as much as possible. But the number of transmissions to new people is also proportional to the time delay. So by making the life of someone who has HIV more endurable, you spread the disease. This is one of the ethical dilemmas faced by the medical profession. There is no easy way out of it, but knowing the numbers is an important part of making an ethical decision. So models are life and death matters in some cases.

A less obviously nightmarish problem arises if the virus is a *meme* replicating in human minds instead of a gene replicating in a cell. Suppose the infection is a new religion. Or a fashion fad like nose rings or wearing jeans. The spread throughout a population is likely to follow the same laws as for the 'flu virus, although the time scale may be different. And two different religions or fashions may be mutually exclusive.

Exercise 6.3.4 *Set up equations for* two *religions in a closed world. Make various assumptions about the degree to which they are mutually exclusive and attractive.*

Some diseases also compete. Vaccination was started when Jenner discovered that people who had cow-pox seldom caught smallpox. So if you caught cow-pox first you recovered relatively unscarred, but if you caught it second you were likely to be scarred. If you died after getting smallpox you were, it is true, saved from cow-pox.

Exercise 6.3.5 *Investigate the propagation of two diseases in the case where one can recover from the first and not from the second.*

Does this have an application to religion?

Some readers may be appalled at the thought that their own sacred beliefs are being treated as a virus instead of divine revelation. If you are among them, consider. You have your beliefs, but you were not born with them you acquired them, you lucky person. In fact it is rather likely that you acquired them from your parents, or at least from some other individual who had the same beliefs. You are not alone in this respect. If we are concerned not with the truth of the beliefs but only on the number of people who have them, then either we regard the whole subject as something not to be thought about, or if we think about it, we shall be supposing that beliefs pass from one mind to another in much the same way as the common cold is passed between noses. Indeed, the spread of *any* belief, sacred or profane, would seem to be describable in the same language. Of course, threatening to burn people at the stake or chop their heads off, or even merely to assure them that backsliders burn in hell, is likely to increase the efficiency of the conversion process, and this has been tried with religions but not, so far as I know, with skate boarding or windsurfing. This may well have an effect on the dynamics by increasing the fraction of converts in a population.

6.4 Zeros of Vector Fields: Stability

The zeros of a vector field are the states of the system which are in equilibrium: if the system starts in such a state, then it stays in it. Of course, this is all a bit idealised. A billiard cue standing on one end is in equilibrium, but nobody sensible would expect it to stay there for long. So the question is, will a small perturbation of the state lead to the system returning to the state or would the system go further from it?

Other things can happen too: as in the case of the chemical kinetics, it is possible to have a whole curve which is an equilibrium. This was a stable equilibrium in a certain sense, because if you started away from it, you moved towards it.

In dimensions one and two, there are not too many possibilities, but in three dimensions and higher things get very nasty very quickly.

I shall not attempt to classify things properly, but instead look at the examples we have been looking at in the last few subsections.

For the species in competition we had

$$\dot{x} = max - mx^2 - cxy \tag{6.24}$$
$$\dot{y} = m'a'y - m'y^2 - c'xy \tag{6.25}$$

If this vector field is to have a zero, we have

$$\begin{aligned} cxy &= max - mx^2 && (6.26) \\ c'xy &= m'a'y - m'y^2 && (6.27) \end{aligned}$$

So $x = 0$ or $y = ma/c - mx/c$ and $y = 0$ or $x = m'a'/c' - m'y/c'$.

Clearly the point $x = 0, y = 0$ is a zero of the vector field. So is $x = 0, y = a'$. So is $x = a, y = 0$. And there is a fourth zero at the point where the line $y = ma/c - mx/c$ cuts the line $x = m'a'/c' - m'y/c'$.

To decide whether an isolated critical point of the field is stable, unstable or neither, again requires some linear algebra which you do not have. Alternatively we can look at the solutions on the *de* program, which is OK if (a) you trust the program and (b) you never intend to look at systems with more than two variables. There appears to be no good answer to this one except to hope that you learn some linear algebra some day so we can discuss stability properly.

We can get a little way by doing some crude approximations and some scruffy arguments which can be made respectable only with more mathematics than you have.

First look at the origin in the above case, and suppose we are going to look in a very close neighbourhood. Then we have

$$\dot{x} = max - mx^2 - cxy$$

Now the last two terms involve products of very small quantities, so are negligible in comparison so we can approximate the equation by

$$\dot{x} = max$$

Similarly the other can be approximated by

$$\dot{y} = m'a'y$$

Now the solutions to these are

$$x(t) = Ae^{mat}; \quad y(t) = Be^{m'a't}$$

and since $m, a, m', a' > 0$, points close to the origin and in the first quadrant are moved away from the origin. So the equilibrium point is unstable.

