Mathematical models and theorems*

Winfried Just†

February 1, 2017

In this module we introduce and compare various types of deterministic and stochastic mathematical models of disease transmission. We then illustrate how one can derive predictions of these models in the form of mathematical theorems.

1 Introduction

When studying models of disease transmission, we would like to answer the following:

Question 1 *(a) What outcomes does a model predict, on average, for a given choice of parameters?*

(b) How, exactly, do these predictions depend on the choice of the parameters?

(c) How much variability in the outcomes should we expect to see between different outbreaks?

(d) How do these predictions scale if we increase the population size N?

In most of the modules at this website[1] we are exploring simulated outbreaks of diseases. These explorations rely on agent-based models of disease transmission that are embodied in the computer program IONTW.

Simulations can lead to meaningful conjectures about the answers to Question 1. But their outcomes are subject to random fluctuations that may distort the predictions. If one runs large batches of them, one can derive estimates on the level of confidence that can be assigned to the answers to parts (a) and (c) for a given choice of parameters. This is often sufficient for practical purposes, but falls short of absolute confidence.

Deriving answers for part (b) of Question 1 from agent-based models is even more problematic. One can run simulations only for a given choice of parameters, or a sample of such choices. But parameter settings that are left unexplored might give very different outcomes. Part (d) of Question 1 is the most difficult one to explore with agent-based models. It may be feasible to run simulations for populations of perhaps a few hundred or a few thousand hosts, but the sizes of real populations of interest may be in the millions.

†Department of Mathematics, Ohio University, Athens, OH 45701 E-mail: mathjust@gmail.com

[1] `https://qubeshub.org/iontw`

No finite number of simulations will give us certainty, will let us explore all possible choices of parameters, or allow us to explore the actual limit $N \to \infty$. If we aim at certainty and complete characterization of the behavior under all possible parameter settings, including the population size N, we need to work with *mathematical models.* The latter are abstract mathematical constructs that can be studied theoretically, using mathematical tools rather than computer simulations.

One type of mathematical models is already described in our module *Differential equation models of disease transmission* at this website.[2] There we also derived some predictions of ODE models that give rigorous answers to parts (a) and (b) of Question 1 under suitable assumptions. But ODE models are not suitable for addressing parts (c) and (d) as they are deterministic and based on the implicit assumption of a "practically infinite" population size.

The remainder of this module is organized as follows: In Section 2 we will briefly revisit ODE models and introduce *difference equations,* a related type of deterministic models where time is assumed to progress in discrete steps. In Section 3 we will then present an entirely different type of models called *stochastic processes.* Some of these processes embody the exact same models as IONTW, but in the form of mathematical constructs instead of computer code.

By analyzing these mathematical constructs, it is sometimes possible to derive rigorous answers to all parts of Question 1 in the form of *mathematical theorems.* We will give examples of such derivations in Section 4, where we will revisit theorems that were stated somewhat informally in Section 9.2 of [4].

This module can only aspire to give you a flavor of the study of mathematical models of disease transmission. Section 5 contains a brief discussion of how to use various types of models together with pointers to the literature for more in-depth study.

2 Deterministic models

Any mathematical model of disease transmission has three ingredients: the variables of the model, a notion of time, and a rule that tells us how the variables change over time.

When time can take only integer values, we talk about a *discrete model;* when time can take all real values or all real values in $[0, \infty)$, we talk about a *continuous-time model* or *continuous model.* The vector of values of the variables at any given time t is the *state* of the model at time t. The state at time $t = 0$ is commonly called the *initial state.* If the rule of change uniquely determines the state at any future time $t \geq 0$ for any given initial state, we talk about a *deterministic model;* if the rule only provides probability distributions, we talk about a *stochastic model* or *stochastic process.*

The purpose of models of disease transmission is to predict what will, or is likely to, happen in a given population of hosts. The models that we are considering here ignore demographics, that is, births, immigration, emigration, and deaths from causes that are

[2] `https://qubeshub.org/iontw`

unrelated to the disease. If each individual host is represented by one or more separate variables, we will talk here about *host-level models.* In the literature, the phrase *individual-based models* is often used instead. Note that the models that are embodied in IONTW are examples of host-level models; albeit not mathematical ones. When instead all variables represent some characteristics of the entire population, such as overall prevalence of the disease at time t, then we speak of *compartment-level models.* The phrase *compartment-based models* is more common in the literature, but slightly misleading, as all known models of disease transmission are based on the notion of compartments.

2.1 Very brief review of ODE models

ODE models of disease transmission are continuous-time compartment-level models. Here we will restrict our attention to models of type SIR. ODE models of this type have variables S, I, and R, which can be interpreted as actual numbers, expected numbers, or (actual or expected) proportions of hosts in the **S**-, **I**-, **R**- compartments. These variables will change over time, and $S(t), I(t), R(t)$ will denote their values at a given time t.

The rule of change is given by a system of ordinary differential equations. In the SIR-model that will be discussed here they take the form

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \alpha I, \\ \frac{dR}{dt} &= \alpha I, \end{aligned} \tag{1}$$

where the parameters α, β are positive constants.

Along any given solution curve, the equation $S(t)+I(t)+R(t) = N$ holds at all times t, where N represents the constant size of the host population.

The use of derivatives makes sense only if N expresses the actual population size in large batches, such as thousands or millions. We can then think of (1) as approximating the behavior of agent-based SIR-models with the uniform mixing assumption for practically infinite populations. Under suitable scaling, the parameters α and β of (1) correspond to the parameters **end-infection-rate** and **infection-rate** of the corresponding continuous-time models that are embodied in IONTW. You may think of the predictions of the ODE model as giving half an answer to part (d) of Question 1 in that they can be interpreted as predicting what happens if $N \to \infty$. They do not, however, shed light on the question how so-called *finite-size effects* might distort these predictions when N is rather small.

The basic reproductive ratio R_0 in the model (1) is

$$R_0 = \frac{\beta N}{\alpha}. \tag{2}$$

Let $s(t) = \frac{S(t)}{N}$ and $s(\infty) = \lim_{t\to\infty} s(t)$. Note that if $R(0) = 0$, then $1 - s(\infty)$ represents

the final size of the outbreak. The model predicts that

$$s(\infty) = s(0)e^{R_0(s(\infty)-1)}. \tag{3}$$

Now consider an initial state with one index case in an otherwise susceptible population. As ODE models implicitly assume a practically infinite (actual) population size, we would have $s(0) \approx 1$, and (3) can be approximated by the equation

$$s(\infty) = e^{R_0(s(\infty)-1)}. \tag{4}$$

When $R_0 > 1$, (4) has exactly one solution in the interval $(0, 1)$; when $R_0 \leq 1$, then $s(\infty) = 1$ is the only solution.

