Markov Chains

Jacob Atkinson Stephen Maschan Constantin Rioland
Ahmed Diksin Ibrahim Topal Hayden Crook

March 21, 2025

Abstract

Markov Chains are stochastic processes that model systems transitioning between states or situations with probabilities dependent only on the current state, and not any previous states. This report aims to explore the fundamental concepts, including transition matrices, classifications, and key applications such as epidemic modeling and forecasting. Special focus has been given to the Reed-Frost model in relation to epidemiology and modifications for other real-world scenarios, in particular, modeling student movement on campus. The predictive power and practical applications of Markov Chains demonstrate their significance in probabilistic forecasting and decision-making across various fields.

1 An Introduction to Markov Chains

Definition 1 (Markov Chain)**.** *A **Markov Chain** is a stochastic process where the next state depends only on the current state, not on any past states. Formally, the sequence of random variables with values in state space S $\{X_i\}_{i=1}^{\infty}$ is called Markov Chain with state space S if:*

$$P(X_{n+1} = s \mid X_n = s_n) = P(X_{n+1} = s \mid X_1 = s_1, \ldots, X_n = s_n),$$

*for all states $s, s_1, \ldots, s_n \in S$. This means the process is **memoryless**, meaning X_{n+1} is conditionally independent of $X_1, \ldots, X_{n-1}$ given X_n.*

1.1 States and Transitions

Definition 2 (State Space)**.** *The **State Space** of a Markov Chain is the set of all possible states.*

Definition 3 (Transition Probability Matrix)**.** *The **Transition Probability Matrix** P describes the likelihood of moving between states, where P_{ij} is the probability of transitioning from state i to state j:*

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

Each row must sum to 1, as the system has to transition to some state at every step.

1.2 Classifications of Markov Chains

Definition 4 (Accessible and Communicating States)**.** *A state j is **accessible** from i (denoted $i \to j$) if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$.*

*Two states i and j **communicate** ($i \leftrightarrow j$) if $i \to j$ and $j \to i$. A Markov chain is **irreducible** if all states communicate.*

Definition 5 (Recurrent and Transient States)**.** *For a state i, let f_{ii} be the probability of returning to i at some $n \geq 1$, given it starts at i:*

$$f_{ii} = P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i).$$

*State i is **recurrent** if $f_{ii} = 1$ and **transient** if $f_{ii} < 1$.*

Proposition 1. *In an irreducible Markov chain, all states are either recurrent or transient.*

Definition 6 (Positive and Null Recurrence)**.** *A recurrent state is:*

- ***Positive recurrent** if the expected return time is finite.*
- ***Null recurrent** if the expected return time is infinite.*

Proposition 2. *If an irreducible Markov chain has at least one positive recurrent state, then all states are positive recurrent.*

Definition 7 (Periodicity)**.** *The **period** of state i is the greatest common divisor of all possible return times:*

$$d(i) = \gcd\{n > 0 \mid p_{ii}^{(n)} > 0\}.$$

*If $d(i) > 1$, state i is **periodic**. If $d(i) = 1$, state i is **aperiodic**.*

Proposition 3. *A Markov chain is aperiodic if and only if all states are aperiodic.*

Definition 8 (Absorbing, Ergodic, and Regular Chains)**.** *A Markov chain is:*

- ***Absorbing** if it contains at least one state that, once entered, cannot be left.*
- ***Ergodic** if it is irreducible, aperiodic, and has at least one positive recurrent state.*
- ***Regular** if some power of the transition matrix has all positive entries, ensuring that every state is reachable with positive probability.*

Proposition 4. *A regular Markov chain converges to a unique stationary distribution.*

1.3 Stationary Distributions

Definition 9 (Stationary Distribution)**.** *A probability distribution π is **stationary** if it remains unchanged under transitions:*

$$\pi P = \pi.$$

This means that if a Markov chain starts in the distribution π, it remains in π after any number of transitions.

Proposition 5. *If a Markov chain is irreducible and aperiodic, it has a unique stationary distribution.*

Proposition 6. *A finite-state Markov chain always has at least one stationary distribution.*

2 Markov Chains: Applications and Future Directions

Markov Chains provide a structured framework for modeling and analyzing stochastic processes, making them indispensable for probabilistic forecasting. Their predictive capabilities extend across multiple disciplines, including finance, artificial intelligence, and epidemiology, allowing industries to optimize decision-making and mitigate uncertainties effectively.

