Mathematical Methods for Molecular Science

John E. Straub



Supplement on Kinetic Models of Infectious Disease

Version 2.0 released June 2020.

This work is licensed under a Creative Commons "Attribution-NonCommercial-ShareAlike 4.0 International" license.



Kinetic models of infectious disease epidemics

A variety of simple models have been developed to predict the progression of infectious disease. The models are also used, after the fact, to analyze the mechanism of disease progression to gain insights into the mechanism of propagation. We will explore two popular models based on a set of coupled first order ordinary differential equations.

In each model, we start with a population of *susceptible individuals* who may become infected with the disease. The fraction of susceptible individuals at a given time t is written s(t). Those individuals may become infected, and the number of *infected individuals* at a given time is i(t). Finally, we track the number of *recovered individuals* at a given time r(t). The rate of change in each of these quantities is described in terms of a first order ordinary differential equation describing the time-dependence of each population from the initial spread of the disease until its conclusion. Let's see how that works.

SIR model of infectious disease

The SIR model divides the total population into susceptible individuals, s(t), infected individuals, i(t), and recovered individuals, r(t).¹ The sum s(t)+i(t)+r(t)=1. How do we expect each of those quantities to change in time?

Let's start with the number of susceptible individuals which initially is taken to be $s(0) \simeq 1.^2$ The probability of a susceptible individual encountering an infected individual is proportional to the number of each type of individual

probability of encounter $\propto s(t)i(t)$

If you double the number of infected individuals, you double the rate of infection. If you cut in half the number of susceptible individuals, the rate of infection is cut in half.³

The rate of infection is proportional to the probability of a susceptible individual encountering an infected individual

rate of infecting susceptible individuals = b s(t)i(t)

where *b* is a rate constant proportional to the number of infectious contacts per day that an infected individual makes with another individual.⁴ As such, the rate of change in the number of susceptible individuals s(t) is

$$\frac{d}{dt}\mathbf{s}(t) = -b\,\mathbf{s}(t)\mathbf{i}(t)$$

and the negative sign leads to a decreasing number of susceptible individuals.

As susceptible individuals are converted into infected individuals, the fraction of infected individuals grows at a rate b s(t)i(t). Assuming no mortality, after some time the infected individual will recover from the illness at a rate proportional to the number of infected individuals⁵

rate of recovery of infected individuals = k i(t)

where $\frac{1}{k}$ is the average number of days required to recover from the illness. As such, the rate of change in the number of infected individuals is

$$\frac{d}{dt}\mathbf{i}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - k\,\mathbf{i}(t)$$

so that the number of infected individuals increases as susceptible individuals are infected and decreases as infected individuals recover.

¹ Originally developed to study the 1666 plague and 1865 cholera epidemics in London. WO Kermack and AG McKendrick, "A Contribution to the Mathematical Theory of Epidemics," *Proc. Roy. Soc. Lond. A* **115**, 700-721 (1927).

² Individuals with immunity or in physical isolation are not considered.

³ This is like a *bimolecular reaction* where the rate of reaction is proportional to the concentration of each species.

⁴ In the SIR model, individuals referred to as infected are infectious and can transmit the disease.

⁵ This is like a *unimolecular reaction* where the rate of reaction is proportional to the concentration of reactant.

4 KINETIC MODELS OF INFECTIOUS DISEASE EPIDEMICS

Finally, the rate of change in the number of recovered individuals is

$$\frac{d}{dt}\mathbf{r}(t) = k\,\mathbf{i}(t)$$

Note that r(t) can also be determined from r(t) = 1 - (s(t)+i(t)). The combination of these three first order ordinary differential equations forms the SIR model of infectious disease

$$\frac{d}{dt}\mathbf{s}(t) = -b\,\mathbf{s}(t)\mathbf{i}(t) \qquad \frac{d}{dt}\mathbf{i}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - k\,\mathbf{i}(t) \qquad \frac{d}{dt}\mathbf{r}(t) = k\,\mathbf{i}(t)$$

By solving these three equations we can determine the time evolution of the conversion of susceptible individuals into infected individuals and finally into recovered individuals.