Thus the model (1) predicts that when $R_0 \leq 1$, *every* outbreak with such an initial state will be minor, and when $R_0 > 1$, *every* outbreak with such an initial state will be major, with final size that can be deduced from (4). The latter prediction is unrealistic, since introduction of one index case into an otherwise susceptible population will not *always* result in a major outbreak; only with a nonzero probability. The model gives a fairly good estimate of the expected final size for a large population *if* a major outbreak does occur. But since it is deterministic, it cannot tell us how likely a major outbreak will be, or give any other information about the expected amount of variability between outbreaks. ODE models cannot address part (c) of Question 1.

More information about ODE models, and in particular, derivations of (2) and (3) can be found in our module *Differential equation models of disease transmission* at this website.[3]

2.2 Models based on difference equations

While ODE models have the advantage that for their study we can rely on well-developed mathematical theory, they have several features that make them a bit difficult to interpret and match to data. One is that the variables S, I, R in such models represent (suitably scaled) integer values, so that their derivatives cannot be treated too literally. Another issue is that time t in these models can take on arbitrary real values, but data on the values of these variables, most notably the disease prevalence I, are usually reported only once for each fixed time interval, such as a day or a week. Both of these problems are absent in models based on *difference equations.* We will introduce these in the context of SIR-models.

In difference equation models, time t can take any nonnegative integer values. As in an ODE model, the variables S, I, R represent actual or expected numbers or proportions of susceptible, infectious, and removed hosts. We will assume here that they represent counts of hosts rather than proportions. As the variables can take values that are not integers, in this view of the models they should be interpreted as estimates. The values of these variables at time t are usually denoted by S_t, I_t, R_t, respectively. The state of the system at time t is represented by the vector (S_t, I_t, R_t). The change of these variables over time is determined by a so-called *updating function* F so that

[3] `https://qubeshub.org/iontw`

$$F(S_t, I_t, R_t) = (S_{t+1}, I_{t+1}, R_{t+1}) \tag{5}$$

gives the state at the next time step.

Let F_S, F_I, R_I denote the components of the vector function F. In this notation, (5) can be written as a system

$$\begin{aligned} S_{t+1} &= F_S(S_t, I_t, R_t), \\ I_{t+1} &= F_I(S_t, I_t, R_t), \\ R_{t+1} &= F_R(S_t, I_t, R_t). \end{aligned} \tag{6}$$

Now suppose we are given an initial state (S_0, I_0, R_0). Then (5) does not immediately give us explicit formulas for (S_t, I_t, R_t) at all times $t > 0$. But it allows us to calculate (S_1, I_1, R_1). We can substitute the result in the left-hand side of (5) and calculate (S_2, I_2, R_2). And so on. Thus (5) provides an *iterative* or *recursive* procedure for calculating (S_t, I_t, R_t) at all times $t > 0$ when the initial state (S_0, I_0, R_0) is given. Similar to ODEs, difference equations give us *deterministic* models.

The description that we have given up to this point applies to all difference-equation models with three variables S, I, R. What particular form should the components of the updating function take in an SIR-model of disease transmission? Let us assume that the actual time interval that is represented by one time step is sufficiently short so that no host can become infectious and then move to the **R**-compartment during the same time step. In other words, let us assume that no host can be susceptible at time t and then already be removed at time $t + 1$. By the assumptions of SIR-models without demographics, at time $t + 1$ the **R**-compartment will contain all hosts who already were in this compartment at time t, plus a certain fraction of hosts who resided in the **R**-compartment at time t. The updated value I_{t+1} must be computed by subtracting this fraction from I_t and adding the number of newly infected hosts. The latter number will depend on S_t and I_t. It must be subtracted from S_t. Therefore the difference equations for an SIR-model will be of the form

$$\begin{aligned} S_{t+1} &= S_t - B(S_t, I_t), \\ I_{t+1} &= I_t + B(S_t, I_t) - A\, I_t, \\ R_{t+1} &= R_t + A\, I_t, \end{aligned} \tag{7}$$

where A is a positive constant and B is a function of two variables.

Note that (7) implies that for all times t

$$S_t + I_t + R_t = S_{t+1} + I_{t+1} + R_{t+1} = N, \tag{8}$$

where N is the size of he population. This is exactly as it should be, since in our construction of the model we have ignored demographics, that is, births, migration, and deaths from unrelated sources, so that the population size remains constant.

Exercise 1 *(a) What form would an SIS-model based on difference equations take? Ignore demographics. in this part of the exercise.*

(b) What form would a SI-model based on difference equations take if we

- *ignore migration,*
- *assume constant per capita birth and death rates that are not influenced by the infection, and*
- *assume that mother-to-infant transmission does not occur so that all hosts are born susceptible?*

(c) What does the model you constructed in part (b) predict about the total population size N_t at time t?

Let us return to the SIR-model (7). The notation A and B is reminiscent of the notation a and b for the probabilities of removal after one time step and effective contact during one time step that were used in [4, 5] when we discussed mathematical aspects of discrete-time agent-based models. In IONTW they are called **end-infection-prob** and **infection-prob**, respectively. Is $A = a$ and $B(S_t, I_t) = b\, S_t\, I_t$? If not, how are A and B related to a and b? We will return to this question in Subsection 3.2.

For now, let us divulge that there is indeed a connection: Both the assumption of homogeneity of hosts and of uniform mixing are implicit in the construction of models based on difference equations. The assumption of homogeneity of hosts is made in all models that are embodied in IONTW, and the uniform mixing assumption can be enforced by choosing the complete graph as the underlying contact network. IONTW actually computes the values of S_t, I_t, R_t at each step of the simulation of a discrete-time SIR-model and even graphically displays them to the screen. However, in the simulations done by IONTW, the current state (S_t, I_t, R_t) does not uniquely determine the next state $(S_{t+1}, I_{t+1}, R_{t+1})$. The software incorporates the stochasticity that is inherent in transmission of actual diseases, while difference equation models are deterministic and do not. But one can interpret $F(S_t, I_t, R_t)$ as the vector of expected values of the variables at the next time step, given the current state. This will allow us in Subsection 3.2 to derive expressions for A and B that correspond to the probabilities a and b.

3 Stochastic process models

The deterministic models of the previous section can be studied with mathematical tools, but they do not account for the stochasticity that is inherent of disease transmission. Agent-based models do account for the randomness, but their predictions can only be studied by running simulations. *Stochastic processes models* give us the best of both worlds in that they incorporate randomness and are mathematical models that allow for the derivation of their predictions in the form of theorems. You can think of a stochastic process model as an agent-based model that is defined in terms of a mathematical construct, called a *stochastic*

process, instead of embodied in computer code. The actual simulations obtained by running the corresponding computer code give so-called *realizations* of the stochastic process. While in practice one can run only a finite number of simulations, in the mathematical framework of stochastic processes one can derive of statistical properties of the entire ensemble of possible realizations.