2.1 Applications of Markov Chains in Prediction

Markov Chains play a critical role in forecasting and decision-making:

- **Stock Market:** Predicting price fluctuations and market trends using probabilistic state transitions.
- **Weather Forecasting:** Modeling transitions between atmospheric states (e.g., sunny, rainy, stormy).
- **Traffic Flow:** Estimating congestion levels based on historical and real-time movement data.
- **Customer Behavior:** Analyzing shopping trends, digital engagement, and retention patterns.
- **Epidemiology:** Predicting the spread of infectious diseases using susceptible-infected-recovered (SIR) models.
- **Artificial Intelligence and Machine Learning:** Reinforcement learning applications for optimal decision-making.

These applications underscore the extensive utility of Markov Chains in real-world predictive modeling.

3 Future Directions

3.1 Advancing Computational Techniques

Enhanced algorithms and computational resources will refine the accuracy and scalability of Markov Chain models, enabling more efficient simulations and optimizations.

3.2 Hybrid Models

The integration of Markov Chains with deep learning architectures and Bayesian inference can yield more powerful predictive frameworks, enhancing accuracy in complex environments.

3.3 Expanding Applications

As personalized recommendations, autonomous systems, and real-time analytics continue to evolve, Markov Chains will remain fundamental in various research and industrial applications.

3.4 Refining Decision-Making

With the increasing reliance on data-driven strategies, refining Markov Chain methodologies will be crucial for understanding and optimizing complex dynamical systems.

4 Chain Binomial Models

4.1 Understanding Chain Binomial Models

A **Chain Binomial Model** is a class of discrete-time stochastic processes specifically developed to simulate and study how infectious diseases propagate through populations over time. These models are powerful tools in epidemiology, enabling researchers to predict epidemic trajectories, assess intervention strategies, and provide intuitive insights into how disease transmission occurs. Central to these models is the inherent randomness in disease spread, captured through probabilistic methods.

At its core, a chain binomial model categorizes individuals into distinct health states—commonly *Susceptible*, *Infected*, and *Recovered*. Over successive, clearly defined time intervals (days, weeks, or other suitable periods), individuals transition between these states according to precise probabilistic rules. A classic example of such models, widely recognized in epidemiological literature, is the **Reed-Frost Model**. The Reed-Frost model elegantly represents the process as follows:

$$I(t+1) \sim \text{Binomial}\Big(S(t),\ 1-(1-p)^{I(t)}\Big),$$

where:

- $S(t)$ represents the number of susceptible individuals—those still at risk—at the start of time step t,

- $I(t)$ denotes the number of infected individuals capable of transmitting the disease at time t,
- and p symbolizes the probability that a given infected individual transmits the infection to a susceptible individual upon contact within one discrete time step.

The intuition behind the above expression is straightforward yet insightful: each susceptible individual faces a chance of infection dependent on the number of currently infected individuals. Specifically, the expression $1 - (1 - p)^{I(t)}$ reflects the probability of at least one successful transmission from any infected individual in the population to a susceptible individual.

4.2 Core Assumptions of Chain Binomial Models

To keep chain binomial models analytically manageable and conceptually transparent, several simplifying assumptions are made. These assumptions ensure clarity, computational simplicity, and a solid theoretical foundation, although they sometimes abstract away from complex real-world scenarios. The main assumptions typically made in chain binomial models, particularly the Reed-Frost type, include:

- **Closed and Constant Population:**
  - The population size remains fixed throughout the epidemic. There are no births, deaths, or migrations into or out of the population. This assumption isolates disease transmission dynamics from demographic influences, ensuring that any changes observed are solely due to the spread of the infection itself.
- **Discrete-Time Framework:**
  - Epidemics are modeled through clearly defined, regular intervals. This discrete-time representation simplifies data collection and aligns naturally with practical epidemiological reporting schedules—such as daily infection counts or weekly health bulletins.
- **Homogeneous Mixing:**
  - Each member of the population interacts randomly and equally with every other member. This implies that there is no structure based on geography, social groups, age clusters, or behavioral differences influencing contacts. In practice, this assumption means everyone has an equal chance of infecting or becoming infected by any other person.
- **Consistent Transmission Probability:**
  - The probability p of transmitting the disease from an infected individual to a susceptible one remains constant throughout the entire duration of the epidemic. This constancy implies that the infectiousness of the pathogen and the susceptibility of individuals do not fluctuate over time—an idealized yet analytically convenient scenario.

- **Finite, Clearly Defined Disease States:**
  - Individuals in the population are distinctly classified into specific health categories, typically Susceptible, Infected, or Recovered (immune or dead). Transitions between these categories follow clear rules, allowing easy tracking and modeling of the epidemic's progression.
- **Binomial Probability of Infection Events:**
  - At each time step, the number of new infections is determined by a binomial random variable. This assumption embodies the inherent randomness of disease spread—while it may be likely that a given number of infections occur, the exact number remains uncertain. The binomial structure captures the stochastic, unpredictable nature of epidemic evolution.

