

Introduction

... while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. (Sherlock Holmes, Sign of the Four).

Complex diseases involve most aspects of population biology, including genetics, demographics, epidemiology, and ecology. Mathematical methods, including differential, difference, and integral equations, numerical analysis, and random processes, have been used effectively in all of these areas. The aim of these notes is to provide sufficient background in such mathematical and computational methods to enable the reader to better understand complex systems in biology, medicine, and the life sciences.

Malaria provides an important example of a complex disease system. Parasite life cycles, described in Figure 6.1, show aspects of the disease process in both the host (humans here) and mosquitoes. In addition to the complex parasite life cycles, these populations are stratified in several ways: There are varieties of genetic structures and immune systems, age structures and geographical distributions, interactions with physical habitats, and complicated timing issues. Overlaid on these biological and physical systems are various social, medical, financial, and governmental structures whose policies bear directly and indirectly on the disease [35]. Many useful mathematical tools used to study various aspects of malaria are described here.

Mathematics is the language of systems, and mathematical methods used for synthesizing and analyzing them are widely applicable. These methods provide:

- systematic listings of the principal components of a system and their interactions with others,
- mathematical tools of analysis including perturbation, bifurcation, stability, computation, and probability methods,
- concise formulation of the overall structure,
- direct implementation in computers for monitoring, control, tracking, prediction, evaluation, and simulation,
- guidance in deriving simplified models that are faithful to the overall structure.

The mathematical background needed to handle this material is difficult to specify — the material is usually taught in the format of a one-room school where students have widely varying levels of preparation. For example, undergraduate engineering students routinely handle linear time-invariant systems (like the renewal equation here), Laplace and Fourier transform methods, delta functions, and

random processes, while many people who finish a PhD in mathematics never encounter any of these ideas in their studies. On the other hand, mathematics students usually end up with a good understanding of calculus and matrices and many of their implications, which students of the nonmathematical sciences and engineering often do not.

Integration, differentiation, matrices, and probability are common threads to almost all mathematical descriptions of the world as we know it. These topics have been extended and refined over the past 400 years, but with the advent of broadly available high-speed computing, they have taken on new meanings and usefulness. Many of the operations formerly requiring advanced knowledge of mathematics, such as Bessel, Legendre, Hermite, Chebyshev, and other special functions, have now largely disappeared behind the computer screen — a metaphor for mathematics being hidden in user-friendly software products. At the same time, the computer has made possible deeper understanding of randomness and intrinsic chaos in nonrandom systems.

While mathematics is the language of systems, many nonmathematical people will find using mathematics to be challenging. Still, it is important for those working in these areas to become familiar with the philosophy, terminology, and methodologies in using mathematics. Not only does mathematics provide a way to interface complex problems with computers for simulation and data processing, it also provides a universal language for describing systems. Even when mathematical aspects have moved behind the computer screen, it is important to understand what computer packages are trying to do.

The mathematician René Thom pointed out a cultural gap between the mathematical and observational sciences: The latter use ordinary language models that are precise where data are known and suitably vague otherwise, and the former use canonical mathematical models that are the simplest mathematical models that capture a particular feature of the phenomenon. Mathematics has an uncanny ability to give insight in a variety of unusual settings: This is due in part to the remarkable utility of canonical models that arise in the same form in many different applications. For example, the mathematics of chemical explosions and the propagation of epidemic diseases in populations have been described using essentially the same mathematical model. The result for explosions is the explosion limit for combinations of pressure and temperature, and the result for epidemics is the tipping point for combinations of infectives and susceptibles. There are only (relatively) few types of bifurcations that can occur when something changes in a system. Newton's implicit function methods and Thom's singularity theory and its generalizations have clarified many aspects of such tipping points.

Population phenomena may be described in terms of discrete-time events or continuous-time events. Both approaches have their strengths and weaknesses, and in many ways are complementary. Discrete-event models often seem more intuitive, and when their time steps are synchronized with the acquisition of data, more practical. They also usually make possible straightforward computer simulations. But they can get messy when many variables are needed, and mathematical analysis for them is difficult — even simple nonlinear discrete-event renewal models are

at the frontier of mathematical research. On the other hand, continuous-time models provide concise mathematical descriptions (e.g., a sequence of discrete-event variables may be replaced by a single function), and they may be analyzed using methods of differential and integral calculus. However, simulating continuous-time models on a computer may require sophisticated numerical analysis. Both approaches are used here.

Charles Darwin, who is a hero to most people working in the life sciences, is generally perceived as having used no mathematics, but in fact it did provide him with something important to his work — a theory. He said in his autobiography (1876):

In October 1838, that is, fifteen months after I had begun my systematic inquiry, I happened to read for amusement *Malthus on Population*, and being well prepared to appreciate the struggle for existence which everywhere goes on from long-continued observation of the habits of animals and plants, it at once struck me that under these circumstances favourable variations would tend to be preserved, and unfavourable ones to be destroyed. The results of this would be the formation of a new species. Here, then I had at last got a theory by which to work.