There is an important constant that can be measured at the end of an epidemic called the *reproduction number*, \mathcal{R}_0 , defined to be the total number of infectious contacts made by individual. In the SIR model the reproduction number is given by a product of two of the model parameters

$$\mathcal{R}_0 = b \frac{1}{k} = (\text{\# contacts per day}) \times (\text{infectious period in days})$$

The reproduction number reflects the virulence or ease of transmission of the disease as well as the social environment and behavior of the population. High reproduction numbers are found in densely populated societies with a higher number of contacts between individuals per day. ⁶

Let's explore the predictions of the SIR model for a number of infectious diseases. The table below lists data for infectious diseases relevant to the SIR model starting with the influenza epidemic of 1918 and reaching the 2019 coronavirus (Covid-19) epidemic.

⁶ High reproduction numbers can also result from promiscuous social behavior. As such, the reproduction number of a disease can be reduced through *social distancing*.

Epidemic model parameters					
disease	daily contacts $= b$	incubation period = $\frac{1}{q}$	infectious period = $\frac{1}{k}$	$\mathcal{R}_0 = \frac{b}{k}$	mortality= \mathcal{M}_0
influenza (1918)	$1 \mathrm{day}^{-1}$	4 days	4 days	2.4 - 4.4	10%
polio (1952)	$1/2 day^{-1}$	5 days	12 days	5 - 7	0.025%
measles (1960)	$2 \mathrm{day}^{-1}$	12 days	8 days	12 - 18	0.1%
Hong Kong flu(1968)	$^{1/2}{ m day}^{-1}$	6 days	5 days	1.2 - 3.6	0.5%
SARS (2003)	$1/12 day^{-1}$	4 days	6 days	0.5	11%
coronavirus (2019)	$^{1/2}{ m day}^{-1}$	5 days	6 days	2.2 - 3.6	3.4%

The reproduction number \mathcal{R}_0 can be determined from a knowledge of the fraction of susceptible individuals who avoid infection over the course of the epidemic⁷

$$\mathcal{R}_0 = \frac{\ln(\mathbf{s}_\infty)}{\mathbf{s}_\infty - 1} = b \frac{1}{k}$$

where $s_{\infty} = s(t = \infty)$. This relation is used to compute \mathcal{R}_0 at the end of an epidemic.

A typical number of daily infectious contacts is taken to be $b = \frac{1}{2} \text{ day}^{-1}$. Exceptions are higher values for the influenza epidemic of 1918, which impacted large confined populations, measles, to accommodate the large reproduction number, and SARS, for which early isolation measures were highly effective. The infectious period is estimated as the product of the assumed number of daily contacts and the inverse of the reproduction number

$$k = b \frac{1}{\mathcal{R}_0}$$

⁷ Using the chain rule we find

$$\frac{d\mathbf{i}}{d\mathbf{s}} = \frac{d\mathbf{i}}{dt} \left(\frac{d\mathbf{s}}{dt}\right)^{-1} = -1 + \frac{k}{b\mathbf{s}(t)}$$

which we solve to find

$$\mathbf{i}(t) + \mathbf{s}(t) - \frac{k}{b} \ln \mathbf{s}(t) = C$$

where *C* is a constant independent of time. Evaluating the expression at t = 0 and $t = \infty$ we can solve for

$$\frac{b}{k} = \mathcal{R}_0 = \frac{\ln s_\infty}{s_\infty - 1}$$

where $s(t=\infty) = s_{\infty}$.

In practice we will vary the parameter *b* over a range of reasonable values for the assumed number of contacts per day. This will allow us to test the sensitivity of our predictions to the value of this variable parameter. The incubation period is the average time between exposure and the onset of detectable symptoms. It is not considered explicitly in the SIR model.

Let's explore the predictions of the SIR model using the model parameters in the table. The results are shown below.



Figure 1: SIR model of time evolution of six epidemics from influenza (1918), polio (1855), measles (1960), Hong Kong flu (1968), SARS (2003), and coronavirus (2019). In the case of influenza (1918) the infection extends to the majority of individuals ($\mathcal{R}_0 > 1$). In the case of SARS (2003) the epidemic is prevented by effective quarantine and social distancing ($\mathcal{R}_0 < 1$).

