STOCHASTIC SIMULATION

Stochastic Epidemic Models

Camille GUIBOURGÉ

Contents

1 Introduction **3**

2 Model and Simulation **3**

2.1 Continuous Time Markov Chain 4

2.1.1 Simulation 5

2.2 Mean-Field Approximation and Diffusion Approximation 6

2.2.1 Simulation 6

2.3 Diffusion Approximation 7

2.3.1 Simulation 7

3 Probability of Extinction **8**

3.1 Method 8

3.2 Results 9

3.3 Variance Reduction 9

3.4 Result 10

3.5 Alternative Starting Population 11

4 Stochastic SIR with demographic effect **11**

4.1 Susceptible 12

4.2 Infected 13

4.3 Recovered 13

4.4 Formulation 14

4.5 Simulation Stochastic SIR-d 14

4.6 Mean Field Approximation 15

5 SIR-d, Extinction and Basic Reproduction Number **16**

5.1 Extinction in SIR-d 17

5.2 Probability Estimation 17

6 Conclusion **19**

1 Introduction

Mathematical models of infectious disease dynamics play a central role in understanding how epidemics emerge, spread, and eventually die out. Among these, the Susceptible–Infected–Recovered (SIR) framework is one of the most widely studied models, as it captures the essential mechanisms of transmission and recovery while remaining analytically and computationally tractable. Depending on the modeling scale and objectives, the SIR process may be described using fully stochastic formulations, deterministic mean-field approximations, or intermediate diffusion-based models.

In this report, we investigate several formulations of the SIR model and compare their behavior through simulation. We begin with a stochastic continuous-time Markov chain (CTMC) description of the epidemic in a closed population, which explicitly accounts for the randomness inherent in individual infection and recovery events. We then study two approximations of this stochastic model: a deterministic mean-field approximation obtained in the large-population limit, and a diffusion approximation that incorporates stochastic fluctuations around the mean-field trajectory.

Beyond trajectory simulation, we focus on the probabilistic question of disease extinction. Using Monte Carlo methods, we estimate the probability that an epidemic dies out within a finite time horizon, and we show that naive Monte Carlo estimators suffer from high variance in this setting. To address this issue, we introduce a variance reduction technique based on control variates and demonstrate substantial efficiency gains. We further explore how extinction probabilities depend on initial conditions and population composition.

Finally, we extend the classical SIR model by incorporating demographic effects, including births, natural deaths, and pathogen-induced mortality. For this extended SIR-d model, we derive the corresponding stochastic transition structure, recover its mean-field approximation, and study extinction behavior in relation to the basic reproduction number. Throughout the report, numerical simulations are used to illustrate theoretical results and to highlight the qualitative differences between stochastic and deterministic epidemic models.

2 Model and Simulation

We consider a closed population of individuals of size N, partitioned into 3 categories, **Susceptible** (S), **Infected** (I), **Recovered** (R), over a finite time horizon [0, T]. Denoting S_t, I_t and R_t the number of susceptible, infected and recovered individuals respectively, we present methods for simulating the trajectories of these populations over time when a disease is spreading

as follows:

(1). Recovered individuals cannot get infected.

(2). Recovery time is an exponential process.

(3). Infection occurs only through the interaction of suceptible and infected.

Let $Y_t = (S_t, I_t, R_t)$ and note that, due to assumption (1). and (2)., Y_t can transition to a different states in only 2 ways. First, an infected individual can recover, in which case we have transition C_1:

$$C_1 : Y_t = (S_t, I_t, R_t) \to (S_t, I_t - 1, R_t + 1) = Y_{t+dt}$$

Second, a susceptible individuals may get infected, in which case, Y_t will transition as follows:

$$C_2 : Y_t = (S_t, I_t, R_t) \to (S_t - 1, I_t + 1, R_t) = Y_{t+dt}$$

Of course, the process could also not transition. In the sections that follows, we present three methods for simulating the SIR process.

2.1 Continuous Time Markov Chain

Note that, since the population is closed, it follows that $R_t = N - S_t - I_t$, and thus we can focus on $X_t = (S_t, I_t)$ to simulate the whole process. Secondly, by assumption (2), we have that at any time $t \in [0, T]$, a recovery occurs with a rate $I_t\gamma$, where γ is the rate of the recovery process. Lastly, by assumption (3), at any time $t \in [0, T]$ we have that, infection occurs at a rate of $S_t I_t \beta$, where β is the per-contact infection rate. By this setup, the process X_t is a **Continuous Time Markov Chain** (CTMC) with transition probability

$$\begin{cases} \mathbb{P}\{C_1\} = \gamma I_t dt + o(dt) \\ \mathbb{P}\{C_2\} = \beta I_t S_t dt + o(dt) \end{cases} \tag{2.1.1}$$

In order to simulate this process using (2.1.1), we use the fact that (2.1.1) describes two Poisson process counting the number of recovery and infection with intensity $\gamma I_t dt$ and $\beta I_t S_t$ respectively. The algorithm used to simulate the process is as follows:

