Epidemiological Models and Process Engineering

Bernt Lie

University of South-Eastern Norway, Porsgrunn, Norway, Bernt.Lie@usn.no

Abstract

The paper discusses the principles behind epidemiology models, with examples taken from the classic SIR (suspectible-infected-recovered) and SEIR (S-exposed-IR) models. Both continuous time deterministic and stochastic models are treated, where the stochastic models are based on Poisson-distributed events/reactions. These models use real approximations to the integers representing the number of people in each of the classes S, (E,) I, R. An alternative stochastic representation is the first reaction time description, where the variables are kept as integers, and where one instead computes the time between each event. The models are presented in a form compatible with standard chemical engineering models.

Based on the model description, the SIR and SEIR models are fitted to a measles case study using the Markov Chain Monte Carlo approach. For the given data, the SIR model appears to give much smaller uncertainty in predicitons. The continuous time stochastic description and the firt reaction time approaches give similar variation in the models.

An important measure of the state of epidemics is the *reproduction number*, R, which tells whether the infection is growing or decreasing from an initial infection. The development of an expression for R is indicated both from eigenvalues and from the Next-Generation Approach, and it is shown that the expression for R is identical for the SIR and the SEIR model.

The principles of epidemiology model development discussed in the paper are used in models ranging from HIV/AIDS to COVID-19.

Keywords: epidemiology, reaction engineering, deterministic models, stochastic models, model fitting, measles case study

1 Introduction

1.1 Background

The COVID-19[1] pandemic spreading in 2020 initially caused fear, irrational hoarding of consumer goods, uncertainty about future food supply, and economic depression. References to historic disasters such as the Black Death and the "Spanish Flu" were drawn. This period also spawned a renewed interest in epidemiology to understand how infections spread, and a massive effort in development of virus medicine and vaccines. Policy making and society saw challenges hardly faced before on how to adapt to the development in real time. The effort to develop vaccines has probably been the largest directed effort since space exploration, and similar to space technology, the medical developments will have wide ranging consequences. The sociological experiment faces serious questions regarding the limits of governing and personal freedom.

Although developed within life sciences, epidemiological models have much in common with chemical reaction engineering. A key difference is that while chemistry operates with particle numbers in the order of 10^{23}, epidemiological models operate with number of people in the range of 10^2– 10^9. Deterministic epidemics models are common, but the "law of large numbers" as used in chemical engineering does not really apply, and stochastic formulations are of interest.

Policy making and epidemics mitigation has much in common with feedback control under uncertainty. It is therefore of interest to relate epidemiological models to formulations and notation from chemical reaction engineering and process control.

1.2 Previous work

Pandemics such as COVID-19 are not new. The (bacterial) bubonic plague[2] that came to Europe ca. 1346-47 became known as the *Black Death* (1346–1353), and killed an estimated 60% of the population in Europe (50-80 million), and some 75–200 million worldwide. Poland closed its borders, and was more or less spared from the plague; Iceland was partially spared since sailors died before the ships arrived to Iceland. An artistic work from this period expressed an almost modern belief in science:

> "Elegant ladies, as I believe you know, the wisdom we mortals possess does not merely consist of remembering things past and apprehending the present, but on the basis of these two activities being able to predict the future, which is considered by serious men to be the highest form of human intelligence."
>
> Boccaccio, Giovanni (1349-1351/-52). *The Decameron*, p. 851. Translated by Wayne A. Rebhorn. W. W. Norton & Company. Kindle Edition.

[1] COVID-19 is the COrona VIrus Disease originating in 2019. The World Health Organization and Wikipedia.com both appear to write COVID-19 in all caps.

[2] https://en.wikipedia.org/wiki/Bubonic_plague

The "Spanish Flue"[3] or the *1918 influenza pandemic*, infected up to 500 million, with 17–100 million deaths. Classical epidemiology models were developed in the decade following the "Spanish Flu".

Historically, the population of N individuals was partitioned into a deterministic 3 compartmental model consisting of those who are susceptible to an infection (S), those who have been infected (I), and those who have recovered (R), (Kermack and McKendrick, 1927). Considering infections as *reactions*, and reaction events following a Poisson distribution, a determistic model can be converted to a stochastic model and solved using stochastic differential equation solvers. Alternatively, the model can be solved using Gillespie's algorithm, which is based on work in the statistics community ca. 1930–1950[4] and also assumes that reactions follow a Poisson distribution (Gillespie, 1976, 1977).

Approximately at the same time as the deterministic Kermack-McKendrick model, the stochastic *Reed-Frost* model was formulated which describes the same SIR types; this model was published much later, (Schwabe et al., 1977). This model is based on SIR compartmentalization, is discrete in time, and assumes a stochastic infection transmission with a fixed probability. To some degree, this model is related to *agent-based* models, where one could envision tracking of individuals within a compartment.

A good, general introduction to epidemiology models is (Keeling and Rohani, 2008). See also (Martcheva, 2015), who gives a more mathematically oriented overview with elements of stability results using Lyapunov theory. In (Brauer et al., 2019), various complex epidemiology models are discussed, with a number of case studies. An important concept related to infection spreading is the so-called *reproduction number*, R, which is related to initial stability of the infection model. An alternative to the above "balance" type models, is models whare "particles" move around in space, and are infected based on some stochastic model due to proximity to other infected "particles", e.g., (Britton and Pardoux, 2019).