Note that we can't solve the equations of the model exactly, so we integrate the equations numerically using a method discovered by *Leonhard Euler* (1707-1783). For the differential equation

$$\frac{dc(t)}{dt} = kc(t)$$

we can say dc(t) = kc(t)dt. Now suppose we replace the infinitesimal step in time dt or concentration dc with a finite step in time Δt or concentration Δc . We then find

$$\Delta c = c(t + \Delta t) - c(t) = kc(t)\Delta t$$

where Δt is the time step. We start at c(0) and find $c(\Delta t) = c(0) + kc(0)\Delta t$. Taking another step we find $c(2\Delta t) = c(\Delta t) + kc(\Delta t)\Delta t$. And so on. Repeating these steps we generate a time series

 $c(0), c(\Delta t), c(2\Delta t), \ldots, c(n\Delta t)$

over *n* time steps out to the desired time $t = n\Delta t$. This method is known as the *forward Euler method*. It was applied to compute the evolution of the populations as a function of time *t* using dt = 1/20 days.

⁸ Parameters for the Covid-19 epidemic were derived from data available in February 2020 before the widespread implementation of social distancing measures in the United States.

⁹ Parameters for the Covid-19 epidemic were derived from data available in February 2020 before the widespread implementation of social distancing measures in the United States.

We can make a number of observations based on a comparison of our results for the six epidemics.⁸ (1) The large reproduction numbers for polio and measles leave few members of the population uninfected. (2) The large reproduction number of the measles leads to a rapid onset of the illness and relatively short duration of the overall epidemic. The long infectious period of polio has the opposite effect. (3) This model of the 2019 coronavirus epidemic most resembles the Hong Kong flu epidemic of 1968.⁹

SEIR model of infectious disease

In the SIR model we assumed that the population could be divided into subgroups of susceptible individuals, infected individuals, and recovered individuals. To improve the model, we add a fourth subgroup of *exposed individuals* at a given time *t* written e(t).¹⁰ Exposed individuals are contacted by an infected individual after which the illness incubates in the absence of symptoms. Following incubation the exposed individual exhibits symptoms and is considered an infected individual.

The population of exposed individuals will increase as susceptible individuals encounter infected individuals and decrease as exposed individuals evolve into infected individuals where

rate of conversion of exposed individuals to infected individuals = q e(t)

and $\frac{1}{q}$ is the average incubation period in days. As such, the rate of change in the number of exposed individuals s(t) is

$$\frac{d}{dt}\mathbf{e}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - q\,\mathbf{e}(t)$$

where the negative sign indicates that the population of exposed individuals decreases as exposed individuals are converted to infected individuals following an incubation period.

We now need to modify the rate of change in the number of infected individuals. The population of infected individuals will increase as exposed individuals are converted to infected individuals and decrease as infected individuals recover

$$\frac{d}{dt}\mathbf{i}(t) = q\,\mathbf{e}(t) - k\,\mathbf{i}(t)$$

The model for the rate of recovery of infected individuals is unchanged. The final result is the SEIR model defined by the first order ordinary differential equations

$$\frac{d}{dt}\mathbf{s}(t) = -b\,\mathbf{s}(t)\mathbf{i}(t)$$
$$\frac{d}{dt}\mathbf{e}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - q\,\mathbf{e}(t)$$
$$\frac{d}{dt}\mathbf{i}(t) = q\,\mathbf{e}(t) - k\,\mathbf{i}(t)$$
$$\frac{d}{dt}\mathbf{r}(t) = k\,\mathbf{i}(t)$$

where s(t)+e(t)+i(t)+r(t)=1 is the total population.¹¹

The incubation period $\frac{1}{q}$ can be measured through the observation of individuals, averaging the time between an infectious contact and the onset of symptoms (see the table). The reproduction number \mathcal{R}_0 can be calculated using the total number of susceptible individuals remaining at the end of the epidemic $s(t=\infty) = s_{\infty}$ where the relation

$$\mathcal{R}_0 = \frac{\ln(\mathbf{s}_\infty)}{\mathbf{s}_\infty - 1} = b\frac{1}{k}$$

is identical to that for the SIR model.

¹⁰ RM May and RM Anderson, "Infectious diseases of humans: dynamics and control," Oxford University Press (1991).

¹¹ We can add detail to the SEIR model by including the birth rate as well as the natural death rate in the population. The *birth rate* is a zeroth order kinetic process, taken to be proportional to the population as a whole. The *death rate* is a first order kinetic process, proportional to the number of individuals of any subgroup independent of the illness.