Algorithm 1 Stochastic SIR

Input: s_0, i_0, β, γ, T
Output: The Susceptible and Infected population size at time T
Procedure:
while $t < T$ **do**
  Generate $J_1 \sim \text{Exp}(i_t s_t \beta)$
  Generate $J_2 \sim \text{Exp}(i_t \gamma)$
  if $J_1 < J_2$ **then**
    Set $t+ = J_1$
    Set $s_t- = 1 \quad \& \quad i_t+ = 1$
  else
    Set $t+ = J_2$
    Set $i_t- = 1$
  end if
end while
Return $(\{s_t\}, \{i_t\})$

We simulate two random times [1] for each Poisson process. We then update X_t according to the event associated to the lowest time, and iterate the process over the time horizon $[0, T]$. Note that we could have also used the exact formula to generate the simulation. However, this would have required to choose an appropriate number of step to ensure that $o(dt)$ is negligible and that, for modeling purposes, the sum of the transition probabilities sum to a number smaller than one, so as to ensure that only one event occurs at a time.

2.1.1 Simulation

Using a population of $N = 100$ and initial conditions $S_0 = 99$ and $I_0 = 1$, we simulate the process over a time horizon $T = 100$ and with $\gamma = 0.4$ and $\beta = 0.02$. The table below shows three simulated trajectories.

[1] These are the waiting times for a new infection or recovery.

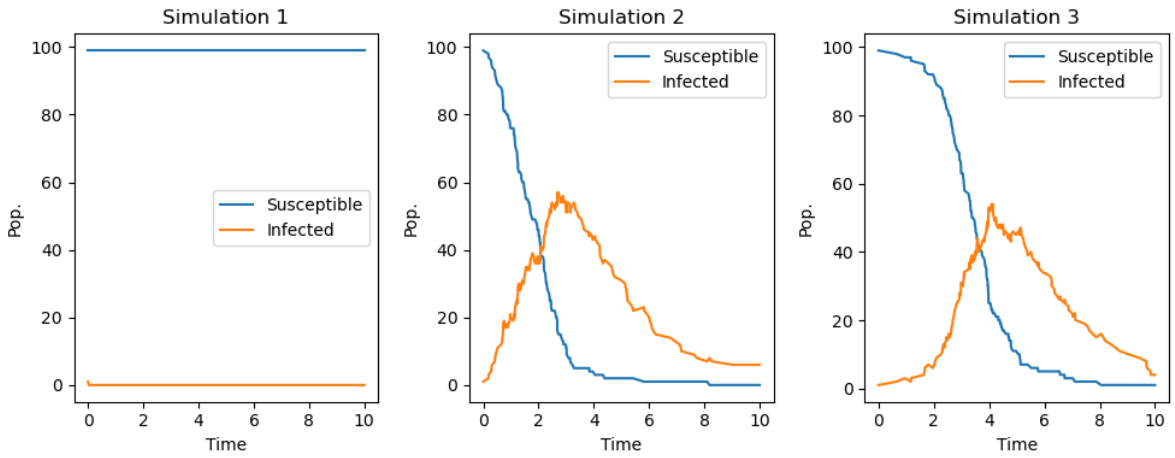


Figure 1: Stochastic SIR Simulation

In the first simulation, the disease went extinct without ever infecting anyone else. In the second and third case, the disease did not go extinct within the time horizon and in both case, the infection peak occurred at around $t = 4$, with a faster ascent to its peak in the second case.

2.2 Mean-Field Approximation and Diffusion Approximation

When the population size N is large, individual infection or recovery events have a negligible impact on the overall state of the system. In this regime, the stochastic SIR process can be approximated by a deterministic dynamical system describing the average evolution of the susceptible and infected populations.

$$\begin{cases} \frac{S_t}{dt} = -\beta I_t S_t \\ \frac{I_t}{dt} = \beta I_t S_t - \gamma I_t \end{cases} \tag{*2}$$

Note that the mean-field system is smooth and Lipschitz on the relevant domain. Therefore the explicit Euler method converges with first-order accuracy. Consequently, the global discretization error scales linearly with the time step Δt. We choose Δt sufficiently small so that further refinement produces negligible changes in the simulated trajectories.

2.2.1 Simulation

Using the same setup as in the previous simulation and running Euler's method for 1000 step with step size $\Delta t = 0.01$ we get:

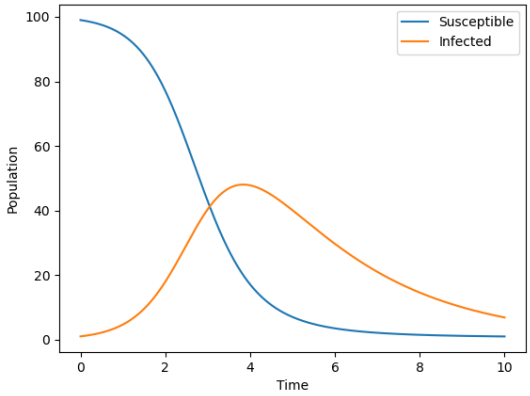


Figure 2: Mean Field Simulation

Here again, we see that the infection peak occurs around $t = 4$. However in this case, the trajectory is fixed, and thus, unlike the previous case there are no mean to model early extinction with the mean field approximation.

