

Review of (Compartmental) Models for Infectious Disease

Evan L. Ray (and anyone else)

1 Introduction

In this document we review basic formulations of models used for infectious disease and inference techniques used to estimate the model parameters and make predictions. These models and estimation strategies can be described by the following five factors:

1. Mechanistic vs. *** (what's a good term?):
 - Mechanistic models describe the mechanisms governing the dynamics of the disease. These models can be further divided into agent-based or compartmental models:
 - Agent-based models describe interactions between individuals within a population.
 - Compartmental models divide the population into large groups according to their disease status, and model the number of people in each disease status category over time.
 - *** models do not explicitly seek to describe mechanics of disease transmission. Instead, they more directly model the number of observed cases at each time function as a function of covariates (possibly including lagged case counts).
2. Temporal framework: Models may work with either
 - continuous time; or
 - discrete time.
3. Spatial scale: Models may describe either
 - dynamics for a single geographic region, such as a country; or
 - dynamics for multiple interconnected geographic regions, such as provinces within a country.
4. Number of serotypes: Models may describe either
 - dynamics for a single serotype; or

- dynamics for multiple interacting serotypes.
5. Measurement process: Models may incorporate aspects of the measurement process such as
- underreporting of case counts
 - delays in reporting.

In this review, we focus on the compartmental variety of mechanistic models. A central challenge for estimation of the parameters in these models is that the number of individuals in each disease status category may be either completely unobserved or censored. This limitation of the observed data is addressed in different ways by different models and estimation strategies. We list several estimation strategies have been developed for these and similar models here:

- two-stage procedures: the first stage estimates the unobserved quantities which are then used as inputs to the second stage. These procedures can be used to obtain approximate maximum likelihood parameter estimates(?)
- nonlinear programming to maximize an approximation of the likelihood in continuous time models
- particle filtering/sequential Monte Carlo: this is a method for performing Monte Carlo integration over all possible values of unobserved variables. This is a computationally demanding method, but can be used to obtain maximum likelihood parameter estimates (?)
- There is an MCMC technique used in maximum likelihood estimation of network model parameters in which the MCMC is used to integrate over all possible network structures – perhaps something like that could be applied in this setting?
- Bayesian or frequentist estimation of splines that approximate the underlying dynamic process – applied to PDE’s from physics context, perhaps could also be applied in this setting?

This review is organized as follows. We discuss variations on the model formulation in Section 2, drawing exclusively from the literature on modeling infectious disease. We then discuss estimation strategies for these models in Section 3; here, we draw from the broader literature on estimation of the parameters in similar dynamical models.

Incompletely categorized papers: These should probably be discussed in more places than they currently are

- Finkenstädt and Grenfell (2000)
- Ionides et al. (2006)
- King et al. (2008)
- Reich et al. (2013)
- Shrestha et al. (2011)
- Word et al. (2010)
- Word et al. (2012)

2 Model Formulation

In this Section we discuss variations that have been proposed for compartmental models of infectious disease. We begin by giving an overview of the Susceptible-Infected-Recovered (SIR) model in Subsection 2.1. As a part of that discussion, we discuss specific formulations of the SIR model for discrete and continuous time. The SIR model is the foundation for many of the other models that have been used for infectious disease. In the following three Subsections, we discuss modifications to the basic SIR model that have been proposed to address the remaining factors that we outlined in Section 1 describing the model formulation. Specifically, we discuss the spatial scale in Subsection 2.2, the number of serotypes in Subsection 2.3, and the measurement process in Subsection 2.4.

2.1 SIR Model

In this Section, we give an overview of the SIR model and discuss its most common formulations in discrete and continuous time. In this model, at each point in time each individual in the population is categorized as either susceptible to the disease, infected with the disease, or recovered from the disease. The model tracks either the total number of individuals or proportion of the population in each disease status category, and allows for either deterministic or stochastic changes in these population totals or proportions over time. The model is illustrated in Figure 1.

Following Keeling and Rohani (2008), we use the following notation:

- $X = X(t)$ = number of susceptible individuals at each point in time
- $Y = Y(t)$ = number of infected individuals at each point in time
- $N = N(t)$ = total population size at each point in time

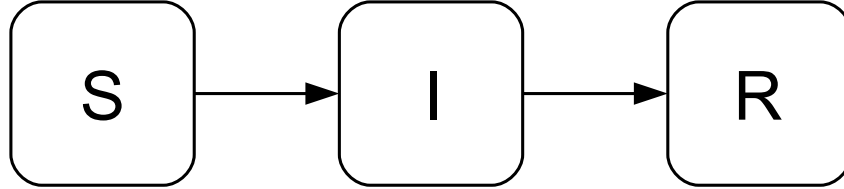


Figure 1: A diagram of the basic SIR model: individuals move from susceptible to infected to recovered.