Classical epidemiological models use a *spatial description* (Euler description) where the focus is on a number of compartments of fixed area/people with fixed attributes (age, immune level, etc.) where people move in and out of the various compartments/attributes. An alternative could be a *material description* (Lagrangian description) where one instead track the status of each individual or a group of individuals of similar attributes, and how these move about in the world, get into proximity with other people, and get infected with a certain probability. Such *material descriptions* are sometimes referred to as *agent-based* models.[5] Agent-based models are not discussed in this paper.

1.3 Scope

The focus in this work is to relate basic epidemiology models to process engineering concepts such as balance equations, "chemical" reactions, and mass action rates. A comparison between deterministic and various stochastic models is given. Some basic analysis is provided, together with methods for model solution and model fitting. As a simple case study, a model of a measles outbreak is used, and model fitting is developed. The paper is organized as follows: in Section 2, some details of the problem scope with materials and methods is given, in Section 3, the methods in Section 2 are applied to a case of measles infection. Section 4 provides basic analysis of models with reference to epidemics growth and reduction: development of expressions for the basic reproduction number. Section 5 provides some conclusions.

All computations in the paper are carried out using language Julia, which has an excellent package for solving both deterministic and stochastic differential equations (`DifferentialEquations.jl`; (Rackauckas and Nie, 2017a), (Rackauckas and Nie, 2017b), (Rackauckas and Nie, 2018)), as well as standard least squares model fitting (package `BlackBoxOptim.jl`) and Markov Chain Monte Carlo methods (`Turing.jl`). Plotting is done with package `Plots.jl`.

2 Materials and methods

2.1 Number balance

In epidemiology models, individuals are categorized into n_{X} types/classes X_j with $j \in \{1,\ldots,n_{\mathrm{X}}\}$. The number of individuals in *class/compartment* X_j is denoted $X_j \in \mathbb{N}_0$, and the total number of individuals in the *population* is $N = \sum_{j=1}^{n_{\mathrm{X}}} X_j$. For efficiency, we collect individuals X_j into vector $X \in \mathbb{N}_0^{n_{\mathrm{X}}}$.

A general (vector) number balance for the individuals in the compartments is

$$X(t+\Delta t) = X(t) + \int_t^{t+\Delta t} \left(\dot{X}_{\mathrm{i}}(\theta) - \dot{X}_{\mathrm{e}}(\theta) + \dot{X}_{\mathrm{g}}(\theta)\right) \mathrm{d}\theta. \tag{1}$$

Here, $\dot{X} \in \mathbb{Z}^{n_{\mathrm{X}}}$ denotes *flow rate*, with subscripts i: immigration, e: emigration, and g: generation, and $\dot{X}_{\mathrm{g}} = N \cdot r_{\mathrm{g}}$, where the per capita rate of generation vector r_{g} for the n_{X} *classes* is

$$r_{\mathrm{g}} = \nu^{\mathsf{T}} r \tag{2}$$

and $\nu \in \mathbb{Q}^{n_{\mathrm{r}} \times n_{\mathrm{X}}}$ is the *stoichiometric reaction matrix*, while r is the rate of reaction *per capita* for the n_{r} *reactions*.

If we make the assumption that the elements of X are "large" numbers so that $\left|\left(X(t+\Delta t) - X(t)\right)_j\right| \ll \left|X(t)_j\right|$,

[3] The disease apparently was first observed in Haskell County, Kansas, USA in January 1918, with first known case being an army cook at Camp Funston in Kansas, March 4, 1918.

[4] https://en.wikipedia.org/wiki/Gillespie_algorithm

[5] Some presentations of epidemiology refer to Lagrangian and Eulerian *movement* in an unconventional way, e.g., Martcheva (2015), pp. 389–392.

we make the approximation $X \in \mathbb{R}_0^{n_X}$. If we also assume that the integrand is a *continuous function of time*[6], the mean value theorem allows the result

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \dot{X}_\mathrm{i} - \dot{X}_\mathrm{e} + N \cdot \left(\nu^\mathsf{T} r\right). \tag{3}$$

With similar assumptions, we can describe the total population as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \dot{N}_\mathrm{i} - \dot{N}_\mathrm{e} + \dot{N}_\mathrm{g}. \tag{4}$$

Introducing the per capita number $\check{X}$, $\check{X} \triangleq X/N$, and similarly $\dot{X}_\mathrm{i} \triangleq \dot{N}_\mathrm{i} \cdot \check{X}_\mathrm{i}$, $\dot{X}_\mathrm{e} \triangleq \dot{N}_\mathrm{e} \cdot \check{X}_\mathrm{e}$, the per capita model is

$$\frac{\mathrm{d}\check{X}}{\mathrm{d}t} = \frac{\dot{N}_\mathrm{i}}{N}\left(\check{X}_\mathrm{i} - \check{X}\right) + \nu^\mathsf{T} r - \frac{\dot{N}_\mathrm{g}}{N}\check{X}. \tag{5}$$

2.2 Assumptions on the total population

Simple epidemiology models consider a single compartment with constant population: either dead people are considered part of the population and births are neglected, or birth rate and death rate are assumed to be equal. If the spread of the infection is slow compared to the birth/death cycle, population growth must be included in the model.

If infection rate and level of illness depends on internal coordinates such as age, underlying illnesses, etc., it may be necessary to pose separate models for each age group, etc. In this case, "birth" and "death" into each age group or group of underlying conditions must be included.

If there is a geographical distribution in the number of infected per capita, some sort of distributed model should be used. Examples could be one compartment per country, per region, per municipality, per suburb, or other natural groupings. In this case, it may be necessary to include immigration and emigration for the individual compartments — this depends on the level of interaction vs. the rate of infection spread.

