Introducing Stochasticity to Basic Epidemic Models

Moitish Majumdar

August 14, 2020

1 Introduction

The basic SIR Model is completely deterministic. If the population size is small, or the disease has just started invading the population, stochastic effects are observed and it becomes necessary to account for this stochasticity in the models. There are three ways of introducing stochasticity. The first method involves introducing stochasticity directly into the population variables. For instance, the recorded number of infected people $Y_{recorded}$ can be written as a sum of two binomial distributions **(1)**:

$$Y_{recorded} = Bin(p_r, Y_{true}) + Bin(p_m, N - Y_{true})$$

Where p_r is the probability of correctly diagnosing an individual as infected and p_m is the probability of misdiagnosing an individual. The second method involves introducing noise directly into the parameters (such as in the transmission rate or the birth rate). The third method involves explicitly modelling the random events through random number generation, and is most prevalent.

Since stochastic models have a close relationship with integer-valued models, X,Y and Z (the total number of susceptibles, infecteds and recovered) are used instead of the proportions S,I and R.

Noise can be added to the parameters in different ways-it can either be constant, or scaled with the magnitude of the parameter. Higher variance in noise causes more negative covariance between the number of susceptibles and number of infecteds, and for large population sizes, the variance in the number of infecteds is expected to be proportional to the mean of the number of cases.

In the event-driven approach, every individual event can be modeled using random number generation, and several algorithms exist which take into consideation the cumulative rates of all possible events. However, in this approach, it is possible that the disease goes "extinct" in some simulations. Given a single infected person entering the population, it is possible to predict the probability of outbreaks using the branching process, which is dependent on the R_0 value of the particular disease. Moreover, critical community size and imports of infected individuals also play a role in determining whether a disease may die out in a population.

Individual based models are highly accurate as they consider the state of every individual in the population, but they require intensive computational power. Noise in the different parameters can interact with other heterogenous

factors, such as spatial structure and temporal forcing, to produce different effects.

To obtain an analytical method of understanding the different approaches, the Fokker-Plank equations, the Master equations and the Moment equations can be used. They give a deterministic description of the stochastic processes involved by considering the distribution over several stochastic simulations. The SIR model can also be realised as a continuous time Markov chain, and thus the Kolmogorov forward (which give the Master equations) and backward equations can be used to model future dynamics. Further, the SIR model can also be observed using stochastic differential equations, which follow a diffusion process, and they can be numerically solved using the Euler-Maruyama method.

2 Adding Noise

If $\xi(t)$ is a time series of random deviates with mean 0 and variance 1, derived from the normal distribution, and considering frequency dependent transmision terms, the SIR equations with noise become **(2)**:

$$\frac{dX}{dt} = \nu N - [\beta XY/N + f(X,Y)\xi] - \mu X$$

$$\frac{dY}{dt} = [\beta XY/N + f(X,Y)\xi] - \gamma Y - \mu Y$$

$$\frac{dZ}{dt} = \gamma Y - \mu Z$$

where ν is the birth rate, μ is the death rate and $f(X,Y)$ can either be constant or scaled according to the magnitude of the absolute values of X and Y.

If f is constant, the variance in the number of infecteds increases with the variance of noise. Moreover, the negative covariance between number of infecteds and number of suscpetibles (Y and X) also increases as the variance in the noise increases. This is because if a particular value of noise drives up the number of infected people, it has the opposite effect on the number of susceptibles and vice versa. Stochastic effects dominate farther away from the deterministic equilibrium point, and if the population size is small and the co-efficient of the noise term is large, the amplitude of oscillation about the deterministic equilibium is large, and a pure random walk is observed.

