Maths Makes A Difference
Making predictions in an uncertain world

This is a summary of the content of the Maths Makes A Difference webinar help on Tuesday 2nd February. Maths Makes A Difference is a collaboration between the outreach teams based at the Maths departments at Oxford and Cambridge.

For more resources linked to this webinar click here.

For more information on Maths Makes A Difference click here.

Introduction

There are many ways in which predictions are important. You could be a buyer for a chain of supermarkets trying to work out how many ice creams you will need in 4 month's time or you could be trying to model how much the Earth's temperature will increase by in the next 100 years. You could work for a council and need to order grit for next winter or you could work in a bank trying to predict what will happen to the price of some stocks over the next few hours (or minutes?).

We have all seen in the news recently how important mathematical models can be, and the predictions from these have had massive implications for our lives. Prediction is very important, but it can also be very difficult! On the 14th January 2021 the BBC weather team published an article entitled Another Beast from the East on the way? in which they talked about the possible effects of a "Sudden Stratospheric Warming" and what the implications might be with one possible outcome being very cold weather and snow for the UK. In the article they stressed how unpredictable the forecasts were, but on the 7th and 8th of February the second "Beast from the East" arrived (it was a more tame beast than the 2018 one, but it did mean some snow for Cambridgeshire!).

Trying to model the weather is difficult, and uses university level maths and a supercomputer (or even a Group of Supercomputers). Instead we are going to investigate a model for the population size of a certain creature. Maths has an important role to play in understanding population dynamics, both analysing data about existing populations and also making predictions. For example, as the climate changes, we need to understand the consequences for populations of different species.

What do we mean by modelling the population? A mathematical model in this context means a system of one or more equations, where we feed in the data we already have, and the equation or equations then tell us how we expect the population to grow or shrink over the next period of time. In this webinar, we modelled the population of a creature that we had made up, called the **Quizzle**.

The trouble with Quizzles

The Quizzle (pictured below) is a guinea-pig sized creature which breeds once a year in the summer.

We are going to model the population of Quizzles with a single equation which shows how the number of Quizzles in one year depends on the number of Quizzles the year before. We might decide that the number of Quizzles is proportional to the number of Quizzles the year before, which kind of makes sense, as we would expect the more Quizzles we have, the more Quizzlets (*Quizzlet = baby Quizzle*) we have the following year. This gives a possible equation:

$$y_{n+1} = k y_n$$

where y_n stands for the number of quizzles in year n, so $y_3 = 2450$ means that after three years we have 2450 Quizzles. The initial population of Quizzles is y_0, so if $y_0 = 1000$ it means that the number of Quizzles that we are starting with is 1000.

The k in the equation is a *parameter* that measures how quickly the population of Quizzles grows (or declines) and could be affected by how many Quizzlets they tend to have, or by how harsh their environment is. If we were trying to model real data then we would probably look at past data on the population of Quizzles and choose a k value that fits this. We do run into problems with this equation though, if we take $k = 2$ then the number of Quizzles doubles each year and eventually the surface of the Earth will be covered in a Quizzle carpet!

To make the model more accurate we are going to add in an extra term which will start to restrict the growth when the number of Quizzles starts to gets large. To do this, it is helpful if we use a new variable, x_n, which represents the proportion of the maximum possible number of Quizzles that we have in a given year. We could work out a theoretical maximum number of Quizzles (perhaps based on their habitats, amount of food etc) and then $x_3 = 0.5$ would mean that in year 3 the number of Quizzles is half of the maximum possible number. Similarly to before, x_0 is the starting value, and represents the proportion of the maximum number of Quizzles that we have at the start.

Our new equation is:

$$x_{n+1} = kx_n(1 - x_n)$$

With this equation, if x_n starts to get close to 1 (so the number of Quizzles is getting close to the theoretical maximum) then $(1 - x_n)$ becomes close to 0 and this works to restrict the growth of the Quizzles. This could be because as the number of Quizzles starts to get large the amount of food starts to run out, or if the number of Quizzles gets large it could lead to an increase in the number of Quizzle-eaters which then start to eat larger numbers of Quizzles. Like before, k is a parameter which could depend on reproduction rate, food availability, predators etc.

This equation looks very simple, but was used by ecologists and biologists in the 1950s to model different populations, for example W E Ricker used the equation to model the population in fisheries before going onto develop the Ricker model. It can produce some surprising effects as we will see later!