2.3 The classical continuous SIR description

In the classical (Kermack and McKendrick, 1927) model, the population of N individuals is divided into 3 groups: individuals of type S are *susceptible* for infection, individuals of type I are *infected*, while individuals of type R have *recovered* from the infection (or died), thus $X = (S, I, R)$. There is a chain of "reactions" $\mathrm{S} \to \mathrm{I} \to \mathrm{R}$, which can be broken down into two independent parallel events $\mathscr{E}_j$ — interaction with infection $\mathscr{E}_\mathrm{i}$ of rate r_i, and recovery $\mathscr{E}_\mathrm{r}$ of rate r_r:

$$\begin{aligned} \mathscr{E}_\mathrm{i} &: \mathrm{S} \overset{\mathrm{I}}{\to}{}^{k_\mathrm{i}} \mathrm{I}, \quad r_\mathrm{i} \\ \mathscr{E}_\mathrm{r} &: \mathrm{I} \to^{k_\mathrm{r}} \mathrm{R}, \quad r_\mathrm{r}. \end{aligned}$$

Here, $\mathrm{S} \overset{\mathrm{I}}{\to}{}^{k_\mathrm{i}} \mathrm{I}$ indicates that an infected individual *catalyses* the transformation of a susceptible S into a new infected individual I without "consuming" the original infecting individual, with a frequency factor k_i or with a mean interval $\tau_\mathrm{i} = 1/k_\mathrm{i}$.[7] The infection rate r_i per capita takes place with a certain probability when a susceptible individual is in proximity of an infected individual, thus the probability of infection depends on the relative concentration of the two types, in accordance with the *law of mass action* in chemical kinetics,

$$r_\mathrm{i} = \frac{1}{\tau_\mathrm{i}}\check{S}\check{I} = k_\mathrm{i}\check{S}\check{I}.$$

At the same time, $\mathrm{I} \to^{k_\mathrm{r}} \mathrm{R}$ indicates a simple recovery (i.e., no catalysis) from infected I to recovered R during a mean time of $\tau_\mathrm{r} = 1/k_\mathrm{r}$. The reaction r_r depends only on the concentration of the infected type, thus

$$r_\mathrm{r} = \frac{1}{\tau_\mathrm{r}}\check{I} = k_\mathrm{r}\check{I}.$$

The stoichiometric reaction is

$$\underbrace{\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}}_{=\nu} \begin{pmatrix} \mathrm{S} \\ \mathrm{I} \\ \mathrm{R} \end{pmatrix} \leftarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where ν is the *stoichiometric matrix*. It follows that

$$\begin{pmatrix} r_{\mathrm{S,g}} \\ r_{\mathrm{I,g}} \\ r_{\mathrm{R,g}} \end{pmatrix} = \nu^T \begin{pmatrix} r_\mathrm{i} \\ r_\mathrm{r} \end{pmatrix} = \begin{pmatrix} -r_\mathrm{i} \\ r_\mathrm{i} - r_\mathrm{r} \\ r_\mathrm{r} \end{pmatrix}.$$

2.4 Extension: the SEIR description

In the SEIR description, individuals are classified as type S (*susceptible*), type E are *exposed* to infection, but the infection is latent, type I (*infected*), and type R (*recovered*). This gives $X = (S, E, I, R)$, and neglecting birth and death, $N = S + E + I + R$. The chain of events $\mathrm{S} \to \mathrm{E} \to \mathrm{I} \to \mathrm{R}$, can be broken down into the three parallel independent events $\mathscr{E}_\mathrm{i}$, $\mathscr{E}_\mathrm{e}$, and $\mathscr{E}_\mathrm{r}$ with rates r_i, r_e, and r_r, respectively:

$$\begin{aligned} \mathscr{E}_\mathrm{i} &: \mathrm{S} \overset{\mathrm{I}}{\to}{}^{k_\mathrm{i}} \mathrm{E}, \qquad r_\mathrm{i} = \frac{1}{\tau_\mathrm{i}}\check{I}\check{S} = k_\mathrm{i}\check{I}\check{S} \\ \mathscr{E}_\mathrm{e} &: \mathrm{E} \to^{k_\mathrm{e}} \mathrm{I}, \qquad r_\mathrm{e} = \frac{1}{\tau_\mathrm{e}}\check{E} = k_\mathrm{e}\check{E} \\ \mathscr{E}_\mathrm{r} &: \mathrm{I} \to^{k_\mathrm{r}} \mathrm{R}, \qquad r_\mathrm{r} = \frac{1}{\tau_\mathrm{r}}\check{I} = k_\mathrm{r}\check{I}. \end{aligned}$$

The stoichiometric reaction is

$$\underbrace{\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}}_{=\nu} \begin{pmatrix} \mathrm{S} \\ \mathrm{E} \\ \mathrm{I} \\ \mathrm{R} \end{pmatrix} \leftarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

[6] Invalid for stochastic processes!