Since variance in the noise term causes negative covariance between X and Y, the product XY decreases, hence this drives down the mean number of infected people.**(3)**

However, in general, the magnitude of variability increases with increase in population size. Thus, if each event (birth, infection or death) is assumed to be a Poisson process, its mean will be equal to variance, hence $f = \sqrt{rate}$. Thus each event can be associated with a noise term: **(4)**

$$\frac{dX}{dt} = [\nu N + \sqrt{\nu N}\xi_1] - [\beta XY/N + \sqrt{\beta XY/N}\xi_2] - [\mu X + \sqrt{\mu X}\xi_3]$$

$$\frac{dY}{dt} = [\beta XY/N + \sqrt{\beta XY/N}\xi_2] - [\gamma Y + \sqrt{\gamma Y}\xi_4] - [\mu Y + \sqrt{\mu Y}\xi_5]$$

$$\frac{dZ}{dt} = [\gamma Y + \sqrt{\gamma Y}\xi_4] - [\mu Z + \sqrt{\mu Z}\xi_6]$$

Hence, there are 6 noise terms, one for each event. Even though the effects of two events may cancel each other out (such as births and deaths), the noise terms add together. Moreover, for a large population size, the effects of the noise terms are small, and the mean number of cases increases linearly with the variance in the number of cases.

Another way of introducing noise is by considering variability in the parameters, specifically the transmission rate. Such variability can occur due to climatic conditions or differences in social interaction etc. which can alter the transmission term β. Thus, the basic equations for the SIR model become: **(5)**

$$\frac{dX}{dt} = \nu N - \beta(1 + F\xi)XY/N - \mu X$$

$$\frac{dY}{dt} = \beta(1 + F\xi)XY/N - \gamma Y - \mu Y$$

Where the variability occurs due to an external force F. If stochasticity is due to heterogeneity in susceptibility, F is proportional to $X^{-0.5}$, and if it is due to variability in population, F is proportional to $Y^{-0.5}$.

3 Event-driven approaches

Event-driven approaches model demographic stochasticity, which are fluctuations in population processes due to the random nature of events at the individual level. Thus, every possible event that causes a change in X or Y must be separately considered. There are three ways to implement this approach.

In the first approach, called Gillespie's Direct Algorithm, the time untill the next event occurs is calculated using the cumulative rates of all possible events and one random number. A second random number is used to estimate *which* event occurs. The pseudocode for the algorithm runs as follows:**(6)**

1) All possible events $E_1, E_2, ..., E_n$ are labelled.

2) For each event, the rate at which it occurs is determined, $R_1, R_2, ..., R_n$.

3) The rate at which any possible event occurs is $R_{total} = \sum_{m=1}^{n} R_m$.

4)The first random number $RAND_1$, taken from a uniform distribution between 0 and 1, gives the time until next event, $\delta t = \frac{-1}{R_{total}} log(RAND_1)$.

5)The second random number from the same distribution $RAND_2$ gives the event which will occur, by setting $P = RAND_2 \times R_{total}$.

6) Event E_p occurs if:

$$\sum_{m=1}^{p-1} R_m < P \leq \sum_{m=1}^{p} R_m$$

7) The time is updated, $t \to t + \delta t$, event E_p occurs, and the algorithm is repeated from step 2 onwards for the total simulation time.

If we consider the SIS model without births or deaths, it has a simplified structure as we consider only two events: transmission $(X \to X-1, Y \to Y+1)$ and recovery, $(Y \to Y-1, X \to X+1)$. If $X + Y = N$ and:

$$\frac{dX}{dt} = -\beta XY/N + \gamma Y,$$

$$\frac{dY}{dt} = \beta XY/N - \gamma Y$$

Consider E_1, E_2 as transmission and recovery respectively, with $R_1 = \beta XY/N$ and $R_2 = \gamma Y$. Thus, $R_{total} = (\beta XY/N + \gamma Y)$. Thus, the time until next event is given by:**(7)**

$$\delta t = \frac{-\log(RAND_1)}{\beta XY/N + \gamma Y}$$

Where $RAND_1$ is derived from the uniform distribution between 0 and 1. If $RAND_2$ is the second random number from the same distribution, then transmission occurs if:

$$RAND_2 < \frac{\beta XY/N}{\beta XY/N + \gamma Y}$$

after time δt, and recovery occurs otherwise. The time is updated and the procedure is repeated.