2.3 Diffusion Approximation

To allow for more exotic trajectories we can improve upon the mean field approximation of the stochastic SIR by using its **Diffusion Approximation** given by

$$\begin{cases} dS = -\beta I_t S_t dt - \sqrt{\beta I_t S_t} dV_t \\ dI = \beta I_t S_t dt - \gamma I_t dt + \sqrt{\beta I_t S_t} dV_t - \sqrt{\gamma I_t} W_t \end{cases} \tag{*3}$$

Where V_t and W_t are standard Brownian motion. To simulate the evolution of the disease using $(*3)$ we use the Euler-Maruyama method and thus, after some time Δt, we define the values of our process at time $t + \Delta t$ as:

$$\begin{aligned} S_{t+\Delta t} &= S_t - S_t I_t \beta \Delta t - \xi_{1,t} \sqrt{S(t) I(t) \beta} \\ I_{t+\Delta t} &= I_t + (S_t I_t \beta - \gamma I_t)\Delta t + \xi_{1,t} \sqrt{S_t I_t \beta} - \xi_{2,t} \sqrt{I_t \gamma} \end{aligned}$$

With $\xi_{i,t} \sim N(0, \Delta t)$.

2.3.1 Simulation

Using the same as set up as before, we simulate the process using 100, 500 and 1000 steps. The trajectories are plotted below:

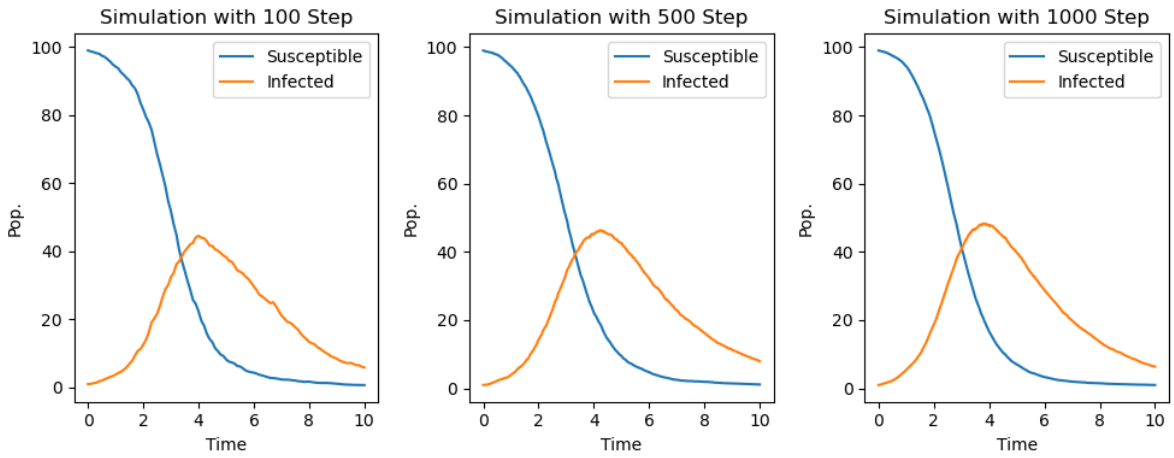


Figure 3: SIR Diffusion Approximation

Here again we observe similar trends regarding the time of the infection peak. Furthermore, note that the diffusion approximation smoothens out as the number of steps increases. This is because in that case $\Delta t \to 0$ and thus the noise we added vanishes, leading to smoother curves.

3 Probability of Extinction

The disease is considered extinct if $I_t = 0$ for some $t \in [0, T]$. Using the stochastic SIR simulation (and the same setup) we estimate the probability of the disease going extinct for $T = 1, 2, 10$.

3.1 Method

To compute the probability of extinction we use a Monte Carlo method. First, we simulate the process N times and define the following random variable:

$$Z_n = \mathbb{I}_{\{I_{t,n}=0 \text{ for } t\in[0,10]\}}$$

Where $I_{t,n}$ denotes the number of infected individual at time t for the n^{th} realization. As defined, the probability that $Z_n = 1$ is exactly the probability that the disease goes extinct in the n^{th} realization. Furthermore we have:

$$\mathbb{P}(Z = 1) = \mathbb{E}[Z] = p \tag{2.1.2}$$

We then use the crude Monte Carlo estimator for the value of p and choose the number of iterations N so that the relative error is smaller than 5%, i.e.,

$$\frac{|\hat{p}_N - p|}{p} \leq 0.05$$

However, since we do not have access to the true value of p we instead use the confidence interval of our estimate to bound the relative error, letting I_α be the confidence interval at the α significance level and $a = \min(I_\alpha)$ we have:

$$\frac{|\hat{p}_N - p|}{p} \leq \frac{|I_\alpha|/2}{a} = \frac{c_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{N}}}{\hat{p}_N - c_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{N}}} \tag{1}$$

for $c_{1-\alpha/2}$ the $1-\alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}$ the sample deviation. We then iterate our population model until the right hand-side of (1) is smaller than 0.05. the method is implemented using a sequential Monte Carlo algorithm with an initial pilot run iterating the population model a thousand time.

