A Short Introduction to Epidemiology Modelling

Paul Simanjuntak

This material is mostly taken from the collection of lecture notes "Mathematical Epidemiology", mostly from Chapter 2 and 3. We will approach the modelling in two ways: first, we will consider the deterministic model governed by some system of ODE. Later, we will use similar framework, but we will replace the ODE system with a random process. Last, we will look at an actual recent paper modelling the recent events.

While pandemics have been with us since the beginning of civilization (think about Moses and the Egyptian plagues), the effort to predict it is relatively recent. For example, facing the Great Plague of 1665 (the one when Isaac Newton stayed at home and, unlike us in the last couple months, made the greatest intelectual progress in human history), the best the society could do back then was to isolate the sick and hope for the best, which only has limited use given that most people live in such unsanitary condition (really, a lot of deaths back in those days could be prevented had they know more about hygiene).

The first great progress, which we're still building upon until today, is made in the early 20th century, when what we call now the compartmental model was introduced. The idea is simple: given a target population (in a city, state, or country, say), divide everyone into mutually exclusive groups and then describe the interaction between these groups. The simplest of such models was first introduced by Kermack and McKendrick in 1927 and is called the SIR model.

We will now describe the SIR model. We begin by classifying everyone into 3 groups: susceptible (S), infected (I), and removed (R). The first two groups are self-explanatory, but we should put a word about the third: when we say removed, we mean either the individual has recovered and is now immune to the disease (at least in short term), the person is isolated and is now of no risk to the target population, or the person is dead. Thus, the pandemic evolution is a one-way street, because one can only moves from one compartment to the next without going back.

Now let's describe the interaction between these compartments. The independent variable in this model is the time t and the rate of interaction between the compartments. If we write $S(t)$, $I(t)$, and $R(t)$ to be the number of individuals in each compartment at time t, then in this model the interaction is governed by

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \alpha I \end{aligned} \tag{1}$$

$$R' = \alpha I$$

where α and β are the interaction coefficients. This model operates under the following assumptions:

(a) An individual can be in contact with anyone else (no local network).

(b) In particular, an average member of the class I makes enough contact to transmit βN others per unit time, where N is the total population.

(c) Infectives leave the class I at the rate αI per unit time.

(d) There is no change in the whole population, except possibly through death.

Let's explore the meaning behind these assumptions. Since anyone can interact with anyone else, the probability that a random infective (in class I) will interact with a susceptible (in class S) is $\frac{S}{N}$. Because we assume that the infectious individual will make enough contact to keep up the rate, then at every unit time we can expect

$$\underbrace{\beta N}_{\text{Number of infections}} \cdot \underbrace{\frac{S}{N}}_{\substack{\text{Probability an infective} \\ \text{meets a susceptible}}} \cdot \underbrace{I}_{\text{Number of infectives}} = \beta SI$$

This explains the first equation in our system. The third assumption simply says of those infected at the same time, the number of this 'cohort' remaining in class I is exponentially decreasing. The last means that the model's timescale is much shorter than the time needed for meaningful demographic change. Later we may discuss what happens if we add some parameter regarding birth or immigration.
Now that we understand the model, we can think about solving it. Of course, by our fourth assumption, once we know S and I we automatically get R. Then we can rewrite (1) as

$$S' = -\beta SI \tag{2}$$
$$I' = (\beta S - \alpha)I \tag{3}$$

Simple as this model is, however, it is not possible to solve this analytically (i.e. there is no closed form solution). That said, we can benefit by analyzing the features of the solution.
If we look at (3), we can see that the pandemic will only progress if $S > \frac{\alpha}{\beta}$. However, S is always decreasing. Thus, if initially we $\frac{\alpha S(0)}{\beta} > 1$, I will increase. Later S will decrease enough that the $S < \frac{\alpha}{\beta}$ and the pandemic winds down. This kind of behaviour is not surprising, as it is what's generally observed (see below for example).

