PROJECT REPORT

Simulation of an SIR Stochastic Process in Continuous Time

David Charry

Benedict Martus

Saad Bin Asad

Compiegne, 2024

1. **transition probability matrix for Markov chain for N = 3 for infinitesimal time-step Δt > 0.**

While CTMCs are typically analyzed with continuous rates, small time steps Δt allow us to approximate transitions as if they occur at discrete intervals. This lets us treat transitions as happening "each step" with a probability based on the rates. With very small Δt, the chance of multiple transitions at once is negligible, so we can interpret the transition matrix similarly to a discrete-time model. In order to construct this matrix, we identified all possible states of the population. This generated 20 possible states in total (All variations where 3, 2, 1, and 0 are alive), including cases where no infectious individuals were present.

Each row in the matrix represents one of these states and the rates at which transitions can occur:

- Infection: A susceptible individual becomes infected at rate $\beta \times \frac{S \times I}{N}$, proportional to the product of susceptible and infectious individuals.
- Recovery: An infectious individual recovers at rate $\gamma \times I$
- Mortality: An infectious individual dies from the disease at rate $\nu \times I$

For each transition possibility, we computed the rate, multiplied it with Δt, and placed it in the appropriate off-diagonal cell in Q. The diagonal elements are set to 1 - the sum of the other probabilities in each row, representing the probability that we stay in the same state, and ensuring that each row sums to one.

	(3, 0, 0)	(2, 1, 0)	(1, 2, 0)	(0, 3, 0)	(2, 0, 1)	(1, 1, 1)	(0, 2, 1)	(1, 0, 2)	(0, 1, 2)	(0, 0, 3)	(2, 0, 0)	(1, 1, 0)	(0, 2, 0)	(1, 0, 1)	(0, 1, 1)	(0, 0, 2)	(1, 0, 0)	(0, 1, 0)	(0, 0, 1)	(0, 0, 0)
(3, 0, 0)	1																			
(2, 1, 0)		1-Σ(2,1,0)	(2*1*β/3)*Δt		1*γ*Δt						1*ν*Δt									
(1, 2, 0)			1-Σ(1,2,0)	(1*2*β/3)*Δt		2*γ*Δt						2*ν*Δt								
(0, 3, 0)				1-Σ(0,3,0)			3*γ*Δt						3*ν*Δt							
(2, 0, 1)					1															
(1, 1, 1)						1-Σ(1,1,1)	(1*1*β/3)*Δt	1*γ*Δt						1*ν*Δt						
(0, 2, 1)							1-Σ(0,2,1)		2*γ*Δt						2*ν*Δt					
(1, 0, 2)								1												
(0, 1, 2)									1-Σ(0,1,2)	1*γ*Δt						1*ν*Δt				
(0, 0, 3)										1										
(2, 0, 0)											1									
(1, 1, 0)												1-Σ(1,1,0)	(1*1*β/2)*Δt	1*γ*Δt			1*ν*Δt			
(0, 2, 0)													1-Σ(0,2,0)		2*γ*Δt			2*ν*Δt		
(1, 0, 1)														1						
(0, 1, 1)															1-Σ(0,1,1)	1*γ*Δt			1*ν*Δt	
(0, 0, 2)																1				
(1, 0, 0)																	1			
(0, 1, 0)																		1-Σ(0,1,0)	1*γ*Δt	1*ν*Δt
(0, 0, 1)																			1	
(0, 0, 0)																				1

Figure 1: Full Transition Rate Matrix

2. **Simulate 100 trajectories of the SIR Markov process up to absorption, with N = 1000, (S(0), I(0), R(0)) = (990, 10, 0), β = 0.2, γ = 0.1, and ν = 0.001. Remember to save the simulated trajectories (values of (S, I, R) over time) so that you can answer the following questions. Explain the algorithm used for the simulations**

To simulate 100 trajectories of the SIR Markov process up to absorption, we used a python code that fundamentally relies on five steps for each of the 100 simulations. Here are the steps of the algorithm each simulation runs through:

2.1 We check if the number of infected individuals is higher than one; otherwise, we stop.
2.2 We calculate the different transition rates for the current state.
2.3 We sample the time to the next event based on a random exponential distribution with:

$$T \sim Exponential(\lambda) \Longrightarrow E[T] = \frac{1}{\lambda}$$

Where λ reflects the total transition rate (Sum of all 3 transition rates), so on average, the time until an event happens is inversely related to the total rate.