Let's explore the predictions of the SEIR model using the model parameters in the table. The results are shown below.

We can make a number of observations based on a comparison of our results for the six epidemics.¹² (1) Explicit inclusion of the incubation period in the kinetic models delays the overall onset of disease. (2) In the case of polio, the peak number of infected individuals, i_{max} , exceeds the peak number of exposed individuals, e_{max} . However, for measles the opposite is true and $e_{max} > i_{max}$. (3) As in the SIR model, our model of the 2019 coronavirus epidemic most resembles the Hong Kong flu epidemic of 1968.

Figure 2: SEIR model of time evolution of six epidemics from influenza (1918), polio (1855), measles (1960), Hong Kong flu (1968), SARS (2003), and coronavirus (2019).

¹² Parameters for the Covid-19 epidemic were derived from data available in February 2020 before the widespread implementation of social distancing measures in the United States.

Exponential and power-law kinetic models of infectious disease

Consider the following data from the World Health Organization (WHO) representing coronavirus related mortality statistics in China during the early days of the epidemic.



Figure 3: The growing number of deaths $\dagger(t)$ attributed to coronavirus in the early phase of the epidemic in China. The data were taken over 25 days spanning 27 January 2020 to 16 February 2020.

The data show rapid growth in the number of coronavirus related deaths over a 25 day period. Does it represent exponential growth? Let's analyze this data using the SEIR model.

Exponential growth of infectious disease

In the figures below are 25 days taken from the SEIR model of the propagation of coronavirus in metropolitan Boston. We have assumed a population of $N_T = 4,375,000$. An early stage in the epidemic has been chosen over which the total death count is predicted to rise to numbers comparable to those on day 25 in China. The prediction of the model is shown below.



Figure 4: Predictions of the SEIR model for a coronavirus epidemic in metropolitan Boston derived from data available in February 2020 before the widespread implementation of social distancing measures in the United States. (Left) Model predictions for the number of deaths over 25 days during an early stage in the epidemic. (Right) The same data on a log-linear plot demonstrating the exponential kinetics predicted by the SEIR model.

On the left, we plot the number of deaths over time. The general pattern of a rapid rise in deaths, starting from relatively few on day 0 to roughly 1400 deaths on day 25, seems to mirror the data from China. On the right, we display the same data on a log-linear plot. The fact that the data fall on a line in a log-linear plot indicates that we have exponential growth in the SEIR model. The data can be fitted to an exponential with a time constant of 4.3 days.

In the SEIR model, we assume that (1) the disease spreads in a closed population of some number of individuals, (2) the rate of exposure to an infected

individual is proportional to the number of susceptible individuals, and (3) recovery from the disease provides immunity from future infection. A probability of contact between a susceptible individual and infected individual is taken to be proportional to the fraction of each type of individual

probability of encounter
$$\propto s(t)i(t)$$

Similarly, the rate of recovery from the infection is taken to be proportional to the number of infected individuals

rate of recovery of infected individuals = k i(t)

When the rate of change in a quantity is proportional to the quantity itself, we expect to find exponential kinetics of growth or decay. This explains the exponential growth observed in the mortality statistics predicted by the SEIR model for a coronavirus epidemic in metropolitan Boston.

Let's perform the same analysis on the data from China by plotting the data on a log-linear plot. The results are shown below.



Figure 5: (Left) Mortality data over 25 days during an early stage in the epidemic. (Right) The same data on a log-linear plot. The data do not fall on a line demonstrating that the growing number of deaths did not follow exponential growth as predicted by the SEIR model.

The data displayed on a log-linear plot does not fall on a line. As such, it fails the basic test for exponential kinetics. While the growth in mortality is rapid it is not exponential. This tells us that one or more of the assumptions of the SEIR model is not valid when applied to the early spread of coronavirus in China.

Alternatives to exponential growth

The most questionable assumption in the model is that the rate of exposure to an infected individual is proportional to the number of susceptible individuals.¹³ That assumption implies that there is a constant probability for an infected individual to encounter any individual in society. The assumption ignores the existence of *social networks* that make it more likely for an infected individual to encounter a member of that individual's immediate social group as opposed to a stranger living at a distance.