At the most general level, a stochastic process is a vector of random variables (r.v.s) that change over time. The *state of the process* at time t is a vector $\vec{x}(t) = (x_1(t), \ldots, x_N(t))$, where $x_i(t)$ is the *state of random variable* x_i at time t. Stochastic processes differ from mere collections of random variables in that they allow for modeling of situations where the distributions of the random variables $x_i(t^+)$ at time t^+ may depend *in interesting ways* on the states $\vec{x}(t)$ of the process at times $t < t^+$. Time t can be be conceptualized as a nonnegative integer or a nonnegative real number, exactly as in agent-based models. Recall that our NetLogo program IONTW has two options for **model-time**. Depending on which one is chosen, the code will generate realizations of discrete-time or continuous-time stochastic processes.

For our purposes, the r.v.s $x_i(t)$ may either represent the states of individual hosts at time t or the counts of hosts in each compartment at time t. In the former case, the stochastic process formalizes a host-level model; in the latter case it formalizes a compartment-level model. Let us look at each of these two possibilities in turn.

3.1 Host-level stochastic process models

Here we will describe stochastic processes that are the exact mathematical counterparts of the agent-based models that are embodied in IONTW.

Consider a population of N hosts. Instead of representing host number i as an agent in a computer program, we model it by a r.v. x_i whose value at time t indicates the state of this host at time t. The set of possible values of each $x_i(t)$ depends on whether we are building an $SEIR$, SIR, SI, or SIS model; in each case it will be a subset of the set $\{S, E, I, R\}$. Thus if host i is susceptible at time t, we will have $x_i(t) = S$, if host i is infectious at time t, we will have $x_i(t) = I$, and so on.

While r.v.s formally are functions that take real numbers as values, it is more intuitive to use symbols here instead of arbitrarily coding states by reals. Moreover, we will somewhat informally refer to ordered N-tuples of the relevant symbols as "vectors." For example, the vector $(IISR)$ would represent the state in a model with $N = 4$ where hosts 1 and 2 are infectious, host 3 is susceptible, and host 4 is removed. The set of all vectors that are possible values of $\vec{x}(t)$ will be denoted by Ω and called the *state space* of the process.

Note that Ω is the same symbol that is normally used for a sample space.[4] In general, you can think of a stochastic process as repeatedly drawing elementary outcomes from a same sample space, one at every time t, where the successive draws are *not* independent, but strongly influenced by previous outcomes.

[4]See Section 1 of *A brief review of basic probability theory* at this website `https://qubeshub.org/iontw`

Exercise 2 *Determine the size* $|\Omega|$ *of the state space, that is, the number of possible states. Assume a population of size* $N = 10$ *and*

(a) an $SEIR$*-model,*

(b) an SIR*-model,*

(c) an SI*-model,*

(d) an SIS*-model.*

Now let us assume that we are formalizing a discrete-time SIR-model as a host-level stochastic process. To keep things transparent, consider $N = 3$ hosts and assume uniform mixing. Let a and b be the parameters of the model that correspond to **infection-prob** and **end-infection-prob** in IONTW. Recall that they represent the probabilities of removal and effective contact over a unit time interval respectively. Consider an initial state $\vec{x}(0) = (x_1(0), x_2(0), x_3(0)) = (ISS)$. What can we say about $\vec{x}(1)$?

We cannot say for sure what $\vec{x}(1)$ will be. The construction of our agent-based models implies that $\vec{x}(1)$ *could* be any of the states

$$(ISS), (IIS), (ISI), (III), (RSS), (RIS), (RSI), (RII). \tag{9}$$

Each of these possibilities will occur with a certain probability strictly between 0 and 1. This is exactly what "stochastic process" means: The future is not rigidly determined, but the observations that we have made so far (at time 0 in our case) influence the *probability distribution* at future times. Most notably, a lot of states in Ω, such as (SIR) and (IRS), are missing from the list (9) and will have probability 0 at time $t = 1$. Our knowledge of $\vec{x}(0)$ gives us *partial information* about $\vec{x}(1)$, but not certainty. This is, in a nutshell, the difference between stochastic process models and the *deterministic* ODE and difference equation models of Subsections 2.1 and 2.2.

While we don't have the certainty offered by deterministic models, it is possible to calculate the probability distribution of $\vec{x}(1)$ by calculating the probability of each of the states in the list (9).[5] Consider, for example, state (RIS). In order for the process to reach this state after one time step, host 1 must become removed, host 2 must become infectious (due to an effective contact with host 1, who is the only infectious host in the given initial state), and host 3 must *not* have had an effective contact with host 1 by time $t = 1$. Since our agent-based models implicitly assume independence of these three events, we get:

$$P(\vec{x}(1) = (RIS)) = P(x_1(1) = R)\, P(x_2(1) = I)\, P(x_3(1) = S) = ab(1 - b). \tag{10}$$

The probability distribution at time $t = 1$ can determined by performing similar calculations for each state on the list (9). This is rather tedious, and we will not pursue it here. Instead, let us consider an initial states for which the calculation of the entire probability distribution at time $t = 1$ becomes more manageable.

[5] We are stretching the definition of probability distribution that is given in *A brief review of basic probability theory* at this website `https://qubeshub.org/iontw` a bit. There it was only discussed in the context of one r.v.; here we are using its generalization to symbolic vectors of r.v.s. This generalization is completely straightforward and fairly intuitive, so we will leave it at the informal level.

Exercise 3 *Assume* $\vec{x}(0) = (SIR)$. *Find the probability distribution of* $\vec{x}(1)$.

Exercise 4 *Assume that* $\vec{x}(3) = (IIS)$. *Calculate* $P(\vec{x}(4) = (RII))$.

A little twist in Exercise 4 is caused by the two infectious hosts in $\vec{x}(3)$. But there is something much more remarkable about the exercise: We didn't tell you anything about the states $\vec{x}(0), \vec{x}(1), \vec{x}(2)$, and you did not need this information you solved the problem. Why not?

As we described in our book chapter [5], the code that simulates outbreaks in IONTW suffers from amnesia. It will determine the next state of the simulation exclusively based on the current state. Mathematicians call this kind of amnesia the *Markov property* of a stochastic process. For discrete-time processes that are called *stationary Markov chains,* the probability distribution of the next state $\vec{x}(t+1)$ depends *only* on the current state $\vec{x}(t)$, but neither on the states $\vec{x}(0), \ldots, \vec{x}(t-1)$ that the process went through before reaching state $\vec{x}(t)$ (this is the Markov property), nor on t itself (this is what "stationary" means).

The author of this module might have chosen a different version of Exercise 4, such as:
Assume that $\vec{x}(0) = (IIS)$. *Calculate* $P(\vec{x}(1) = (RII))$.
Or: *Assume that* $\vec{x}(t) = (IIS)$. *Calculate* $P(\vec{x}(t+1) = (RII))$.
Or: *Find* $P(\vec{x}(7) = (RII) \,|\, \vec{x}(6) = (IIS))$.
Or: *Find* $P(\vec{x}(t+1) = (RII) \,|\, \vec{x}(t) = (IIS))$.

