

Chapter 2

Differential Equations: Notation and Architecture

2.1 Introduction and Chapter Overview

This and the next three chapters are resources, designed to get the neophyte started, to recall and extend the skills of the initiated, and to call the expert's attention to the relative importance of what is already known. In this chapter we first introduce and motivate the notation that we use throughout the book, and specifically the *operator notation* Dx for the derivative of a function x . That is, we view a derivative as a *transformation* of a function into a new function whose value at any location t is the slope of x at that point.

Differential equations are divided into various classes, and we first describe the autonomous/forced dichotomy defined by whether or not the system has an external or exogenous input. Then we consider linear versus nonlinear systems, a distinction that is already familiar to statistical readers in regression analysis.

The data are assumed to be distributed over the whole interval of observation, and to be subject to measurement error to other types of random disturbances. This contrasts with the errorless initial and boundary value data configurations that apply to so many models.

2.2 Notation for Dynamical Systems

2.2.1 Dynamical System Variables

Dynamical systems concern observations distributed over a substrate, such as time, space, space/time, wavelength, frequency and so forth. We mostly focus on continuous dynamical systems, where the substrate has the properties of a line, plane or higher dimensional vector space. We tend to index positions on the substrate by t for derivatives with respect to one-dimensional continua such as time, and we reserve s for planes and volumes and other higher dimensional structures. Thus, we leave out

systems defined explicitly over discrete countable and equally spaced points such as days, quarters, years and etc.; that is, systems dealt with in the time series literature. We are aiming to keep the mathematical level as low as possible, and we therefore assume that few readers will mind that we also ignore data distributed over curved surfaces and other manifolds, but it is worth stating here that such situations are both important and not rare.

These chapters are also restricted to the description of *deterministic* dynamic systems where all the elements in the system are nonrandom entities. By contrast *stochastic* dynamic systems explicitly incorporate random elements, such as randomly varying functional inputs, randomly varying parameters or random initial values. But stochastic systems are typically modifications of systems that have a deterministic core, so that it is wise to get a firm grip on deterministic systems first.

A variable x in a dynamical system is a function of position on the substrate, and the corresponding values of the functions at specific substrate values are $x(t)$ and so on. We will also refer to these functions as “states”, “paths” and “trajectories.” But we must also remember that certain dynamic systems are so familiar in particular fields that specialized letter allocations are used. An example is the SIR system in epidemiology where the letters S , I and R refer to “susceptible”, “infected” and “recovered” populations, respectively. Although we will define these symbols before using them, readers will have to be adaptable.

Most situations in applied statistical settings involve more than one variable, and we follow the usual practice of using bold face, as in \mathbf{x} , to refer to a vector of functions. Occasionally the data configurations and the dynamic models themselves call for rectangular and higher dimensional arrays, in which case capital boldface letters such as \mathbf{X} are employed.

Statisticians are familiar with linear and nonlinear regression models where some variables are designated as covariate, independent, causal or input, and one or more others as dependent or output. Consequently, we will use tend to use u , v and w as well as their boldface versions for these input variables, which may vary over the substrate t or s . Typically the dynamical systems literature refers to these variables as “forcing” functions, and especially when they affect the dependent variable in an additive manner. Additive forcing tends to change the location of x in function space but not alter its fundamental behavior. By contrast, inputs which affect output variables in a multiplicative way or have more complex effects will usually change more drastically the shape characteristics of an output trajectory.

2.2.2 Dynamical System Parameters

Models almost always involve parameters that define the dynamical system, and which typically must be estimated from the data at hand or from data collected in collateral experiments. Here we follow standard practice in using the Greek alphabet for parameters, with boldface versions where vectors of parameters are assumed, and $\boldsymbol{\theta}$ will be the usual choice.

A fundamental goal is to estimate the values of these parameters from data, and also estimate the precision of these estimates given the noise level in the data. We may also conjecture that parameter values will vary randomly, typically in some structured way, from one realization of the system to another, and such parameters are called *random effects*. We will naturally then want to estimate characteristics of their variation, such as their means, variances and covariances.