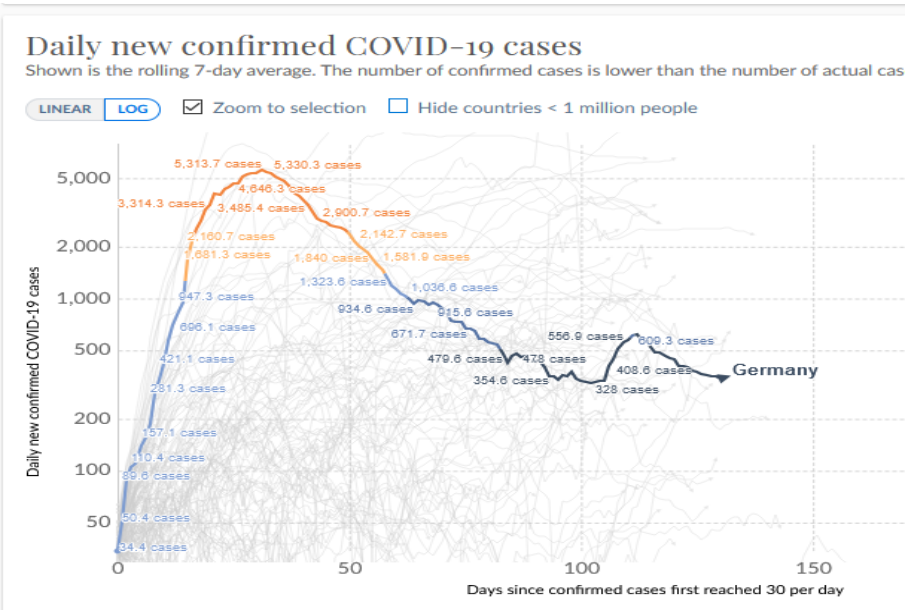

Figure 1: Number of Cases in Germany up to 14 July 2020. Source: Our World in Data

As a side note, the number $\frac{\alpha S}{\beta}$ is what is called the reproduction number $\mathcal{R}$ (initially it is the basic reproduction number $\mathcal{R}_0$). Recently, we often see it cited as a measure of how successful the surpression methods are, but really it is a measure of how fast the epidemic is growing: if we measure introduce a single infective in the population, how many secondary infection can we expect to occur?

Instead of solving S and I as a function of time, what we can do divide (3) with (2) and try to get a relation between S and I. Thus we will get

$$\frac{I'}{S'} = \frac{dI}{dS} = \frac{(\beta S - \alpha)I}{-\beta SI} = -1 + \frac{\alpha}{\beta S}$$

Integrating this equation with respect to S gives us

$$I = -S + \frac{\alpha}{\beta} \log S$$

In other words, we have a family of solution given by the level curves of the function

$$V(S, I) = S + I - \frac{\alpha}{\beta} \log S$$

In the theory of ODE, this function V is called the Lyapunov function of the system. The existence of this function gives us an easy'ish way to check if we are guaranteed some sort of stable solution.

Theorem 0.1. *Suppose the system*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \qquad x \in \mathbb{R}^n$$

has an isolated critical system at the origin. If there exists a continuous scalar function V with continuous first partial derivatives, and there is a bounded domain D containing the origin where V is is bounded, positive definite, and $\dot{V}$ is negative definite on D, then every solution of the system that starts at D is asymptotically stable.

In other words, given a solution (S, I), we can always get a solution as we vary the initial condition (so long $S > 0$ and $I > 0$) and $(S, I) \to 0$ in the limit $t \to \infty$.
To verify that we have a stable solution is now simple. The Lyapunov function is known and it's not hard to see that $V > 0$ (for values of S in our simulation). It remains to show that V is negative definite. Indeed, using (2) and (3),

$$-\nabla V \cdot (S', I') = -2\beta SI < 0$$

Thus theorem guarantees us stable solutions that approaches equilibrium as $t \to \infty$. Now consider that when in a population of size K, we introduce a small number of infectives, i.e. $S(0) \approx K$, $I(0) \approx 0$, and $\mathcal{R}_0 = \dfrac{\beta K}{\alpha}$. Call S_∞ to be the limit of $S(t)$ as $t \to \infty$. Because $I(t) \to 0$ and $V(S(0), I(0)) = V(S_\infty, 0)$ (because they belong to the same curve of solution), we must have

$$K - \frac{\alpha}{\beta} \log S(0) = S_\infty - \frac{\alpha}{\beta} \log S_\infty$$

Thus

$$\frac{\beta}{\alpha} = \frac{\log S(0) - \log S_\infty}{K - S_\infty}$$

which we can also write as

$$\log S(0) - \log S_\infty = \mathcal{R}_0 \left(1 - \frac{S_\infty}{K}\right)$$

As the right side is finite (because $\mathcal{R}_0 > 0$ and $0 \leq S_\infty \leq K$), we have the left side is also finite, so $S_\infty > 0$. In other words, we can expect that some portion of the population will escape the pandemic, regardless of the dynamics.