Now look at the point $x = 0, y = a'$. Since

$$\dot{x} = max - mx^2 - cxy$$

and $x << 1$ we have that this can be approximated by

$$\dot{x} = x(ma - a'c)$$

to first order. If $ma < a'c$ then this is stable in the x direction.

Since

$$\dot{y} = m'a'y - m'y^2 - c'xy$$

at $x = 0, y = a + \Delta$, for positive small Δ, we get that $\dot{y}$ is negative, and at $x = 0y = a - \Delta$ we find that $\dot{y}$ is positive; so the vector is a 'restoring' vector in the Y-direction.

This suggests that $(0, a')$ is stable provided $ma < a'c$.

Exercise 6.4.1 *Explore this using a suitable computer program. You should find that there are four qualitatively different kinds of vector field. One has a stable fixed point of the flow (zero of the vector field) somewhere where both populations have non-zero stable values. This will happen when both interactions are small. Another has a solution with only one stable fixed point of the flow when one species is driven to extinction, and a third is symmetric with the other species suffering inevitable extinct. And the last has two fixed points and the possibility of either being extinct and the other surviving, depending on the initial conditions. Draw pictures of all four cases. Explain what is happening in terms of restaurants or goldfish.*

The exploration of the other possibilities through algebra can be done, but is rather hard for you in your present state of innocence.

Exercise 6.4.2 *Examine the zeros, if any, for the case of the chemical kinetics and also the epidemics.*

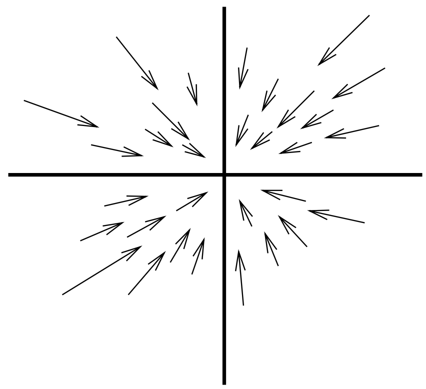


Figure 6.6: A stable fixed point of the flow

There are several distinct cases when we are looking at a zero of a vector field and trying to determine its stability properties:

Case 1: Stable

This is shown in figure 6.6.

Example 6.4.1

$$\dot{x} = -x \tag{6.28}$$
$$\dot{y} = -y \tag{6.29}$$

Case2: Totally Unstable

(The time reversed system is stable: figure 6.7)

Example 6.4.2

$$\dot{x} = x \tag{6.30}$$
$$\dot{y} = y \tag{6.31}$$

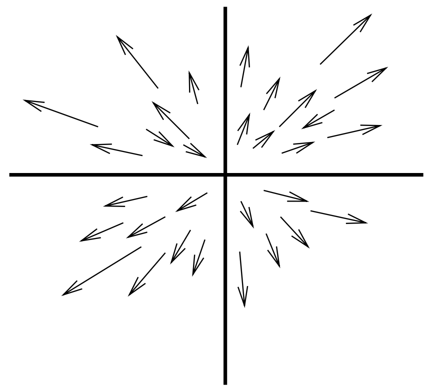


Figure 6.7: A very unstable fixed point of the flow

Case 3: A Hyperbolic Fixed Point- unstable

This is shown in figure 6.8. In one direction the flow is expanding and the vectors point out, in a transverse direction they are coming in and the flow is contracting.

Example 6.4.3

$$\dot{x} = x \tag{6.32}$$
$$\dot{y} = -y \tag{6.33}$$

Case 4: A Centre

Shown in figure 6.9.