The Logistic Map

The equation

$$x_{n+1} = kx_n(1 - x_n)$$

is called the *logistic map*. Let's run through an example of how to use it.

We will take $k = 1.5$ and $x_0 = 0.8$. The first few years are:

$$x_0 = 0.8$$
$$x_1 = 1.5 \times 0.8 \times (1 - 0.8) = 0.24$$
$$x_2 = 1.5 \times 0.24 \times (1 - 0.24) = 0.2736$$
$$x_3 = 1.5 \times 0.2736 \times (1 - 0.2736) = 0.29811456$$

If you keep repeating this process, it looks as if the population eventually settles down to $\frac{1}{3}$ of the maximum population.

Try yourself: if $k = 2$ and $x_0 = 0.3$, what is the value of x_3? What happens as the number of years increases? What happens if you use a different x_0?

Working out values using a calculator gets tedious after a few iterations. You could use a spreadsheet to help you investigate populations, or you could use the GeoGebra pages we have created.

GeoGebra page with a slider for k. This page shows us graphically what happens to a population over 30 years. You can input an initial value x_0 and use a slider to change k.

GeoGebra page with an input box for k. This page is the same as the previous one, but this time you type in k.

Have a play with the above GeoGebra pages. Below are some questions you might like to think about:

1. Can you find a parameter so that the population always ends up dying out? What can you say about parameters with this property?

2. Can you find a parameter so that the population settles to a non-zero constant value? What can you say about parameters with this property?

3. Can you find a parameter so that the population eventually oscillates between two values? Or eventually cycles between three or four values?

4. Why have we set the slider for parameter k so that it can only go between 0 and 4?

Probably my favourite pair of values are $k = 3.68$ and $x_0 = 0.7$, which shows some really odd behaviour!

Measurement Errors

With the model that we have for the Quizzle population, if we know exactly how many Quizzles there were at the start then we can accurately tell how many Quizzles there will be at any time in the future. However we might need to use an estimate for the initial size of the population. When predicting complicated systems such as weather patterns there may be a lot of initial conditions (wind speeds, temperature, humidity etc) which will not be measured exactly.

This GeoGebra page allows you to input three initial populations at once, which can be useful if you want to investigate the effect of errors in the initial values.

Try investigating the effect of errors for the following models:

- $k =$1.5 with initial populations in the range $0.6 \leqslant x_0 \leqslant 0.8$ (this means take any values between 0.6 and 0.8 inclusive).

- $k =$3.5 with initial populations in the range $0.6 \leqslant x_0 \leqslant 0.8$

- $k =$3.7 with initial populations in the range $0.6 \leqslant x_0 \leqslant 0.8$

You will probably have found that in the first case, no matter what your initial value is (as long as it is not 0) the population settles down to $\frac{1}{3}$ of the maximum population.
For the second case, it looks at if the population ends up oscillating between two values, and the initial condition determines which of the two values you end up on after 30 years.
In the third case it looks as if you can end up almost anywhere after 30 years. Perhaps a more accurate measurement of the initial condition might help?
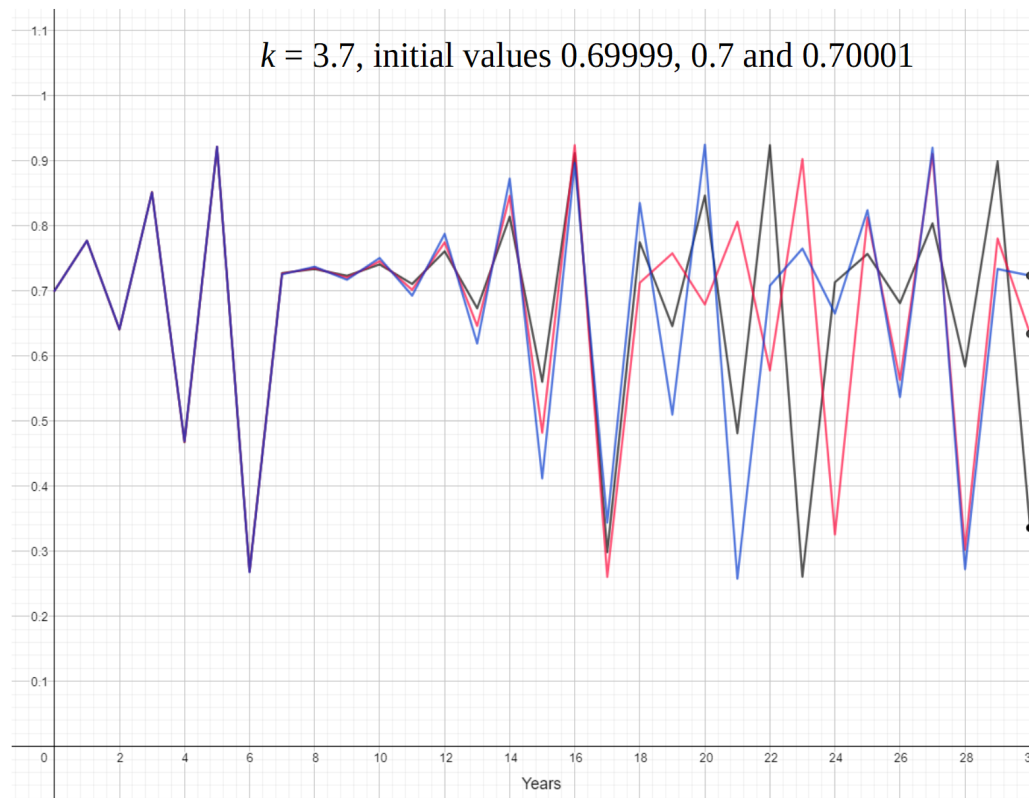
Chaos and the Butterfly Effect