2.2.3 *Dynamical System Data Configurations*

Since we approach dynamic systems as models for data, we will consider data configurations that are seldom considered in textbooks on the mathematics of dynamical systems, which tend to focus heavily on the *initial value* problems arising in science and engineering where data are only available at time $t = 0$, and are assumed to be virtually free of error. They also consider *boundary value* problems, where the system is also constrained to be in a specific state at some other fixed time t_1 . By contrast, the statistician is seldom willing to assume error-free data, and we focus more on *distributed data* problems where data are available at a considerable number of time points for at least some subset of the variables involved in a dynamic system.

Data configurations are apt to vary from one variable to another. For example, some measurements are inexpensive and available on line, such as temperature, but others such as chemical assays are expensive and available only some time after the experiment is completed. It is typical that some variables will not be observed at all, so that these variables have some of the characteristics of the latent variables often used in various areas of statistics to model known but unobservable sources of variation.

2.2.4 *Mathematical Background*

We do need to know some of the fundamental theorems to be found in mathematical texts, and we shall try to state these as clearly as we can, but leave the reader to turn to textbooks for the proofs of these results. Moreover, a fair part of a college text on differential equations is devoted to techniques for calculating algebraic solutions to certain classic equations and classes of equations. But these constitute a tiny fraction of the differential equations in actual use. Thus, although it certainly worthwhile to derive solutions in a few simple cases, much more reliance will be placed in this book on numerical methods for the accurate approximation of solution values given both the values of parameters defining the system and, in most cases, initial states of the systems. Chapter 5 is devoted to this approximation technology.

Fortunately for both the readers and writers of this book, some fine texts have been written for first-timers and more advanced students of dynamical systems. We especially like Tennenbaum and Pollard (1963), a classic text that is now available as

an inexpensive paperback, and which has received the highest reader rating among the texts distributed by a well-known on-line bookseller. It is organized into 65 lessons, each corresponding roughly to a single lecture in a two-semester course, and each lesson's topic is illustrated with examples of historical and practical interest. Another source now in paperback is Coddington (1961). Engineering texts often have readable accounts of dynamical systems in the context of specific topics such as electrical circuit theory or chemical process control. Textbooks such as Barnes and Fulford (2009), Blanchard et al. (2006) and Borrelli and Coleman (2004) emphasize how differential equation systems are constructed in order to capture the behaviour of specific observed systems.

2.3 The Architecture of Dynamic Systems

2.3.1 Elementary Autonomous or Unforced Systems

Dynamical systems are defined by equations involving one or more derivatives. The most elementary of these define polynomials; the solution to $D^m x = 0$ is the space of all polynomials of degree $m - 1$. But most differential equations have expressions on the right side, and we will give considerable attention to the prototypical

$$\text{Diffusion equation or first order buffer: } Dx = -\beta x, \quad (2.1)$$

where β is a nonzero positive constant. The solution to this system is $x(t) = C \exp(-\beta t)$, constant C being arbitrary. Notice here that (2.1) hides the role of t , which we could bring back in by expressing the equation as $Dx(t) = -\beta x(t)$ or, more obsessively, as $(Dx)(t) = -\beta x(t)$. The result of removing t is a nice clean expression that works well when we already understand what the substrate variable t is.

The *order* of a differential equation is the highest order of derivative involved, so that, for example,

$$\text{Harmonic equation: } D^2 x = -\omega^2 x, \quad (2.2)$$

is of order two. Its solution is $x(t) = C_1 \sin(\omega t) + C_2 \cos(\omega t)$, where both C_1 and C_2 are arbitrary constants. The wavelength of the sinusoid is $2\pi/\omega$ and $2\pi/\omega$ is its period. The harmonic equation in turn is a special case of

$$\text{Second order buffer: } D^2 x = -\beta_0 x - \beta_1 Dx. \quad (2.3)$$

It is common and convenient notation to isolate the highest order derivative in this way on the left side of the equation without any multiplier, and to move all other elements of the equation to the right side. According to this convention, a single differential equation of order m can be expressed as

$$D^m x = f(x, Dx, \dots, D^{m-1}x) . \quad (2.4)$$

Models of practical value in the real world typically require an account of how several variables interact to produce the behavior of a system, as we saw for the SIR and handwriting models in Chap. 1. Hence we need to gather several variables x_i , $i = 1, \dots, d$, into a vector of variables \mathbf{x} of dimension d . Letter d carries a heavy load in this book, but as a rule it will be clear from the context when we use d to indicate the *dimension* of a system, rather than a differential dx or dt . This another reason for preferring the operator notation Dx as our standard notation for a derivative.