The required calculations and and correct answer would have remained exactly the same.

The Markov property may not always be biologically realistic, but it greatly simplifies calculations and modeling. To see how these calculations work, consider the initial state $\vec{x}(0) = (x_1(0), x_2(0), x_3(0)) = (SIR)$ of Exercise 3. How can we calculate the probability $P(\vec{x}(2) = (IIR))$? According to the Markov property this probability depends only on $\vec{x}(1)$, but we don't know with certainty what the state $\vec{x}(1)$ will be. We know, however, that it must be one of the states on the list $\{(SIR), (SRR), (IIR), (IRR)\}$ of our sample solution. Neither of the states $(SRR), (IRR)$ can be followed by state (IIR) one time step later. Thus we know that $\vec{x}(1) \in \{(SIR), (IIR)\}$. Now we can use the formula for the total probability and the Markov property to calculate $P = P(\vec{x}(2) = (IIR) \,|\, \vec{x}(0) = (SIR))$ as follows:

$$\begin{aligned} P &= P(\vec{x}(1) = (SIR) \;\&\; \vec{x}(2) = (IIR)) + P(\vec{x}(1) = (IIR) \;\&\; \vec{x}(2) = (IIR)) \\ &= P(\vec{x}(1) = (SIR) \,|\, \vec{x}(0) = (SIR))\, P(\vec{x}(2) = (IIR) \,|\, \vec{x}(1) = (SIR))) \\ &\quad + P(\vec{x}(1) = (IIR) \,|\, \vec{x}(0) = (SIR))\, P(\vec{x}(2) = (IIR) \,|\, \vec{x}(1) = (IIR))) \\ &= (1-b)(1-a)b(1-a) + b(1-a)(1-b)^2. \end{aligned} \tag{11}$$

The principle behind these calculations stays the same when we are looking for $P(\vec{x}(t) = \vec{y} \,|\, \vec{x}(0) = \vec{z})$ for any given $\vec{y}, \vec{z} \in \Omega$ and $t > 0$: The Markov property and the formula for the total probability allows, at least in principle, to successively compute the probabilities $P(\vec{x}(t^-) = \vec{y} \,|\, \vec{x}(0) = \vec{z})$ for $t^- = 1, 2, \ldots, t-1, t$ and all $\vec{y} \in \Omega$. Performing such calculations by hand could be extremely tedious, but for moderately large examples a computer will happily perform them, using powers of the so-called *matrix of transition probabilities.* You

can see that the initial state $\vec{x}(0)$ *uniquely determines the probability distribution* of $\vec{x}(t)$ for all $t \geq 0$, albeit not usually the actual value $\vec{x}(t)$ with certainty.

In the example that we have considered so far we have assumed uniform mixing. This assumption simplified the description of the model, but the translation of our agent-based models into host-level stochastic processes remains essentially the same if we assume that effective contacts can only occur between two hosts i, j that are connected by an edge $\{i, j\}$ of a given contact network G. Only the calculations of the actual probabilities will be different.

Let us illustrate these similarities and differences with the following example of a discrete-time SIR-model. We assume that the population consists of N hosts, where N will be left unspecified. Furthermore, we assume that $a = 1$, so that we have a next-generation model, and $0 < b < 1$. The edges of the contact network G will be all pairs of the form $\{i, i+1\}$ for $i = 1, 2, \ldots, N-1$. When N is a prime number like 17, you can easily set up such models in IONTW by choosing, for example,

model-time $\rightarrow$ **Discrete**
infection-prob: 0.7
end-infection-prob: 1
gain-immunity: On
latent-period: Off
network-type $\rightarrow$ **Nearest-neighbor 2**
num-nodes: 17
d: 1

In the visualization of G the hosts appear as evenly spaced dots on a line segment. When you run simulations with initial states that have exactly one infectious host in an otherwise susceptible population, at the end you will always see a set of consecutive host that have experienced infection.

Now consider an initial state $\vec{x}(0) = (\cdots SSISS \cdots)$ where host i is infectious and all other hosts are susceptible. We want to find the probability distributions of $\vec{x}(1)$ and $\vec{x}(2)$ for this given initial state. The construction of our agent-based models implies that $\vec{x}(1)$ could be any of the states

$$(\cdots SSRSS \cdots), (\cdots SIRSS \cdots), (\cdots SSRIS \cdots), (\cdots SIRIS \cdots). \tag{12}$$

The probabilities of reaching one of these states after one time step are

$$\begin{aligned} P(\vec{x}(1) = (\cdots SSRSS \cdots)) &= (1-b)^2, & P(\vec{x}(2) = (\cdots SIRIS \cdots)) &= b^2, \\ P(\vec{x}(1) = (\cdots SIRSS \cdots)) &= b(1-b), & P(\vec{x}(2) = (\cdots SSRIS \cdots)) &= (1-b)b. \end{aligned} \tag{13}$$

For all other states $\vec{y} \in \Omega$ we have $P(\vec{x}(1) = \vec{y}) = 0$.

Exercise 5 *(a) Find the probability distribution of* $\vec{x}(1)$ *in this model given that the initial state is* $\vec{x}(0) = (ISS \cdots)$.

(b) Find the basic reproductive ratio R_0 *for this model.*
How does this value behave if $N \rightarrow \infty$*?*

Now let us revisit the initial state $\vec{x}(0) = (\cdots SSISS \cdots)$ where host i with $3 \leq i \leq N-2$ is infectious and all other hosts are susceptible. We want to find the probability distribution of $\vec{x}(2)$ for this given initial state. The construction of our agent-based models implies that $\vec{x}(2)$ could be any of the states

$$\begin{array}{l}(\cdots SSRSS \cdots), (\cdots SRRSS \cdots), (\cdots SSRRS \cdots), (\cdots SRRRS \cdots), \\ (\cdots IRRSS \cdots), (\cdots SSRRI \cdots), (\cdots IRRRS \cdots), (\cdots SRRRI \cdots), (\cdots IRRRI \cdots).\end{array} \tag{14}$$

There are as many as 9 possibilities, but due to the particularly simple structure of the model we have the rare luxury of being able to determine with certainty state $\vec{x}(1)$ based on the observation of state $\vec{x}(2)$ and the assumption about $\vec{x}(0)$. For $\vec{x}(2)$ listed in the order of (14), the preceding states $\vec{x}(1)$ must have been:

$$\begin{array}{l}(\cdots SSRSS \cdots), (\cdots SIRSS \cdots), (\cdots SSRIS \cdots), (\cdots SIRIS \cdots), \\ (\cdots SIRSS \cdots), (\cdots SSRIS \cdots), (\cdots SIRIS \cdots), (\cdots SIRIS \cdots), (\cdots SIRIS \cdots).\end{array} \tag{15}$$

Thus the probabilities of reaching the states in (14) after two steps are

$$\begin{array}{ll}P(\vec{x}(2) = (\cdots SSRSS \cdots)) = (1-b)^2, & P(\vec{x}(2) = (\cdots SRRSS \cdots)) = b(b-1)^2, \\ P(\vec{x}(2) = (\cdots SSRRS \cdots)) = b(1-b)^2, & P(\vec{x}(2) = (\cdots SRRRS \cdots)) = b(1-b)^2, \\ P(\vec{x}(2) = (\cdots IRRRS \cdots)) = b^3(1-b), & P(\vec{x}(2) = (\cdots SRRRI \cdots)) = b^3(1-b), \\ P(\vec{x}(2) = (\cdots IRRRI \cdots)) = b^4. & \end{array}$$

For all other states $\vec{y} \in \Omega$ we have $P(\vec{x}(2) = \vec{y}) = 0$.