Rate models that take into account the structure of social networks lead to different predictions for the rate of growth of infectious disease. Studies have shown that the network of social contacts can have a *fractal* structure.¹⁴ In the context of social networks, the fractal structure implies that the probability of encountering another individual in society diminishes with social distance. This is sometimes called a *small world network* after the seminal work of Milgram.¹⁵ That idea was generalized to many phenomena including the spread of infectious disease by Barabasi and Albert¹⁶ and is now widely applied to model kinetics of growth in the physical and social sciences.

¹³ When we consider the spread of the disease outside of Wuhan the assumption that the disease spreads in a closed population would be invalid. However, we will assume that it is a reasonable assumption when applying the model to the initial spread of the disease.

¹⁴ Fractal objects have *self-similar structure* that is observed on many length scales. Such structures are found in dendritic objects like snowflakes, trees, ferns, shorelines, and branched molecules. The form of a tree repeats itself in each branch, which takes the form of the tree, as does each branch of a branch. This is what we refer to as self-similarity. A piece of the object shares the form of the object as a whole.

¹⁵ S Milgram, *Psychology Today* **1**, 61-67 (1967).

¹⁶ A-L Barabasi and R Albert, *Science* **286**, 509-512 (1999).

The fractal nature of social networks informs the probability of encounters between individuals in a way that leads to *power-law kinetics* modeled as

$$\dagger(t) \propto t^{\alpha}$$

where α is the power-law exponent that is greater than one. While the derivation of the power-law time dependence is challenging, the assessment of whether we have power-law kinetics is quite simple. In evaluating data for exponential growth, we plot the data on a log-linear scale. If the data falls on a line it follows *exponential kinetics*. In evaluating the data for power-law growth, we plot the data on a log-log scale. If the data falls on a line it follows *power-law kinetics*. Let's apply this analysis to the China data.¹⁷

Shown below are the coronavirus related mortality statistics from China displayed on a log-log scale. The points fall on a line showing a clear power-law time-dependence for the growth in mortality.



Our analysis of data from the early rise in coronavirus related mortality suggests that predictions of exponential growth resulting from the SIR and SEIR models fail to capture essential features of the epidemic. An accurate kinetic model of the spread of the disease and the resulting mortality must reflect the underlying network of contacts between susceptible and infected individuals.

Growth kinetics reflect the social network

To get more insight into the essential features that the SEIR model is missing, let's consider a network of social contacts. Below is a plot of a *social network*.



Each red dot is an individual. Each line connecting two red dots represents a physical contact between that pair of individuals. The distance between

¹⁷ AL Ziff and RM Ziff, medRxiv 2020.02.16.20023820; doi: https://doi.org/10.1101/2020.02.16.20023820.

Figure 6: Mortality data over 25 days during an early stage in the epidemic shown on a log-log plot. The data fall on a line demonstrating that the growing number of deaths follows power-law kinetics with an exponent of $\alpha = 2.2$. Note that in the case of the exponential growth model, the fitting parameter was a time constant with units of days. In the case of power-law kinetics, the exponent is dimensionless. This is a property of *scale-free networks*.

Figure 7: A social network of 114 individuals (red dots) forming 7 clusters. The distance between dots represents social distance. Lines represent encounters between individuals formed randomly (left) or with a diminishing probability with increasing social distance (right). points represents social distance. Pairs of points that are close together indicate relationships of friends and family living in near proximity. Pairs of points that are far separated indicate stranger relationships or physical distance that makes frequent contact less likely. The points form clusters that represent groups of family members, friends, or coworkers.

Each graph shows a network of 114 individuals forming 7 clusters. The graph on the left shows 1,000 encounters between randomly chosen pairs of individuals. The graph on the right shows the same 114 individuals. However, the lines represent 1,000 encounters chosen from a gaussian distribution that favors near contacts and makes more distant contacts less probable. Intuitively, the network on the right represents a more realistic pattern of interactions between individuals reflecting the underlying structure of the social network.

Shown below are histograms of the probability distribution p(d) of contacts of varying length *d*.



Figure 8: Distributions reflecting the probability p(d) of encounter between two individuals have a social distance of *d*. The SEIR model assumes an encounter probability that is independent of social distance. The random and uniform connectivity across a population gives rise to exponential kinetics.