An equation of order m can always be reduced to a set of first order equations by defining artificial variables. For example, if we define $y = Dx$, then the second order linear equation (2.3) is split into the two order one equations $Dx = y$ and $Dy = -\beta_0 x - \beta_1 y$. Of course, the result is often neither natural nor easy to read, so that keeping multiple orders explicitly in a single equation is often to be preferred for clarity. But this does allow us to express any dynamical system, whether of order 1 or of order $m > 1$, as the differential equation system

$$D\mathbf{x} = \mathbf{f}(\mathbf{x}) \quad (2.5)$$

where right-hand function \mathbf{f} has a vector-valued argument and maps the values of \mathbf{x} into a vector of dimension d , and where, moreover, some of the variables in \mathbf{x} may be dummy variables coding one or more derivatives of one or more state variables. However, while this level of abstraction may have some appeal and utility in describing theoretical results, in practice we prefer to express dynamical systems in a manner that keeps the special structure of the system visible.

Parameters, like β in the simple first order diffusion system above, ω in the harmonic equation and (β_0, β_1) in the second order buffer equation play a prominent role in dynamic models, and we often like to keep them up front in our expressions. But if we opt for abstraction, then we can collect all the n parameters into a parameter vector $\boldsymbol{\theta}$ in the expression

$$D\mathbf{x} = \mathbf{f}(\mathbf{x}|\boldsymbol{\theta}) . \quad (2.6)$$

Occasionally, there may even be remaining time-variant aspects to a system, in which case we can add a t itself as an additional argument in f so that

$$D\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}|\boldsymbol{\theta}) . \quad (2.7)$$

2.3.2 Forced or Non-autonomous Systems

Now we turn to the critical role of input or forcing variables, such as force F in

$$\text{Newton's second law: } D^2x = F/M , \quad (2.8)$$

where x is position and M is the mass of the body on which the force is acting. Dynamical systems model change, and are essential where the response of an system to a change in an input is not direct or straight-through. Often, for example, inputs can change abruptly, or arrive as short but intense bursts, and the system does not have the sensitivity or the energy resources to respond immediately to these nearly instantaneous changes. Cars can't stop instantly, so that the yellow traffic light permits the driver to spread the stopping response over an acceptable time interval. We use the term *buffer* to describe how a dynamical system spreads its response over a longer time interval than that over which the input changes.

Let \mathbf{u} be a vector of ℓ external “forces” that change the behavior of a system in some way. We want to distinguish the behavior of a system isolated from any such inputs from one where they apply. Forced dynamic systems are expressed as

$$D\mathbf{x} = \mathbf{f}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{u}) . \quad (2.9)$$

An important special case, that is often assumed in the term “force”, is forcing of an additive nature, expressible as

$$D\mathbf{x} = \mathbf{f}(\mathbf{x}|\boldsymbol{\theta}) + \mathbf{g}(\mathbf{u}) . \quad (2.10)$$

If a system is relatively linear in the sense defined in Sect. 2.4.1, and the input is additive, the system will tend to respond in a more or less mild and predictable manner. For example, a step change in input will produce a more gradual approach to a new level in the output. A simple example that we will explore in more detail is the equation defining the concentration C of a chemical species in the output of a chemical reactor with volume $V(m^3)$ and input flow rate $F(V/min)$

$$DC = -\beta C + FC_0/V \quad (2.11)$$

where the speed-of-reaction parameter $\beta = (F + Vk)/V$ and k is the rate constant for the reaction. Here C_0 in the second forcing term on the right side is a fixed equilibrium concentration, and we will see in the next chapter that C will approach C_0 at an exponential rate no matter what the initial concentration $C(0)$.

But external values can also affect other aspects of the system, and specifically an external variable may change one or more parameter values of an otherwise autonomous system. When this happens, and it does often, the results can be much more dramatic. Stable systems can become unstable, relatively predictable systems can exhibit surprising variation, or even the seemingly unpredictable shapes called *chaotic* in the dynamical systems literature.