Exercise 6 *(a) Find the probability distribution of $\vec{x}(2)$ in this model given that the initial state is $\vec{x}(0) = (ISS \cdots)$.*

(b) Generalize the result of part (a) by finding the distribution of $\vec{x}(t)$ for arbitrary $0 < t < N$ for the given initial state.

(c) Assume $\vec{x}(0) = (\cdots SSISS \cdots)$ is an initial state where host i is infectious and all other hosts are susceptible. Find the probability distribution of $\vec{x}(t)$ for $0 < t < \min\{i, N-i+1\}$.

(d) Generalize the results of points (b) and (c) to arbitrary times $t \geq 0$.

Hint: *For parts (b)–(d), first invent a more convenient notation for all states that could possibly be reached from the given initial state.*

So far, we have assumed that time is discrete. Continuous-time host-level stochastic processes are constructed similarly, with the difference that the relevant events (removal of a given host i or an effective contact between hosts i and j for an SIR model) are assumed to occur at random times. In the processes that formalize the continuous-time agent-based models that are embodied in IONTW, the time until removal of a host i after onset of infectiousness is an exponentially distributed r.v. with parameter α, which corresponds to the parameter **end-infection-rate** of IONTW. The time until the next effective contact between hosts i and j that are adjacent in the contact network is an exponentially distributed

r.v. with parameter β, which corresponds to the parameter **infection-rate** of IONTW. This description, together with the assumption that the r.v.s for the times until removal and next effective contact for different (pairs of) hosts are independent, specifies a continuous-time host-level stochastic process for any given choice of a contact network and of the parameter values α and β. Casting this description in formal mathematical notation allows for sophisticated derivations of theorems, but we will not need it here.

3.2 Compartment-level stochastic process models

Host-level stochastic processes have a number of useful properties: They directly formalize the kind of agent-based models that we have been exploring with IONTW. In contrast to the latter, they are mathematical constructs and allow, at least in principle, to derive rigorous answers to all parts of Question 1. They are stationary Markov processes, which is usually very helpful in such derivations. However, as you have seen in Exercise 2, these processes tend to have huge state spaces Ω. Unfortunately, this feature may cause great difficulties in mathematical arguments.

The deterministic models of Section 2 do not suffer from the latter defect: For SIR-models of this type, each state can be represented as a three-dimensional vector. But unfortunately, these models don't account for stochastic effects. Can we have the best of both worlds, stochastic process models with state spaces of manageable complexity and simple mathematical structure? Yes we can. Up to a point.

Consider a host-level stochastic process model of type SIR, and let $S(t), I(t), R(t)$ denote the numbers of hosts in the **S**-, **I**-, and **R**-compartments, respectively. The notation is the same as in Subsection 2.1, but there these numbers where uniquely determined by the initial state, whereas here they are r.v.s. In other words, for a given initial state $(S(0), I(0), R(0))$ we will in general no longer know the vector $(S(t), I(t), R(t))$ for $t > 0$ with certainty; the model only implies a probability distribution of the latter vectors. This probability distribution will usually depend on the initial state though. More generally, the distribution of $(S(t), I(t), R(t))$ will depend, *in interesting ways,* on the states $(S(t^-), I(t^-), R(t^-))$ at times $t^- < t$. Thus we have defined another stochastic process. It gives compartment-level summaries of the behavior of the underlying host-level process. We will call it a *compartment-level stochastic process.* For discrete-time compartment-level processes we will sometimes use the notation (S_t, I_t, R_t) for the states; similarly to the one that we used in Subsection 2.2.

Do compartment-level stochastic process models give us the best of both worlds? Obviously, they account for the randomness inherent in disease transmission. For an SIR-model with population of size N the state space Ω consists of all three-dimensional vectors (s, i, r) of nonnegative integers with $s + i + r = N$. For moderately large N this will not be an astronomically large number of states.

Exercise 7 *Find the total number $|\Omega|$ of states for an SIR-model as above if $N = 10$.*

OK, the probability distribution of $(S(t), I(t), R(t))$ will depend *in interesting ways* on

the states $(S(t^-), I(t^-), R(t^-))$ at earlier times $t^- < t$. But does it depend on them also in *mathematically nice ways?*

Not necessarily. Consider the example of a the host-level process for the network-based model that was discussed in the previous subsection. If the corresponding compartment-level process were to have the Markov property, then, in particular, $P((S_{t+1}, I_{t+1}, R_{t+1}) = (1, 2, N-2))$ should depend *only* on the state (S_t, I_t, R_t) and nothing else. But if you recall our analysis of this model, then you will notice the following anomaly:

$$\begin{aligned} &P((S_1, I_1, R_1) = (1, 2, N-2) \,|\, (S_0, I_0, R_0) = (1, 0, N-1)) > 0, \\ &P((S_2, I_2, R_2) = (1, 2, N-2) \,|\, (S_1, I_1, R_1) = (1, 0, N-1)) = 0. \end{aligned} \tag{16}$$

This follows from the fact that while for an initial state of this model the index case could be flanked by two susceptible hosts who could both become infectious at time $t = 1$, at later times an infectious host must have one removed neighbor and could infect at most one other host. The compartment-level process fails to be a stationary Markov chain! The probability of the next state may depend not only on the current state, but also on prior history; in this case, on whether or not there was any prior history.

However, as the next proposition will show, if we assume homogeneity of hosts and uniform mixing, then the compartment-level processes that summarize underlying host-level processes will be stationary and have the Markov property. In this case they do give us the best of both worlds: Mathematical models that incorporate stochasticity and are relatively easy to study. They might enable us to quantify the expected amount of variability between outbreaks and the distortions that occur for small N due to finite-size effects.

