Exploring modeling with data and differential equations using R

John M. Zobitz

2021-02-02

Contents

Chapter 1

Welcome

Chapter 2

Creating this from Bookdown to Github

I stumbled on how to do this by going to this website and built this out from here. I am glad that I will be able to work on this.

Part I

Models with Differential Equations

Chapter 3

Models of rates with data

3.1 What is the book about?

The focus of this textbook is understanding *rates of change* and how you can apply them to model real-world phenomena. In addition this textbook focuses on *using* equations with data, building both your competence and confidence to construct a mathematical model from data and a context.

I can imagine that your first sustained encounter of rates of change was in your calculus course, perhaps answering the following types of questions:

- If $y = xe^{-x}$, determine the derivative function $f'(x)$.
- Where is the graph of $\sin(x)$ increasing at an increasing rate?
- If you release a ball from the top of a skyscraper 500 meters above the ground, what is its speed when it impacts the ground?

Some of these mathematical questions derive from models from physical phenomena, such as the ball falling off a skyscraper which assumes that acceleration of the ball is constant. This assumption is typically a starting point to build a mathematical model. Using acceleration, the velocity (or the antiderivative of acceleration) can be found, from which the position function can be calculated through antidifferentiation.

However many times we observe data and *then* construct a mathematical model to corroborate our observations. Some of the these models come from well-understood physical phenomena (such as the case of the falling ball).

3.2 Modeling in context: the spread of a disease

To see what the first steps would be, consider the data in Figure 3.1, which come from an Ebola outbreak in Sierra Leone in 2014.
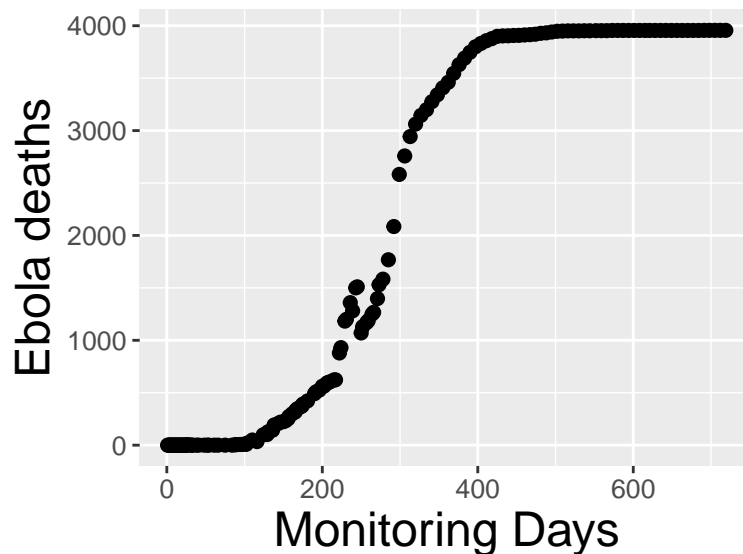
Figure 3.1: An Ebola outbreak in Sierra Leone

These data represent the deaths due to Ebola in Sierra Leone. Constructing a model from disease dynamics is part of the field of mathematical epidemiology. How we construct a mathematical model of the spread of this outbreak largely depends on the assumptions underlying the dynamics of the disease, such as considering the rate of spread of Ebola. This model also depends on the spatial scale studied as well - how a model is constructed depends if we wish to examine the spread of the disease in an individual immune response, or understand general principles for the spread through a population.

Basic questions that one can ask is if the rate of deaths due to Ebola proportional to:

- The number of people infected?
- The number of people not infected?
- The number of infected people coming into contact with those not infected?

Let's see what each of these mathematical models would look like if we wrote down an equation. Since we are discussing *rates* of infection, this means we will need a *rate of change* or derivative. Let's use the letter I to represent the number of people that are infected.

3.2.1   Model 1: Infection rate proportional to number infected.

In the first case (the rate of infection proportional to the number of people infected), to translate that statement into an equation would be the following:

$$\frac{dI}{dt} = kI \qquad (3.1)$$

where k can be thought of as a proportionality constant, with units of time^{-1} for consistency. This is an example of a *differential equation*, which is just a mathematical equation with rates of change.

Differential equations may look different because what we are solving for is the function $I(t)$.[1] Later on we will examine how to "solve" a differential equation (which means we determine the family of functions consistent with our rate equation).