[7] Note that $\tau_\mathrm{i} = 1/k_\mathrm{i}$ is *not* the length of time an infected person stays infected (which probably is more or less constant), but rather the mean interval between each time one infected individual causes a susceptible to become infected. In other words, if a freshly infected person is removed from society (e.g., locked up, or otherwise put away), then $\tau_\mathrm{i} \to \infty$, alternatively $k_\mathrm{i} \to 0$, but the same infected person will still recover (or die) in finite time.

where ν is the *stoichiometric matrix*. It follows that

$$\begin{pmatrix} r_{\mathrm{S,g}} \\ r_{\mathrm{E,g}} \\ r_{\mathrm{I,g}} \\ r_{\mathrm{R,g}} \end{pmatrix} = \nu^T \begin{pmatrix} r_{\mathrm{i}} \\ r_{\mathrm{e}} \\ r_{\mathrm{r}} \end{pmatrix} = \begin{pmatrix} -r_{\mathrm{i}} \\ r_{\mathrm{i}} - r_{\mathrm{e}} \\ r_{\mathrm{e}} - r_{\mathrm{r}} \\ r_{\mathrm{r}} \end{pmatrix}.$$

2.5 Poisson distribution in events

Let N_t be the (random) number of events that already have occurred at time t, and let $N_{t+\Delta t}$ denote the number of events that have occurred at time $t+\Delta t$. Introducing $\Delta N_t \triangleq N_{t+\Delta t} - N_{\mathrm{t}}$, it follows that in the interval $[t, t+\Delta t]$, we have $\Delta N_t \in \mathbb{N}_0$ new occurrences. When occurances appear with a constant mean rate λ, random variable ΔN_t is Poisson distributed, $\Delta N_t \sim \mathbf{Pois}(\lambda)$, with

$$\mathbb{E}(\Delta N_t) = \lambda$$
$$\mathbb{V}(\Delta N_t) = \lambda.$$

Next, consider the random *generation rate* $R = \frac{\Delta N_t}{\Delta t}$ in a relatively short time interval Δt, and an average rate of generation $\bar{r}$ so that $\lambda = \bar{r} \cdot \Delta t$. There is no guarantee that realization $r = \frac{\Delta n}{\Delta t}$ is an integer, but the random variable R is distributed according to a quasi continuous Poisson distribution, $R \sim \mathbf{cPois}(\bar{r})$, where for "large" $\bar{r}$,

$$\mathbb{E}(R) \approx \bar{r}$$
$$\mathbb{V}(R) \approx \bar{r}.$$

Furthermore, for realistic values of $\bar{r}$, the random rate R will approach a normal distribution $R \sim \mathbf{N}\left(\bar{r}, \sqrt{\bar{r}}\right)$.

2.5.1 Stochastic differential equation

When the reaction rate is random (r in Eq. 3 becomes R), the mean value theorem is invalid, and Eq. 3 must be rephrased as a stochastic differential equation (SDE)

$$\mathrm{d}X = \dot{X}_{\mathrm{i}}\mathrm{d}t - \dot{X}_{\mathrm{e}}\mathrm{d}t + N \cdot \left(\nu^{\mathsf{T}} R\right)\mathrm{d}t.$$

With $R \sim \mathbf{N}\left(\bar{r}, \sqrt{\bar{r}}\right)$ and introducing $Z \sim \mathbf{N}(0,1)$, this can be rephrased as

$$\mathrm{d}X = \left(\dot{X}_{\mathrm{i}} - \dot{X}_{\mathrm{e}} + N \cdot \left(\nu^{\mathsf{T}} \bar{r}\right)\right)\mathrm{d}t + N \cdot \left(\nu^{\mathsf{T}} \sqrt{\bar{r}}\right) Z \mathrm{d}t.$$

Using an SDE solver, a number of realizations are found. Then based on these realizations, statistics (mean, etc., including the distribution) can be computed for each time instance. Alternatively, an associated determistic Fokker-Planck equation can be posed, and solved to find the probability distribution directly.

When formulating the SDE, the approximation $X \in \mathbb{R}^{n_{\mathrm{X}}}$ is used.

2.5.2 First reaction time

Let N_t be the random number of arrivals accumulated at time t. Let ΔT_t be the random time it takes for the event of one additional arrival, assuming that someone arrived at time t.

By definition, the following two events are equivalent:

$$(\Delta T_t > \Delta t) \equiv (N_t = N_{t+\Delta t}).$$

With $\Delta N_t \sim \mathbf{Pois}(\lambda)$, it can be shown that ΔT is *Exponentially* distributed, $\Delta T \sim \mathbf{Exp}(\bar{r})$, or alternatively with U uniformly distributed $U \sim \mathbf{U}_{[0,1)}$ we can use

$$\Delta T = -\frac{1}{\bar{r}} \ln(U).$$

In the simplest version of Gillespie's algorithm, a three step procedure is used: (i) first, a uniform random number generator is used to find which event takes place (i.e., which of the reaction takes place), (ii) secondly, the time ΔT until next event is computed by drawing from an Exponential distribution. (iii) Then only the event found from (i) is carried out, and the time index is updated to $t_{i+1} = t_i + \Delta T$.

A number of realizations of the first reaction time model can be carried out, and it is possible to compute statistics for each time point.

Alternatively, similarly as for SDEs and the Fokker-Planck equation, a deterministic *master equation* can be posed to describe the probability distribution of the solution for the first reaction time model.

When using the first reaction time formulation, we maintain the fact that the number of people are integers, $X \in \mathbb{N}_0^{n_{\mathrm{X}}}$.

2.6 Reproduction number

The *reproduction* number of the disease is the average number of persons that an individual infects before recovering. The *basic reproduction* number R_0 is the reproduction number when (i) starting from a disease-free state, (ii) for the zeroth generation, i.e., the natural reproduction at initial time when everyone are susceptible, prior to invoking any mitigation policy.