Proposition 2 *Consider a host-level stochastic process model of disease transmission that is constructed under the assumptions of homogeneity of hosts and uniform mixing and is a stationary Markov chain. Then the corresponding compartment-level process whose states represent the counts of hosts in the compartments of the host-level process is also a stationary Markov process.*

This is a curious assertion. When you think about it for a moment, you will notice that the information that we have provided so far in this module is not sufficient to rigorously delineate the assumptions of Proposition 2. To be sure, the assumptions of homogeneity of hosts and uniform mixing were extensively discussed in [4] and also reviewed in the document *Network-based models of transmission of infectious diseases: a brief overview* at this website.[6] But host-level stochastic process model of disease transmission were only *illustrated* in the previous subsection, not rigorously defined in full generality. The ones we did illustrated were models of type SIR that are stationary Markov processes. But the wording of the assumption suggests that the class of such models is much broader and might include nonstationary processes and processes without the Markov property. It certainly includes models of types SI, SIS, $SEIR$, and more. If this were a mathematical research

[6] `https://qubeshub.org/iontw`

paper, we first would need to formally define the entire class before stating and proving the proposition.

At this semiformal introductory level though, the full mathematical rigor would only oscure the main ideas. The beauty of Proposition 2 is that in order to prove it, you only need to use the following properties of the two stochastic processes that it refers to:

- The model has finitely many compartments.
- The host-level process has r.v.s $x_i(t)$ that indicate the compartment that host i resides in at time t.
- The compartment-level process has the same timeline as the host-level process. Its state at each time t counts the number of hosts in each compartment at time t.
- A stochastic process is a *stationary Markov process* if, and only if, for any given state $\vec{x}(t)$ at time t, the distribution of $\vec{x}(t^+)$ at any future time $t^+ > t$ depends only on $\vec{x}(t)$ and the time difference $\Delta t = t^+ - t$, and not on t itself or any values $\vec{x}(t^-)$ for $t^- < t$.

Exercise 8 *Outline a proof of Proposition 2.*

Note that compartment-level processes make prediction about the same features of the spread of an infection as the deterministic models of Section 2. Stochastic processes are more accurate as they will predict also the amount of inherent variability. For small population sizes, the predictions of of deterministic models may be far off the mark. However, for large populations and cases where the assumptions of homogeneity of hosts and uniform mixing are justified, deterministic models can give reasonably good approximations of the average behavior of a corresponding compartment-level stochastic process model. This will work for initial states that represent fairly substantial numbers of hosts in each compartment. And, of course, the parameters in both types of models need to be chosen in such a way that the two models do in fact approximate, each in its own way, the same underlying host-level stochastic process.

This brings us back to the question of how A and B of the difference equation version (7) of the SIR-model are related to the probabilities a and b of removal and effective contact after one time step. If the model based on difference equations is supposed to approximate average behavior of the compartment-level stochastic process, then the next state $(S_{t+1}, I_{t+1}, R_{t+1})$ must represent the mean values of hosts in the three compartments at the next step given that the current state is (S_t, I_t, R_t). From the second and third lines of (7) it follows that the parameter A must represent the proportion of currently infectious hosts that get removed after one time step. Recall that a is the *probability* of removal after one time step for *each individual* host. This will also be the expected *proportion* of infectious hosts that enter the **R**-compartment at time $t + 1$. Thus we can conclude that $A = a$.

The relation between b and B is more subtle. Recall that b represents the probability of an effective contact between any two *given* hosts $i \neq j$ by the next time step. If host i

is infectious, then the exact same argument as in the previous paragraph shows that the expected number of susceptible hosts j that have an effective contact with host i by the next time step is $b\,S_t$. Similarly, since there are I_t infectious hosts total, the expected total number of pairs (j, i) of susceptible hosts j and infectious hosts i that will have an effective contact by time t must be $b\,S_t\,I_t$. Before reading on, take a couple of minutes to think about the following problem:

Exercise 9 *Can we conclude that $b\,S_t\,I_t$ also represents the expected number of new infections at the next time step, so that it will be equal to $B(S_t\,I_t)$? Why or why not?*

Well, by definition, if host j is susceptible at time t and has an effective contact with an infectious host, then j will be infectious at the next time step. If not, then j will remain susceptible at time $t + 1$. But note that "with *an* infectious host" means "with *at least one* infectious host." When $I_t > 1$, some susceptible hosts may have effective contacts with several infectious hosts, and only the first such contact of these will lead to moving one host into the **I**-compartment. Thus, in general, the product $b\,S_t\,I_t$ will be larger than the expected number of new infections.

Let b^+ denote the probability that a given susceptible host j has an effective contact with *at least one* infectious host. By the assumptions of homogeneity of hosts and uniform mixing, b^+ will be the same for each susceptible host j, and the same argument as previously shows that the expected number of new infections at time step $t + 1$ will be $b^+\,S_t$.

How can we estimate b^+? It will obviously depend on I_t, as host j could become infectious due to an effective contact with any one of the hosts that are infectious at time t. But the dependence will not be linear. It will be easier to determine the probability $1 - b^+$ that host j has *no* effective contact with *any* of the hosts i that are infectious at time t. For a given host i the probability that no effective contact between j and i occurs by the next time step is $1 - b$, and for a fixed j and different hosts i these events are assumed to be independent in the models that we discuss here. Project 8.1 of the online appendix of [5] gives a detailed illustration of this assumption. The assumption of independence allows us to multiply probabilities, so that

$$b^+ = 1 - (1 - b)^{I_t} \tag{17}$$

For large population sizes, that is, for the kind of populations where a deterministic model might give us good approximations, b is usually very small so that $1 - b \approx e^{-b}$ and $b+$ can be approximated by:

$$b^+ = 1 - (1 - b)^{I_t} \approx 1 - e^{-b\,I_t}. \tag{18}$$

It follows that either of the following would be an appropriate choice for the function B in (7):

$$\begin{aligned} &B(S_t, I_t) = S_t \left(1 - (1 - b)^{I_t}\right), \quad \text{or} \\ &B(S_t, I_t) = S_t \left(1 - e^{-b\,I_t}\right). \end{aligned} \tag{19}$$

The second option is usually more convenient in calculations. When $b\,I_t$ is very small, then $1 - e^{-b\,I_t} \approx b\,I_t$ and the right-hand-side of (19) will be very close to $b\,S_t\,I_t$. This makes sense, as in this case the probability of effective contact with multiple infectious hosts becomes negligible. In general,

$$B(S_t, I_t) < b\,S_t\,I_t. \tag{20}$$

When I_t comprises a significant fraction of the population, the difference between the two sides of (20) may be substantial.

4 Theorems

Mathematical models of disease transmission may allow us to derive answers to all parts of Question 1 in the form of *mathematical theorems.* Let us revisit here three theorems that were stated in Section 2 of [4]. Equipped with the conceptual framework of the preceding sections, we will be able to analyze their meaning in more depth than was possible in [4]. We will even be able to include some proofs.