Notice the proportionality constant k - we call this a *parameter*. We can always try to solve an equation without specifying the parameter - and then if we wanted to plot a solution the parameter would also be specified. In some situations we may not be as concerned with the particular *value* of the parameter but rather its influence on the long-term behavior of the system (this is one aspect of bifurcation theory). Otherwise we can use the collected data shown above with the given model to determine the value for k. This combination of a mathematical model with data is called *data assimilation* or *model-data fusion*. How exciting!

Before we think about possible solutions let's try to reason out if the first model would be plausible. This model states that the rate of change (the amount of increase) gets larger the more people that are sick. That does seem reasonable as a model, but perhaps unreasonable in real life. In the case of Ebola or any other infection disease, stringent public health measures would be put in place before large numbers of people die. We would expect that the rate of infection would decrease and the number of deaths to slow. So perhaps the second model might be a little more plausible. At some point the number of people who are *not* sick will reach zero, making the rate of infection be zero (or no increase).

### 3.2.2 Model 2: Infection rate proportional to number NOT infected.

In this description notice how we are talking about people who are sick (which we have denoted as I) and people who are *not* sick. This looks like we might need to introduce another variable for the "not sick" people, which we will call S, or susceptible. So the differential equation we would write down would be:

$$\frac{dI}{dt} = kS \qquad (3.2)$$

We are still using the parameter k as with the previous model. Also note we introduced the second variable S is in Equation (3.2). Because we have

---

[1]You may be used to working with *algebraic equations* (e.g. solve $x^2 - 4 = 0$ for x). In that case the solution can be points (for our example, $x = \pm 2$).

introduced another variable S we should also include a differential equation for how S changes as well. One way that we can do this is by considering our entire population as consisting of two groups of people: S and I. Infection brings someone over from S to I, which we have in this diagram:

$$\boxed{S} \xrightarrow{k} \boxed{I}$$
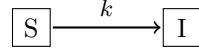
There are three reasons why I like to use diagrams like these: (1) they help organize my thinking about a mathematical model (2) any assumed parameters are listed, and (3) they help me to see that rates can be conserved. In other words, if I enter into the box for I, then someone is leaving S. In other words, $\frac{dS}{dt} = -kS$. So the two equations together can be represented as:

$$\frac{dS}{dt} = -kS$$
$$\frac{dI}{dt} = kS$$

This differential equation is what we would call a *coupled differential equation.* In order to "solve" the system we need to determine functions for S and I. This coupled set of equations looks a little clunky, but we do notice something cool. Algebriacally we have:

$$\frac{dS}{dt} + \frac{dI}{dt} = \frac{d(S+I)}{dt} = 0 \tag{3.3}$$

Recall from calculus that if a rate of change equals zero then the function is constant. In this case, the variable $S+I$ is constant, or we can also call $S+I = N$, the number of people in the population. This means that $S = N - I$, so we can re-write our differential equation in one equation:

$$\frac{dI}{dt} = k(N - I) \tag{3.4}$$

This second model does have some limiting behavior to this model as well. As the number of infected people reaches N (the total population size), the values of $\frac{dI}{dt}$ approaches zero, meaning I doesn't change. There is one caveat to this - if there are no infected people around ($I = 0$) *the disease can still be transmitted*, which might make not good biological sense.

3.2.3 Model 3: Infection rate proportional to infected meeting not infected.

The third model rectifies some of the shortcomings of the second model (which rectified the shortcomings of the first model). This model states that the rate of

infection is due to those who are sick, actually infecting those who are not sick. This would sort of scenario would also make some sense, as it focused on that *transmission* of the disease are between susceptibles and infected people. So if nobody is sick ($I = 0$) then the disease is not spread. Likewise if there are no susceptibles ($S = 0$), the disease is not spread as well.

In this case the diagram outlining this approach looks something like this:

$$\boxed{\text{S}} \xrightarrow{k} \boxed{\text{I}}$$
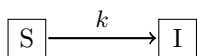
The differential equations that describe this scenario are the following:

$$\frac{dS}{dt} = -kSI$$
$$\frac{dI}{dt} = kSI$$

Just like before for Model 2 we can combine the two equations to yield a single differential equation:

$$\frac{dI}{dt} = k \cdot I \cdot (N - I) \tag{3.5}$$

Look's pretty similar to model 2, doesn't it? In this case notice the variable I outside the expression. Notice this seems to be appropriate - if $I = 0$, then there is no increase in infection. If $I = N$ (the total population size) then there is no increase in the infection.