3.2 Results

For the 99% confidence level the results obtained are summarized in the table below:

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	16.57%	$14,734$	0.1381	$[15.78\%, 17.36\%]$
$T = 2$	16.64%	$14,664$	0.1386	$[15.84\%, 17.42\%]$
$T = 10$	16.9%	$14,383$	0.1404	$[16.1\%, 17.7\%]$

Our estimates are quite poor. The deviation is quite high (on the order 10^{-2}) relative to our estimates. Furthermore from one trials to the other, the estimated probability do not always make sense. It is indeed possible for the probability of extinction in $T = 1$ to be higher than that in $T = 10$. Increasing the accuracy of our estimates appears to be quite desirable. However the value seem to consistently hover around the 16%-18% mark for all values of T.

3.3 Variance Reduction

So far we have estimated the probability of extinction using a crude Monte Carlo estimator on the indicator variable 2.1.1.. In order to improve the Mean Square Error (MSE) of our estimate we should minimize the variance of the variable we use within our estimator. For this purpose, we use a control variate Y with known mean $\mathbb{E}[Y]$ and estimate the probability of

extinction using:

$$\hat{p} = \frac{1}{N}\sum_{n=1}^{N} Z_n + \alpha(Y_n - \mathbb{E}[Y]) \tag{3.1.1}$$

The variance achieved in this case is given by:

$$\mathbb{V}\text{ar}(Z_\alpha) = \mathbb{V}\text{ar}(Z)(1 - \rho_{Z,Y}^2) \tag{3.1.2}$$

For $\rho_{Z,Y}$ the correlation between Z and Y. We define our control variate Y as follows:

$$Y = \mathbb{I}_{\{\text{Recovery occurs first}\}} \tag{3.1.3}$$

The expected value of Y is simply the probability that, within our simulation, the first event to occur is a recovery. In the setting of part 1 and 2, this leads to the disease going extinct from the get go. Since the two events are independent we let $U \sim \text{exp}(99\beta)$ and $W \sim \text{exp}(\gamma)$ and recover the probability as:

$$\mathbb{P}(W < U) = \int_0^\infty 99\beta e^{-99\beta u} \int_0^u \gamma e^{-\gamma x} dx du = 0.1681 = \mathbb{E}[Y] \tag{3.1.4}$$

We also show that the covariance is non-zero:

$$Cov(Y, Z) = \mathbb{E}[ZY] - \mathbb{E}[Y]\mathbb{E}[Z] = \mathbb{E}[Y] - p\mathbb{E}[Y] \neq 0 \tag{3.1.5}$$

Since the probability of the disease going extinct in any of the time horizon is never 1. Furthermore:

$$\alpha = \frac{\text{Cov}(Y, Z)}{\mathbb{V}\text{ar}(Y)} = \frac{\mathbb{E}[Y] - p\mathbb{E}[Y]}{\mathbb{E}[Y](1 - \mathbb{E}[Y])} = \frac{1 - \mathbb{E}[Z]}{1 - \mathbb{E}[Y]} \tag{3.1.6}$$

We use formula 3.1.5 to estimate the covariance between Y and Z. This requires us to only estimate $E[Z]$. We use the algorithm developed in part 2 to do so. Keeping the relative error condition, our algorithm thus uses a pilot run with a sufficient number of iteration to ensure a 5% relative error for the value $alpha$.

3.4 Result

The results are summarized in the table below:

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	17.31%	1001	0.005	$[16.74\%, 17.89\%]$
$T = 2$	17.9%	1001	0.0106	$[17.07\%, 18.74\%]$
$T = 10$	18.02%	1073	0.0119	$[17.17\%, 18.88\%]$

Our method has proven itself quite effective! This is likely due to a high degree of correlation between our control variate and the output Z. In fact when comparing the value obtained in part 2 with $E[Y]$ we see that they are rather close. This suggest that extinction, when it occurs, happens due to the first event being a recovery, and that extinction occurring towards the end of the time horizon are rather rare.

3.5 Alternative Starting Population

We used a similar method (adjusting the probability of $Y = 1$) to estimate the probability of extinction in a population starting with 95 susceptible individual and 5 infected. Given that the probability of extinction is rather small we relaxed the relative error condition in order to avoid excessive computation time. We ran the simulation 10^4 times. Note also that formula 3.1.5 is no longer valid as the first occurrence being a recovery no longer guarantees extinction. We thus estimate the covariance using a pilot run with 1.5×10^4 iteration. We dropped the relative error condition for this simulation as the probability of extinction is rather small (on the order of 10^{-4}) and thus would've required significant run time to achieve it.