Theorem 1 *Assume homogeneity of hosts and uniform mixing in an SIR- or SIS-model. Consider two times t, t^+ with $0 \le t < t^+$. If $R_0 < 1$ and $0 < I(t)$, then $I(t^+) < I(t)$.*

There is something important missing in the statement of Theorem 1: What kind of model are we talking about? In theorems that you will find in research papers or textbooks this information is usually implied by the context and rarely stated explicitly. But in this module the context does not help. "A model" could be based on ODEs, difference equations, could be a compartment-level stochastic process, or built in yet another mathematical framework.

For the ODE-based model (1) the proof of Theorem 1 is easy: If we factor out I in the second line of (1), then we get

$$\frac{dI}{dt} = I\,(\beta S - \alpha) \le I\,(\beta N - \alpha) = \alpha\,I\,(R_0 - 1). \tag{21}$$

The last equality in (21) follows from the expression (2) for R_0. Under the assumptions of Theorem 1 the inequality is strict, and we conclude that the function $I(t)$ is strictly decreasing in this case. This is exactly what the theorem says.

Exercise 10 *Prove Theorem 1 for the difference equation model* (7) *under the additional assumptions that $a = 1$ and N is sufficiently large so that we can ignore the difference between N and $N-1$.* ***Hint:*** *Use the version of the model where the function B is given by the first line of* (19)*. Start by deriving an expression for R_0 and then use the results that we presented at the end of the previous section. Keep in mind that the notation $I(t)$ in the statement of the theorem needs to be changed to I_t.*

For stochastic process models Theorem 1 appears to be wrong: Due to random fluctuations, the actual number of infectious hosts may increase, even when R_0 is very small. This point was already illustrated in [4]. But we need to keep in mind that the theorem was quoted here *verbatim* from our book chapter [4]. In the book chapter, the variable $I(t)$ denotes the *expected number* of infectious hosts, while in this module the same notation is used for actual numbers in compartment-level stochastic processes. When $I(t)$ is interpreted as a mean value, the theorem remains true for stochastic processes.

Unfortunately, the supply of letter symbols is very limited compared with the wealth of different mathematical objects, and mathematicians will use the same symbol in different meanings. Moreover, we mathematicians sometimes use different notations for essentially the same notion; see the hint for Exercise 10. Whenever you want to rely on a result that is quoted from a different source of mathematical writing, you need to carefully check what the notation in that source actually stands for.

Now let us take a look at the second theorem of [4].

Theorem 2 *Assume homogeneity of hosts and uniform mixing in an SEIR-, SIR- or SIS-model. For any given probability $p < 1$ there exists a constant $B(p)$ such that whenever $R_0 \leq 1$, then with probability at least p the number of hosts who will experience infection at some time during the outbreak will not exceed $B(p)$, regardless of the population size N. Thus if the population size is large, then with probability very close to 1, introduction of a single index case into an otherwise susceptible population will result only in a minor outbreak.*

This theorem does not make sense in deterministic models. Period. Its very wording implies that the models for which it is true are stochastic processes. Notice also that in order to get any meaningful upper bound on the number of hosts that will experience infection, one must assume an upper bound on the number of initially infectious hosts. The theorem implicitly assumes that the use of the word "outbreak" in the first part is restricted to an outbreak that is caused by a single index case. This information is implicit in the context of [4], but not quite clearly stated in the theorem itself. If you want to use a theorem that is quoted from another source, you always need to be on the alert for such additional information that is given by the context. In an ideal world the wording of each theorem would be entirely self-contained, but this cannot always be achieved in reasonably succinct mathematical writing.

A complete formal proof of Theorem 2 in the framework of stochastic processes is beyond the scope of this module. But at least when $R_0 < 1$, the gist of the proof is fairly easy to understand if we think about *generations of infection* (see online appendix of [5]): The index case constitutes generation 0, and generation $n + 1$ will consist of all hosts that got infected by a host in generation n. Under the assumptions of homogeneity of hosts and uniform mixing, the mean number of hosts that any given host will infect prior to recovery or removal is always $\leq R_0$. If there is exactly one index case, it follows that the mean number of hosts in generation n is at most R_0^n. Let $R(\infty)$ be the total number of hosts that experience infection during the outbreak. Since every host that experiences infection must

belong to *some* generation, the mean of $R(\infty)$ satisfies

$$\mu(R(\infty)) = \sum_{k=1}^{\infty} kP(R(\infty) = k) \leq \sum_{n=0}^{\infty} R_0^n. \tag{22}$$

When $R_0 < 1$, the right-hand side of (22) is finite. Now fix $p < 1$, and let B be such that $\frac{\mu(R(\infty))}{B+1} < 1 - p$. Then

$$P(R(\infty) > B) = \sum_{k=B+1}^{\infty} P(R(\infty) = k) \leq \sum_{k=1}^{\infty} \frac{k}{B+1} P(R(\infty) = k) = \frac{\mu(R(\infty))}{B+1} < 1 - p,$$

and Theorem 2 follows for the case $R_0 < 1$.

Theorem 3 *Assume homogeneity of hosts and uniform mixing in an SIR- or $SEIR$-model. If $R_0 > 1$, then there are numbers $r(\infty)$ and z_∞ that satisfy the inequalities $0 < r(\infty), z_\infty < 1$ such that as long as the population size is large, with probability very close to $1 - z_\infty$, introduction of a single index case into an otherwise susceptible population will result in a major outbreak with final size close to $r(\infty)$.*

The number $r(\infty)$ will be larger for larger values of R_0 and the number z_∞ will be smaller for larger values of R_0.

The number z_∞ represents the probability of a minor outbreak for very large population sizes. This probability is supposed to depend only on R_0, but not (much) on the actual population size N. Thus for very large population sizes N, major outbreaks should occur with approximately the same probability $1-z_\infty$. The number $r(\infty)$ represents the predicted approximate final size of major outbreaks in large populations.

Theorem 3 is meaningful only for stochastic process models. It should be understood that once we fix the type of the stochastic process, in particular, whether it is discrete or continuous, the numbers z_∞ and $r(\infty)$ will be determined by R_0 alone. The theorem leaves open the possibility that the actual values of $r(\infty)$ and z_∞ may depend on the type of model as well as on R_0. We will show below how one can prove the theorem and derive formulas for $r(\infty)$ and z_∞ for one particularly simple type of stochastic process models.

Before we do so, let us show that the assumption of uniform mixing is very important in this theorem. Assume that we have a class of stochastic process models for which the expected number of hosts who will eventually experience infection in an outbreak that is caused by a single index case has a finite upper bound that is valid for all population sizes. Then the inequalities $1 - z_\infty > 0$ and $r(\infty) > 0$ cannot hold simultaneously. Now consider the network-based models that were discussed at the end of Subsection 3.1. By the solution to Exercise 5, for large population sizes we will have $R_0 \approx 2b$ in these models, and for $b > 0.5$ we get $R_0 > 1$. Thus the next exercise implies that the conclusion of Theorem 3 fails for these models. This, of course, does not invalidate the theorem, as the uniform mixing assumption is not satisfied in these network-based models.

