Components of Modeling

Part 3: Parameters

Understanding the Parameters in Your Model

Many parameters arise naturally as you develop the model. This was the case in all the models we have seen so far.

When it comes time to estimate the values of these parameters it is crucial that you understand *exactly* what quantities the parameters measure. In particular, if a quantity is a proportion, you should have clarity on what it is a proportion *of*.

Example: Consider the parameter p in the Reed-Frost model of an epidemic.

Recall that p appeared in the distribution of the number of people that passed from the susceptible group to the infected group X_n; that number of people had a binomial distribution with parameters S_n and $1-(1-p)^{I_n}$ where S_n is the number of susceptible people and I_n is the number of infected people at time n.

What does p measure? Is it:

- The probability in one time step that an infected person meets a susceptible person meet and the disease is passed.
- The probability in one time step that a particular susceptible person meets a particular infected person and the disease is passed.
- The probability in one time step that a particular individual meets any other particular individual and the nature of the meeting is such that the disease would be passed if one of them had the disease and the other did not.

The third one is the correct answer.

Notice the very subtle but very important differences in the meanings. If you didn't have clarity on the meaning of the parameter you could:

- use the same parameter in two different ways in two parts of the model and/or
- estimate a value for the parameter that is way off based on a misinterpretation of its meaning.

To illustrate the latter, consider trying to estimate p from observations of a disease.

The following are all quantities that you might be able to estimate from data.

Let's think about which (if any) might be a reasonable estimate of p.

a) The proportion of people that acquire the disease each time step, averaged over many time steps.

b) The proportion of all susceptible people that acquire the disease each time step, averaged over many time steps.

c) The proportion of all potential meetings between a susceptible person and an infected person that result in the disease being passed.

Discussion of a): According to the model the proportion of people that acquire the disease in one time-step is X_n/N.

In a) we average this quantity over many time steps in a single sample path of the process. In other words, the estimate being proposed is the value of

$$E_a = \frac{X_m + X_{m+1} + X_{m+2} + \ldots + X_{m+k-1}}{kN}$$

where m is some number (that we probably don't know) and k is large and these are values taken from a single sample path of the process.

Since the average of X_n (given the state of the system at time n) is $S_n(1-(1-p)^{I_n}) \approx S_n(1-(1-I_np)) = S_nI_np$ (if p is small), this estimate will look something like:

$$\frac{p}{kN}\left(S_mI_m + S_{m+1}I_{m+1} + \ldots + S_{m+k-1}I_{m+k-1}\right).$$

Now, the S_i's and I_i's are varying and on the order of N, so S_iI_i is on the order of N^2, so this estimate could be on the order of a factor of N too big.

Discussion of b): The quantity in b) is similar, except, rather than taking the proportion of all people, we take the proportion of all susceptible people. In other words, we average X_n/S_n over many time steps. Thus the second estimate is the value of

$$E_b = \frac{1}{k}\left(\frac{X_m}{S_m} + \frac{X_{m+1}}{S_{m+1}} + \ldots \frac{X_{m+k-1}}{S_{m+k-1}}\right)$$

Reasoning as above, this estimate would look something like

$$\frac{p}{k}\left(I_m + I_{m+1} + \ldots + I_{m+k-1}\right)$$

Again - this estimate looks like it is on the order of a factor of N out.

The analysis above suggests another estimate (if this is something we can measure) namely the value of X_n/S_nI_n taken over many time steps. This estimate would be

$$\frac{1}{k}\left(\frac{X_m}{S_mI_m}+\frac{X_{m+1}}{S_{m+1}I_{m+1}}+\ldots+\frac{X_{m+k-1}}{S_{m+k-1}I_{m+k-1}}\right).$$

If p small, each X_i has an average of approximately pS_iI_i, so if the variations of X_i from their average cancel each other out, then this would be a reasonable estimate of p.

In fact, the variances of the X_i's differ, so the averages won't typically cancel out. That would be a serious disadvantage to this as an estimator.