The physicist R. Feynman said (paraphrased) that mathematics produces correct results that take many additional years of work to “understand” in physical terms. An example is the development and application of quantum mechanics, which is even now not fully “understood.” Much of this “understanding” is rote memorization of what mathematics tells until it seems natural. Similarly, our understanding of electromagnetic theory is shaped by knowing the solutions to Maxwell’s equations, not by direct physical observation; our understanding of fluid mechanics is largely based on mathematical solutions of the Navier-Stokes equations, and our understanding of the mechanical universe is based on Newton’s and Einstein’s mathematical theories.

Mathematics plays comparable roles in population biology. Our understanding of population genetics is based in large part on mathematical models by Fisher, Wright, and Haldane. Renewal theory, built largely on work of Euler and Lagrange, describes how populations age and reproduce. The Kermack-McKendrick, Reed-Frost, and Ross-Macdonald theories of epidemics describe the propagation of diseases in small groups, and they have been useful in formulating optimal vaccination strategies and identifying tipping points for epidemics that measure likelihood of severe disease outbreaks. Mathematics is also used in working with microbial ecologies, such as describing nutrient uptake, that are central to food and drug production.

These notes build partly on the text *Modeling and Simulation in Medicine and the Life Sciences* [23]. Here we introduce and use slightly more advanced concepts in mathematics to study population phenomena with the goal of describing complicated aspects of a disease involving several species. But the mathematical level is kept to essentially advanced undergraduate mathematics.

The culmination of this work is in Chapter 6, where population stratifications involve two kinds of age classes (chronological and infection age) for both the mosquito and human populations, geographic distributions and movements of species, and the microbial ecology of the parasites. Parasite dynamics in a human host, parasite genetics in a mosquito, and the influences of these microfeatures on the dynamics of human and mosquito populations are also described by these models. For the most part, the components of the models are derived and tested in the first five chapters. As pointed out in the earlier chapters, models may be modified to include random perturbations, multiple time scales, and particular bifurcation features. Better understanding of the disease results when the cumulative models are analyzed by using the methods in the first five chapters. Still, one is in awe of the ways in which the malaria parasite has evolved to thrive in a very challenging environment, and the models here bring better appreciation of nature at work.

Some general principles of modeling are guided by Einstein's remark: "Any fool can make things bigger, more complex Make everything as simple as possible, not simpler." While mathematical population models may seem to far outstrip present-day knowledge and data, they should be based on real mechanisms of important phenomena, and should not be impeded by only what is known or assumed today. In fact, describing a system as fully as possible guides acquisition of data, evaluation of control efforts, prediction of likely outcomes when using reasonable surrogate data, and identification of critical dimensionless parameters that may characterize important aspects of a disease. They also guide how to derive simpler models that capture critical aspects of the phenomenon; for example, using quasi-static state methods in the spirit of the Michaelis-Menten and Jacob-Monod approximations studied in Chapter 3. It is in this spirit that the malaria models in this chapter are presented.

Notwithstanding the lack of data, there remain significant mathematical challenges in understanding these "simplest" models, which involve combinations of time delays, geographical distributions, and nonlinear dynamics. The earlier chapters introduce methods for time delays by introducing agelike variables, for random forces by replacing parameters by stochastic processes, and for nonlinearities by using mathematical analysis (gradient and perturbation methods) and computer simulations.

Ordinary-language models may describe a disease in ways that are "precise" where data are known and vague otherwise. For example, an ordinary-language model of an $S \rightarrow I \rightarrow R$ system, which is typical of diseases like measles and chickenpox, might say "susceptibles are infected by contact with infectious people; contact may vary with mixing group (family, play group, school, social organization, etc.); communicability may depend on age; infectiousness may depend on time since first infection; removal of infectives may be through treatment, immune response, isolation, or death; recovered individuals may not again become susceptible to the disease; the disease process is faster than arrival of new susceptibles; there may be no carriers for the disease; and the disease is local."

Kermack and McKendrick derived a mathematical model of such a disease process to provide the simplest mathematical model that fits the "most important"

of these restrictions: There is pairwise contact between susceptibles and infectives; infectives may be removed; there is a single mixing group; removed individuals are immune to further infection by the disease; chronological and infection ages are ignored; population dynamics (birth, death) are ignored; carriers are ignored; and, geographical movements are ignored. The result is the model (4.2). It is shown in Chapter 4, equation (4.3), that the Kermack-McKendrick number

$$\kappa = \frac{rS(0)}{\lambda},$$

where r is the contact rate, λ is the removal rate of infectives, and $S(0)$ is the initial susceptible population size, defines a critical dimensionless parameter: If $\kappa < 1$, infectives will be removed without increasing their numbers. If $\kappa > 1$, a severe epidemic may occur. No matter what other features we add to the model, the parameter κ will still be relevant. In this sense, the Kermack-McKendrick model is a canonical model of disease propagation.

This manuscript summarizes notes developed in a graduate course in computational biology and applied mathematics at the Courant Institute of Mathematical Sciences in fall 2010. The models and results here should be taken as metaphors of real-life phenomena, not as exact quantitative or qualitative descriptions of them. In particular, these results are intended to provide readers with tools for performing more in-depth analysis of population phenomena and make their own contributions to resolving problems. I thank Charles Peskin, Andy Sundstrom, Don Chen, and Karen Day for input to this work. They added valuable information for some of the models at critical times, but only I can lay claim to the errors made in this book, which I hope are few.

Frank C. Hoppensteadt
 Courant Institute of Mathematical Sciences
 New York University
 New York, New York
 April 29, 2011