Intuitively, in a large population, all events becomes more frequent (transmission, recovery etc) even though the per capita rates remain constant. However, in Gillespie's direct method, simulation time increases with population size. Gillespie's first reaction method reflects the fact that increase in transition rates leads to a decrease in time to next event (δt). The pseudocode for the algorithm runs as follows:**(8)**

1) Label all possible events $E_1, E_2, ...E_n$

2)For each event, rate is determined $R_1, R_2...R_n$

3)For each event m, the time until it next occurs is given by $\delta t_m = \frac{-1}{R_m} \log(RAND_m)$

4)The event E_m with the smallest δt_m is determined, and this event occurs.

5) The time is updated and event m occurs, and the procedure is repeated from step 2.

This is a slower but more intuitive method of modeling demographic stochasticity. This approach required the generation of m random numbers at every iteration, which is why it is substantially slower.

Gillespie's "Next Reaction" or "τ-leap" method provides a substantially faster algorithm, and simulation time is less affected by population size. The pseudocode for the algorithm is:**(9)**

1) Let the time increment between two events be δt, sufficiently small, and fixed. For small δt, the increments in X and Y are approximately Poisson.

2) Defining $M_T(t)$ and $M_R(t)$ as the number of transmissions and number of recoveries by time 't', let $\delta M_i = M_i(t + \delta t) - M_i(t)$, where i=T,R.

3)Then,

$$P(\delta M_T = 1|X, Y) = \beta XY/N\delta + o(\delta t)$$

$$P(\delta M_R = 1|Y) = \gamma Y \delta t + o(\delta t)$$

4) Since δM_i's are approximately Poisson, δM_T is Poisson with mean $\frac{\beta XY}{N}\delta t$ and δM_R is Poisson with mean $\gamma Y \delta t$.

5) The variables X and Y are updated:

$$X(t + \delta t) = X(t) - \delta M_T + \delta M_R$$

and

$$Y(t + \delta t) = Y(T) - \delta M_R + \delta M_T$$

The time is also updated, and the process is repeated from step 4. As depicted in (figure 1) **(10)**, the τ-leap method is substantially faster than the previous methods.

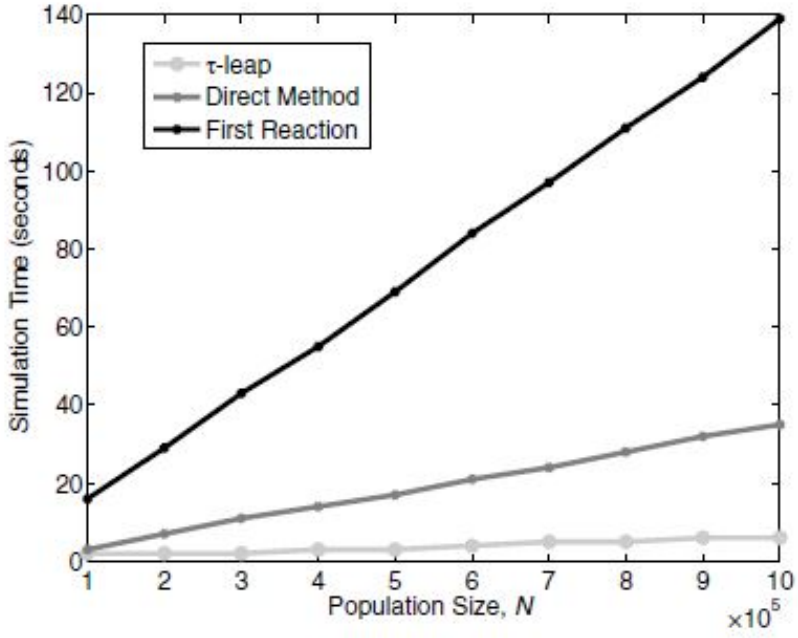

Figure 1: Time for simulation by different algorithms

4 Critical Community Size and Imports

Near the disease-free equilibrium, the branching process can be used to study the the stochastic behaviour of the disease. For the number of infectious individuals I, the disease is said to be extinct if I reaches zero and remains zero, which is considered to be an absorbing state. If the pgf of this continuous branching process is given by: **(11)**