However, we might be able to understand the variance of this quantity over different sample paths and thereby get an understanding of the accuracy of our estimate. We might be able to do this analytically, or we could investigate it numerically. (We'll come back to this a little later.)

Discussion of c):

Here we proposed looking at the proportion of all meetings between a susceptible person and an infected person that resulted in the disease being passed.

This might conceivably be measured by observing an infected person, making a note of the people with whom they come into contact, and seeing how many of those people acquire the disease. Our estimate would be

$$\frac{\text{total number of cases where disease was passed}}{\text{totatl number of people met}}$$

Since the model assumes that the value of p is uniform over all people, you might improve the estimate by doing this for many different people.

This might look good at first, but in fact it's probably way too high. The value of p should be the probability that any two given people meet and the nature of the meeting is such that the disease is passed.

The value computed here is conditional on the susceptible and infected person meeting.

We could fix this if we get and estimate of the probability that any two given people meet.

If the population under study is a small family and the time interval not too small, this number might be close to 1. If the population is, say, a school, this might be estimated using social network data, school schedules, etc.

The estimate of p would then be the product of

- this probability and
- the conditional probability above of the disease being passed given that they meet.

Message: Think very carefully about the meaning of the parameters when developing the model and devising methods to estimate them.

Units and Dimensional Analysis

You should always determine the *units* of each parameter (and the variables, for that matter!). Determining the units will help you understand

- the meanings of the parameters,
- whether proposed estimators are reasonable (an estimator should have the same units as the parameter it is estimating!),
- how the value of one parameter may be related to the values of other parameters,
- how the values of the parameters in one model relate to the values of the parameters in another model,
- how the values of the parameters will change if you change the units of the variables.

An investigation of units can also uncover some hidden or overlooked parameters.

For example, suppose you have the term e^t in your model where t measures the passage of time.

The variable t has units and it doesn't make sense to raise e to the power of something that has units, so that means you should really have e^{kt} where the units of the parameter k are units of time^{-1}.

Dimensional Analysis: an investigation of units to reveal the relationship between parameters.

The classic example of this is an investigation into the period of a pendulum.

The period of a pendulum has units of time.

What parameters might determine the period of a pendulum?

Here are some candidates.

- The amplitude of oscillation (which is related to the initial conditions), a. This is unitless.
- The length of the pendulum, l. Units of this are length.
- The mass of the pendulum, m. Units of this are mass.
- The acceleration due to gravity, g. Units of this are length/time^2.

How can we combine these quantities and get something whose units are time?

We'll have to have the acceleration due to gravity in our expression, since it's the only thing that depends on time. To get time we'll need to have $\sqrt{\frac{1}{g}}$. The units of this are time/$\sqrt{\text{length}}$.

Now we need an expression whose units are length in order to balance the units. This is l. Thus, the units of $\sqrt{l/g}$ are the same as the units of the period of oscillation of the pendulum.

Thus, we might expect the period of oscillation of the pendulum to be proportional to

$$\sqrt{\frac{l}{g}}.$$

A more careful analysis of the pendulum reveals this to be the case.

Notice: Since the amplitude of the pendulum is unit less, the amplitude could appear in this expression in any form. A careful analysis reveals that the period of the pendulum does not depend on its amplitude.

What this analysis does reveal that is, perhaps, surprising, is that the period of the pendulum does not depend on its mass.

Interdependency Between the Parameters in Your Model

For modifying your model and estimating the values of the parameters, it is important to ask yourself if the value of any of the parameters depends on the values of any other parameters in the model.

We saw how important this was in the models of an epidemic we developed above; it was crucial to understand how a and b depended on the time step Δt when we developed the continuous time deterministic model.

Ease of Interpretation and Measurability

You usually don't have much choice about the parameters in your model. However, in so far as you do, use a model where the parameters can be interpreted easily and where you can get data to estimate their values.

It is well-known that one of the biggest issues in modeling is the estimation of the parameters.