2.4 We choose the next event based on the transition-rates following a random number of a Uniform distribution with U ~ [0,1].

- The random number is in the range $\left[0, \frac{Infection\ Rate}{\lambda}\right]$ → An infection occurs
- Falls within $[\frac{Infection\ Rate}{\lambda}, \frac{Infection\ Rate + Recovery\ Rate}{\lambda})$ → A recovery occurs
- Otherwise → death occurs

2.6 We update the state (S, I, R) and go back to step 1 for the next iteration.

Here is the corresponding python code (beforehand we define N = 1000, beta = 0.2, gamma = 0.1, nu = 0.001, S0 = 990, I0 = 10, R0 = 0, num_simulations = 100):

```
# Simulation loop
for sim in range(num_simulations):
    # Initialize the current state
    S, I, R = S0, I0, R0
    t = 0  # Start time
    trajectory = [(t, S, I, R)]  # Record initial state

    # Run until absorbing state (I = 0)
    while I > 0:
        # Calculate rates of each possible transition
        infection_rate = beta * S * I / N
        recovery_rate = gamma * I
        mortality_rate = nu * I
        total_rate = infection_rate + recovery_rate + mortality_rate

        # Sample time to next event from an exponential distribution
        time_to_next_event = np.random.exponential(1 / total_rate)
        t += time_to_next_event  # Advance time by this sampled amount

        # Choose which event occurs
        event_prob = np.random.uniform(0, 1)

        # Infection event
        if event_prob < infection_rate / total_rate:
            S -= 1
            I += 1
        # Recovery event
        elif event_prob < (infection_rate + recovery_rate) / total_rate:
            I -= 1
            R += 1
        # Mortality event
        else:
            I -= 1

        # Record the new state after the event
        trajectory.append((t, S, I, R))

    # Save the trajectory for this simulation
    trajectories.append(trajectory)
```

Figure 2: Simulation of Trajectories

3. **Empirical mean and variance of the epidemic duration (defined as Text = inf{t ≥ 0 : I(t) = 0}).**

In order to calculate the empirical mean and variance of the epidemic duration, we simply loop through each of the simulations and each layer of the simulation to check the occurrence where I (t) = 0. We save the time at which I (t) = 0 occurs in an array and calculate the empirical mean and variance with that.

```
# Calculate the epidemic duration (T_ext) for each trajectory
epidemic_durations = []

for trajectory in trajectories:
    # Find the step where I(t) became zero -> Ending the pandemic
    for (time, S, I, R) in trajectory:
        if I == 0:  # Epidemic is ended the first time I(t) hits 0
            epidemic_durations.append(time)
            break
```

```
epidemic_durations = np.array(epidemic_durations)

# Calculate the empirical mean and variance of the epidemic durations
mean_duration = np.mean(epidemic_durations)
variance_duration = np.var(epidemic_durations)
```

Figure 4: Empirical Mean and Variance of the FInal Size of the Epidemic

Empirical Mean of Epidemic Duration (T_ext) = **149.93**

Empirical Variance of Epidemic Duration (T_ext) = **432.76**

4. Calculate the empirical mean and the empirical variance of the final size of the epidemic, which is equal to $N - ST_{ext}$ (corresponding to the number of individuals who were infected between the start of the epidemic at time 0 and the end of the epidemic at time T_{ext}).