$$f(u) = \sum_{j=0}^{\infty} p_j u^j$$

Where p_j is the probability that an infectious individual produces "j" more infectious individuals. If "u" is considered to be the probability of extinction (alternatively called the probability of a "minor outbreak"), then:

$$u = \frac{\gamma}{\gamma + \beta} + \frac{\beta}{\gamma + \beta} u^2$$

Starting with a single infectious individual, the probability that they infect nobody else is the probability of recovery, γ, and this branch dies out. However, the probability that another person is infected, is the rate of transmission, β and two infectious individuals are produced. Solving the above equation, we obtain $u = \gamma/\beta = \frac{1}{R_0}$ Thus, from a single infectious person, the probability that the disease goes extinct (or causes only a minor outbreak) is $\frac{1}{R_0}$. If "n" infectious individuals are introduced, it can be shown that the same probability becomes $\frac{1}{R^n}$ where $R = R_0 X/N$.

Thus, the risk of extinction is much lesser when R_0 is high. For endemic diseases (not the disease-free equilibrium), the only way to determine the probability of extinctions is by carrying out multiple simulations of the stochastic model. Every stochastic model thus tries to capture the Critical Community Size (CCS), which is the smallest population size not to suffer disease extinctions. Theoretically, a population of any size can suffer from disease extinction, but in reality, large population sizes do not suffer these extinctions and the disease persists.

The infectious period $1/\gamma$, has the greatest effect on CCS, and R_0 has a lesser impact. From (figure 2) **(12)**, it can be seen that for large infectious periods, the CCS is low even for small R_0 values. For smaller infectious periods, large R_0 values are required to maintain a small CCS.

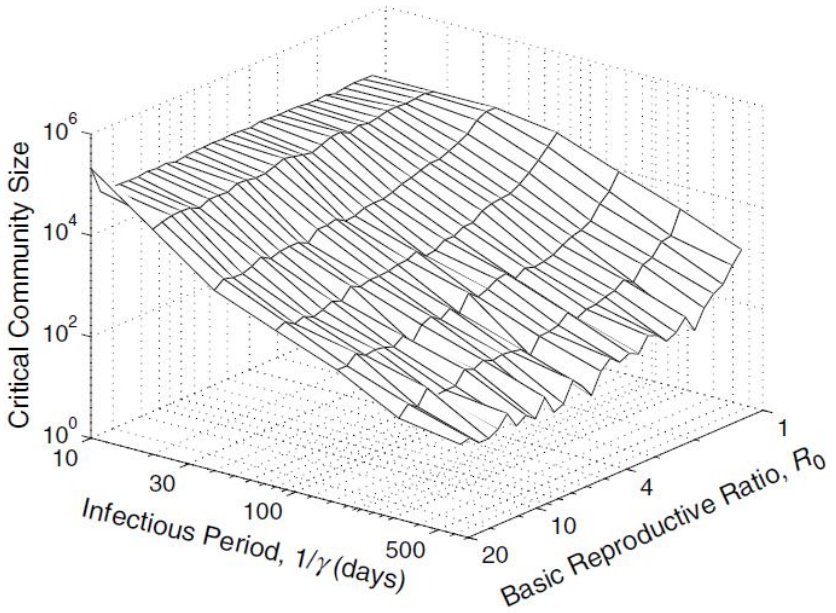

Figure 2: CCS plotted as a function of $1/\gamma$ and R_0

In general, a disease can be stopped from dying out if there are frequent imports of infectious individuals. There can be two ways imports can affect a population. Either an infectious individual can join the population, hence $Y \to Y + 1$ or a commuter/traveller can introduce the disease and leave the population, which does not affect the overall size N, but $X \to X - 1$ and $Y \to Y + 1$.