Not Too Many, Not Too Few

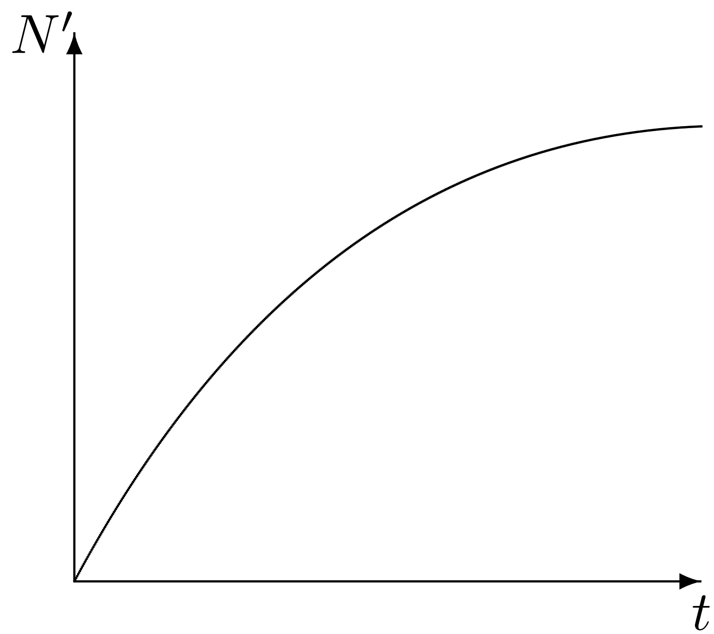
Often parameters appear when you are modeling the net relationship between two variables rather than the mechanism that determines that relationship. Notice, in these cases it may be hard to give precise physical meaning to the parameters but you will still need to estimate their values from data.

It is important here to include the correct number of parameters; not too many and not too few.

Example: Consider a line of cars stopped at a traffic light. Let N denote the number of cars that have made it through the light after time t where $t = 0$ is the time at which the light turns green.

Notice that the rate at which cars go through the light is initially equally to 0 but then increases over time with a limiting value that is related to the speed limit on the road.

In other words, we expect N' as a function of t should have value 0 when $t = 0$ and be increasing and concave down with a limiting value.


N'
t

The models below are all based on this observation. Compare the models in terms of their use of parameters. What are the units of the parameters involved in each case? Do any of the models have too few parameters? Do any have too many?

a) $N' = \dfrac{t}{1+t}$

b) $N' = \dfrac{At}{1+t}$

c) $N' = \dfrac{At}{k+t}$

d) $N' = \dfrac{At}{1+kt}$

e) $N' = \dfrac{At}{c+kt}$

In this example we are not attempting to model the *mechanism* determining the rate at which the cars pass through the light. Instead, we are are simply attempting to produce a rate function that agrees with an observed phenomenon.

The observed phenomenon is the *shape* of the rate function. The exact details such as:

- how fast the function is increasing at 0,
- how fast the derivative of the function decreases, and
- what the limiting value is

are things we might hope to fit to data.

The first function $N' = t/(1+t)$ has no parameters. This means that we are imposing the three values above and we will almost certainly impose them incorrectly.

Notice, we can also see that this isn't a good choice by an investigation of units.

The left-hand side is measured in cars per second (if t is measured in seconds).

On the right-hand side we have the expression $1+t$ in the denominator; that doesn't make sense since t has units. There are two natural ways to change this.

- We could change the denominator on the right to be $k + t$ where the units of k are seconds. Then the right-hand side will be $t/(k + t)$ which is unit-less, so we'll need another parameter A out front whose units are cars/second. This is model c), $N' = At/(k + t)$.
- Alternatively, we could change the denominator to $1 + kt$ where k has units of second^{-1}. In this case the denominator is unit-less, so we will need another parameter A out front whose units are cars/second^2. This is model d), $N' = At/(1 + kt)$.

In the second model $N' = At/(1+t)$, the parameter A controls the asymptotic value of N' as $t \to \infty$ but we aren't able to independently control the slope of the rate function at the origin. So this has too few parameters.

Again, we can see that a parameter is missing by the fact that $1+t$ in the denominator doesn't make sense.