2.3.3 Differential Operator Notation

We can go further in emphasizing the input/output nature of a specific dynamical system. As an example, we can re-express the diffusion equation with forcing by input u , $Dx = -\beta x + \alpha u$, as

$$Dx + \beta x = \alpha u, \quad (2.12)$$

where parameter α modulates the impact of changes in forcing function u on the system. Here we have put the autonomous part of the equation on the left side, and have effectively recast the model into regression format familiar to statisticians.

We use the notation $Lx = \beta x + Dx$, or more generally, $Lx = f(x|\theta) + Dx$ for (2.6), and call L the *differential operator* associated with the dynamic system. In this operator notation, the differential operator for the forced diffusion equation is $L = D + \beta I$, and for the second order buffer equation it would be $L = D^2 + \beta_0 I + \beta_1 D$. It is useful to think of L as a generalization of the notion of a derivative, and especially since the Taylor series machinery for approximating functions at specific values of t , as well as other concepts associated with derivatives in elementary calculus courses, have versions in this more general setting. The autonomous differential equation in differential operator notation becomes

$$Lx = 0. \quad (2.13)$$

In the next chapters, and elsewhere in the book, we explore how systems defined by specific differential operators respond to various prototypical inputs such as localized spikes, step functions, oscillating functions and random random noise processes.

Other differential operators can play a large role in scientific systems. For example, the curl and divergence operators, curl and div , are used as a compact way of expressing the four fundamental equations of electromagnetism, and are now used in many other contexts.

2.4 Types of Differential Equations

Between the simple diffusion equation and the general formulation (2.10) are a whole range of specialized classes of differential equations, each with their own special properties. It is essential at this point to learn the terminology for these classes before we take up their special characteristics in the next chapter.

2.4.1 Linear Differential Equations

Statisticians will not be surprised that linearity in differential equations, as in regression models, enables a host of specialized mathematical tools and results. We have

already seen three linear equations: the diffusion, harmonic and second order equations.

Here by *linear* we mean the linearity of the right side function f with respect to its arguments x and x 's derivatives. The single unforced linear equation of order m is

$$\text{Order } m \text{ buffer: } D^m x = \sum_{j=0}^{m-1} \beta_j D^j x \quad \text{or} \quad L = D^m - \sum_{j=0}^{m-1} \beta_j D^j . \quad (2.14)$$

The *coefficient functions* β_j may themselves vary over t , but must not depend on the values of x or any of its derivatives.

Equation (2.14) may be extended in various ways that still preserve linearity. The highest order term D^m may also be multiplied by a coefficient function β_m , and if this coefficient goes to zero at a specific value t over an interval, the equation then loses dimensionality as a system but retains linearity as a property.

Specific linear systems, such as the diffusion and harmonic equations, often have specific signs associated with their terms, such as the minus sign for the x -coefficient in the diffusion and harmonic equations. When this happens, it is often assumed that the sign of the coefficient does not change and, usually, the coefficient preceded by a minus sign is assumed positive. For example, in the second order buffer equation (2.3) it is common in many applications to assume that β_0 , but not necessarily β_1 , is always positive.

When all the coefficients functions β_j are constants, we call the system *stationary* as well as linear, because that the fundamental structure of the dynamic system does not vary over t . Stationary linear systems have additional useful analytical techniques, especially, as we shall see later, the Laplace transform, and stationary systems are useful for a great many applications where the characteristics of a process do not substantially change within the time scale over which the system is observed. An example is the dynamics of electrical circuits where their physical characteristics have been rendered stable by cooling systems.

Be warned that many treatments of differential equations, and especially in the engineering literature, use the term “linear” to mean by default stationary as well. This unfortunate lack of clarity is also common in the statistical time series literature.

2.4.2 Nonlinear Dynamical Systems

Even seemingly minor departures from the linearity of the relation of right side function f to its arguments can have dramatic consequences. For example,

$$\text{Catalytic equation: } Dx = \beta x(K - x) , \quad (2.15)$$

where β is a positive constant has a solution

$$x(t) = K / \left[1 + \frac{K - x_0}{x_0} \exp(-\beta t) \right]$$

where x_0 is the value of x at time zero and $0 < x_0 < K$. Solutions of this form are widely used in statistics for constraining function values to be within the interval $[0, K]$. We see that this happens because the slope of the function goes to zero as the value of x approaches either zero or K .