Example 6.4.4

$$\dot{x} = -y \tag{6.34}$$
$$\dot{y} = x \tag{6.35}$$

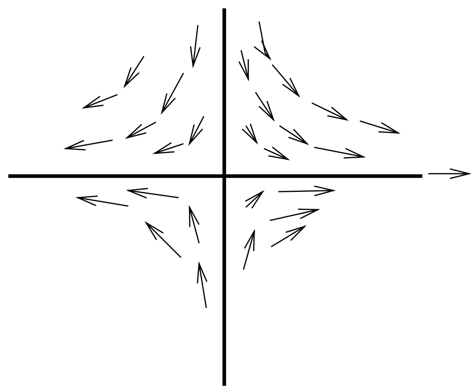


Figure 6.8: A hyperbolic fixed point of the flow

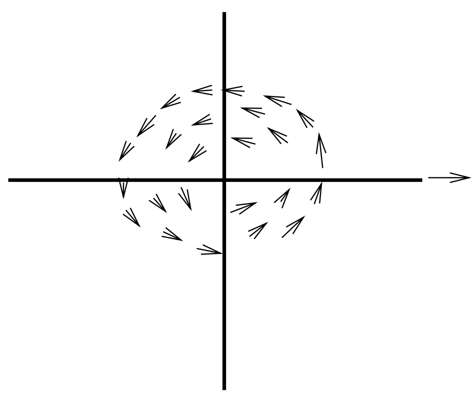


Figure 6.9: A centre point of the flow

We can allow the vectors in a stable fixed point to have a bit of curl to them so that instead of heading straight for the fixed point they spiral in towards it, and similarly for the others.

We can also have curves which are fixed, we can have limit cycles, we can have many different directions in which we are expanding and intermediate directions in which we are contracting, and in higher dimensions things horrible enough to give nightmares to all but the toughest. You can see some of the pictures in [19]. The book [20] is included as a reference just so that you can see some of the wilder ideas in the area. Personally I think Ralph is a bit of a whacko, but he isn't dull. If only his brother had been a physicist instead of a psychologist!

The fact that we can change the parameters in a vector field/ system of ordinary differential equations and wobble the vector field without changing it too much (sometimes) is of some importance. The qualitative theory is concerned with the kinds of *singularities* i.e. fixed points of the flows or zeros of the vector fields, and how different bits link up to other bits. This is very important given the natural tendency of not very well educated people to just bung a formula into a computer and not stop to ask if the answer is reasonable.

Chapter 7

Serious Modelling

7.1 Why Models Matter

The references [21], and [22], were pointed out to me by colleagues in Zoology and Psychology respectively. You should be able to see that although in this course we have only dealt with simple cases (because (a) you don't know much Mathematics, and (b) it is necessary to start with the simpler things before you move onto the complicated), there are many applications of enormous importance.

The paper [21] deals with the eminently practical problem of assessing the extent to which species threatened by extinction are at risk. There are some computer programs which lead the ecologist to measure things which have been found to be important and produce figures which help in an ecological management program. Getting quantitative ensures that scarce resources can be used as efficiently as possible. The delightful book by Anderson, [22] deals with a model for intelligence. This is not a dynamical model, and it is still premathematical, as is to be expected in a young science. Making it into a mathematical model will probably require some new mathematics. We are working on it.

The importance of models in the physical sciences is taken for granted by anyone who has studied any physics, chemistry or engineering. One I rather like was a study of power failures in the United States: an overloaded system can cause a switch to trip, increasing the load on other systems, causing more switches to break the circuits, and other systems to overload. Result, chaos over large areas of the planet. But the loads are usually a sign of bad weather, and it isn't economical to put in more power lines which are hardly ever used. The best solution is to have an elaborate computer control over the system which diverts the power supply more intelligently. Another system I had some involvement with was the optimum frequency to run fluorescent lamps. This involves some attempt to explain why the higher frequencies gave more light out for a given amount of power in, and the collection of data to estimate the parameters reasonably accurately. You must understand that whenever you switch a light on or strike a match, a huge amount of careful work went into designing the object, whether a light globe or the power system that delivers the electricity or the chemical plant that produced the inflammable material at the head of the match, or the design of the saw teeth that cuts the wood of the match with minimum wastage, and it is more than likely that mathematical and computer models were used to give you the finished product. Milk does not, despite the impressions of many simple folk, originate in supermarkets. It started off as grass and got operated on by a cow. It comes in paper-plastic laminates which involve lots of chemistry to make sure it doesn't dissolve in the milk and is cheap to produce. And the chemistry was not a matter of mixing things up at random to see what comes out. Molecules are designed these days. The feed for the cow has to be optimised, which requires study of cows. One of my colleagues has been involved with the job of estimating from sheep droppings the effective diet for sheep, a problem involving some statistical modelling. The milk was extracted from the cow by machinery which is carefully designed to get the most out for the least pain to the cow. The people who designed the machinery had to study cows and flow of milk both. Your morning breakfast cereal comes to you by courtesy of a lot of hard work, some of it mathematical.