- $S = S(t) = X(t)/N(t)$ = proportion of the population who are susceptible at each point in time
- $I = I(t) = Y(t)/N(t)$ = proportion of the population who are infected at each point in time

The proportion of the population who have recovered from an infection at time t is $1 - S(t) - I(t)$.

To do:

- Diagram of SIR
- Diagrams of common minor extensions still operating on one geographical unit and one serotype – SEIR, births, etc.
- Define/distinguish between discrete and continuous time models, say these will be discussed in Subsections
- cite Keeling and Rohani (2008)

2.1.1 Discrete Time Formulation of the SIR Model

To do:

- everything
- justify form of equation for I and discuss alternatives
- cite Finkenstädt and Grenfell (2000)

Finkenstädt and Grenfell (2000) model transmission of measles and base their work on the SEIR model. This model has four components: **S**usceptible, **E**xposed, **I**nfected, and **R**ecovered. The assumed transitions between these

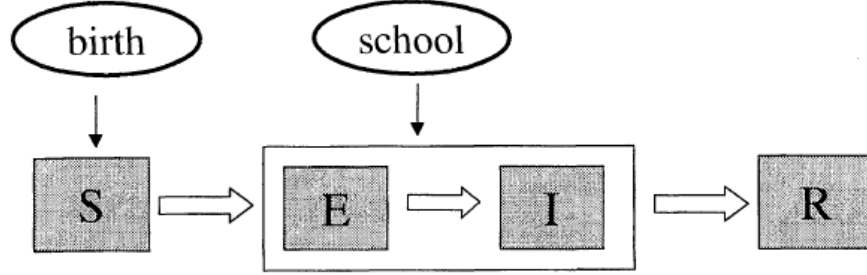


Figure 2: A reproduction of Figure 1 from Finkenstädt and Grenfell (2000). Original caption: “Flow diagram of the SEIR compartmental model: S, susceptible; E, exposed; I, infected; R, recovered”

categories are shown in Figure 2, which reproduces Figure 1 from their work. They use the following set of equations to model the infected and susceptible populations over time:

$$I_t = r_t I_{t-1}^{\alpha_1} S_{t-1}^{\alpha_2} \epsilon_t \quad (1)$$

$$S_t = B_{t-d} + S_{t-1} - I_t + u_t \quad (2)$$

The model works with data aggregated to a biweekly scale, and previous work indicates that it takes 2 weeks for an infected individual to recover from measles. This means that (by assumption), all infected individuals at time $t - 1$ are recovered at time t . The authors also incorporate a mechanism for underreporting of the number of infected individuals at each time point, with a time-varying reporting rate.

2.1.2 Continuous Time Formulation of the SIR Model

In continuous time, the SIR model consists of a system of (stochastic) differential equations specifying the rate of change of the quantities S and I as a function of their current values and some unknown parameters. There are two commonly used formulations with different infection rates, based on different assumptions: frequency-dependent transmission and density-dependent transmission. We begin with a discussion of frequency-dependent transmission, and then turn to density-dependent transmission.

The frequency-dependent transmission structure can be obtained by making the following assumptions:

1. **Homogeneous mixing:** All individuals in the population contact each other with equal probability
2. **Homogeneous contact rate:** The rate of contacts with other individuals is constant across individuals and time

3. **Homogeneous transmission probability:** The probability of transmission of the disease from infected individuals to susceptible individuals per contact is constant across individuals and over time
4. **Independence of disease transmission:** Transmission of the disease from different contacts is independent.

Under these assumptions, we can derive the rate at which individuals move from the susceptible population to the infected population in a deterministic model of the disease dynamics. Our discussion follows that in Keeling and Rohani (2008) closely. We begin by setting κ equal to the number of contacts per unit time that each individual experiences; this quantity is constant across individuals and time according to Assumption 2 above. We next set c equal to the probability of disease transmission to a susceptible individual following a single contact; this quantity is constant across individuals and time according to Assumption 3.

Our overall strategy for finding $\frac{dS(t)}{dt}$ follows four steps:

1. First, we find the probability ρ_δ that one particular susceptible individual becomes infected over a short time interval $[t^*, t^* + \delta]$.
2. Second, we extend this probability to the population of all susceptible individuals at time t^* to obtain an expected change in the number of infected individuals over the given time interval.
3. Third, we divide this expected change in the number of infected individuals by the width of the time interval to obtain an expected average rate of change in the number of infected individuals per unit time over the given time interval.
4. Finally, we let the width δ of the time interval go to 0 to obtain the expected instantaneous rate of change in the number of infected individuals at the time t^* .