The fifth model $N' = At/(c + kt)$ has too many parameters.

In general, it's harder to detect that a model has too many parameters.

In this case we can see it because of the following.

- The units of the parameters are not forced on us. We could have k be unit-less, c measured in seconds, and A measured in cars/second or we could have k be measured in seconds^{-1}, c be unit-less and A be measured in cars/second^2 or etc. This is an indication that there are too many parameters.
- If we divide top and bottom by k we get $N' = \tilde{A}t/(\tilde{c} + t)$ where $\tilde{A} = A/k$ and $\tilde{c} = c/k$. In other words, we have obtained the exact same function using only two parameters instead of three.

When a model has too many parameters it is hard to make physical sense of the parameters.

So, it's clear that a) and b) have too few parameters and e) has too many parameters. What about c) and d)?

c) $N' = \dfrac{At}{k+t}$

d) $N' = \dfrac{At}{1+kt}$

If we take c) and divide the top and bottom of the right-hand side by k we get

$$N' = \frac{\tilde{A}t}{1+\tilde{k}t}$$

where $\tilde{A} = A/k$ and $\tilde{k} = 1/k$. Thus, the family of functions described by c) is exactly the same as the family of functions described by d). So these are equivalent models.

For this family of functions, two parameters are the right number of parameters to describe the complete family without duplication (i.e. different sets of parameters produce different functions).

The two parameters control the slope at the origin and the height of the horizontal asymptote. For this family of functions, the rate at which the slope decreases is determined by the slope at the origin and the height of the horizontal asymptote.

If there is no function in this family that fits the data well, another family of functions to try that has the same general shape (starts at 0 and is increasing and concave down, with a horizontal asymptote) is

$$N' = A\left(1 - e^{-kt}\right).$$

Location-Scale Families of Functions

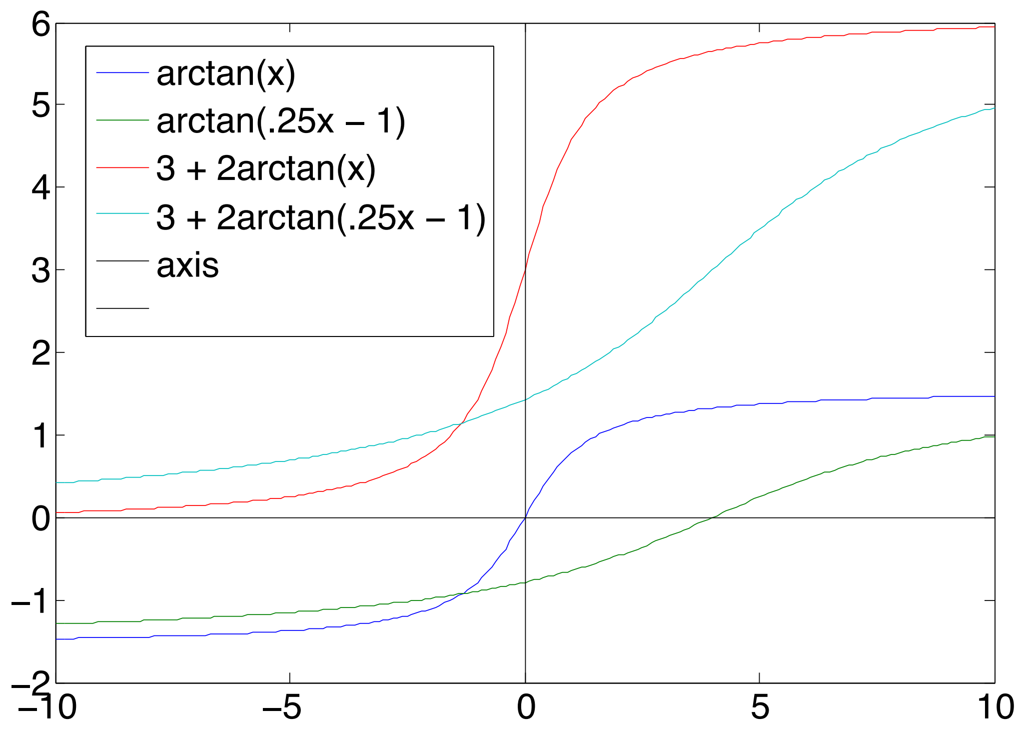
Suppose, as in the example above, you know that you want a function of a particular shape (in the example above we wanted a function that was increasing and concave down and had a horizontal asymptote) but you don't the details of that shape (like the slope or the height of the asymptote).