Frequent imports may prevent extinctions, and long delays between extinctions and imports may improve the chance that an import successfully triggers a large pandemic, due to an increase in the number of susceptibles.

Extinctions can give us a measure of persistence of the disease, and simulation can be started at the deterministic equilibrium. The proportion of simulations that have gone extinct after a given time can be measured. Alternatively, populations can be simulated for several generations and those simulations where the pathogen dies out can be discarded and the remaining sim-

ulations can be further simulated to ascertain the rate of extinction. The advantage of this method is that it does not have artificial starting conditions and the calculations start out due to the stochastic dynamics.

Stochastic extinction effects can also benefit vaccination strategies. The critical level of vaccination to eradicate an infectious disease

$$V_C = 1 - \frac{1}{R_0}$$

However, vaccination below the eradication threshold can reduce the average level of infection within the population, and this increases the risk of extinction.

5 Individual-based approaches and heterogeneities

In individual-based models, we know the state of each individual and hence which individuals are infectious. Thus, if the infection and recovery rates are determined for all individuals, then using the Gillespie direct algorithm, the time until the next event is:**(13)**

$$\delta t = \frac{-log(RAND_1)}{\sum_{i=S} T_i + \sum_{i=I} G_i}$$

The next event is transmission if:

$$RAND_2 < \frac{\sum_{i=S} T_i}{\sum_{i=S} T_i + \sum_{i=I} G_i}$$

and it is infection otherwise. If the next event is infection, then individual 'p' is infectious if:

$$\sum_{i=S}^{i \leq p-1} T_i < RAND_3 \times \left(\sum_{i=S} T_i \right) \leq \sum_{i=S}^{i \leq p} T_i$$

Another way of individual modeling is deciding a cumulative level of pathogen, C_i, beyond which a transmission even takes place. For every individual at birth, let:

$$C_i = -log(RAND)$$

and at every time step δt let these values decrease by $\beta Y \delta t / N$. The transmission occurs when an individual's C_i value drops below zero. The advantage of this method is that if the fate of an individual is set at birth, the random numbers will be independent of epidemic dynamics and the effect of different parameters or controls can be seen on the same stochastic epidemic process.

Noise can interact with different heterogeneities in the population. Temporal forcing occurs due to climatic conditions or due to seasonal variations like the school year. The inclusion of stochasticity can often cause the disease to resonate at its natural frequency, as the short-term dynamics dominate. Thus there is a clash between the deterministic period due to forcing and the natural period due to stochasticity, the latter dominating in smaller populations.

Another interaction is that of stochasticity with risk or age structured models, where a particular part of the population is more susceptible to the disease due to compromised immunity or behavioral patterns. These smaller groups experience high stochasticity, but extinction is rare due to a large R_0 value.

Spatially structured models also interact with stochasticity, and often a large well mixed population may experience lesser stochastic effects than a spatially structured population with smaller sub-units. Each spatial sub-unit experiences a high level of stochasticity due to its small size.

6 Analytical Methods

There are three analytical formulations that provide an intuitive understanding of the stochastic behviour of diseases.

The **Fokker-Plank equations**, linked to the addition of noise to the standard equations, replaces noise with diffusion in the form of a partial differential equation.

Suppose $P(X, Y, t)$ is the probability of a stochastic simulation having X susceptible individuals and Y infecteds at time t. In deterministic models, the probability distribution is zero everywhere except at a single point. For a general differential equation along with noise: **(14)**

$$\frac{dx}{dt} = F(x) + f(x)\xi$$

The probability distribution $P(x, t)$ is given by the partial differential equation:

$$\frac{\partial P(x,t)}{\partial t} = \frac{-\partial}{\partial t}(F(x)P(x,t)) + \frac{1\partial^2}{2\partial x^2}(f(x)^2 P(X,t))$$