For the SIR model, the initial time Jacobian $J_{\check{X}}$ of the model vector field is

$$J_{\check{X}} = \begin{pmatrix} -k_{\mathrm{i}}\check{I} & -k_{\mathrm{i}}\check{S} & 0 \\ k_{\mathrm{i}}\check{I} & k_{\mathrm{i}}\check{S} - k_{\mathrm{r}} & 0 \\ 0 & k_{\mathrm{r}} & 0 \end{pmatrix}.$$

Starting from a disease-free state, $\check{I}(0) \equiv 0$,

$$J_{\check{X}} = \begin{pmatrix} 0 & -k_{\mathrm{i}}\check{S}(0) & 0 \\ 0 & k_{\mathrm{i}}\check{S}(0) - k_{\mathrm{r}} & 0 \\ 0 & k_{\mathrm{r}} & 0 \end{pmatrix}.$$

$J_{\check{X}}$ has two eigenvalues in the origin, and one eigenvalue at

$$\lambda = k_{\mathrm{i}}\check{S}(0) - k_{\mathrm{r}}.$$

Table 1. Daily number of measles infected at boarding school with 763 boys in Northern England, January-February 1978. Taken from (Martcheva, 2015).

Day	# infected	Day	# infected
3	25	9	192
4	75	10	126
5	227	11	71
6	296	12	28
7	258	13	11
8	236	14	7

Stability requires $\lambda < 0$,

$$\frac{k_\mathrm{i}\check{S}(0)}{k_\mathrm{r}} < 1.$$

The infection rate constant k_i varies with mitigation policy, etc. Let k_i^0 denote the natural rate constant *without* mitigation. The *basic reproduction number* R_0 is assessed when $\check{S}(0) \equiv 1$ and without mitigation, hence

$$\mathsf{R}_0 = \frac{k_\mathrm{i}^0}{k_\mathrm{r}}.$$

With mitigation, we could define an *effective reproduction number* $\mathsf{R} \triangleq k_\mathrm{i}/k_\mathrm{r}$.[8]

2.7 Model fitting

In simple model fitting, model parameters and initial states are posed as unknowns in a loss function that measures the difference between experimental data and a shooting (ballistic) formulation of the model. More sophisticated methods for Bayes estimation are conveniently solved using Markov Chain Monte Carlo methods (MCMC), see, e.g., (Evensen, 2009) for a practical introduction with his Ensemble Kalman Filter approach. Other methods and tools exist.

2.8 Measles case study

At a boarding school in North England, January–February 1978, measles infection was observed among the 763 pupils, with 25 infections on assumed day 3 of the epidemic. The evolution of observed infections was as recorded in Table 1.

In (Martcheva, 2015), a SIR model is used to model the infection, assuming that $N = 763$ is constant, and that all pupils are locked up in the school. The basic reproduction number for measles is estimated to be in the range $\mathsf{R}_0 \in [16, 18]$, (Keeling and Rohani, 2008), which should be contrasted to the range $[3, 4]$ for seasonal flu.

[8] R_0 and R has nothing to do with the number R of recovered.

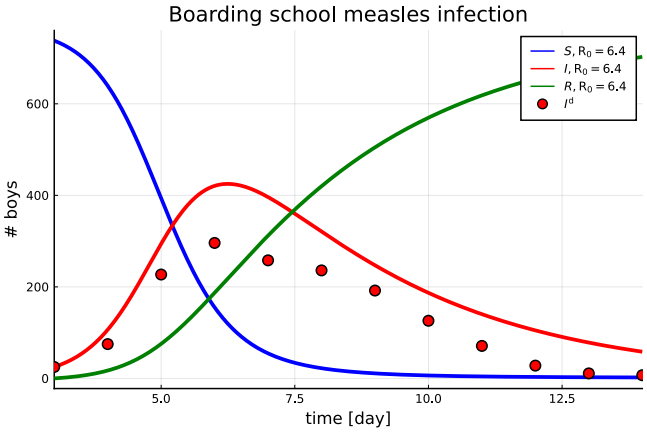

Figure 1. Comparing SIR model with original parameters vs. registered infection data for boarding school in North England, January–February 1978.

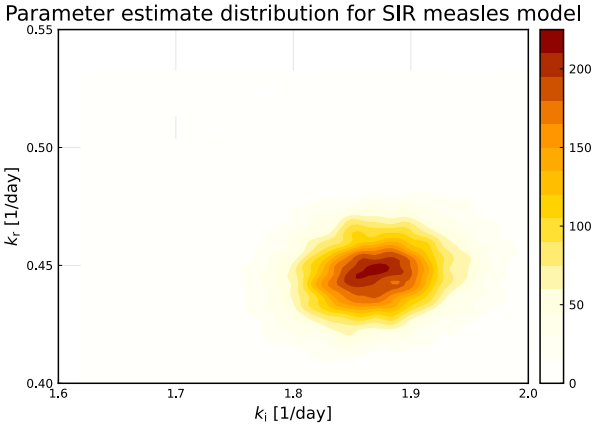

Figure 2. Distribution in Bayes estimates of parameters $\tau_\mathrm{r} = 1/k_\mathrm{i}$ and $\tau_\mathrm{i} = 1/k_\mathrm{r}$ in SIR model based on infection data for boarding school in North England, January–February 1978. Dr. Tamas Papp, TU Wien, provided the function of doing this plot.