In the model assuming random connections, encounters between pairs of individuals within a social group are underweighted while encounters between pairs of socially distant individuals are exaggerated. In contrast, in the small world model the probability of encounter reasonably reflects the social distance between individuals. Encounters between individuals in close proximity within a social cluster are more probable than contact between individuals who are distant by being strangers or through physical distance.

The solid line represents the predictions of the SEIR model that assumes a constant probability of encounter between any two individuals. Both the SEIR model and random model assume an equal probability of encounter between any two individuals. The difference in the contact probability distribution results from the fact that the random encounter model reflects the underlying structure of the social network, consisting of separate social groups forming separate clusters, and clusters of clusters, and clusters of clusters. The self-similar fractal structure of the small world networks, combined with encounter probabilities reflecting the social distance between individuals, can give rise to power-law kinetics.

The widely applied SIR and SEIR epidemiological models predict exponential time dependence in the spread of infectious disease. However, the assumption of a constant probability of encounter between any two individuals, on which the SIR and SEIR models are based, fails to reflect the heterogeneous structure underlying social networks. In contrast, small-world networks can give rise to power-law behavior reflected in infectious disease epidemics including the initial spread of the coronavirus epidemic of 2019.

End-of-supplement exercises

E.1 Which parameters in the SIR model can be controlled by changes in behavior? Which parameters are largely independent of behavior?

E.2 Consider the data below showing predictions of the SEIR model for the time evolution of coronavirus for four values of *b* modeling the average number of contacts per day. The initial number of infected individuals is taken to be $N_I = 10$ in a population of $N_T = 4,875,390$ representing metropolitan Boston. We assume $\frac{1}{a} = 5$ days and $\frac{1}{k} = 5$ days.



Explain the observed dependence of the onset of the disease on *b*.

E.3 The reproduction number $\mathcal{R}_0 = \frac{b}{k}$ is an important measure of the growth or recession of disease. In our analysis of the SIR model we found that

$$\mathcal{R}_0 = \frac{\ln(\mathbf{s}_\infty)}{\mathbf{s}_\infty - 1} = b\frac{1}{k}$$

where $s_{\infty} = s(t = \infty)$. Below \mathcal{R}_0 is plotted over a range of s_{∞} on linear and linear-log scales. (a) Starting from the first order differential equations defining the SIR model, derive the expression above relating the reproduction number \mathcal{R}_0 to the fraction of susceptible individuals s_{∞} remaining uninfected at the end of the epidemic.

(b) As individuals become immune to a disease through exposure or vaccination, the fraction of susceptible individuals s(t) can diminish below a critical value s_c after which the disease will recede. This is known as *herd immunity*. Derive the critical fraction s_c as a function of \mathcal{R}_0 .

(c) The reproduction number \mathcal{R}_0 is shown above as a function of s_∞ using linear (left) and linear-log (right) scales. What does this dependence of \mathcal{R}_0 on the $\log_{10}(s_\infty)$ imply about the uncertainty in published values of the reproduction number \mathcal{R}_0 ? For the same relative uncertainty in s_∞ do you expect to find a larger or smaller relative uncertainty in \mathcal{R}_0 ?

E.4 Consider the data below showing predictions of the SEIR model for the time evolution of coronavirus for four values of N_I defined as the number of initially infected individuals in an overall population taken to be $N_T = 4,875,390$ representing metropolitan Boston. We assume $b = \frac{1}{2} \text{ day}^{-1}$, $\frac{1}{q} = 5 \text{ days}$, and $\frac{1}{k} = 5 \text{ days}$.



Explain the observed dependence of the onset of the disease on N_I .

E.5 The reproduction number $\mathcal{R}_0 = \frac{b}{k}$ is often calculated using the relation

$$\mathcal{R}_0 = \frac{\ln(\mathbf{s}_\infty)}{\mathbf{s}_\infty - 1} = b\frac{1}{k}$$

where $s_{\infty} = s(t = \infty)$. This requires a knowledge of the fraction of susceptible individuals who resist infection over the course of the epidemic. Recall that in the SIR model the rate of change in the fraction of infected individuals is

$$\frac{d}{dt}\mathbf{i}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - k\,\mathbf{i}(t)$$

Evaluate this expression at t^* defined as the time at which the number of infected individuals reaches a maximum. Use your result to form an alternative definition of \mathcal{R}_0 . Test the validity of your expression using data in Figure 1.