To calculate the empirical mean and variance of the final size of the epidemic (Number of people that got infected and either died or recovered), we simply loop through each of the simulations and each layer of the simulation to check the occurrence where I (t) = 0. We save the final_size = N - final_S at which I (t) = 0 occurs in an array and calculate the empirical mean and variance with that.

```
# Calculate the final size of the epidemic for each simulation
final_sizes = []

for trajectory in trajectories:
    # Find the last time step in the trajectory where I(t) became zero
    for (time, S, I, R) in trajectory:
        if I == 0:
            final_sizes.append(N-S)
            break

final_sizes = np.array(final_sizes)

# Calculate the empirical mean and variance of the final sizes
mean_final_size = np.mean(final_sizes)
variance_final_size = np.var(final_sizes)
```

Figure 4: Empirical Mean and Variance of the FInal Size of the Epidemic

Empirical Mean of Final Size of the Epidemic: **795.37**

Empirical Variance of Final Size of the Epidemic: **956.39**

5. Graph 10 simulated trajectories of (S, I, R, D) in the same graph.

To display all 4 states for the first 10 simulations, we have to calculate the amount of dead people, since this was not required in Exercise 2. For that, we simply substrate the Number of susceptible, infected or recovered people from the initial population size, for each timestep for each of the 10 simulations. Afterwards we plot it on a logarithmic scale for added visibility of the number of dead people.

```
for i in range(10):  # Loop through the first 10 simulations
    times = [state[0] for state in trajectories[i]]
    S_values = [state[1] for state in trajectories[i]]
    I_values = [state[2] for state in trajectories[i]]
    R_values = [state[3] for state in trajectories[i]]
    D_values = [N - state[1] - state[2] - state[3] for state in trajectories[i]]
```

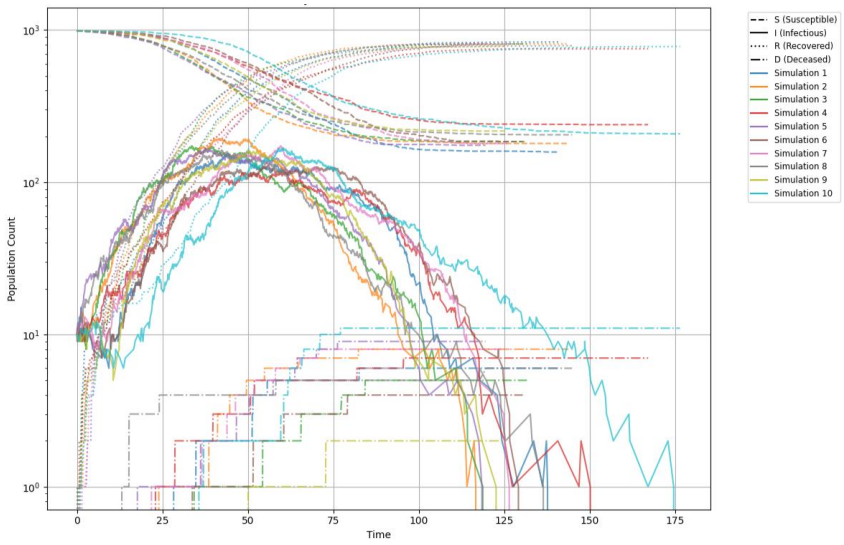

Figure 5: Simulated Trajectories of (S,I,R,D)

6. Propose a method for estimating the value of R0 from contact real data, and support your choice with a discussion of data collection in the context of an epidemic.

The basic reproduction number, R_0, quantifies the average number of secondary infections caused by one infectious individual in a fully susceptible population. To estimate R_0, this approach combines simulation-based data with real-world epidemiological data for accuracy and relevance.

6.1. Data Collection:

- Simulation Data: Using the SIR Markovian model, we track the susceptible (S) and infected (I) populations over time. At each time step, values are recorded for analysis.
- Real-World Data: Collected data includes:
- Contact Rate (CR): The average number of close interactions per individual.
- Infection Probability (IP): The likelihood of transmission during a contact.
- Average Infectious Period (AIP): The duration for which an individual remains contagious.