3 Measles case study

3.1 SIR model

3.1.1 Deterministic model with model fitting

With the data in Table 1, and using parameters ((Martcheva, 2015), p. 127)

$$k_\mathrm{i} = \beta N = 0.0025 \cdot 763 = 1.9075$$
$$k_\mathrm{r} = \alpha = 0.3,$$

the deterministic model fits data reasonably well, Figure 1.

It is of interest to find better model parameters. Without going into details, the `Turing.jl` package for Julia supports Markov Chain Monte Carlo (MCMC) estimation of the model parameters, with parameter probability distribution as in Figure 2.

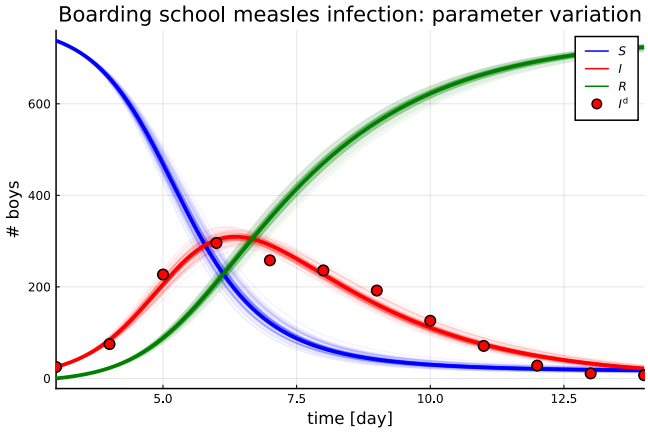

Figure 3. Comparing SIR model *data retrodiction* vs. registered infection data for boarding school in North England, January–February 1978. Dark/thick lines are based on the point estimates.

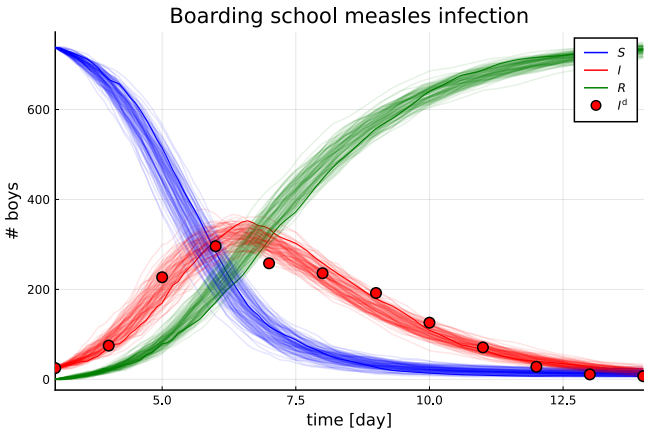

Figure 4. Stochastic realizations (trajectories) for an ensemble of 100 possible scenarios.

Point estimators (mean values) of the parameters are found to be:

$$k_\mathrm{i} \approx 1.817$$
$$k_\mathrm{r} \approx 0.4618,$$

which gives an *effective* reproduction number of $\mathrm{R} \approx 1.817/0.4618 \approx 3.9$. MCMC provides a number of possible parameter values, and this parameter uncertainty translates into prediction uncertainty, known as *data retrodiction*; results are given in Figure 3 in pale/thin lines, together with the simulation when using the point estimates in dark/thick lines.

From the data retrodiction, we see that there is relatively little uncertainty in the model even with varying parameters. This indicates that the model has quite good predictive properties.

3.1.2 SDE model

Using basic data from the deterministic simulation, we expand the model to a set of stochastic differential equations as described Section 2.5.1. The results are shown in Figure 4.

3.1.3 First reaction event model

Instead of formulating the model as Stochastic Differential Equations, we can write integer difference equations with first reaction event description for changes, see Section 2.5.2. The results are shown in Figure 5, and should be compared to the results in Figure 4.

3.2 SEIR model

It is of interest to consider the SEIR model for the measles case, where we assume that I_3 is known, that $R_3 = 0$, that N is known, but that S_3 or E_3 are unknown; we choose to estimate S_3. Reusing the model parameters for k_i and k_r from the SIR model, we initially assume that $k_\mathrm{e} = 2$; see Figure 6.

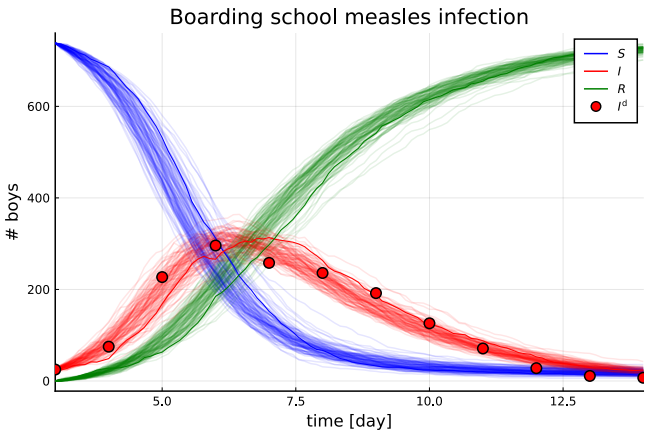

Figure 5. Stochastic realizations (trajectories) for an ensemble of 100 possible scenarios using Gillespie's First Reaction Method.