3.3   The qualitative nature of solution curves

So far we have primarily been focused on the qualitative understanding of the different models. One way we can look at how these different models work together is by plotting $\frac{dI}{dt}$ versus I. I know we have the parameters k and N to specify, but let's just set them to be $k = 1$ and $N = 10$ respectively. Plots of these functions are shown in Figure 3.2.
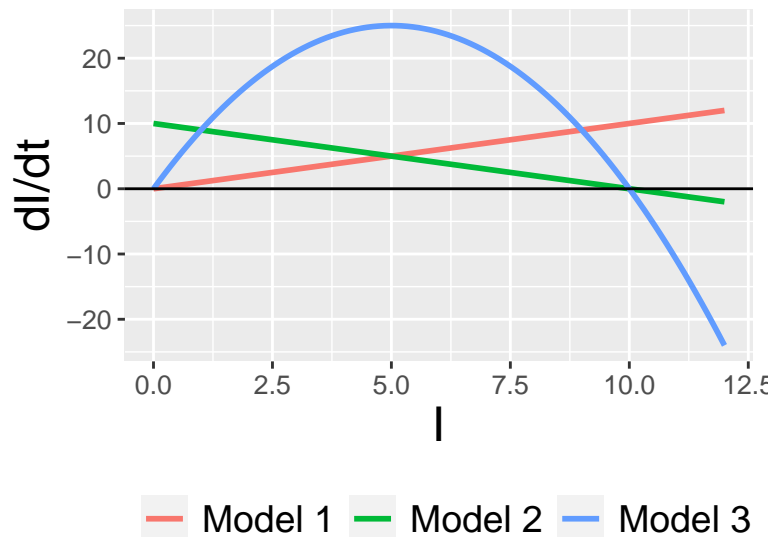
Figure 3.2: Comparing rates of change for three models

There is a lot that we can tell from this figure. Notice how the sign of $\dfrac{dI}{dt}$ is always positive for Model 1, indicating that the solution (I) is always increasing. For Models 2 and 3, $\dfrac{dI}{dt}$ equals zero when $I = 10$, which also is the value for N After that case, $\dfrac{dI}{dt}$ turns negative, meaning that I is decreasing.

In summary, we can tell a lot about the *qualitative behavior* of a solution to a differential equation even without the solution.

3.4   Simulating a differential equation

Let's talk solutions. One thing to note is that usually a differential equation also has a starting, or an initial value that actualizes the solution. When we state a differential equation with a starting value we have an **initial value problem**. For our case here we will assume that $I(0) = I_0$, where this is also another parameter at our disposal.

With that assumption, we can (and will solve later!) the following solutions for these models:

Model 1 (Exponential): $I(t) = I_0 e^{kt}$

Model 2 (Saturating): $I(t) = N - (N - I_0)e^{-kt}$

Model 3 (Logistic): $I(t) = \dfrac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}}$

```
## Warning: Removed 63 row(s) containing missing values (geom_path).
```
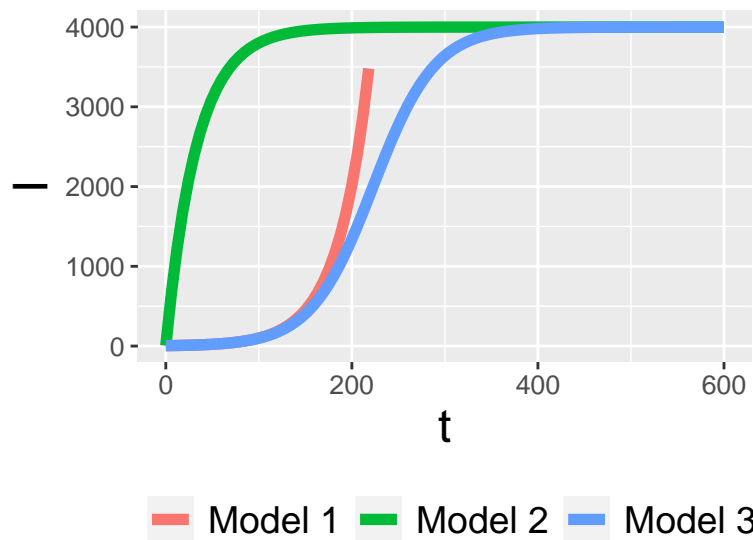

Figure 3.3: Three models compared

In Figure 3.3, I plot these solutions over the course of several days, using $k = 0.03$ and $N = 4000$ and $I_0 = 5$. Notice how Model 1 increases quickly - it actually grows without bound off the chart! Model 2 and Model 3 have saturating behavior, but it looks like Model 3 might be the one that actually captures the trend of the data. Models 2 and 3 are more commonly known as the **saturating** and **logistic** models respectively.

