Abstract

This module gives a brief introduction to models of disease transmission on contact networks. These models allow for exploration of stochastic effects and incorporation of more biological detail than the classical compartment-based ODE models. The module contains both conceptual and simulation-based exercises and projects. The former mainly illustrate some of the tradeoffs inherent in choosing a mathematical model, while the latter focus on exploring how properties of the underlying contact network influence the expected dynamics.

1 Introduction

We are interested in diseases that are triggered when infectious agents such as viruses or bacteria (called microparasites) enter the organism of a host (human, animal, plant). We want to model how the disease spreads between hosts of a given population. We focus on diseases whose transmission requires direct contact (of a certain type) between hosts.

At the most coarse-grained level, disease transmission can be modeled by partitioning the host population into compartments. The most basic models consider up to four compartments, defined as follows:

- $S$ comprises all susceptible hosts,
- $E$ comprises all exposed hosts that will eventually develop the disease but are not yet infectious,
- $I$ comprises all infectious hosts, and
- $R$ comprises all removed hosts that either have recovered and acquired permanent immunity or have died from the disease.
A model with all four compartments is called an \textit{SEIR-model}. The simplifying assumption that hosts become infectious as soon as they are infected eliminates the $E$-compartment and leads to the \textit{SIR-model} of [6] which was the first compartment-based model of disease transmission described in the literature. If in addition no recovery or death from the disease is considered, then the $R$-compartment becomes redundant and we get an \textit{SI-model}. If we assume instead that hosts eventually recover without acquiring immunity to reinfection, then we get an \textit{SIS-model}.

There exists an extensive literature on these and related models, see for example the textbooks [2, 3, 5, 10] or the survey [4]. In particular, [3] contains many excellent and well-structured exercises. Most introductory courses of mathematical biology discuss at least the ODE versions of some of these models.

However, compartment-based models ignore \textit{heterogeneities between hosts} in the same compartment and are based on the assumption of \textit{uniform mixing} between hosts in any given two compartments. ODE models ignore the stochastic nature of disease transmission. The former deficiency can be addressed by considering \textit{individual-based models}, the latter by \textit{stochastic models}. In principle, stochastic individual-based models allow for incorporation of any level of biological detail that might be relevant, but may be mathematically intractable and may have more parameters than can realistically be estimated from available data.

The disease transmission network models described here are one type of stochastic individual-based models that strike a compromise between mathematical tractability and biological realism. Hosts are still assumed to be identical in their propensity to get infected during a given contact and to recover from the disease during a given time interval, but possible contacts are restricted to the edges of a graph which is assumed to represent the contact network. The study of this type of model is a vibrant area of recent and current research, see for example Chapter 9 of [1], Chapter 17 of [7] or Chapter 7 of [8]. The materials posted here aim at providing an entry point into this approach to modeling disease dynamics at a level suitable for active exploration by advanced undergraduates.

\textit{When is a model good enough?}

This is the central question of our proposed explorations. We want to build models that are simple enough to be tractable, either by mathematical analysis or computer simulations, and have parameters that can be estimated from data that we actually can collect. On the other hand, the model should take into account sufficiently many biological details so as to make reasonably correct predictions, in particular, about the \textit{probability of an epidemic}, its \textit{final size}, and the effectiveness of possible \textit{control measures}.

Tractability and limits on the number of parameters can only be achieved by suitable simplifying assumptions. Many of our conceptual exercises aim at initiating open-ended discussions about how and when a given simplifying assumption might distort the predictions of a model. The explorations of specified network models by means of simulations aim at elucidating properties of the contact network that appear to most significantly influence the predictions of the model. Such properties can often be expressed in graph-theoretic language, for example, in terms of degree distribution, and knowing how they influence the expected disease dynamics may allow us to make realistic predictions in the typical situa-
tion where the full contact network is unknown, but some of its properties can be estimated from available data.

2 Conceptual exercises

Read the lecture given by the lead author at Lecture.pdf. The exercises below are best attempted immediately after watching or reading the corresponding slides. Many of them are open-ended and admit more than one correct answer. Our sample solutions are only meant as suggestions for possible answers.

**Exercise 1** *(Attempt after slide 5)* We have assumed that host number \(i\) will always be initially susceptible, may be exposed to the disease at time \(T_{E_i}\), will subsequently become infectious at time \(T_{I_i} \geq T_{E_i}\), may cease to be infectious at time \(T_{R_i} \geq T_{I_i}\), and will never subsequently become susceptible or infectious. What kind of diseases do not satisfy the above assumptions?

Sample Solution

**Exercise 2** *(Attempt after slide 11)* Can you think of meaningful compartment models of types other than SEIR, SIR, SI and SIS?