A common source of nonlinearity is the appearance of products of variables in systems of two or more equations. For example, the equations of the SIR model for the spread of disease are

$$\begin{aligned} DS &= -\beta SI \\ DI &= \beta SI - \nu I \\ DR &= \gamma I \end{aligned} \tag{2.16}$$

where S is the number of susceptible but uninfected members, I is the number of infected members, and R is the number of recovered (including dead) of a population. Here the appearance of the product SI renders these equations nonlinear, even though the equations are linear in their parameters β and ν . Equations of this type can be properly called quadratic in their variables, and constitute the largest subclass of nonlinear systems in applications. In our terminology for forcing, S is rate-forced by I , I in turn is rate-forced by S (but with the opposite sign), and R is add-forced by I . In the last R equation, there is no term involving R , so that the change in R is simply proportional to input I .

If the susceptible population S is large, the number of infected individuals will exhibit exponential increase, so that the I -equation behaves as an *anti-buffer* that turns a small increase in infected individuals into an epidemic, as we saw with the smallpox data in Chap. 1. The great importance of the SIR model and its many variants demands that we return often to this example.

2.4.3 Partial Differential Equations

Partial differential equations arise when more than one substrate is involved or when a substrate is multidimensional. The most common examples involve variation over space or over space and time. Partial differential equations play a huge role in climate modelling, the spatial aspects of spread of disease and pollution, fluid and airflow dynamics, two- and three-dimensional mechanical systems and a great many other areas in engineering and science.

Methods for working with these equations are beyond the scope of this book, since the mathematical technology involved can be substantially more advanced than for the ordinary differential equations that we consider. Nevertheless, they can retain aspects of ordinary linear systems. The Laplace equation

$$\frac{\partial x}{\partial t} = - \left(\frac{\partial^2 x}{\partial s_1^2} + \frac{\partial^2 x}{\partial s_2^2} \right) = -\Delta x \quad (2.17)$$

describes how spatial irregularities in the concentration of a solution diffuse over time to finally reach a state of equilibrium. The equation be viewed as a first order buffer in time forced by the negative spatial curvature of function $x(t, s)$ with respect to the two spatial coordinates s_1 and s_2 . This forcing implies that over time hills (negative curvature) will be levelled and valleys (positive curvature) will be filled in.

2.4.4 Algebraic and Other Equations

Mixtures of conventional algebraic equations, expressed as $g(x, u) = 0$, and differential equations arise in many contexts. These equations often define constraints on the solutions, of which a most common example is the initial value constraints $\mathbf{x}(0) = \xi$ where ξ is a vector of real-valued constants.

In addition to algebraic equations, it is not unusual to have integral equations applying to the system as well. Moreover, delay-differential equations often are required where $Dx(t) = f[x(t - \delta)]$ for some nonzero lag δ .

2.5 Data Configurations

The data available for estimating both the parameters defining a dynamic system and a function \hat{x}_i can come from many different sources, and be related to the values $x_i(t)$ in many different ways. We specify some of the data configurations that are commonly encountered. We use y as a rule to refer to data, and y_{ij} , $i = 1, \dots, d$; $j = 1 \dots, n_i$ to indicate data specific to variable x_i and substrate value t_{ij} . Additional subscripts may also be needed to identify replicated data for specific pairs (i, j) .

2.5.1 Initial and Boundary Value Configurations

By far the largest amount of mathematical material on dynamical systems is for *initial value problems* where the state of system $\mathbf{x}(t_0)$ at an initial time t_0 is known and the goal is to predict the behavior of a system at times $t > t_0$. Also included in initial value systems are those in which values of certain derivatives are also known, but we have seen that the use of dummy variables to label derivatives can convert these to the simpler initial state formulation.

Initial value systems are also of great interest to statisticians when the initial values are random, so that the trajectories or space curves with values $\mathbf{x}(t)$ are themselves random objects, possibly because initial values are measured with substantial

imprecision. In these cases, statisticians and probabilists find themselves teaming up with numerical analysts specializing in approximating solutions to the system given specific initial states and differential geometers bring their expertise on the analysis of structures on manifolds.

To initial value systems we can add boundary value systems where something is also known about the system at terminal time t_1 . For example, a Brownian bridge is a trajectory that begins and ends at the same fixed value, such as a periodic process with fixed period.

2.5.2 *Distributed Data Configurations*

As a rule, we hope to work with measurements that are distributed over the complete range of observation.