One problem (among others) in this basic SIR model is about the contact rate. In Assumption (b), we assume that an infective will make enough contact to keep the rate up. This is not always realistic because we tend to meet the same small number of people, unless we work in a people-facing jobs such as a grocery store cashier. It should make more sense now to scale the contact rate β with N: suppose an average

member of the population makes $C(N)$ contacts per unit time and $C'(N) \geq 0$. Define then the new contact rate

$$\beta(N) = \frac{C(N)}{N}$$

Because we consider when an individual's contacts hardly change even when the population is large (i.e. we have about the same number of friends and families regardless of where we live), we assume that $\beta'(N) \leq 0$. Note that the first model is a simple case when $C(N) = \beta N$, while in actuality we're probably closer to $C(N) = \lambda$. Empirical data suggest that in a moderate population size (e.g. Columbia), $C(N) = \lambda N^{0.05}$.

Because the model now includes explicit expression of N, we must then separate the possibilities in the removed class. Suppose that of those in the class I, a fraction f of them will die. Then the system now can be written as

$$\begin{aligned} S' &= -\beta(N)\ S\ I \\ I' &= \beta(N)\ S\ I - \alpha I \\ R' &= -f\alpha I \\ N' &= -(1-f)\alpha I \end{aligned} \tag{4}$$

Of course either the third or fourth equation say sthe same thing because $N = S + I + R$. Again, we can define the reproduction number at a given time by

$$\mathcal{R} = \frac{\beta(N) \cdot S}{\alpha}$$

The observation is the same as before: if $\mathcal{R} < 1$ for all large t, then the epidemics will wind down. Let us now calculate the rate of change of $\mathcal{R}$: by product rule,

$$\begin{aligned} \frac{d}{dt}\mathcal{R} &= \frac{S'(t) \cdot \beta(N) + S(t) \cdot \beta'(N) \cdot N'(t)}{\alpha} \\ &= \frac{(-\beta(N))^2 \cdot SI - S \cdot (1-f)\alpha I \cdot \beta'(N)}{\alpha} \\ &\leq \frac{\beta(N) \cdot SI}{\alpha}\left(\beta(N) - \frac{(1-f)\alpha}{N}\right) \end{aligned}$$

because $\beta(N) + N \cdot \beta'(N) \geq 0$ (this is just $C'(N) \geq 0$). Thus $\mathcal{R}' < 0$ is $N\beta(N) < \alpha(1-f)$, or, after plugging back in the expression for R, if $\mathcal{R} < 1$. In other words, once $\mathcal{R}$ dips below 1, it will stay there and the pandemic will die off. This suggests that even on this model, we should expect to see similar wave shape as in the simpler case.

We can of course make the model more complicated. Chapter 2 in the source material contains, for example, models that incorporate incubation time, quarantine and

isolation, and demographic features such as birth and death. We will not cover those models here.

Now that we have a basic understanding of the deterministic model, we can move on to the probabilistic models. To simplify matters, we will consider an even easier compartmental model than before. We'll take what is called the SIS model, where there are only two compartments: susceptible (S) and infected (I). An individual may get an in contact with an infective, become infected, recover, but is not conferred a lasting immunity. This model has been used to study sexually transmitted diseases (which, before the arrival of current pandemic, is where a lot of energy in epidemiology is spent).

In a non-probabilistic world, we can write (analogous to the previous SIR model) the ODEs governing this model by

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \alpha I \\ \frac{dI}{dt} &= \beta SI - \alpha I \end{aligned} \tag{5}$$

where $\beta > 0$ is the contact rate and $\alpha > 0$ is the recovery rate. Our assumptions are similar as before:

(a) An individual can be in contact with anyone else (no local network).