The book *Chaos* by Gleick, [23] gives a popular introduction to Chaos Theory, and on pages 281-284 he discusses heart fibrillation as chaos. Gleick gives references to papers on the dynamics of physiological systems, including the heart, on page 337.

The beautiful book *Does God Play Dice?* by a well known Mathematician Ian

Stewart, [24] discusses heart models on page 277, where he points out that the Van der Pol equation[1] This is a system of differential equations, as in Chapter Six. You can plot them using a suitable program. If you do so, you will discover that there is a new type of behaviour, a *limit cycle*, The system was originally intended for the study of electronic oscillators, and has a bearing on the functioning of the heart. Notice how the same underlying equations keep turning up. Both books are fascinating, and warmly recommended if you would like to know more about Chaos. Gleick is a journalist and Stewart is a mathematician, and it shows. The mathematician writes much better prose.

Your life without technology would be nasty, brutish and short. And the technology requires science to underpin it, and the science requires mathematics. It requires other things as well (like huge amounts of patience and honesty), but the mathematics is indispensable.

7.2 How To Model

The steps to modelling some system by Difference or Differential Equations have already been mentioned, but it is worth repeating them now you have seen some more of them:

1. First decide if you are interested in the changing state of some system in time. If not, it may be that you don't need ODEs or difference equations. This doesn't mean that you don't have the option on modelling the system, just that you may need others sorts of mathematics.

2. Try to picture the system clearly and work out what properties of the system you wish to measure in order to describe it. This may involve

[1]The simplest form of the Van der Pol equation is:

$$\begin{aligned}\dot{x} &= y - \frac{x^3}{3} + x \\ \dot{y} &= -x\end{aligned}$$

some choices, and it may not be at all clear which things to choose. Keep it tentative, and be prepared to go back and try again.

3. Decide whether you prefer a discrete time system or a continuous time system

4. work out which are the measured quantities, the *state variables* which change in time. Write down any rules you can formulate which link the rate of change of the variables, or maybe higher derivatives, to other derivatives or to the state variables themselves. A qualitative description of the direction of change of variables should come first (wolves eat sheep and produce more wolves, sheep breed but get eaten by wolves). Then you need to write down some equations making these ideas precise:

$$S'(t) = \alpha S(t) - \beta S(t)W(t)$$

for example. This translates into algebra the statement: 'Sheep breed but get eaten by wolves'. Note that in this phase, you may have to make some simplifying choices. If you feel that the simplifications go too far, make a note of the need to improve the model, but keep it simple on your first pass. Make a list of the factors that increase the rate of change of each variable, and those that decrease it.

5. Check to see if all you know about the system has been translated into algebra. Pay careful attention to anything you know about the possible range of the variables, e.g. always positive.

6. Do not bother about the problem of solving the equations until you are pretty sure they are the right ones to solve. Go back and forward a few times to see if everything that can be said in English is translated correctly into algebra.

7. Make any estimates you can of plausible parameter values.

8. For two variable systems, examine the phase plane and use a program to investigate the time development from various starting places.

9. When you have explored the initial value dependencies, check to see if the actual system is going to be in a state which is sensitive to initial conditions, i.e. when nearby points go to different places.

10. Check to see if you have sensitivity to parameters. Does wobbling the parameters produce reasonably similar outcomes? Do this numerically using computer graphics when possible.

11. Check on the fixed points by algebra. Would you expect to have fixed points here? Are the fixed points stable or unstable? Can you sketch the solution using a little algebra, so you can check on whether the numerical program is behaving properly?

12. Check your model against reality wherever possible. If it isn't possible at all, why did you waste time on this nonsense?

13. Are discrepancies in the model likely to arise from the simplifications you introduced? Try developing the model to see if it can be made a better model by bringing in more reality, a bit at a time.

14. If the equations are *really* simple, as when you have put in some simplifications, look them up in a good book to see if they have a closed form solution. There are many excellent books giving solvable equations, although a preferred alternative these days is to use a computer package such as MATLAB or **Mathematica**. Do not however put your faith in a computer package. Sometimes they give daft answers.

There is nothing particularly obvious to the untrained mind about most of these steps, and it has taken a few centuries of thought by some of the best minds in the world to conclude that this approach works well. So don't think the ideas are obvious, and *practise* a lot to be able to do it yourself.

7.3 One Last Model: Days of Empire

Let me finish by giving one last example of how to put these ideas into practice. Just for fun, I look at the rise and fall of empires.