	Prob. of Extinction ($\hat{p}$)	Variance	99% Conf. Interval	relative error
$T = 2$	0.07%	0.0007	$[0\%, 0.14\%]$	$3,600\%$

Our estimates are quite unreliable. The relative error is quite high. As mentioned earlier, our control variable Y is a lot less correlated to the target variable Z as the outcome $Y = 1$ no longer guarantees the outcome $Z = 1$. One could use a similar control variates, for example 5 recovery in row, to increase the correlation between the two.

The bigger problem with our estimation however, is that the probability of extinction is pretty small (on the order of 10^{-3}) and thus simulating "enough" extinction is tough. To solve this problem we could either increase the number of iteration or use an importance-sampling-ish method whereby occurrence of extinction are simulated more often leading to a more accurate estimate of the desired probability. One could perhaps derive the distribution of the chain after 5 iteration within T unit of time in order to compute the weight for the importance sampling method, and draw from a distribution where after the 5th occurrence the disease is much more likely to be extinct.

4 Stochastic SIR with demographic effect

We now simulate the SIR process including demographic effects, that is we also take into account the birth and death of individuals within each groups of the population. In this case,

the mean-field approximation of the stochastic SIR-d is given by:

$$\begin{cases} \frac{dS}{dt} & = m(S_t + I_t + R_t) - mS_t - \beta S_t I_t dt \\ \frac{dI}{dt} & = \beta S_t I_t - \gamma I_t - (m+v) I_t \\ \frac{dR}{dt} & = \gamma I_t - m R_t \end{cases}$$

Where m is the natural death/birth rate and v is the pathogen induced death rate. We recover the transition probabilities of $X_t = (S_t, I_t, R_t)$ and show that $(*4)$ is indeed the mean field approximation of the stochastic SIR-d. Supposing that the number of infections, deaths, and recoveries are Poisson process, we investigate how each of the infected, susceptible and recovered population evolve over time.

4.1 Susceptible

As described by the mean field approximation, there are three event affecting the evolution of the susceptible population. A birth in the population will lead to a new susceptible individual, a death within the susceptible population will reduce the number of susceptible individual and similarly an infection will reduce the number of susceptible individual. Note the latter will also increase the infected population.

If a birth occur, the the number of susceptible individual goes up by one. More formally letting $X(t) \in \mathbb{R}^3$ denote the state of the population at time t, this event can be written as going from $X(t)$ to $X(t) + (1, 0, 0)$. We thus seek the probability of:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (1,0,0) | X(t) = (s,i,r)) \tag{4.1.1}$$

Since birth occur within the population with a rate m and the number of birth within the whole population is captured by a Poisson process the probability in 4.1.1 should be the same as the probability of a Poisson process with rate $m(i+s+r)$ going up by 1. We have:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (1,0,0) | X(t) = (s,i,r)) = m(s+i+r)dt + o(dt) \tag{4.1.2}$$

Following the same reasoning we recover the probability of a death within the susceptible population:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (-1,0,0) | X(t) = (s,i,r)) = msdt + o(dt) \tag{4.1.3}$$

and of an infection:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (-1,1,0) | X(t) = (s,i,r)) = \beta sidt + o(dt) \tag{4.1.4}$$

4.2 Infected

The infected population is affected by four events. An infection will increase its number, a pathogen-induced death will decrease it and so will a non-pathogen-induced death, lastly a recovery could also occur, decreasing the infected population and increasing the recovered population. Note the probability of the first event (infection) has already been recovered in part 4.1 by the equation 4.1.4. Following the same reasoning as before we recover the probabilities of the last three events. If a pathogen-induced death occurs we have:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,0)|X(t) = (s,i,r)) = vidt + o(dt) \tag{4.2.1}$$

For a non-pathogen death:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,0)|X(t) = (s,i,r)) = midt + o(dt) \tag{4.2.2}$$

And for a recovery:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,1)|X(t) = (s,i,r)) = \gamma idt + o(dt) \tag{4.2.3}$$

Note, since 4.2.1 and 4.2.2 describe distinct event with the same effect on the infected population the total probability of a death within the infected becomes:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,0)|X(t) = (s,i,r)) = (m+v)idt + o(dt) \tag{4.2.4}$$

4.3 Recovered

The recovered population is affected by 2 events. A recovery will increase its number, and a death within the recovered will decrease it. The probability of the former has already been described by equation 4.2.3. Therefore a death occurs within the recovered with probability:

$$\mathbb{P}(X(t+dt) = (s,i,r) + (0,0,-1)|X(t) = (s,i,r)) = mrdt + o(dt) \tag{4.3.1}$$

4.4 Formulation

Putting all this together we have that, $X(t) = (s, i, r)$ is a Markov chain with transition probabilities

$$\begin{aligned}
&\mathbb{P}(X(t+dt) = (s,i,r) + (1,0,0) | X(t) = (s,i,r)) = m(s+i+r)dt + o(dt) \\
&\mathbb{P}(X(t+dt) = (s,i,r) + (-1,0,0) | X(t) = (s,i,r)) = msdt + o(dt) \\
&\mathbb{P}(X(t+dt) = (s,i,r) + (-1,1,0) | X(t) = (s,i,r)) = \beta sidt + o(dt) \\
&\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,1) | X(t) = (s,i,r)) = \gamma idt + o(dt) \\
&\mathbb{P}(X(t+dt) = (s,i,r) + (0,-1,0) | X(t) = (s,i,r)) = (m+v)idt + o(dt) \\
&\mathbb{P}(X(t+dt) = (s,i,r) + (0,0,-1) | X(t) = (s,i,r)) = mrdt + o(dt)
\end{aligned} \tag{4.4.1}$$