6.2. Estimation Formula:

- From the data, the initial growth rate (r) is calculated as: $r = \frac{1}{t}\ln\left(\frac{I_t}{I_{t_0}}\right)$
- Using r, R_0 is estimated with the formula: $R_0 = 1 + \frac{r}{\gamma}$
- where γ is the recovery rate.

6.3. Interpretation and Limitations:

$R_0 > 1$ signals an epidemic is likely to grow, while $R_0 < 1$ suggests containment. Limitations include assumptions of constant γ and exclusion of recovered (R) individuals.

These calculations help us estimate how the disease will spread and understand how different actions can help control it in real-life situations.

It is a helpful method for estimating (R_0) and understanding how an epidemic might grow. However, it makes some simple assumptions and uses idealized data, which might not reflect real-world situations. To get more accurate results, we could use more detailed models, like SEIR, that include different stages of the disease and account for variations in populations.

7. **Propose a modification to the SIR model, by adding a class, to be able to take into account a delay between the moment when an individual is contagious and the moment from which he can transmit. Give the state space and transition rate matrix Qe corresponding to this new model.**

In order to take into account the delay between infection and contagiousness, the SEIR model introduces an Exposed (E) compartment. This better represents diseases with incubation periods.

7.1. Compartment Definitions:

- Susceptible (*S*): Individuals who are at risk of infection.
- Exposed (*E*): Infected but not yet contagious.
- Infectious (*I*): Individuals who can transmit the disease.
- Recovered (*R*): Individuals who have recovered or died and are no longer infectious.

7.2. State Space and Transition Rates:

- Total population size is constant: $N = S(t) + E(t) + I(t) + R(t)$
- The transitions are governed by the following probabilities:

$$P_{(s,e,i),(s-1,e+1,i)}(\Delta t) = \frac{\beta si}{N}\Delta t \qquad P_{(s,e,i),(s,e,i-1)}(\Delta t) = \gamma i \Delta t$$

$$P_{(s,e,i),(s,e-1,i+1)}(\Delta t) = \sigma e \Delta t \qquad P_{(s,e,i),(s,e,i)}(\Delta t) = 1 - \left(\frac{\beta si}{N} + \sigma e + \gamma i\right)\Delta t$$

β is the transmission rate, **σ** is the rate at which exposed individuals become contagious, **γ** is the recovery rate, and **Δt** time step.

7.3. Transition matrix for (SEIR model for $N = 2$):

$$(2,0,0), (1,1,0), (0,2,0), (1,0,1), (0,1,1), (0,0,2).$$

$$Q = \begin{bmatrix} -q_1 & \frac{\beta si}{N} & 0 & 0 & 0 & 0 \\ 0 & -q_2 & \sigma e & 0 & 0 & 0 \\ 0 & 0 & -q_3 & \gamma i & 0 & 0 \\ 0 & 0 & 0 & -q_4 & \sigma e & 0 \\ 0 & 0 & 0 & 0 & -q_5 & \gamma i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Possible states include the following:

Diagonal Entries ($-q_k$), which represent the negative sums of all outgoing rates from each state. Such as **(1,1,0): q2 = βsi / n + αe** and the off-diagonal entries which are listed as follows:

- **βsi / n:** Rate of infection (from S to E)
- **αe:** Rate of exposed individuals becoming infectious (from E to I)
- **γi:** Rate of infectious individuals recovering (from I to R)

And a final state (0,0,0), which is an absorbing state. No transition occur once the entire population is in the recovered state.

The SEIR model show how people move through different stages of a disease: from being susceptible (S) to getting exposed (E), then infectious (I), and finally recovered (R). It adds an "Exposed" stage to show the time between an infected one and when they can spread the disease. This helps us understand diseases that have a delay before people can infect others, like illnesses with an incubation period.