(b) In particular, an average member of the class I makes enough contact to transmit βN others per unit time, where N is the total population.

(c) Infectives leave the class I at the rate αI per unit time.

(d) The total population is constant at N.

with the basic reproduction number defined as

$$\mathcal{R}_0 = \frac{\beta}{\alpha}$$

Arguing similarly as before, we know that this number $\mathcal{R}_0$ controls the course of the epidemic: if $\mathcal{R}, \leq 1$, then in the long run (i.e. when $t \to \infty$) there will be no infective left. On the other hand, if $\mathcal{R}_0 > 1$, we'll approach an endemic equilibrium (and yes, it'll be an equilibrium by the previous reasoning) where

$$S(\infty) = \frac{N}{\mathcal{R}_0} \qquad I(\infty) = N\left(1 - \frac{1}{\mathcal{R}_0}\right)$$

We'll now consider when we consider the same SIS model probabilistically. Again, consider discrete time steps Δt. We'll assume that Δt is small enough that at any time step, the number of infectives change by at most 1. The number of infectives at time t, $I(t)$, is now a random variable with density

$$p_i(t) = \mathbb{P}\ (I(t) = i)$$

Because $p_i(t)$ is a probability density, at each time t we have

$$\sum_{i=0}^{N} p_i(t) = 1$$

When we move by a step Δt, by assumption 3 things can happen:

$$i \mapsto i+1 \qquad i \mapsto i-1 \qquad i \mapsto i$$

We can assign the probability of the random variable $I(t) = i$ transitions to $I(t+\Delta t) = j$ as $p_{ji}(\Delta t)$, which is given

$$p_{ij}(\Delta t) = \begin{cases} \beta i(N-i)\Delta t & j = i+1 \\ \alpha i \Delta t & j = i-1 \\ 1 - [\beta i(N-i) + \alpha i]\Delta & j = i \\ 0 & \text{otherwise} \end{cases}$$

So here we reinterpret β and α as a probability of being infected and recovering, respectively, on the individual level. Here the transition probability is the same for all time because the ODE system is autonomous (no explicit dependence on t). To simplify the notation, we'll write the probability of new infection as $b(i)$ and of recovery as $d(i)$ (they stand for birth and death, because we can see this process as a birth-death dynamic).

Again, note that the sum of $p_{ji}(\Delta t)$ for all j has to be equal to 1. Moreover, the probability of each case has to be in $[0, 1]$, so we have yet another constraint on our time step Δt.

For those in the audience that has taken any probability course, they might suspect that we are heading for the Markov process. That is definitely the case here. For the benefit of everyone, we'll define the Markov chain:

Definition 0.2. Let $X = \{X_0, X_1, \cdots\}$ be a sequence of $(\Omega, \mathcal{F}, \mathbb{P})$- measurable discrete random variables which take values in finite set S, called the state space. The process X is called a Markove chain if it satisfies the Markov condition:

$$\mathbb{P}\ (X_n = s \,|\, X_0 = x_0,\, X_1 = x_1, \cdots, X_{n-1} = x_{n-1}) = \mathbb{P}\ (X_n = s \,|\, X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and all $s, x_1, \cdots, x_{n-1} \in S$.

In other words, a random process is a Markov chain if its current position is determined solely by its position on the previous step. Since the structure is now much simplified, we can encode the evolution of a Markov chain in what is called the transition matrix $P = (p_{ij})_{i,j=1}^{N}$, where $p_{ij} = \mathbb{P}(X_{n+1} = j \,|X_n = i)$. This matrix is a stochastic matrix. That is,

(a) P has only non-negative entries.

(b) P has row sums equal to one.