To recover the provided mean field approximation of the SIR-d system we look at the expected rate of change of each sub-population. This suffice because:

$$\begin{aligned}
s(t) - s(t+dt) &\approx \mathbb{E}[\Delta s \text{ over } dt] \\
i(t) - i(t+dt) &\approx \mathbb{E}[\Delta i \text{ over } dt] \\
r(t) - r(t+dt) &\approx \mathbb{E}[\Delta r \text{ over } dt]
\end{aligned} \tag{4.4.2}$$

Taking dt sufficiently small so that $o(dt)$ is negligible the expected changes in each sub-population is the sum of the expected value of the processes affecting it, that is:

$$\begin{aligned}
\mathbb{E}[\Delta s \text{ over } dt] &= (m(s+i+r) - ms - \beta is)dt \\
\mathbb{E}[\Delta i \text{ over } dt] &= (\beta is - (m+v)i - \gamma is)dt \\
\mathbb{E}[\Delta r \text{ over } dt] &= (\gamma i - mr)dt
\end{aligned} \tag{4.4.3}$$

And thus dividing by dt and taking the limit as dt goes to 0, 4.4.2 and 4.4.3 together yield:

$$\begin{aligned}
\frac{dS}{dt} &= m(s+i+r) - ms - \beta is \\
\frac{dI}{dt} &= \beta is - (m+v)i - \gamma is \\
\frac{dR}{dt} &= \gamma i - mr
\end{aligned} \tag{4.4.4}$$

As needed.

4.5 Simulation Stochastic SIR-d

The algorithm used to simulate the stochastic SIR-d process is essentially the same as that presented at the inception of the report. We run the program for the following (m, v) pairs:

$(10^{-4}, 10^{-3})$, $(10^{-4}, 10^{-2})$, $(10^{-3}, 10^{-3})$, $(10^{-3}, 10^{-2})$. The results are shown below:

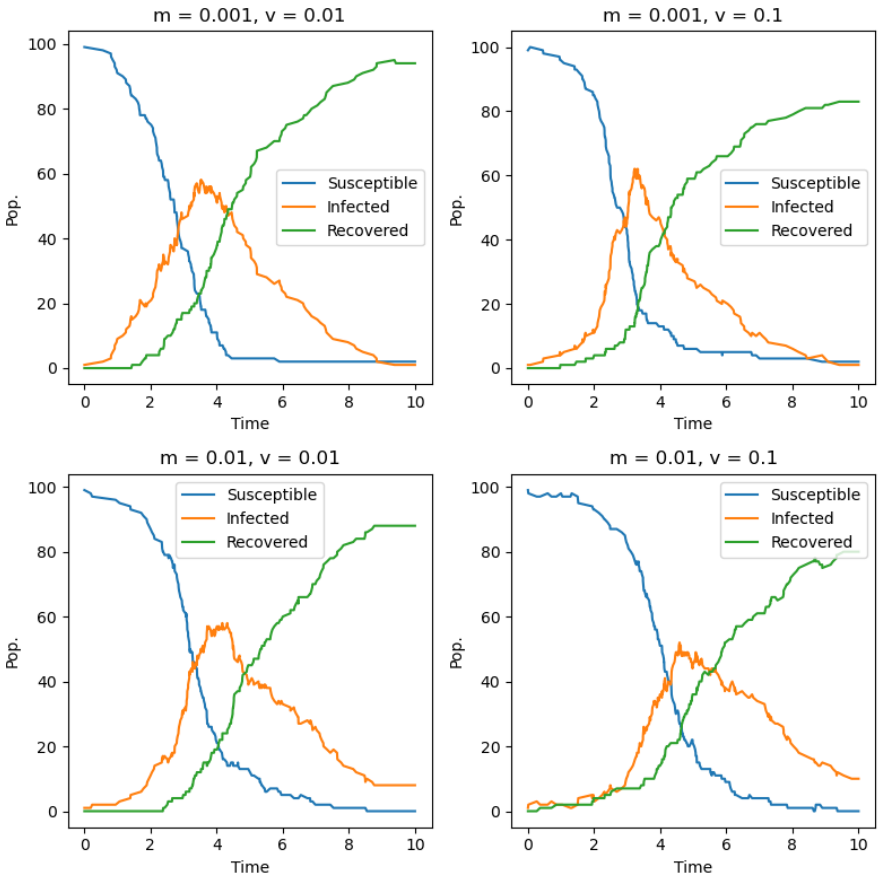
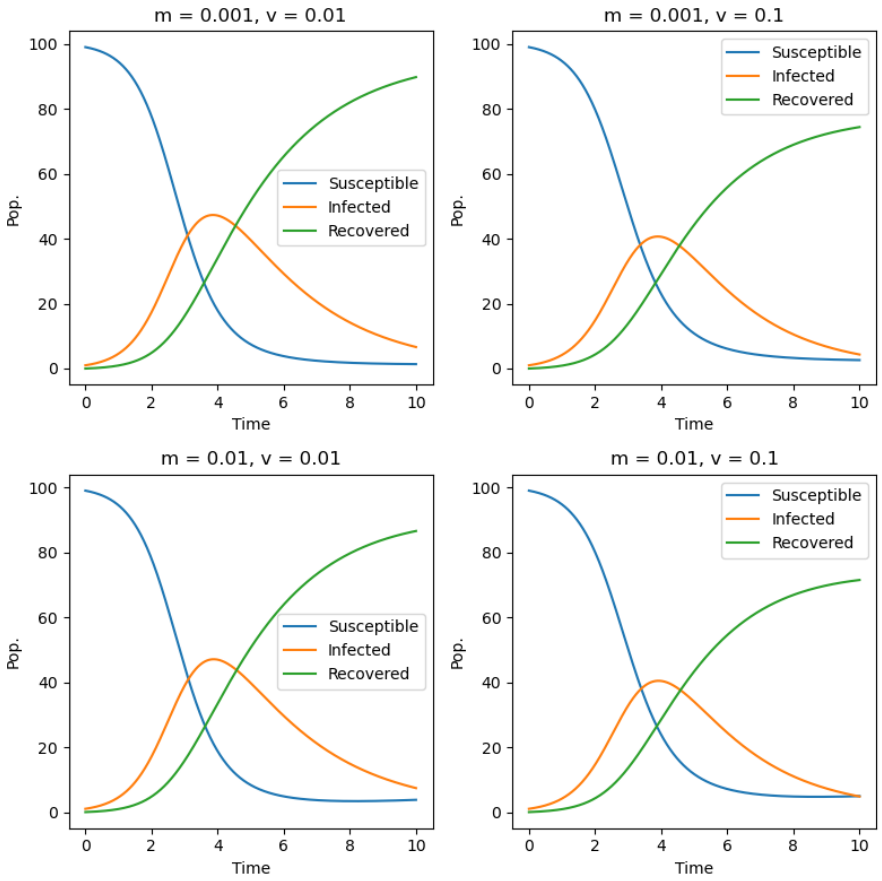