Instead of having initial values in the range $0.6 \leqslant x_0 \leqslant 0.8$, lets try having a more accurate estimate of the initial conditions.

Try:

- $k =$3.7 with initial populations in the range $0.69 \leqslant x_0 \leqslant 0.71$

- $k =$3.7 with initial populations in the range $0.699 \leqslant x_0 \leqslant 0.701$

- $k =$3.7 with initial populations in the range $0.6999 \leqslant x_0 \leqslant 0.7001$

You will probably find that as the initial population estimate gets better, the model agrees for the first few years, but as the number of years increases the populations diverge. Even with an error of less than 0.1% (as in the last case) the population models only agree for the first few years. The picture below shows what happens when we take $k = 3.7$ and initial conditions $x_0 = 0.69999, 0.7$ and 0.70001.

$k = 3.7$, initial values 0.69999, 0.7 and 0.70001

Chaos theory is the study of systems where the rules are well defined (as in our population model!), but very small changes in initial conditions can have very big differences in outcomes. This is sometimes refereed to as *The Butterfly Effect*, from the title of Edward Lorenz's 1972 paper "Predictability: Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" (by which is meant that tiny differences in initial conditions, such as whether a butterfly flaps its wings or not, can lead to very large differences in outcomes, such as whether there is a tornado or not). Lorenz was a meteorologist who worked on simulating weather patterns, and in 1961 discovered that if he replaced input values with rounded versions (e.g. 0.506 rather than 0.506127) then his simulation produced very different results.

During the webinar we saw that sometimes a model can exhibit chaotic behaviour. We saw with the logistic map that for values of k around 3.7, a tiny difference in the input can lead to a huge difference in the output. This doesn't have to be a failure of the model: sometimes this is just how the real world is! We see this happening with the weather. In theory, if we knew the exact atmospheric conditions right now, we could confidently predict how the weather would evolve. But it's impossible to know the atmospheric conditions really precisely. Sometimes that doesn't matter: the behaviour is quite stable, and the uncertainty about the starting values doesn't make too much difference to whether it'll rain next Tuesday. But sometimes it's a real problem. You might have heard weather forecasters commenting that they're unsure about the weather over the next couple of days, and that's because these uncertainties about the starting values then propagate into major differences in the weather. Weather forecasters handle this by using ensemble methods . . . but that's a story for another time!

7

Where next?

The logistic map is a 'discrete-time' model: we jump from the population in one year to the population in the next, rather than going smoothly between them. In practice, a lot of real-world situations need continuous models rather than discrete, and this is where calculus comes in, in the form of differential equations. You might have already studied differentiation, and as you progress through your studies you'll meet differential equations. If you study maths at university (or various other subjects that use maths), you'll study differential equations in much more detail, because they are so crucial for modelling the real world: weather, populations, epidemiology, the motion of the planets, finance, ... Mathematical biology is a whole branch of maths, which you might be able to study during a maths degree, and where there's lots of current exciting research. And using maths to make predictions is relevant for lots of careers outside of universities too.

Appendix - Cobweb Diagrams

A *cobweb diagram* is another way that you can explore what is happening with the logistic map.