If you can find a single function f that has that shape then consider the 4-parameter family of functions that is given by:

$$a + bf(cx + d)$$

This family will include all functions that have the same shape as f but are stretched and shifted by different amounts in the horizontal and vertical directions.

For example, here are graphs of some of the functions in the family when $f(x) = \arctan x$.



Sometimes this family will include too many parameters. This can happen for two reasons:

- You may want to fix some particular feature (for instance you might want your function to go through the origin).
- For some functions, one of these parameters might have the same effect as another one.

Example: If $f(x) = x$ then a has the same effect as d and b has the same effect as c. The family

$$a + bf(cx + d) = a + b(cx + d) = (a + d) + bcx$$

is the set of all straight lines. This is a two-parameter family of functions (determined by, say, the y-intercept and the slope) and not a four parameter family.

Example: If $f(x) = x^2$ then we can achieve the same effect using either the parameter b or c since $f(cx) = (cx)^2 = c^2 x^x$. The family

$$a + bf(cx + d) = a + b(cx + d)^2 = bc^2x^2 + 2bcdx + (a + d^2)$$

is the set of all parabolae. This is a three-parameter family determined by, say, the location of the vertex (2 parameters) and the steepness of the parabola.

In the example above we were looking for a function that went through the origin, was increasing and concave down and had a horizontal asymptote. We started by noticing that

$$f(t) = \frac{t}{1+t}$$

had these properties. The location-scale family associated with this functions is

$$\begin{aligned} a + bf(ct+d) &= a + b\frac{ct+d}{1+ct+d} \\ &= (a+b) - \frac{b}{(1+d)+ct} \\ &= (a+b) - \frac{1}{\frac{1+d}{b} + \frac{c}{b}t} \end{aligned}$$

We can see that this is really a 3-parameter family of functions and not a 4-parameter family; the parameters are $a+b$, $(1+d)/b$ and c/b.

Since we want the functions to go through the origin, we require

$$0 = (a+b) - \frac{b}{1+d}.$$

This gives us a 2-parameter family of functions:

$$\frac{1}{\tilde{b}} - \frac{1}{\tilde{b} + \tilde{c}t}.$$

Non-dimensionalizing the model

This is a process you can go through to get rid of extraneous parameters (as in e) above) and to highlight the characteristic units of the problem.

It is most commonly used on models that consist of one or more ordinary or partial differential equations.

The idea is to write the variables and parameters in units that are somehow characteristic to the problem. This can simplify the model considerably and can delete unnecessary parameters.

The process works as follows:

- Consider each variable in the model (both independent and dependent) and the quantities that they measure. Change variables so that each quantity is measured in units that are to be determined. The new variables will be scalar multiples of the old variables.
- Determine the relationships between the new variables in the problem.
- Divide through by the coefficient of the highest order polynomial or derivative term.
- Choose the units of each variable so that the coefficients of as many terms as possible are equal to 1 (there may be more than one way to do this).

Example: Suppose we have the following model for how x and y vary with time t.

$$x' = ax + bxy$$
$$y' = cy - dxy + e\,(1 + \sin(\alpha t))$$

Notice, this might be a non-autonomous predator-prey model in which x and y are populations of two different species.

- In the absence of the other, both species grow exponentially at rates a and c.
- Species x preys on species y, so y declines at a rate that is proportional to the number of possible meetings between the two species and x increases at this rate. The constants of proportionality between y declining and x increasing are not the same because one unit of y does not produce one unit of x (these are different species, so it would be surprising if it did).
- Species y increases because an external source of y is being added and removed from the system; this external source is being added periodically with period equal to $2\pi/\alpha$.