Let \mathcal{S}_{t^*} be the set of all susceptible individuals at time t^* , which has cardinality $X(t^*)$. For each individual $i \in \mathcal{S}_{t^*}$, let

$$Z_{i,t^*,\delta} = \begin{cases} 0 & \text{if individual } i \text{ is still uninfected at time } t^* + \delta, \\ 1 & \text{if individual } i \text{ is infected at time } t^* + \delta. \end{cases}$$

We are interested in finding $E(Z_{i,t^*,\delta}) = P(Z_{i,t^*,\delta} = 1)$. As a preliminary computation, we find the number of contacts with infected individuals that a susceptible individual engages in during the specified time interval from t^* to $t^* + \delta$:

$$\kappa \frac{\text{contacts}}{\text{unit time}} \times \delta \text{ units of time} \times \frac{\{I(t^*) + \varepsilon(t^*, \delta)\} \text{ contacts with infected individuals}}{\text{contact}}$$

This computation makes use of assumption 1 above to obtain the third term equating the proportion of the population who are infected with the number of contacts with infected individuals per contact. The error $\varepsilon(t^*, \delta)$ is an adjustment for the fact that the proportion of the population who are infected does not remain constant over the whole time span from t^* to $t^* + \delta$.

Suppose that the number of contacts with infected individuals, $\delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\}$, is an integer. Then by the independence of Assumption 4, we can calculate

$$\begin{aligned}
P(Z_{i,t^*,\delta} = 0) &= P(\text{Individual } i \text{ not infected after } \delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\} \\
&\quad \text{contacts with infected individuals}) \\
&= \prod_{k=1}^{\delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\}} P(\text{Individual } i \text{ not infected after } k\text{th} \\
&\quad \text{contact with infected individual}) \\
&= \prod_{k=1}^{\delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\}} (1 - c) \\
&= (1 - c)^{\delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\}} \tag{3}
\end{aligned}$$

By analogy with the integer case, we also use the result in Equation (3) if $\delta\kappa\{I(t^*) + \varepsilon(t^*, \delta)\}$ is non-integer. This is intuitively plausible, but disagrees with the results we might get if we were a little more careful about thinking about randomness. We return to this issue later.

Now we define $\beta = -\kappa \log(1 - c)$. Then, continuing from Equation (3), we have

$$\begin{aligned}
P(Z_{i,t^*,\delta} = 0) &= \{(1 - c)^\kappa\}^{\delta\{I(t^*) + \varepsilon(t^*, \delta)\}} \\
&= \{\exp(-\beta)\}^{\delta\{I(t^*) + \varepsilon(t^*, \delta)\}} \\
&= \exp[-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}] \tag{4}
\end{aligned}$$

This concludes step 1 of the derivation.

For step 2, we note that the total number of susceptible individuals at time t^* who are infected by time $t^* + \delta$ is given by $\sum_{i \in S_{t^*}^*} Z_{i,t^*,\delta}$. Thus, the expected

number of susceptible individuals who become infected by time $t^* + \delta$ is

$$\begin{aligned}
E\left(\sum_{i \in \mathcal{S}_t^*} Z_{i,t^*,\delta}\right) &= \sum_{i \in \mathcal{S}_t^*} E(Z_{i,t^*,\delta}) \\
&= \sum_{i \in \mathcal{S}_t^*} P(Z_{i,t^*,\delta} = 1) \\
&= \sum_{i \in \mathcal{S}_t^*} \{1 - P(Z_{i,t^*,\delta} = 0)\} \\
&= \sum_{i \in \mathcal{S}_t^*} [1 - \exp[-\beta\delta\{sI(t^*) + \varepsilon(t^*, \delta)\}]] \\
&= X(t^*) [1 - \exp[-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}]]
\end{aligned}$$

For step 3, we divide this expected change in the number of infected individuals by the width of the time interval to obtain an expected average rate of change in the number of infected individuals per unit time over the given time interval:

$$\frac{X(t^*) [1 - \exp[-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}]]}{\delta} \quad (5)$$

For step 4, we take the limit of (5) as the interval width δ goes to 0:

$$\begin{aligned}
&\lim_{\delta \rightarrow 0} \frac{X(t^*) [1 - \exp[-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}]]}{\delta} \\
&= \lim_{\delta \rightarrow 0} \frac{X(t^*) \left[1 - \sum_{k=0}^{\infty} [-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}]^k / k!\right]}{\delta} \\
&= \lim_{\delta \rightarrow 0} \frac{X(t^*) \left[-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\} - \sum_{k=2}^{\infty} [-\beta\delta\{I(t^*) + \varepsilon(t^*, \delta)\}]^k / k!\right]}{\delta}
\end{aligned}$$

If $\varepsilon(t^*, \delta) \in o(\delta)$, in the limit we obtain

$$-\beta X(t^*) I(t^*) = \frac{-\beta X(t^*) Y(t^*)}{N(t^*)}$$

, By assumption 4, the expected number of susceptible individuals who become infected over that time interval is that probability times $X(t^*)$

We consider the probability

To do:

- everything
- cite Word et al. (2010, 2012)

2.2 Spatial Scale

To do:

- think about the name of this section? I think what I want is number of geographic units modeled, not the size of those units.
- find and discuss papers about this?