This raises the critical question, “How many observations do I need?” What matters is not so much the number of measurements as how they are allocated along the line. Solutions to differential equations can have sharp localized features, often as a consequence of being forced by functions u with abrupt changes. By a “feature” one means something like a change in level, a change in slope, a peak or valley, a crossing of some fixed threshold, or even a point of inflection. Using peak as an example, which needs to be defined in terms of its location, its amplitude and its curvature at the peak, three points carefully placed are the minimal configuration required to convey this information. But if the measurements are subject to any error, an accurate characterization can require anywhere from five to eleven points within the range of the peak depending on the amount of imprecision involved.

We call this the *resolution* of the data. Of course over regions in which nothing of interest happens, rather fewer observations will do, with the number depending on whether these regions are flat, tilted or subject to mild curvature. The refinery data in Sect. 1.1.3 are high resolution data by any standard, in contrast to the smallpox data where there are only a couple of observations defining the rate of change after vaccination.

2.5.3 *Unobserved or Lightly Observed Variables*

It is quite common to have some variables that are not observed or only sporadically measured. The spread of disease SIR model is a prototypical example. Fairly comprehensive records are often kept of the number of cases presenting with an infection, with daily or weekly observation times for variable I being typical. On other hand, the number of susceptibles S in a population may be available only annually, and then only approximately since the property may not apply to the whole population. A common situation is that S is only observed once, at a time close to the beginning of the outbreak. The recovered variable R may not be measured at all.

Having none or only a few observations for a variable can have a great effect on what can be estimated in a model. Certain parameters will typically be only determined by information on one or two variables, and therefore may be inestimable if these are not measured. The relationship holding between a measured and an unmeasured variable can also matter. If a variable is causal with respect to another, such as S with respect to I , it will matter greatly that it is observed; but a downstream variable like R can be well defined by a well defined upstream variable like I . If only $S(0)$ is observed, it can matter a great deal that the measurement error is small.

2.5.4 *Observational Data and Measurement Models*

Whether data are available at only time t_0 , or distributed over a range of t -values, it may be that the relationship of the data to a function value $x_i(t)$ may be nontrivial, rather than the usual straight-up additive error relation $E[y_{ij}] = x_i(t_j)$. For example, a set of 0's and 1's can indicate a set of binary outcomes occurring for variable i with a probability $P[x_i(t)]$, in which case a logistic link generalized linear model is a likely choice for estimating function P that connects the model to the data. Another example occurs when the data available relate to the sum, or some other combination, of two or more of the variables that cannot in practice be separated at the observational level. We may see this when we conceptually differentiate between reproducing organisms and the same organism in a non-reproductive phase of their life cycle, and it only the reproducing individuals that are affecting other variables.

It may happen, too, that the data for various variables differ enormously in their scale, as can happen for population and spread of disease models. In such cases we may consider either a model on the log scale with an exponential transformation of model values to fit the data, or logging the data prior to an analysis, which also requires a log-scale model, but with a straightforward connection to the data. That is, transforming variables can be important. In the log case, for example $Dx = f(x)$ is equivalent to $D \ln x = f(x)/x$ or, if $y = \ln x$, $Dy = f(\exp y) \exp(-y)$.

The nature of a measurement can impact the data resolution in a distributed data configuration. For example binary variables convey much less information per measurement than directly connected continuous measurements with a high-signal-to-noise ratio. A useful rule of thumb is that it takes at least five binary measurements at or near a location to yield the same resolution as a single low-noise continuous measure.

It happens often that the data do not provide enough information about the values of certain parameters to yield a useful estimate. For example, Chap. 10 looks at an example where a parameter specifying an upper limit cannot be estimated because there are no observations anywhere near such a limit. Similarly, estimating what a variable does near a lower limit such as zero requires that there be observations in that zone. Again, if a pair of parameters always appear in a product, then they are not individually identifiable.

The term *experimental design* in statistics refers to the the relationship between the data and elements of a model designed to fit them. Good experimental designs ensure that there are data informing the estimation of all elements of a model. Bad designs should usually be seen as an imperative to simplify the model before attempting any analysis.