This is a system that has six parameters a, b, c d, e and α.

- a is the natural growth rate of the predator measured in days^{-1}.
- c is the natural growth rate of the prey measured in days^{-1}.
- b is the rate of growth of the predator due to consumption of the prey measured in days^{-1}prey^{-1}.
- d is the rate of decay of the prey due to consumption by the predator measured in days^{-1}predator^{-1}.
- e is the amplitude of the oscillation of the amount of prey being added to the system measured in prey day^{-1}.
- α is determined by the period of oscillation of the external source of prey and is measured in days^{-1}.

The variables in this system are x, y, and time t.

If we changed the units in which we measured these variables we could produce a model that has fewer parameters. For example, we could scale time so that the natural growth rate of the predator is equal to 1 or we could measure the prey y in units such that one unit of y produces one unit of x.

To explore this a little more systematically we will *non-dimensionalize* the system.

To non-dimensionalize the system we consider new units for measuring time, the size of the predator, and the size of the prey. We will choose these units to simplify the system as much as possible. Let

- τ be time measured in terms of our new characteristic unit to be determined,
- χ be the size of the predator measured in our new characteristic unit to be determined, and
- γ be the size of the prey measured in our new characteristic unit to be determined.

Notice we'll have

$$t = t_c \tau$$

$$x = x_c \chi$$

$$y = y_c \gamma$$

for some constants t_c, x_c and y_c. Choosing new units means we need to to choose the values of t_c, x_c and y_c.

Normally we'd think of t being a variable measured in one unit (in our case days) and τ being the same variable measured in another unit (say hours) and t_c being the conversion constant, in this case 24 *days per hour.*

However, when we non-dimensionalize a system we think of t and t_c as both being measured in the old units and τ being dimensionless.

Here are the units of the conversion constants:

- t_c is measured in days
- x_c is measured in predator
- y_c is measured in prey

We need to take the differential equations and write them in the new units. We have:

$$\frac{dx}{dt} = x_c \frac{d\chi}{dt} = x_c \frac{d\chi}{d\tau} \frac{d\tau}{dt} = \left(\frac{x_c}{t_c} \right) \frac{d\chi}{d\tau}.$$

Similarly

$$\frac{dy}{dt} = \left(\frac{y_c}{t_c} \right) \frac{d\gamma}{d\tau}.$$

Notice $\dfrac{d\chi}{d\tau}$ and $\dfrac{d\gamma}{d\tau}$ are both dimensionless quantities.

Thus, the new system of equations becomes

$$\left(\frac{x_c}{t_c}\right)\chi' = ax_c\chi + bx_cy_c\chi\gamma$$
$$\left(\frac{y_c}{t_c}\right)\gamma' = cy_c\gamma - dx_cy_c\chi\gamma + e\left(1+\sin\left(\alpha t_c\tau\right)\right)$$

The units on the left and the right of the first equation are predator per day and of the second equation are prey per day.

Now we divide each equation through by the coefficient of the derivative to make both equations dimensionless:

$$\chi' = at_c\chi + bt_cy_c\chi\gamma$$
$$\gamma' = ct_c\gamma - dt_cx_c\gamma\chi + \left(\frac{et_c}{y_c}\right)\left(1+\sin\left(\alpha t_c\tau\right)\right)$$

Now we choose x_c, y_c and τ_c to make as many things disappear as we can. There are different ways to do this. One way might be to choose:

$$t_c = \frac{1}{a} \qquad x_c = \frac{a}{d} \qquad y_c = \frac{a}{b}$$

This gives us the system

$$\begin{aligned} \chi' &= \chi + \chi\gamma \\ \gamma' &= A\gamma - \chi\gamma + B(1 + \sin\omega\tau) \end{aligned}$$

where $A = c/a$, $B = eb/a^2$ and $\omega = \alpha/a$. These are all dimensionless quantities.

Now we have a system with only 3 parameters instead of 6.