Figure 4: Stochastic SIR-d with different birth-death rate and pathogen-induced death rate

The trajectories are similar in shapes to those simulated for the process without demographic effects, however, note that the recovered population is smaller at $T = 10$ when the pathogen induced death-rate v is higher. This is expected as, if infected individuals die out, they can no longer recover.

4.6 Mean Field Approximation

To plot the mean field approximation we use an explicit Euler method with the same number steps as in section 2.2. The trajectories are plotter below under the different parameters values.



Again, the shapes of the trajectories are comparable with infections peaks at around $T = 4$, and as in the previous case, we see that for the larger values of v, the recovered population is smaller at $T = t$ than otherwise. Furthermore, while the "infection peaks" have no particular trend in the stochastic simulation, we observe that their height (largest number of infected individual at once) is larger in the mean field approximation when the pathogen-induced death rate is smallest (left column of figure 6, $v = 0.01$). This corroborates with intuition as in this case death due to the disease are less likely to occur, and thus "less constraints" are put on the evolution of the infected.

5 SIR-d, Extinction and Basic Reproduction Number

In a deterministic model, a disease will go extinct ($I_t \to 0$ as $t \to \infty$) if the basic reproduction number ρ is strictly smaller than 1.

$$\rho = \frac{\beta(S_0 + I_0 + R_0)}{m + v + \gamma}$$

We investigate how the stochastic $SIR-d$ handles extinction for values of $\rho > 1$. In particular we consider a population of size $N = 100$ and initial condition $X_0 = (95, 5, 0)$ and compute

the probability of extinction over different time horizon.

5.1 Extinction in SIR-d

For the disease to go extinct we simply need $I_t = 0$ for some $t \in (0, \infty]$. One way for this to occur in the stochastic SIR-d model is if within the first N events there are at least I_0 recoveries or deaths of infected individuals and 0 infection. Denoting by $J_{R,n}$ and $J_{D,n}$ the jumps time after n events of the Poisson processes capturing the number of recovery and death of the infected respectively, and by $J_{I,n}$, $J_{B,n}$ $J_{S,n}$ and $J_{M,n}$ that of the number of infection, birth of susceptible, death of susceptible and death of recovered respectively. Letting $N = I(0)$, the disease will go extinct if the following probability is non-zero:

$$\prod_{n=1}^{N} [\mathbb{P}(J_{R,n} < \min_{j \in \{I,B,S,M\}} J_{j,n}) + \mathbb{P}(J_{D,n} < \min_{j \in \{I,B,S,M\}} J_{j,n})] \tag{6.1.1}$$