The Roman Empire was built upon the ruins of the Roman Republic, which in turn was built on some bad experiences the Romans had with their kings. The Roman Republic started out as a small city of people little distinguished from their neighbours. But they had a common purpose and a common threat of domination by superior forces. Over about four hundred years they grew more powerful, spreading out over the Italian peninsula and eventually taking on the great sea power of Carthage. Power struggles broke out between important men, Julius Caesar came to power (after conquering Gaul and taking a close look at Britain and deciding it was too cold and foggy) and was assassinated. His adopted son Octavian took over and became the first emperor, changing his name to Augustus. The emperors continued to rule Rome until it was conquered by the barbarians in the fifth century. The Eastern Empire continued in Byzantium until 1452 and the Turks took it using the new-fangled gunpowder to blow down the walls. Some say that the Renaissance was started by the scholars leaving Byzantium for the West around this time. Others have a quite different explanation. Altogether, from humble beginnings to final collapse there is a duration of about two thousand years, although the city of Rome was the focus of power for only about half that time.

There have been many empires, before and since. How do they start from small beginnings, grow large and then wither away? What factors are operating? If, as some say, we are taking part in the collapse of Western Civilisation, then there may be lessons to be learnt from thinking about these things. And then again, there may not.

Since we are concerned with the state of something (an empire) changing in time, it is suitable for modelling with difference or differential equations, so far as we can tell at this stage.

The first item having been ticked, we proceed to the second and try to get a clear picture of the subject matter. Building models is much concerned with clarification and trying to figure out what is relevant and what can be discarded.

There is clearly a huge amount to be understood here, all of history some might say, but let's try to clarify matters by abstracting something simple. Instead of asking for a philosophical kind of explanation of why people do what they do, let's look at it the way a scientist does and ask what we can measure.

There are a lot of possibilities; the area under Roman control is one. It can be assessed fairly well at different dates. A historian might object that this does not go to the underlying causes of why it came under Roman control. A military historian might say it was because the Romans won a lot of battles and expanded and then started losing them so contracted. Another might say that the Romans won because they had superior military organisation and the foreigners eventually caught on and copied it. Some historians like to argue that the Romans started out with high ideals and conquered other countries because they deserved to, but then became corrupted by power; Gibbon, [25], says their military valour was corrupted (by Christianity among other things), which is a more believable version of the moral superiority argument[2]. It is easy to measure the area under Roman control, rather harder to measure their moral virtue or superiority in military technology. Since the territory was conquered by force of arms, then using the terms in their broadest sense to include superior strategic capacity as well as the training of the legions, we could argue that the measure of land area was in fact related to the military advantage; if nation A has a military edge on nation B, and if they go to war, then the amount of land which A takes off B per annum might reasonably be supposed to be proportional to the military advantage. If the advantage is zero, then there will be a lot of drawn battles, or more or less alternate victories and defeats, so there will not be much net change. While there is any superiority, there will be an expansion, but this might slow down as the area held gets bigger, from the practical difficulties of keeping control of a frontier which is a long way away. Much of China's history consists of warring states, and Europe was very much the same, with only rare periods of any kind of total central control, and never any very complete control in Europe.

Let us suppose then that a small group of people somewhere hits on a piece of military technology; this may be the invention of bronze or iron, or it may be the invention of the phalanx or the legion. Let us suppose that the probability of winning a battle is proportional to the 'military edge', the difference in military competence of the two sides in any battle, and that each battle results in some fixed amount of territory coming into the possession

[2]I find it hard to understand why moral superiority should be regarded as a good reason for killing those who argue with you. On the other hand I can see that if you have gone out and killed a lot of people, some folk might figure they can get places by explaining to you that you were entitled to kill the poor devils because you are morally superior.

of the victors[3].

We cannot measure the 'military edge' directly, but we can infer that a country that keeps defeating its neighbours and taking territory off them must have it. In particular cases we can observe it directly, as when Britain invaded Australia. In other cases, such as when, fifty thousand years earlier, the present aborigines invaded Australia[4], it may be too far ago to detect directly.

If this were the only factor, then empires once started would continue until the entire planet had been conquered by the most advanced group of people. Two factors spring to my mind which would tend to stop this happening; one is the difficulty of holding territory that is too big for you to control, the other is that given time, the other side may learn your methods and copy them. If more than two factors spring to your mind, put them in the model. Other matters that spring to mind are that a small country might be superior to a large country in terms of its 'military edge' but get clobbered by sheer weight of numbers. So let's assume that the continent on which we perform our symbolic simulation of conquest is composed of warring tribes which are not organised[5].