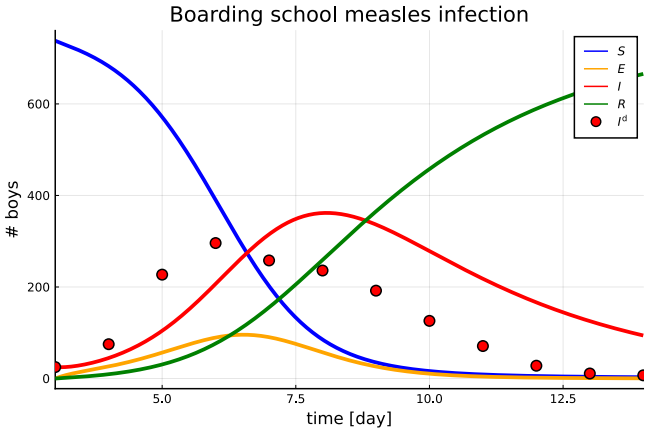

Figure 6. SEIR model with original parameters vs. registered infection data for boarding school in North England, January–February 1978.

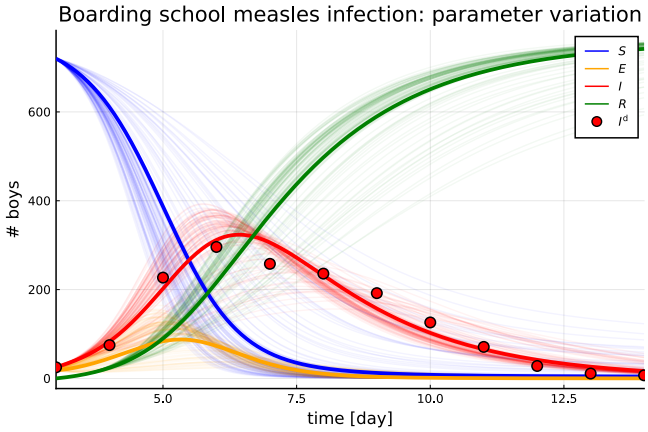

Figure 7. Comparing SEIR model data *retrodiction* vs. registered infection data for boarding school in North England in January–February 1978. Dark lines are based on the point estimates.

We can also fit the parameters of the SEIR model, including the initial value of S_3. The result are the following point estimates (mean value):

$$\begin{aligned} k_\mathrm{i} &\approx 2.50 \\ k_\mathrm{e} &\approx 2.826 \\ k_\mathrm{r} &\approx 0.477 \\ S_3 &\approx 720 \end{aligned}$$

If we draw estimates of these parameters and initial value and redo simulations, the data retrodiction is as in Figure 7. As we see, the uncertainties in the retrodiction of the SEIR model, Figure 7, are far larger than the uncertainties in the retrodiction of the SIR model, Figure 3. This may indicate that the measles infection is best modeled by a SIR model, but it must be remembered that we have few data points and the SEIR model has almost twice as many parameters as the SIR model[9].

4 Analysis of epidemiology models

4.1 Condition for infection growth

With model

$$\frac{\mathrm{d}\check{X}}{\mathrm{d}t} = F\left(\check{X}\right)$$

in per capita variables $\check{X}$, a standard procedure for finding conditions for infection growth is to analyze the Jacobian

$$J_{\check{X}_0} = \left.\frac{\partial F}{\partial \check{X}}\right|_{\check{X}_0},$$

where $\check{X}_0$ is the operating point. In general, the operating point will depend on the current states.

[9]Estimated parameters for the SIR model: k_i, k_r, and standard deviation in output error. Estimated parameters for the SEIR model: k_i, k_e, k_r, initial value for S_3, and standard deviation in output error.

In a standard way, stability can be assessed based on eigenvalues, or related Routh-Hurwitz criteria. However, eigenvalues are often difficult to compute for realistic models.

An alternative is the *Next-Generation Approach* (van den Driessche, 2017; Martcheva, 2015). With model

$$\frac{\mathrm{d}\check{X}}{\mathrm{d}t} = \nu^\mathsf{T} r$$

where $r = r\left(\check{X}\right)$, assume that $\check{X}$ has been sorted such that $\check{X} = (x,y)$ and x contains the *infected* compartments while y contains all other compartments. Thus, we have the differential equations

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= f(x,y) \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= g(x,y) \end{aligned}$$

where

$$\begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix} = \nu^\mathsf{T} r(x,y).$$

We split $f(x,y)$ into two terms,

$$f(x,y) = \mathscr{F}(x,y) - \mathscr{V}(x,y)$$

where $\mathscr{F}_i(x,y)$ contains the rate of *appearance* of *new infections* in compartment i, while $\mathscr{V}_i(x,y)$ include all other terms: births, deaths, disease progression, recovery. Observe that $\mathscr{F}(x,y)$ and $\mathscr{V}(x,y)$ are not necessarily unique.

Let F and V be the Jacobians of $\mathscr{F}$ and $\mathscr{V}$ in the disease-free case given by x^0, y^0, where $x^0 = 0$ and $y^0 \geq 0$:

$$\begin{aligned} F &= \left.\frac{\partial \mathscr{F}(x,y)}{\partial x}\right|_0 \\ V &= \left.\frac{\partial \mathscr{V}(x,y)}{\partial x}\right|_0. \end{aligned}$$

Then $N = FV^{-1}$ is the *Next-Generation Matrix*. Let $\rho(N)$ denote the *spectral radius* of matrix N: with λ_i the eigenvalues of N, $\rho(N) = \max_i |\lambda_i|$. We can then define the reproduction number R as

$$\mathsf{R} \triangleq \rho^m\left(FV^{-1}\right)$$

where m is some integer to make $\mathsf{R} \propto k_\mathrm{i}$. The infection is under control/decreases if $\mathsf{R} < 1$, and is out of control/grows if $\mathsf{R} > 1$.