2.6 Differential Equation Transformations

We noted above that variables can often vary over large scales, suggesting that it would be better to work with a logarithm of the data, whether with base e or 10. Of course, that would imply that it would be convenient to know the differential equation in the log scale that is equivalent to the original. Letting $y = \log x$, we see that

$$D \log x = Dy = \frac{f(x|\theta)}{x} .$$

That is, substitute $x \rightarrow e^y$ into the raw-scale equation and divide by e^y to obtain the equivalent equation in the Y or log-scale.

Similarly, a linearization of a differential model for a variable bounded below by 0 and above by K would imply the transformation $y = \log[x/(K - x)]$ and therefore the multiplication of the right side function by the factor $K/(K - x)$.

2.7 A Notation Glossary

Now that we have introduced the basic set of models that will be employed in this book, we will collate our notation before continuing. The reader will have already seen some of these, but we will formalize it here. This book will largely be concerned with the combination of two models, although we have so far only formally discussed one. The first of these is the *process model* which describes the way a system changes over time. Formally, we will use

- x_i is used to designate a single variable among a set of d variables.
- $\mathbf{x}(t)$ is used to represent the state vector of the system at time t . In the case of the SIR models above we have $\mathbf{x}(t) = (S(t), I(t), R(t))$ and we will continue to refer to the constituent parts of the state vector by other names or by using subscripts as in $\mathbf{x}(t) = (x_S(t), x_I(t), x_R(t))$. When the state vector \mathbf{x} is viewed as a function of time, it will be described as the state *trajectory*.
- $D\mathbf{x}(t)$ is the vector of time derivatives of $\mathbf{x}(t)$. The differential operator D will be used exclusively for time.
- $\mathbf{f}(\mathbf{x}|\theta)$ represents a vector-valued function of \mathbf{x} that depends on a parameter vector θ . Generally this is used to represent the right hand side of a differential equation so that the SIR differential equation can be expressed as $D\mathbf{x} =$

- $\mathbf{f}(\mathbf{x}|\boldsymbol{\theta})$. Again, the components of \mathbf{f} can be indexed by named subscripts. In the SIR model we might have $\mathbf{f} = (f_S, f_I, f_R)$.
- $\mathbf{u}(t)$ will be used to describe a vector of external inputs into the system that are not affected by system dynamics. If we imagine an external source of new susceptibles in the SIR spread of disease differential equation which increases S with rate $u(t)$, the S equation could be modified to $DS = u(t) - \beta SI$. In a generic ODE model, we will include these by writing $D\mathbf{x} = \mathbf{f}(\mathbf{x}|\boldsymbol{\theta}, \mathbf{u})$.
- \mathbf{x}_0 will refer to the starting point of the trajectory. We distinguish this from $\mathbf{x}(0)$ because it will often need to be estimated as additional parameters. When this is the case, we will use the augmented parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\theta}, \mathbf{x}_0)$.

Along with these notations, we need a second model to describe the measurement of the system, or the *observation model*. We may observed several quantities about a system and so describe

- \mathbf{y}_j will be used to represent the vector of measurements of length d of a set of d variables at time t_j . Sometimes it will be useful to write down a matrix of observation vectors over time. For this we will use \mathbf{Y}_j in which rows denote time points and columns the dimension of the observation vector.
- $P(\mathbf{y}_t|\mathbf{x}(t), \theta)$ gives a probabilistic model for the observations. Frequently, we will model $y_{it} = x_i(t) + \varepsilon_{it}$ — that is we will measure at least some components of the state vector directly and assume additive noise. However, it will often be useful to be more complex than this. In the case of the SIR models above, we might assume binomial sampling of the infected individuals, for example. Other times, the error variance will scale with the observations. There are also occasions in which the sum of two state variables is observed. This model can, of course, depend on $\boldsymbol{\theta}$, although we will mostly be concerned with parameters that govern the process model.

Finally, in order to avoid confusion, we have used D to represent solely a derivative with respect to time. In many places, we will need to differentiate a likelihood with respect to parameters or other quantities and we will refer to the vector of derivatives by $\partial_{\boldsymbol{\theta}}$ where the subscript indicates which derivatives are being taken.



<http://www.springer.com/978-1-4939-7188-6>

Dynamic Data Analysis

Modeling Data with Differential Equations

Ramsay, J.; Hooker, G.

2017, XVII, 230 p. 84 illus., 50 illus. in color., Hardcover

ISBN: 978-1-4939-7188-6