The two factors which will limit the extent of empire act quite differently; the first would lead to an empire growing until it reached some equilibrium size and then holding onto it indefinitely. The second would mean that the passage of time along would ensure that eventually the edge would disappear. Human nature being what it is, one might cynically argue that the edge would be reversed somewhere, and some other group would take over an empire, since the first group might not do anything creative, having turned from a military empire into a bureaucracy run by eunuchs at some point. And even if the edge were reduced to zero, the extent of the empire which could be

[3]It is worth pointing out that a military edge might arise from something completely unconnected with weapons. Such as a well constructed clock, which in the past allowed some countries to navigate with more accuracy, prepare better charts, find less technologically advanced people, steal their wealth off them, and use it to buy more guns to defeat other, militarily equivalent, countries which didn't have quite such good clocks.

[4]according to some anthropologists

[5]Cortes and Pizarro took out huge chunks of real estate with ludicrously small armies working against large empires, so the assumption may not be so necessary if the 'edge' is big enough.

effectively governed might be expected to shrink. An acknowledged military edge would be a useful source of authority; take it away and people might be less respectful to the civil administration. You are only obedient to the politicians because you don't want to be put in jail by the cops.

We are working towards four quantities: first is the military edge, second is the controllability of territory, third is the rate of learning of what the military edge is based upon, and last is the amount of territory held. This represents a considerable effort of abstraction of what seem to me to be most likely significant, but many people would disagree with my choices. Pick your own. See what happens when they are put into a model. In other words, if you don't like my model, provide a better one.

So far, it is still not a quantitative model, but it is coming along.

Let us give the variables names, and let's choose single letters. Let M denote the 'military edge', and since it is going to change in time, it is a function, $M(t)$. Let $E(t)$ denote the size of the empire at time t. This is at least observable.

Let m denote the rate of collapse of the military edge in time due to the foreigners picking up imperial ways. It makes sense to have this as a constant. We can therefore suppose

$$\frac{dM}{dt} = -m$$

at least until $M(t) = 0$. So the military edge starts off from some value and falls off in time, fairly slowly in ancient times, much faster nowadays, until it gets to zero.

Now let's worry about the question of ease of maintenance of an empire. We can conjecture that there is some increase in the difficulty of adding extra territory which goes up as the amount of territory goes up. Let us suppose that when the military edge is positive, it is easy to add territory in proportion to what you already hold (more resources to exploit) but that when there is zero military edge you tend to lose territory in proportion to what you already have. This gives us the equation:

$$\dot{E}(t) = (aM(t) - b)E(t)$$

You might, again, want to set up a different system. Go ahead. It's fun, and there's nothing at this stage to say that I am right and you are wrong. So think hard, work out what seems a reasonable description of what the variables and their rates of change depend upon, and start writing equations. You will find it rather entertaining.

Since we have that $M(t)$ is simply a steady decrease from some initial value until it gets to zero when it stops, we can take constants m and c and write

$$M(t) = c - mt$$

remembering that this only works for $M(t) \geq 0$, and that after this, $M(t)$ stays at zero.

This gives us the observable equation:

$$\dot{E}(t) = (ac - amt - b)E(t)$$

which with some names of constants changed becomes:

$$\dot{E}(t) = (A - kt)E(t) \tag{7.1}$$

I show the graph of $E(t)$ against time in figure 7.1 for the case $A = 1, k = 0.15$, and the starting point $y = 0.3$. The edge of the square in which I drew the curve is visible on the right hand side, but the shape is roughly symmetrical and returns to zero.

One interesting result of this model is the strong dependency on the parameters A, k.

k represents the rate at which the foreigners catch up on your technology. Quite small changes in k lead to considerable changes in the size of the empire and its duration. Bigger empires generally last longer when they arise from a relatively small military edge as with the Roman Empire. The British Empire had a much larger military edge, representing about a century of technological advance in the case of India and China and about twelve thousand years in the case of Australia. Figures 7.1 and 7.2 represent a change from $k = 0.15$ to $k = 0.2$. Figure 7.3 has an increase of A to 1.03 while $k = 0.15$ still. So a