In the diagram below, the purple line is the graph of $y = kx(1-x)$ and the blue line is the line $y = x$. The initial population proportion is $x_0 = 0.2$. To find the value of x_1 move vertically up to the line $y = x(1-x)$ — the coordinates of this point will be (x_0, x_1). Then move across to $y = x$, the coordinates of this point will be (x_1, x_1). Then keep repeating the process!
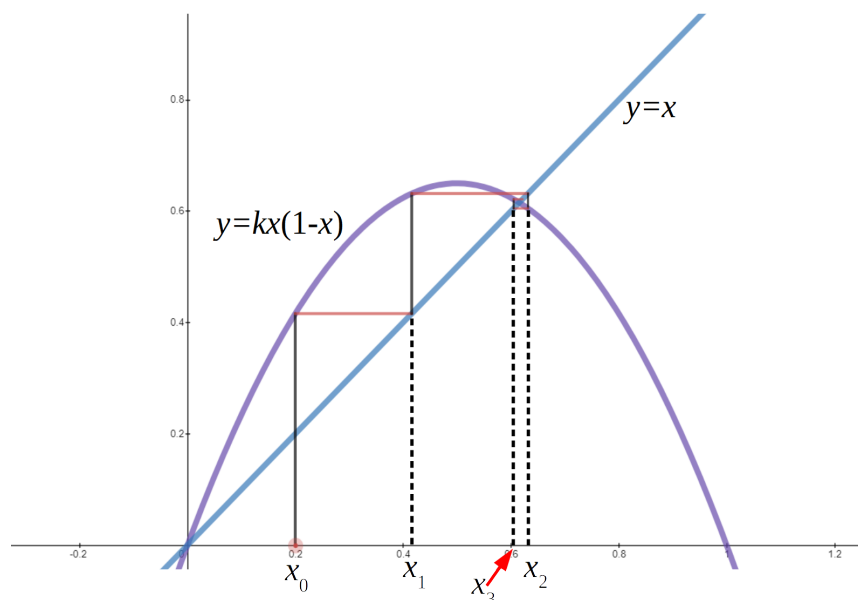
For the situation in the graph above, as the process is repeated the population proportion converges to a fixed point, which is where the graphs of $y = kx(1-x)$ and $y = x$ intersect. In other situations the process gets "stuck" in a square about the fixed point (which is when the population proportion oscillates between two values), or the population proportion behaves chaotically. For other values of k there will be oscillations about more than 2 points.
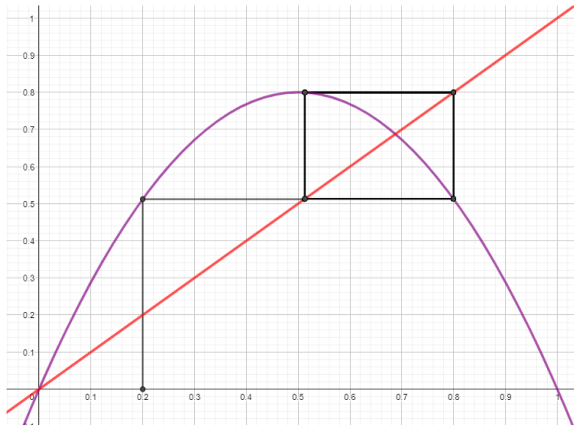
Cobweb diagram showing oscillation
between two values
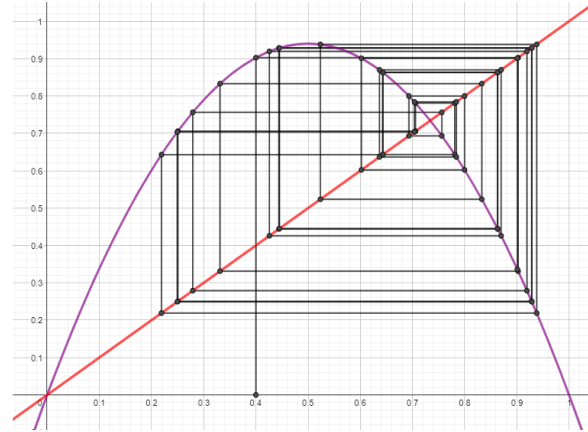
Cobweb diagram showing chaos

You can find two interactive cobweb diagrams here:

- Desmos Cobweb Diagram — this is the one used in the webinar - it only shows a few iterations

- Geogebra Cobweb Diagram — this shows 30 iterations (like the Geogebra graphs). As well as inputting an initial value you can drag the value on the x-axis along to change it.