E.6 Modify the differential equations in the SEIR model to include the natural birthrate, defined by the constant p, and the natural death rate, defined by the constant d. The birthrate only impacts the number of susceptible individuals while the natural death rate impacts the populations of each of the four subgroups. Assume the birthrate is a zeroth order kinetic process proportional to the overall population while the death rate is a first order kinetic process proportional to the population of each subgroup.

E.7 It is possible to use the SIR model or SEIR model to determine the overall mortality resulting from the disease. In the SIR and SEIR models, an infected individual is assumed to recover with a rate *k*. To introduce mortality, we can assume that a fraction of those individuals who would otherwise have recovered ultimately die. We take that fraction to be

 $\mathcal{M}_0 = fraction of individuals$ who die rather than recover

where the fraction of infected individuals who recover is $1 - M_0$.

(a) Derive an expression for the cumulative number of deaths $\dagger(t)$ resulting from the disease as a function of time *t*. Your expression should depend on $\mathbf{r}(t)$, \mathcal{M}_0 , and the total population N_T .

(b) Derive an expression for the number of individuals dying per day as a function of time $\Delta(t)$. Your expression should depend on i(t).

(c) Derive an expression relating $\Delta(t)$ to $\dagger(t)$.

E.8 In the SEIR model, the number of cumulative deaths shows exponential growth

 $t(t) \propto e^{t/\tau}$

where τ is the time constant. Consider data for cumulative deaths $\dagger(t)$ plotted on a log-linear plot using the base-10 logarithm. The data can be fit to a line with slope *m*. Derive an expression relating τ to *m*.



E.9 Power-law kinetics converges more slowly than exponential kinetics. Since power-law growth continues to evolve in time after exponential growth has reached its conclusion, power-law kinetics is said to be characterized by *heavy tails* or *long time tails*. An example is shown below.

What influence will the presence of long time tails have on the progression of an infectious disease epidemic?

E.10 In a small world network, diminishing long distance contacts through practices like *social distancing* can isolate clusters of individuals from the overall network. Above the so-called *percolation threshold* there is one connected cluster of all individuals. It is possible to connect any two points through a series of encounters. Below the percolation threshold it is no longer possible to connect any two points in the network through a series of encounters. Discuss how the propagation of disease varies above and below the percolation threshold.

E.11 Consider a chemical reaction where the reactant S undergoes reaction with E to create product P. The rate of change in S is written

$$\frac{d}{dt}[\mathbf{S}](t) = -k[\mathbf{S}](t)[\mathbf{E}](t)$$

where [S] and [E] are concentrations of S and E. We assume $[E](0) \gg [S](0)$ so that $[E](t) \simeq [E](0)$ is approximately constant in time. The solution for the time dependence of the concentration of S is

$$[S](t) = [S](0)e^{-k't}$$

where k' = k[E](0). This so-called *pseudo-first order kinetics* is often observed in chemical reactions.

The mechanism of reaction involves collisional encounters between S and E that lead to reaction, much like contact encounters between susceptible and infected individuals leads to reaction and a decrease in the number of susceptible individuals. Why might one expect exponential kinetics to be valid for chemical kinetics, in which reactions occur through collisional encounters between molecules, but not for the early spread of coronavirus in China?

E.12 The equations defining the SIR model

$$\frac{d}{dt}\mathbf{s}(t) = -b\,\mathbf{s}(t)\mathbf{i}(t) \qquad \frac{d}{dt}\mathbf{i}(t) = b\,\mathbf{s}(t)\mathbf{i}(t) - k\,\mathbf{i}(t) \qquad \frac{d}{dt}\mathbf{r}(t) = k\,\mathbf{i}(t)$$



Figure 9: Comparison of the convergence of models of exponential and power-law growth. Note the rapid convergence of the exponential model compared with the long time tail of the power-law.

define the rate of change in the fraction of susceptible s(t), infected i(t), and recovered r(t) individuals at a given time t. Imagine a molecular reaction where s(t), i(t), and r(t) represent molecular species. Define the molecular species.

E.13 The models we have considered capture essential qualitative features of an infectious disease epidemic. At the very least, they allow us to organize our thinking about critical factors in the spread of infectious disease. Nevertheless, they are qualitative models at best. Identify a short-coming of one of the models and suggest how the model can be improved.