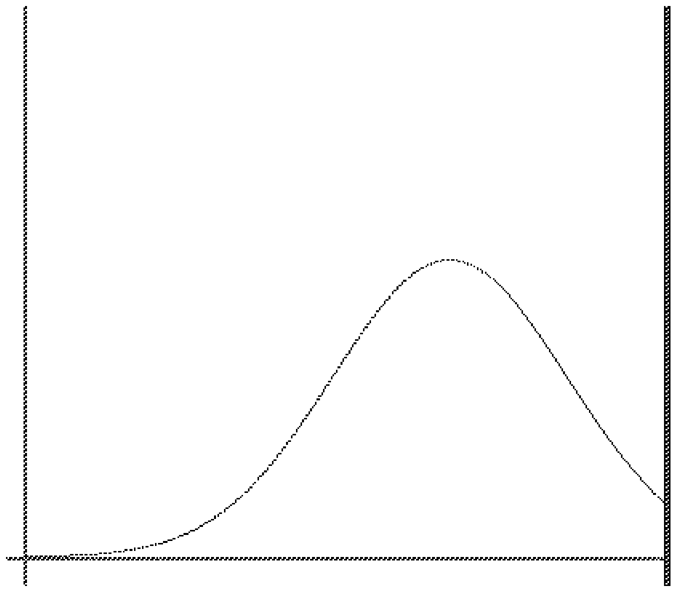


Figure 7.1: The rise and fall of Empire

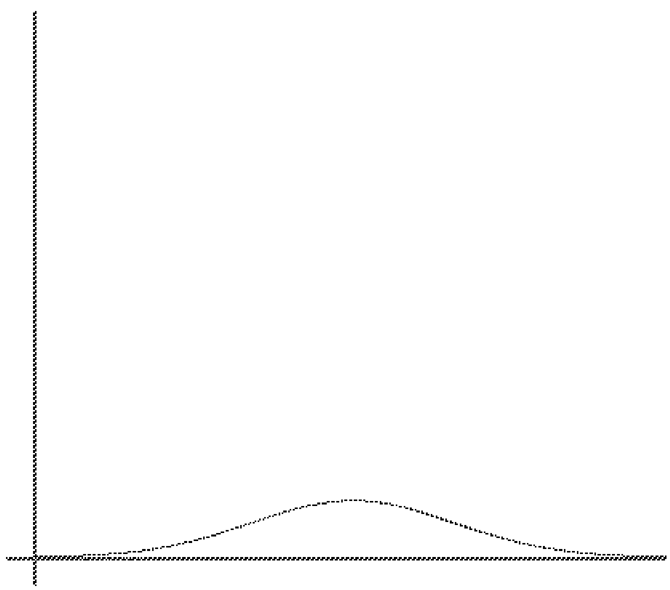


Figure 7.2: Faster Learning

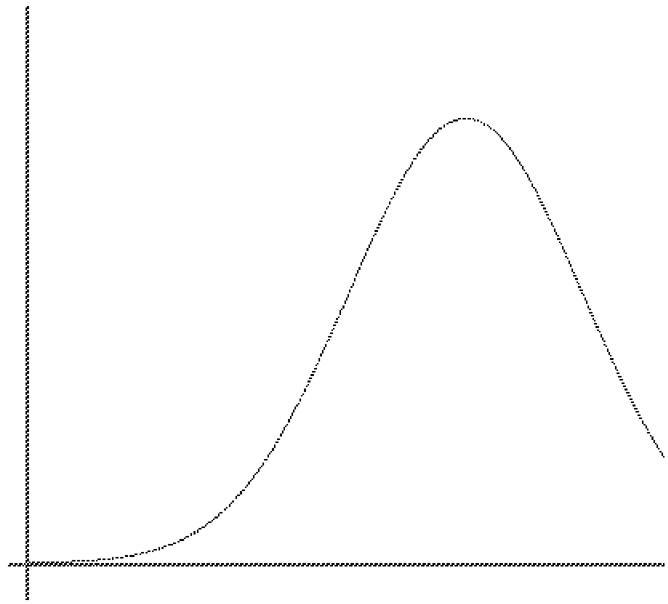


Figure 7.3: Bigger initial edge

small increase in the military edge can have a profound effect on the size of the empire.

It is not too hard to see that the solution to the equation 7.1 is a scaled normal or Gaussian function. Old empires never die, they only fade away. Or get eaten by later empires, a matter we have not considered. India was in the last stages of decay of the Moghul Empire just before it got incorporated into the British Empire.