It is not hard to see that our description of the SIS model is a Markov chain with the sequence given by $\{I(0), I(\Delta t), I(2\Delta t), \cdots\}$. The transition matrix $P(\Delta t)$ is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ d(1)\Delta t & 1-(b+d)(1)\Delta t & b(1)\Delta t & 0 & \dots & 0 & 0 & 0 \\ 0 & d(2)\Delta t & 1-(b+d)(2)\Delta t & b(2)\Delta t & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & d(N-1)\Delta t & 1-(b+d)(N-1)\Delta t & b(N-1)\Delta t \\ 0 & 0 & 0 & 0 & \dots & 0 & d(N)\Delta t & 1-d(N)\Delta t \end{pmatrix}$$

and thus if we take the distribution at time t to be the vector $p(t) = (p_0(t), \cdots, p_n(t))$, we have

$$p(t+\Delta t) = P(\Delta t)p(t) = P^{n+1}(\Delta t)p(0)$$

The SIS epidemic process has then now been completely described. Let's now consider some of its consequences. First we'll consider what happen with the moment of $I(t)$. Since

$$\begin{aligned} \mathbb{E}\,[I(t+\Delta t)] &= \sum_{i=0}^{N} i p_i(t+\Delta t) \\ &= \sum_{i=1}^{N} i p_{i-1}(t) b(i-1)\Delta t + \sum_{i=0}^{N-1} i p_{i+1}(t) d(i+1)\Delta t \\ &\quad + \sum_{i=0}^{N} i p_i(t) - \sum_{i=0}^{N} i p_i(t) b(i)\Delta t - \sum_{i=0}^{N} i p_i(t) d(i)\Delta t \end{aligned}$$

Simplifying and substituting the expressions for the b's and the d's give

$$\begin{aligned} \mathbb{E}\, I(t+\Delta t) &= \mathbb{E}\, I(t) + \sum_{i=1}^{N} p_{i-1}(t)\beta(i-1)[N-(i+1)]\Delta t \\ &\quad - \sum_{i=0}^{N-1} p_{i+1}(t)\alpha(i+1)\Delta t \\ &= \mathbb{E}\, I(t) + (\beta N - \alpha)\Delta t\, \mathbb{E}\, I(t) - \beta \Delta t\, \mathbb{E}\, I^2(t) \end{aligned}$$

where $\mathbb{E}\, I^2(t) = \sum_{i=0}^{N} i^2 p_i(t)$. It seems here that we've hit a wall, because we see that we'll need the second moment to compute the first moment. This means we might need some extra assumption on higher moments. However note that because $\mathbb{E}\, I^2(t) \geq [\mathbb{E}\, I(t)]^2$ (Lyapunov inequality), we can bound the difference equation above by

$$\frac{1}{\Delta t}(\mathbb{E}\, I(t+\Delta t) - \mathbb{E}\, I(t)) \leq (\beta N - \alpha)\mathbb{E}\, I(t) - \beta[\mathbb{E}\, I(t)]^2$$

Taking the limit as $\Delta t \to 0$, we get

$$\begin{aligned} \frac{d\mathbb{E}\, I(t)}{dt} &\leq (\beta N - \alpha)\mathbb{E}\, I(t) - \beta[\mathbb{E}\, I(t)]^2 \\ &= \beta[N - \mathbb{E}\, I(t)]\mathbb{E}\, I(t) - \alpha\mathbb{E}\, I(t) \end{aligned}$$

But note that $N - \mathbb{E}\, I(t) = \mathbb{E}\, S(t)$, so really the last expression is the right side of the differential equation (5) in our deterministic SIS model. Thus, the expectation of I is lower than predicted by the deterministic model.

We'll now make a further observation about our process. Since we can only jump one step at a time, our description of the SIS model falls under a very special (and interesting) case, the random walk model. In probability class, the story usually goes like this: suppose we have a drunk mathematician that just left the bar and cannot walk straight. How likely is it that he will get home safely (or, for a more cynical interpretation, get stuck in the gutter)?

In this story, we see that once this poor character reaches home, he'll be safe and the chain will practically end. This kind of state is called the *recurrent state*. On the other hand, if he's still out there finding his way home, he'll keep on moving. This kind of state is called the *transient state*.

In our SIS model, which state is transient and which state is recurrent? The general Markov chain theory gives the following two propositions:

Proposition 0.3. *In a Markov chain with transition matrix P, let $p_{ij}(n)$ be the ij-th entry of P^n. Then,*

(a) State i is persistent if $\sum_n p_{ii}(n) = \infty$.

(b) State i is transient if $\sum_n p_{ii}(n) < \infty$.