2.3 Number of Serotypes

To do:

- everything
- cite Shrestha et al. (2011); Reich et al. (2013)

2.4 Measurement Process

To do:

- everything
- underreporting – cite Finkenstädt and Grenfell (2000), Krzysztof?

3 Estimation

In this Section, we describe estimation strategies for the SIR and related models. We discuss two-stage estimation procedures in Subsection 3.1, nonlinear programming in Subsection 3.2, sequential Monte Carlo in Subsection 3.3,

3.1 Two-Stage Estimation

In order to perform estimation, Finkenstädt and Grenfell (2000) use a two-step process:

1. First, they use Equation (2) to obtain an estimate of the deviations of the size of the susceptible population from its long-term mean at each time point. This step works by rearranging Equation (2) to express the cumulative births at each time point as a function of the cumulative observed cases, deviations in the reporting rate from the overall average reporting rate, deviations of the size of the susceptible population at time t from its overall mean, and a random noise term. They then use locally linear

regression to approximate this function locally in t . The terms in this approximation do not uniquely identify all of the quantities in the model; in particular, they use the residuals from this regression as estimates of the desired deviations, but it appears that the deviations of the size of the susceptible population from their overall mean are conflated with the random noise term. This would mean that the estimates used are valid only when the random noise term is approximately 0 (or has very low variance).

2. Second, they combine the estimated deviations from step 1 with Equation (1) to estimate the parameters governing infection dynamics over time.

In addition to merging together the noise term with fluctuations in the state variable (perhaps inevitable?), it seems like this method loses track of uncertainty and possibly introduces bias by using a single value for the size of the state variable rather than integrating over its distribution?

3.2 Nonlinear Programming (may rename this?)

To do:

- everything
- cite Word et al. (2010, 2012)

3.3 Sequential Monte Carlo

To do:

- everything
- cite Ionides et al. (2006); Shrestha et al. (2011) (?)

3.4 Maximum Likelihood MCMC

To do:

- Read Geyer and Thompson (1992), Section 5 of <http://www.jstatsoft.org/v24/i03/paper> and <http://arxiv.org/abs/1208.0121>
- Decide if it's worth including this method in this review; if so, write it up.

3.5 Estimation of Splines used to Represent Dynamic Process

To do:

- Rename this Subsection?
- Discuss Xun et al. (2013) (and references it contains?)

References

- Finkenstädt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49(2), 187–205.
- Geyer, C. J. and E. A. Thompson (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 657–699.
- Ionides, E., C. Bretó, and A. King (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 103(49), 18438–18443.
- Keeling, M. J. and P. Rohani (2008). *Modeling infectious diseases in humans and animals*. Princeton, New Jersey: Princeton University Press.
- King, A. A., E. L. Ionides, M. Pascual, and M. J. Bouma (2008). Inapparent infections and cholera dynamics. *Nature* 454(7206), 877–880.
- Reich, N. G., S. Shrestha, A. A. King, P. Rohani, J. Lessler, S. Kalayanarooj, I.-K. Yoon, R. V. Gibbons, D. S. Burke, and D. A. Cummings (2013). Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of The Royal Society Interface* 10(86), 20130414.
- Shrestha, S., A. A. King, and P. Rohani (2011). Statistical inference for multi-pathogen systems. *PLoS computational biology* 7(8), e1002135.
- Word, D. P., D. A. Cummings, D. S. Burke, S. Iamsirithaworn, and C. D. Laird (2012). A nonlinear programming approach for estimation of transmission parameters in childhood infectious disease using a continuous time model. *Journal of The Royal Society Interface* 9(73), 1983–1997.
- Word, D. P., J. K. Young, D. Cummings, and C. D. Laird (2010). Estimation of seasonal transmission parameters in childhood infectious disease using a stochastic continuous time model. *Computer Aided Chemical Engineering* 28, 229–234.

Xun, X., J. Cao, B. Mallick, A. Maity, and R. J. Carroll (2013). Parameter estimation of partial differential equation models. *Journal of the American Statistical Association* 108(503), 1009–1020.