4.2 Stability from SEIR model

For the SEIR model with $\check{S}(0) = 1$, the disease-free Jacobian is

$$J_{\check{X}(0)} = \begin{pmatrix} 0 & 0 & -k_\mathrm{i} & 0 \\ 0 & -k_\mathrm{e} & k_\mathrm{i} & 0 \\ 0 & k_\mathrm{e} & -k_\mathrm{r} & 0 \\ 0 & 0 & k_\mathrm{r} & 0 \end{pmatrix},$$

and it is feasible to find R from eigenvalue analysis. However, for illustration, we consider the Next-Generation Approach.

Here, $x = (\check{E}, \check{I})$, and we need to extract elements $(2:3, 2:3)$, i.e.,

$$J_0 = \begin{pmatrix} -k_\mathrm{e} & k_\mathrm{i} \\ k_\mathrm{e} & -k_\mathrm{r} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & k_\mathrm{i} \\ k_\mathrm{e} & 0 \end{pmatrix}}_{=F} - \underbrace{\begin{pmatrix} k_\mathrm{e} & 0 \\ 0 & k_\mathrm{r} \end{pmatrix}}_{=V},$$

and the Next-Generation Matrix N is

$$N = FV^{-1} = \begin{pmatrix} 0 & k_\mathrm{i} \\ k_\mathrm{e} & 0 \end{pmatrix} \begin{pmatrix} k_\mathrm{e} & 0 \\ 0 & k_\mathrm{r} \end{pmatrix}^{-1} = \begin{pmatrix} 0 & \frac{k_\mathrm{i}}{k_\mathrm{r}} \\ 1 & 0 \end{pmatrix},$$

with eigenvalues

$$\det \begin{pmatrix} \lambda & -\frac{k_\mathrm{i}}{k_\mathrm{r}} \\ -1 & \lambda \end{pmatrix} = \lambda^2 - \frac{k_\mathrm{i}}{k_\mathrm{r}} = 0$$

and spectral radius

$$\mathsf{R} = \rho^m(N) = \sqrt[m]{\frac{k_\mathrm{i}}{k_\mathrm{r}}};$$

where we choose $m = 2$ so that $\mathsf{R} \propto k_\mathrm{i}$; $\mathsf{R} = \frac{k_\mathrm{i}}{k_\mathrm{r}}$ — just as for the SIR model, see Section 2.6.

5 Conclusions

An overview of principles for formulating epidemiological models has been given. Here, models based on balance laws are treated; the principle is identical to what is used in chemical engineering and other process engineering fields. A deterministic model is the starting point, and it is discussed how the assumption of Poisson distribution in the reactions events leads to either a stochastic differential equation or the first reaction time model/Gillespie formulation. These ideas of Poisson distribution carries over to chemical reactions in general.

Based on the published measles infection data, a susceptible-infected-recovered (SIR) epidemiological model is fitted to data using a Markov Chain Monte Carlos approach (MCMC), and appears to be adequate. It is also possible to fit an extended model with an exposed class (SEIR) to the measles data, but this gives much wider uncertainty in model parameters, with resulting large uncertainty in model predictions.

In an analysis part, several ways of finding an expression for the reproduction number R are discussed; R is used to describe the stability of an infection. The Next-Generation Approach is probably the method that is simplest to use for complex models.

This paper lays out the fundamental ideas behind epidemiology models, as used, e.g., in COVID-19 studies.

References

Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*. Number 69 in Texts in Applied Mathematics. Springer, New York, 2019. ISBN 978-1-4939-9826-5.

Tom Britton and Etienne Pardoux, editors. *Stochastic Epidemic Models with Inference*. Number 2255 in Lecture Notes in Mathematics. Springer, Springer Nature, Switzerland, 2019. ISBN 978-3030308995.

Geir Evensen. *Data Assimilation. The Ensemble Kalman Filter*. Springer, Berlin, 2nd edition, 2009. ISBN 978-3-642-03710-8.

Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976. doi:https://doi.org/10.1016/0021-9991(76)90041-3.

Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81 (25):2340–2361, 1977. doi:10.1021/j100540a008.

Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 41 William Street, Princeton, New Jersey 08540, 2008. ISBN 978-0-691-11617-4.

W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A.*, 115(772):700–721, 1927. doi:https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118.

Maia Martcheva. *An Introduction to Mathematical Epidemiology*, volume 61 of *Texts in Applied Mathematics*. Springer, New York, 2015. ISBN 978-1-4899-7611-6.

Christopher Rackauckas and Qing Nie. DifferentialEquations.jl — A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of Open Research Software*, 5(15), 2017a. doi:10.5334/jors.151.

Christopher Rackauckas and Qing Nie. Adaptive methods for stochastic differential equations via natural embeddings and rejection sampling with memory. *Discrete and continuous dynamical systems. Series B*, 22(7):2731, 2017b.

Christopher Rackauckas and Qing Nie. Stability-Optimized High Order Methods and Stiffness Detection for Pathwise Stiff Stochastic Differential Equations. *arXiv:1804.04344 [math]*, 2018. URL `http://arxiv.org/abs/1804.04344`.

C.W. Schwabe, H.P. Riemann, and C.E. Franti. *Epidemiology in Veterinary Practice*. Lea & Febiger, 1977. pp. 258–260.

Pauline van den Driessche. Reproduction numbers of infectious disease models. *Infectious Disease Modelling*, 2(3):288–303, aug 2017. doi:10.1016/j.idm.2017.06.002.