By setting $\frac{\partial P}{\partial t} = 0$, we can obtain a long-term stationary distribution of values. Thus, for the SIS model:

$$\frac{dY}{dt} = [\beta XY/N + \sqrt{\beta XY/N}\xi_1] - [\gamma Y + \sqrt{\gamma Y}\xi_2]$$

$$= \beta XY/N - \gamma Y + \sqrt{\beta XY/N + \gamma Y}\xi$$

The two noise terms can be combined as the sum of two normal distributions with variances v_1 and v_2 is another normal distribution with variance $v_1 + v_2$. Thus the probability P, of having Y infectious individuals at time t is given by:

$$\frac{\partial P(Y,t)}{\partial t} = \frac{-\partial}{\partial Y}([\beta XY/N - \gamma Y]P(Y,t)) + \frac{1\partial^2}{2\partial Y^2}([\beta XY/N + \gamma Y])P(Y,t))$$

To obtain the long-term behaviour of the pathogen, an analytical expression for the equilibrium distribution $P^*(Y)$ can be found by setting $\frac{\partial P}{\partial t} = 0$

The **Master equations** (also called the Ensemble or Kolmogorov forward equations) are the integer-valued equivalent of the Fokker-Plank equations.

Let $P_Y(t)$ be the probability that there are Y infectious individuals at time t, and for the SIS model: $X + Y = N$. Thus, we do not need to explicitly keep track of the number of suscetibles. Alternatively, one can think of P_Y to the proportion of simulations that have Y infecteds (out of a large number of simulations). Four processes can occur that can alter the state of the system:

1) A simulation in state Y can have an infected individual recover at the rate γY, and now there are Y-1 infecteds.

2) A simulation in state Y can have a susceptible individual become infected at the rate $\beta(N - Y)Y/N$ and now have Y+1 infecteds.

3) A simulation in state Y+1 can have an infected individual recover at the rate $\gamma(Y+1)$ such that Y infecteds remain.

4) A simulation in state Y-1 can have a susceptible individual getting infected at the rate of $\beta(N-Y+1)(Y-1)/N$ such that there are now Y infecteds.

Thus, to obtain an explicit equation for these processes:**(15)**

$$\frac{dP_Y}{dt} = -P_Y[\gamma Y] - P_Y[\beta(N-Y)Y/N] + P_{Y+1}[\gamma(Y+1)] + P_{Y-1}[\beta(N_Y+1)(Y-1)/N]$$

This generates N+1 differential equations, with P_{-1} and P_{N+1} being zero. For the SIR model, 1/2(N+1)(N+2) equations are generated. To determine the final distribution of cases, we can set $\frac{dP_Y}{dt} = 0$. Alternatively, one can equate the "proportion of simulations" moving from Y to Y+1 with the proportion moving from Y+1 to Y. If the model is at equilibirum, these two "movements" must be equal, else some of the P_Y's would be changing.

The **moment equations** are a convenient method to calculate the effect of stochasticity on the dynamics of a disease. The process starts by considering the effects of second-order moments (variances and covariances) on first order moments (mean values): **(16)**

$$\begin{aligned}\frac{d[Y]}{dt} &= [\beta XY/N - \gamma Y] \\ &= \beta[XY]/N - \gamma[Y] \\ &= \beta[X][Y]/N + \beta Cov_{XY}/N - \gamma[Y]\end{aligned}$$

where [,] denotes the average over many simulations.

7 References

1) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **193**

2) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **194** Equation 6.1

3) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **195-197**

4) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **197** Equation 6.2

5) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **199** Equation 6.3

6) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **201**

7) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **202**

8) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **203**

9) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **204**

10) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **205** Figure 6.4

11)Allen, Linda J.S. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis *Infectious Disease Modelling* (2017)

12) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **208** Figure 6.6

13) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **217**

14) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **222**

15) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **225**

16) Keeling, Matt J., Rohani, Pejman *Modeling Infectious Diseases in Humans and Animals* Princeton University Press (2007) pp. **228**