Since the jump times follow an exponential distribution and that:

$$\mathbb{P}(J < a) = \int_0^a \lambda e^{-\lambda x} dx > 0 \tag{6.1.2}$$

For any fixed a arbitrary small, it follows that the probability laid out in 6.1.1 is never 0, and thus it is possible for the disease to go extinct in the stochastic SIR-d model regardless of the value of R_0 although this event may be very unlikely. Note however that the probability we have formulated is that of the disease going extinct at some point and not within T units of time. To recover the latter it suffices to replace:

$$\min_{j \in \{I,B,S,M\}} (J_{j,n}) \text{ by } \min \left(\frac{T}{I(0)}, \min_{j \in \{I,B,S,M\}} (J_{j,n}) \right) \tag{6.1.3}$$

This will ensure that:

$$\sum_{n=1}^{N} J_n < T \tag{6.1.4}$$

and thus that the extinction occurs within T unit of time.

5.2 Probability Estimation

To estimate the probability of extinction in the given settings we use the same method as for the stochastic SIR, adapted to the stochastic SIR-d model. Furthermore we set $\gamma = 0.4$, $m = 10^{-4}$, $v = 10^{-2}$ and pick β so that $\rho = 1.01, 1.05, 1.1, 1.5$. We also weaken the relative

error condition given in 2.1.4 and use the following instead:

$$\frac{c_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{N}}}{\hat{p}_N} < 0.05 \tag{6.2.1}$$

We assumed that $\gamma = 0.4$ in order to recover the value of β using the formula for R_0. The estimate are summarized below:

For $\beta = 0.00414201$ and $R_0 = 1.01$

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	0.49%	$537,233$	0.0049	$[0.47\%, 0.51\%]$
$T = 2$	4.03%	$63,200$	0.0387	$[3.83\%, 4.23\%]$
$T = 10$	62.95%	$10,000$	0.2333	$[61.71\%, 64.19\%]$

For $\beta = 0.00430605$ and $R_0 = 1.05$

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	0.48%	$549,707$	0.0048	$[0.46\%, 0.5\%]$
$T = 2$	3.86%	$65,996$	0.0372	$[3.67\%, 4.06\%]$
$T = 10$	60.67%	$10,000$	0.2386	$[59.41\%, 61.93\%]$

For $\beta = 0.0045111$ and $R_0 = 1.1$

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	0.47%	$566,265$	0.0046	$[0.44\%, 0.49\%]$
$T = 2$	3.86%	$66,003$	0.0372	$[3.67\%, 4.06\%]$
$T = 10$	58.49%	$10,000$	0.2428	$[57.22\%, 59.76\%]$

For $\beta = 0.0061515$ and $R_0 = 1.5$

	Prob. of Extinction ($\hat{p}$)	N	Variance	99% Conf. Interval
$T = 1$	0.36%	$740,194$	0.0036	$[0.34\%, 0.37\%]$
$T = 2$	2.62%	$98,537$	0.0255	$[2.49\%, 2.75\%]$
$T = 10$	35.93%	$10,000$	0.2302	$[34.69\%, 37.16\%]$

As can be expected, the probability of extinction for any of the time horizon considered goes down as the R_0 of the model goes up. This is because, as R_0 goes up, the infection rate goes up. Furthermore, contrary to what was observed in part 2, the time horizon affect the probability of extinction drastically. Comparing across the R_0, we also note that the probability of extinction decreases by a larger amount when the time horizon is longer.

6 Conclusion

In this report, we examined the SIR epidemic model through a sequence of increasingly refined stochastic and deterministic formulations. Starting from a continuous-time Markov chain description, we demonstrated how individual-level randomness gives rise to a wide range of epidemic trajectories, including early extinction events that cannot be captured by deterministic models. The mean-field approximation was shown to provide an accurate description of average behavior in large populations, while the diffusion approximation offered a useful compromise by retaining stochastic variability around the mean-field dynamics.

A central focus of this study was the estimation of disease extinction probabilities over finite time horizons. Using crude Monte Carlo methods, we observed that extinction probability estimates can exhibit high variance, particularly when extinction events are rare. By introducing a control variate based on the first-event structure of the stochastic process, we achieved a significant reduction in variance and computational cost. Our results suggest that, for the parameter regimes considered, extinction is most likely to occur very early in the epidemic, often due to an initial recovery event.

We further extended the analysis to an SIR model with demographic effects, incorporating births, natural deaths, and pathogen-induced mortality. For this SIR-d model, we derived the stochastic transition probabilities and recovered the corresponding mean-field equations. Simulations confirmed that demographic effects alter both the long-term population composition and the likelihood of extinction, particularly through the pathogen-induced death rate. Finally, we showed that, unlike deterministic models where extinction is governed strictly by the basic reproduction number, stochastic models allow for extinction with positive probability even when the reproduction number exceeds one.

Overall, this work highlights the importance of stochastic modeling in epidemic dynamics, especially when studying extinction events and early outbreak behavior. While deterministic approximations remain valuable for capturing average trends, stochastic effects play a crucial role in realistic epidemic scenarios and should be carefully accounted for in both modeling and simulation.