Exercise 7.3.1 *Calculate the height, mean and variance of the function which solves equation 7.1.*

7.4 Concluding Thoughts

I do not present the above example (which I made up, for fun) as a serious model of empires, but it is no dafter than many other models which may be found in textbooks, and more interesting (to me) than most. It is notable that my model has the peak in the middle of the empire, which didn't seem to hold for the British Empire unless you reckon that Australia and Canada are still part of it, which is not as daft as it may sound[6]. I do not know of any good studies showing a graph of the extent of the Roman Empire as a function of time, but it was at its greatest extent around the time of Trajan, around 100 A.D, If you go from the start of the republic around 350 B.C to the fall of the City to the Visigoths in 410 AD, you are not too far out.

The trouble with all these numbers is that you can argue about the values all night long. Undergraduates often find this fun, but after a while it seems unsatisfying. Do you feel that the Eastern Empire was part of the Roman Empire or not? Are Canada and Australia going to be regarded as currently part of the British Empire by a historian in a thousand years from now writing on Ganymede? To make the model more testable, one would have to define the term 'empire' a good deal more carefully. There is a case for saying that the above model is not susceptible of being checked against reality and is hence totally pointless.

Well, not totally. It gave me some innocent fun, and it may have helped show you how one can go about constructing models just for the hell of it. Rather a lot of the creative things that have changed the world have been done for the sheer fun of playing around. Bright people like playing with ideas, dumb people like playing with balls. Really dumb people like watching other people play with balls.

[6] There is a case for including the United States also.

Bibliography

[1] Richard Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 1976.

[2] Richard Dawkins. *The Blind Watchmaker*. Longmans Scientific, 1986.

[3] Richard Dawkins. *The Extended Phenotype*. Oxford University Press, 1989. 575 1989 EXT.

[4] Goldstein, Lay, Schneider. *Calculus and It's Applications*. Prentice-Hall, New York, seventh edition, 1996.

[5] B.F. Skinner. *Beyond freedom and dignity*. Cape, London, 1972. 150.19434 1972 BEY.

[6] B.F. Skinner. *About Behaviourism*. Cape, London, 1974. 150.19434 1974 ABO.

[7] B.F. Skinner. *Upon Further Reflection*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987. 150.19434 1987 UPO.

[8] Hans Eysenck. *Sense and Nonsense in Psychology*. Penguin, 1957.

[9] Hans Eysenck. *Uses and Abuses of Psychology*. Penguin, 1953.

[10] H.J. Eysenck. *Fact and Fiction in Psychology*. Penguin, London, 1965. 150 1965 FAC.

[11] David Marr. *Vision*. W.H.Freeman, New York; San Francisco, 1982. 152.1402854 1982 VIS.

[12] Walter C. McCrone. Light Microscopical Study of the Turin "Shroud". *Microscope*, 28:105–128, 1980.

[13] Walter C. McCrone. Light Microscopical Study of the Turin "Shroud". *Microscope*, 29:19–38, 1981.

[14] Damon, P.E. et al. Radiocarbon Dating of the Shroud of Turin. *Nature*, (337):611–615, 1989.

[15] James T Sandefur. *Discrete Dynamical Systems*. Clarendon Press, Oxford, 1990.

[16] James R. Newman. *The world of mathematics*, volume I-IV. Simon and Schuster, New York, 1956. 510.9 V. 1.

[17] F.D. Abraham. *A Visual Introduction to Dynamical Systems Theory for Psychology*. Aerial Press, inc, PO Box1360 Santa Cruz, CA., 1990.

[18] M.S. Burghes, D.N. Borrie. *Modelling with Differential Equations*. Ellis Horwood, 1981.

[19] Abraham R; Shaw C. *Dynamics, The Geometry of Behaviour*, volume 1,2,3,4 of *The Visual Mathematics Library*. Aerial Press Inc, Santa Cruz, 1990.

[20] R Abraham. *On Morphodynamics*. The Science Frontier Express Series. Aerial Press, Santa Cruz, 1985.

[21] T.W. Possingham H.P Lindenmayer, D.B Norton. A framework for the improved management of threatened species based on population viability analysis. *Pacific Conservation Biology*, 1:39–45, 1993.

[22] Mike Anderson. *Intelligence and Development*. Oxford, 1992. 153.9 1992 INT.

[23] James Gleick. *Chaos*. Cardinal, New York, 1987.

[24] Ian Stewart. *Does God Play Dice?* Basil Blackwell, Oxford, 1989.

[25] Edward Gibbon. *Decline and Fall of the Roman Empire*. Various, 1776.