Sample Solution

**Exercise 3** *(Attempt after slide 12)* Assume an SIS model instead of an SIR model. How would vaccination at time \(T_v < T_E\) of a fraction \(r\) of hosts translate into compartmentalese?

Sample Solution

**Exercise 4** *(Attempt after slide 12)* Does compartmentalese allow us to make a distinction between vaccination, culling, and quarantine? Consider both SIR models and SIS models.

Sample Solution

**Exercise 5** *(Attempt after slide 57)* Is there a problem with the proof on this slide?

Sample Solution

**Exercise 6** *(Attempt after slide 58)* How would one collect, for a real population of hosts of plants, animals, or humans, data on each of the following?

- The pattern of mixing between susceptible and infectious hosts.
- The probability that a given contact between an infectious and a susceptible host results in a “successful” transmission.
• The distribution of times $T^i_I - T^i_E$ from exposure to onset of infectiousness and $T^i_R - T^i_I$ of duration of infectiousness.

When would obtaining such data be relatively easy, when would it be difficult?

Sample Solution

Exercise 7 (Attempt after slide 58) Can you think of a population of real hosts for which it would be possible to collect all the relevant data?

Sample Solution

Exercise 8 (Attempt after slide 61) Our model ignores demographics, the latency period between exposure and onset of infectiousness, and the possibility of multiple below-threshold exposures adding up to an infection.

How distorting are these simplifying assumptions likely to be?

How could we incorporate the ignored details into our model?

Sample Solution

Exercise 9 (Attempt after slide 62) We assume that for any given state $\vec{x}(t)$ the relevant transition times $T^i_I$ and $T^i_R$ are all independent.

How realistic is this assumption?

Sample Solution

Exercise 10 (Attempt after slide 65) Can you think of a scenario where we would want $\alpha_i(\vec{x})$ to actually depend on $\vec{x}$? In other words, when would the expected time to removal of infectious host number $i$, which is determined by this parameter, depend on the states of the other hosts?

Sample Solution

Exercise 11 (Attempt after slide 66) The distribution of removal times $T^i_R - T^i_I$ is usually closer to a normal than to an exponential distribution.

How can we modify the model so that the distribution of recovery times becomes more realistic without sacrificing the Markov Property of the process?

Sample Solution

Exercise 12 (Attempt after slide 68) We have assumed the following: there is only one type of contact that lasts an instant, transmission probabilities per contact are fixed, and waiting times for the next contact between any given pair of hosts are exponentially distributed. We have also ignored heterogeneities in individual immune response.

What are some potential problems with these assumptions and how could we address them without adding too many parameters to our model?

Sample Solution
3 Projects and software

The projects are suitable for exploration by students on their own or in groups and based on computer simulations. Projects 1–5 are geared towards advanced undergraduate or beginning graduate students in mathematical biology or mathematics. They aim at leading to a deeper understanding of how mathematical properties of the network dynamics are related to graph-theoretic properties of the underlying network. The instructions for these projects are written for the MatLab version of the software that has been developed for this module. The documentation for the MatLab software as well as the projects themselves can be found in the file MatLabProjects.pdf. Here we give brief descriptions of these projects.

- Project 1 explores the meaning of the parameters $\alpha$ and $\beta$ of our models and how they influence the expected dynamics.

- Project 2 compares the predictions of the ODE and stochastic versions of the SIR model with uniform mixing. In particular, students will investigate the probability of an epidemic vs. a minor outbreak.

- Project 3 explores potential problems with estimating $R_0$ from epidemiological data.

- Project 4 explores the relationship between the uniform mixing assumption and the assumption that the underlying contact network is random, more specifically, is either an Erdős-Rényi or a regular random graph. In particular, students will explore whether $R_0$ is still a good predictor of an epidemic under the latter two assumptions.

- Project 5 addresses the precise meaning of the word “repeat” in simulation studies and contains some MatLab coding exercises.

Subsequently we developed a version of the simulation software under the NetLogo platform [11]. The software NetLogo is free, has better visualization capabilities, and may be preferable for most students. Our code, called virus_on_network3.nlogo and partly based on [9, 12], is included in this package. The file NetLogoProject.docx contains a sample project that is introductory in nature and geared towards undergraduate students in the life sciences.

Acknowledgements

Development of the materials presented here is in part an outcome of fruitful discussions during two hands-on sessions of the MBI workshop Teaching Discrete and Algebraic Mathematical Biology to Undergraduates. We wish to express our gratitude to the MBI and the organizers of the workshop for providing such a productive environment and the participants of the discussions, especially David Dorman, Jennifer Galovich, and Talitha Washington, for important contributions. We are greatly indebted to Dmitry Kondrashov of the University of Chicago whose contributions and initiative were crucial for catalyzing this collaborative effort.
References