Exercise 11 *Let $b < 1$ be fixed. Show that in the class of models described above the expected value of the number of hosts that will experience infection in an outbreak that is caused by introduction of a single index case into an otherwise susceptible population is bounded from above by a fixed constant that does not depend on N.* ***Hint:*** *Use the results from your solution of Exercise 6.*

The standard formal proof of Theorem 3 relies on techniques that are beyond the scope of our exposition. However, an argument that we already presented in our module *Exploring Erdős-Rényi random graphs with IONTW* gives a proof for stochastic processes that formalize next-generation models, that is, discrete-time agent-based models with $a = 1$. For convenience, let us repeat the gist of this argument here.

Consider a next-generation compartment-based SIR-model with probability b of effective contact until the next time step. This can be interpreted as a network-based model on a complete graph K_N. Consider a pair (i, j) of distinct hosts. Suppose host i becomes infectious at some time step t and host j will still be susceptible at time t. The probability that these hosts will have an effective contact during the interval of infectiousness of host i, that is, until step $t + 1$ in terms of the model, must be equal to b. The same will be true if host j becomes infectious at time step t and host i is still infectious at time t.

Now construct a random subgraph G of K_N as follows: Consider a simulation of a model as described in the previous paragraph that starts with one index case j^* in an otherwise susceptible population. One can think of the simulation as being performed by tossing biased coins sequentially as the simulation progresses, for all edges that have one endpoint that corresponds to a host who is infectious at the current time step, and whose other endpoint represents a host who is still susceptible at the current time step. The coin will need to be biased in such a way that it comes up heads with probability b. In terms of the simulation, the hosts that will be infectious at the next time step are exactly the ones represented by endpoints of those edges for which a coin is tossed and comes up heads. Include these edges in $E(G)$, and don't include those edges for which the coin comes up tails.

The "simulations" that are mentioned in the above description are of course realizations of the host-level stochastic process that formalizes the agent-based model. In the construction given above, the decision about inclusion of a given potential edge $\{i, j\}$ is based on exactly one coin toss, the probability b of the coin coming up heads is fixed, and, most importantly, the coin tosses are all independent. Thus the construction gives an instance $G = G_{ER}(N, \lambda)$ of an Erdős-Rényi random graph, where $\lambda = b(N - 1) = R_0$.

Moreover, the set of hosts that will experience infection in this realization of the stochastic process will be the connected component of the index case in G. For large N, with probability $z_\infty \approx 1 - \varrho(\lambda)$ it will be a small component of negligible relative size, which corresponds to a minor outbreak. And with probability $1 - z_\infty \approx \varrho(\lambda)$ it will be the giant component, with relative size $r(\infty) \approx \varrho(\lambda)$. For $\lambda > 1$ we get $\varrho(\lambda)$ as the unique solution ϱ of the equation

$$1 - \varrho = e^{-\lambda\varrho} \tag{23}$$

in the interval $(0, 1)$.

Exercise 12 *Show that the predictions that are summarized in the preceding paragraph give the same estimate of the final size as the ODE-based SIR model* (1) *predicts.*

As we already remarked in the module on Erdős-Rényi random graphs, the same argument does not work for other types of models though. We can extract a construction of a random graph from a given realization, but we lose independence in our decisions about whether or not to include edges with a common endpoint i, because the probability of an effective contact along these edges will depend on the time that node i stays infectious, which is no longer fixed. As was illustrated with simulations in the online part of book chapter [4], this will affect z_∞. But, curiously enough, it will not alter $r(\infty)$. For continuous-time stochastic processes this follows from Exercise 12, as for large population sizes the ODE model (1) will give fairly good approximations of the expected final size under the condition that a major outbreak did occur.

5 Review and further readings

This module has given you a bird's-eye-view of the toolbox that is available in mathematical epidemiology. It may have left you wondering: *Which type of model would be the right one?*

This question does not have one right answer. All models are just that: Models. Simplified versions of the real world where infections spread. They differ in what they take into account and what they ignore. And they differ in how difficult it is to derive actual predictions from them.

Basically, the choice of model depends on what kind of questions you want to answer and how detailed information about the spread of the infection in actual population is available. Consider, for example, Theorem 3. It addresses all parts of Question 1: The theorem (or rather its more specific versions like the one that we derived at the end of the previous section) tells you what kind of outbreaks to expect and how their final sizes depend on the parameter R_0. It quantifies the variability between the two possible scenarios in terms of the probability z_∞, and it tells you what happens when the population size increases without bound. Theorems of this kind can only be obtained in the framework of stochastic processes. Unfortunately, studying stochastic processes is usually more difficult than studying deterministic models like the ones of Section 2. Similarly to the constructions of the latter models, Theorem 3 relies on the assumptions of homogeneity of hosts and uniform mixing. Depending on the real populations of interest, these may or may not be appropriate. Some analogues of Theorem 3 for certain types of network-based models can be proved, but are technically much more challenging.

A simpler model may give you very different answers to your questions than a more complex one that incorporates more realistic assumptions. In this is the case, the simpler model cannot be trusted. If there is reasonable agreement between the predictions of the simpler model and the more complex one, using the former is usually preferable.

As you saw in several places of this module, it is often enlightening to look at properties of one type of model from the vantage point of a corresponding model that is constructed

within a different mathematical framework. Successful research in mathematical epidemiology requires the ability to look at a biological problem through multiple theoretical lenses, that is, by studying and comparing multiple types of models. Achieving this versatility requires in-depth study of modeling within each framework. There are a number of good introductory texts on epidemic modeling. ODE-based modeling is covered in detail in the survey article [3] and the textbook [6]. For an easy-to-follow and applications-oriented introduction of several modeling frameworks, including models based on difference equations, a good choice would be [7]. The textbook [1] gives the most comprehensive treatment, at a more advanced mathematical level, and extensively covers stochastic process models. It also contains many excellently structured exercises with hints and worked-out solutions. Network-based models are covered to a limited extent in the textbooks that we listed above. A more extensive and mathematically more advanced treatment of them can be found in [2].

References

[1] Odo Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical tools for understanding infectious disease dynamics.* Princeton University Press, 2012.

[2] Moez Draief and Laurent Massoulié. *Epidemics and Rumors in Complex Networks*, volume 369 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, 2010.

[3] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

[4] Winfried Just, Hannah Callender, and M Drew LaMar. Disease transmission dynamics on networks: Network structure *vs.* disease dynamics. In Raina Robeva, editor, *Algebraic and Discrete Mathematical Methods for Modern Biology.* Academic Press, 2015.

[5] Winfried Just, Hannah Callender, M Drew LaMar, and Natalia Toporikova. Transmission of infectious diseases: Data, models, and simulations. In Raina Robeva, editor, *Algebraic and Discrete Mathematical Methods for Modern Biology.* Academic Press, 2015.

[6] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals.* Princeton University Press, 2008.

[7] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling.* Oxford University Press, 2010.