3.5   Which model is best?

All three of these scenarios describe different modeling scenarios. While we haven't solved these differential equations, we do have some intuitive sense of what could occur. With the saturating and logistic models (Models 2 and 3) we have some limiting behavior the possibility that the the rate of infecion slows. There are several possible models that on the surface seem plausible, but which one is the *best* one? We will also address that question later on in this textbook when we discuss *model selection*.

Model selection is one key part of the modeling hypothesis - where we investigate the implications of a particular model analyzed. If we don't do this, we don't have an opportunity to test out what is plausible and what is believeable in our models.

Exercise 3.1. Solutions to an outbreak model of the flu are the following:

$$\text{Saturating model: } I(t) = 3000 - (2990)e^{-.1t}$$
$$\text{Logistic model: } I(t) = \frac{30000}{10 + (2990)e^{-.15t}},$$

where t is in days. Make a plot of both of these models for $0 \leq t \leq 100$. How would you describe the growth of the outbreak as t increases? How many people will be infected overall? Finally, evaluate $\lim_{t \to \infty} I(t)$. How do these results compare to values found on your graph?

Exercise 3.2. The general solution for the saturating and the logistic models are:

$$\text{Saturating model: } I(t) = N - (N - I_0)e^{-kt}$$
$$\text{Logistic model: } I(t) = \frac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}},$$

where I_0 is the initial number of people infected and N is the overall population size. Using the functions from the previous exercise, for both models, what are N and I_0?

Exercise 3.3. The general solution for the saturating and the logistic models are:

$$\text{Saturating model: } I(t) = N - (N - I_0)e^{-kt}$$
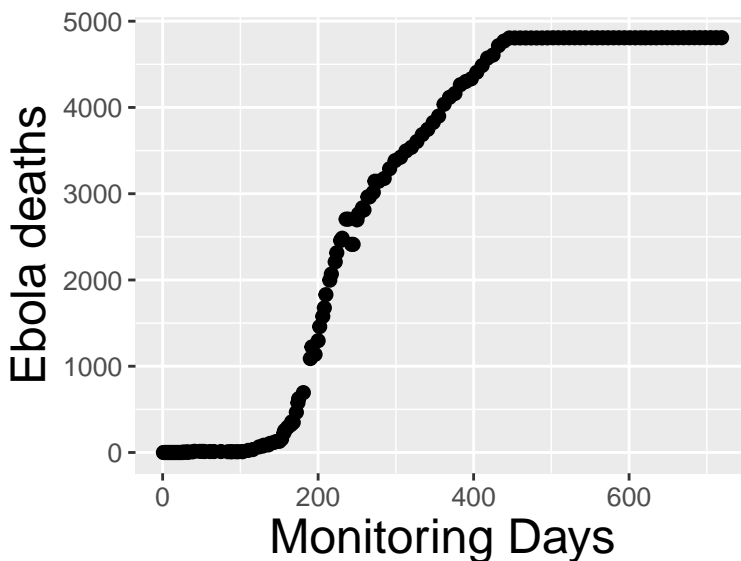$$\text{Logistic model: } I(t) = \frac{N \cdot I_0}{I_0 + (N - I_0)e^{-kt}},$$

where I_0 is the initial number of people infected and N is the overall population size. For both models carefully evaluate the limits to show $\lim_{t \to \infty} I(t) = N$. How do these compare to the steady-state values you found for Models 2 and 3 of the outbreak data?

Exercise 3.4. Figure **??** shows the Ebola outbreak for the country of Liberia in 2014. If we were to apply the logistic model based on this graphic what would be your estimate for N?

Exercise 3.5. A model that describes the growth of sales of a product in response to advertising is the following:

$$\frac{dS}{dt} = .55\sqrt{1 - S} - S,$$

where S is the product's share of the market (scaled between 0 and 1). Make a plot of the function $f(S) = .55\sqrt{1 - S} - S$. for $0 \leq S \leq 1$. Interpret your plot to predict when the market share will be increasing and decreasing. At what value is $\frac{dS}{dt} = 0$? (This is called the **steady-state** value.).