Proposition 0.4. *A set C of states is closed iff $p_{ij} = 0$ for all $i \in C$ and $j \notin C$. Then if C is a set of recurrent states, then C is closed.*

Because the first row of our matrix $P(\Delta t)$ has 1 as the first entry and 0 otherwise, $p_{00}(n) = 1$ for all n, so 0 is a persistent state. Moreover, because $p_{0j} > 0$ for all $j \geq 1$, the set of states $\{1, \cdots, N\}$ is not closed, it is not a set of recurrent states. In fact, none of them is, so the all non-zero states are transient.
What happens then in the limit as $t \to \infty$? Where should be land? The answer is coming from a corollary of 0.3:

Corollary 0.5. *If j is transient, $p_{ij}(n) \to 0$ for all state i.*

The proof is simple but requires generating function method, so we will omit it. In any case, combining all the fact above, we can guess (correctly) that from any starting condition $p(t)$,

$$p(t) \to (1, 0, \cdots, 0) \qquad \text{as } t = n\Delta t \to \infty$$

This result implies that in the stochastic SIS model, in contrast to the deterministic model, we always end up with the disease-free equilibrium regardless of the magnitude of $\mathcal{R}_0$. Of course, this result is asymptotic and tells us nothing about the rate of convergence (which in general is a much harder problem!). Thus we may have to wait for a long time, longer than the expected time horizon of our simulation.

In principle, quantifying such time is not very hard, as it is a well-known problem in the birth-death process. Let us consider the T_i to be the random variable that represents the time until extinction of disease if $I(0) = i$ and $\tau_i = \mathbb{E}\, T_i$. In time Δt, either i increases with probability $b(i)$, decreases with probability $d(i)$, or stays the same with the remaining probability. Thus the average time to extinction is either $\tau_{i+1} + \Delta t$ (if there is an initial increase), $\tau_{i-1} + \Delta$, or $\tau_i + \Delta t$ (with corresponding probabilities). Written as an equation, we have a difference equation

$$\tau_i = (b(i)\Delta t)(\tau_{i+1} + \Delta t) + (d(i) + \Delta t)(\tau_{i-1} + \Delta t) + [1 - (b(i) + d(i))\Delta t](\tau_i + \Delta t)$$

for $i = 1, \cdots, N$. We can rewrite the equation as

$$d(i)\tau_{i-1} - (b(i) + d(i))\tau_i + b(i)\tau_{i+1} = -1 \tag{6}$$

If we then write the matrix D as

$$\begin{pmatrix} -(b(1)+d(1)) & b(1) & 0 & \dots & 0 & 0 \\ d(2) & -(b(2)+d(2)) & b(2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & d(N) & -d(N) \end{pmatrix}$$

and $\tau = (\tau_1, \tau_2, \cdots, \tau_N)^T$, then the system of linear equations (6) can be written as $D\tau = -\vec{1}$. Since the matrix D is non-singular and diagonally dominant (see the

definition of $b(i)$ and $d(i)$), then the system has an unique solution $\tau = D^{-1}\vec{1}$. For example,

$$\tau_1 = \frac{1}{d(1)} + \sum_{i=2}^{N} \frac{b(1)\cdots b(i-1)}{d(1)\cdots d(i)}$$

As an example, if $N = 25$, $I(0) = 1$, and $\mathcal{R}_0 = 2$ ($\beta = 1$ and $\alpha = 1/2$), then $\tau \approx 300$ (as τ_i saturates for large i). If $N = 50$, the $\tau \approx 25000$. Of course here it depends on our time scale. If Δt is a day, then we expect the disease will die in ≈ 25000 days ≈ 68 years, much longer than any reasonable prediction. Thus we may see that dependending on $\mathcal{R}_0$, for all practical purpose, the disease may become endemic.

This completes our discussion. In conclusion, we have seen that we may use a compartmental model to get a reasonable prediction for the course of an epidemic. In our compartmental model, we can approach it deterministically using a system of ODE or stochastically using the theory of Markov chain. The first will perform better on the latter stage of the epidemic and with larger population size, but the first gives a closer prediction early on. There is still a lot to the theory, but to see these methods in action in the recent epidemic, an interesting example is the following paper with the SIR model.