A second campaign is has the following differential equation:

$$\frac{dS}{dt} = .2\sqrt{1 - S} - S$$

What is the steady-state value and how does it compare to the previous one?

Exercise 3.6. A more general form of the advertising model is the following:

$$\frac{dS}{dt} = r\sqrt{1 - S} - S,$$

where S is the product's share of the market (scaled between 0 and 1). The parameter r is related to the effectiveness of the advertising (between 0 and 1).

Solve this equation for the steady state value (where $\frac{dS}{dt} = 0$). Make a plot of the steady state value as a function of r, where $0 \leq r \leq 1$. What can you conclude about the steady state value as the effectiveness of the advertising increases?

Exercise 3.7. A common saying is "You are what you eat." This saying is mostly true and can be related in a mathematical model! An equation that relates a consumer's nutrient content (denoted as y) to the nutrient content of food (denoted as x) is given by:

$$y = cx^{1/\theta}, \tag{3.6}$$

where $\theta \geq 1$ and c are both constants is a constant. Units on x and y are expressed as a proportion of a given nutrient (such as nitrogen or carbon).

Let's start with an example: $y = x$. In this case the point $(0.05, 0.05)$ would say that if an animal ate food that was 5% nitrogen, their body composition would be 5% as well.

Let's just assume that $c = 1$. How does the nutrient content of the consumer compare to the food when $\theta = 2$? $\theta = 5$? $\theta \to \infty$? Draw some sample curves to help illustrate your findings.

Exercise 3.8. A model for the outbreak of a cold virus assumes that the rate people get infected is proportional to infected people contacting susceptible people, as in the logistic model 3. However people who are infected can also recover and become susceptible again with rate α. Construct a diagram similar Model 3 for this scenario and also write down the system of differential equations.

Exercise 3.9. A model for the outbreak of the flu assumes that the rate people get infected is proportional to infected people contacting susceptible people, as in Model 3. However people also account for recovering from the flu, denoted with the variable R. Assume that the rate of recovery is proportional to the number of infected people with parameter β. Construct a diagram like Model 3 for this scenario, and also write down the system of differential equations.

Exercise 3.10. Organisms that live in a saline environment biochemically maintain the amount of salt in their blood stream. An equation that represents the level of S in the blood is the following:

$$\frac{dS}{dt} = I + p \cdot (W - S),$$

where the parameter I represents the active uptake of salt, p is the permeability of the skin, and W is the salinity in the water. What is that value of S at *steady state*, or when $\dfrac{dS}{dt} = 0$?

Exercise 3.11. Use your steady state solution from the last exercise to determine what parameters (I, p, or W) cause the steady state value S to increase?

Exercise 3.12. The immigration rate of bird species (species per time) from a mainland to an offshore island is $I_m \cdot (1 - S/P)$, where I_m is the maximum immigration rate, P is the size of the source pool of species on the mainland, and S is the number of species already occupying the island. Further, the extinction rate is $E \cdot S/P$, where E is the maximum extinction rate. The growth rate of the number of species on the island is the immigration rate minus the extinction rate.

1. Make representative plots of the of the immigration and the extinction rates as a function of S. You may set I_m, P, and E all equal to 1.

2. Determine the number of species for which the net growth rate is zero, or the number of species is in equilibrium. Express your answer as S as a function of I_m, P, and E.

3. Suppose that two islands of the same size are at different distances from the mainland. Birds arrive from the source pool and they have the same extinction rate on each island. However the maximum immigration rate is larger for the island farther away. Which of the two islands will have the larger number of species at equilibrium?

Exercise 3.13. This problem relates to animal size and volume. Assume that an animal assimilates nutrients at a rate R proportional to its surface area. Also assume that it uses nutrients at a rate proportional to its volume. You may assume that the size of the animal is implicitly a function of the nutrient intake and usage. Determine the size of the animal if its intake and use rates were in balance (meaning R is set to zero), assuming the animal is the following shapes:

1. A sphere (assume size is measured with radius r) *Note:* first determine the geometric formulas for surface area and volume.

2. A cube (assume size is measured with length l)

For both of these problems your goal is to determine a numeric value of r and l.