4.3 Why Adopt These Simplifications?

While each of these assumptions might appear to oversimplify complex real-world conditions, they serve vital roles in both research and practical epidemiology:

- **Analytical Tractability:** These simplifying assumptions allow mathematicians and epidemiologists to derive explicit analytical results and theoretical insights, such as calculating epidemic thresholds, estimating peak infection sizes, and understanding conditions necessary for epidemics to occur or to fade away.
- **Conceptual Clarity:** By reducing the complexity of real-world interactions, these assumptions illuminate fundamental epidemic dynamics, making the underlying mechanisms clearer and more intuitive.
- **Foundation for More Complex Models:** Simple models like the Reed-Frost provide a baseline from which more sophisticated and realistic models can be developed. Researchers often begin with these foundational models and progressively introduce complexities such as age structures, varying infectiousness, geographical spread, or behavioral responses.
- **Alignment with Empirical Data:** The discrete-time steps assumption naturally aligns with how data are routinely collected in real-world epidemics. Epidemiological data commonly appears in discrete intervals (days or weeks), making such models immediately useful for practical data analysis and public health planning.

Thus, despite their simplifications, chain binomial models remain an essential cornerstone of theoretical epidemiology, providing critical insights and robust starting points for more comprehensive modeling strategies.

5 Theory behind Chain Binomial (Reed-Frost)

5.1 Basics

First, let's make some notations. Let S_t, I_t, R_t be the numbers of susceptible, infected and recovered individuals, let p be the probability of infection per contact between susceptible and infected, and let $q = 1 - p$ be the probability of avoiding the infection per contanct.

Recall that in R-F, there are three possible states: Susceptible, Infected and Recovered. Also, contacts between infected and susceptible individuals are considered independent events. Formally, for each susceptible individual s and for each pair of infected individuals i_1, i_2:

$$P(s \text{ infected after contact with } i_1) = P(s \text{ infected after contact with } i_2) = p$$

and events

$$\{s \text{ infected after contact with } i_1\}, \{s \text{ infected after contact with } i_2\}$$

are independent.

Therefore, for each susceptible individual, the probability of being not infected at timestep $t+1$ would be q^{I_t}. Consequently, probability of being infected would be $1 - q^{I_t}$.

Finally, if at timestep t we have S_t susceptibles and I_t infected, we can calculate the probability of each value of I_{t+1}.

$$P(I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t) = \begin{cases} \binom{s_t}{i_{t+1}} (1 - q^{i_t})^{i_{t+1}} (q^{i_t})^{s_t - i_{t+1}}, & i_{t+1} \leq s_t \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Proposition 7. *The number of infected individuals at timestep t follows a binomial distribution.*

$$I_t \sim Binom(S_t, 1 - q^{I_t})$$

Since $s_{t+1} + i_{t+1} = s_t$, we can write transition probabilities between any two pairs $\{s_t, i_t\}, \{s_{t+1}, i_{t+1}\}$:

Proposition 8. *The sequence $\{S_t, I_t\}_{t \geq 0}$ is a two-dimensional Markov Chain with transition probabilities:*

$$P(\{s_{t+1}, i_{t+1}\} | \{s_t, i_t\}) = \begin{cases} \binom{s_t}{i_{t+1}} (1 - q^{i_t})^{i_{t+1}} (q^{i_t})^{s_t - i_{t+1}}, & i_{t+1} + s_{t+1} = s_t \\ 0, & \textit{otherwise} \end{cases}$$

Now we are able to identify absorbing states of this Markov Chains. That would be $\{s_t, i_t\}$ s.t. $P(\{s_t, i_t\} | \{s_t, i_t\}) = 1$. We need to make three observations:

1. Obviously, if s_t and i_t are not equal to zero, then $\{s_t, i_t\}$ is not an absorbing state.

2. If at some point $s_t = 0$, then $s_{t+1} = i_{t+1} = 0$.

3. If $i_t = 0$, then $\{s_t, i_t\}$ is an absorbing state. Indeed:

$$P(\{s_t, 0\}|\{s_t, 0\}) = \binom{s_t}{0}(1-q^0)^0(q^0)^{s_t} = 1$$

Therefore, the only absorbing states are states of type $\{s_t, 0\}$.

The state space is a $\{\{a, b\}|a + b \leq n\}$ and contains $\frac{(n+1)(n+2)}{2}$ states, so the size of the transition matrix is $O(n^4)$, which makes transition matrices of Reed-Frost model unusable. It worth noting that transition matrix is highly sparse.

Before moving further, we make an important definition:

Definition 10 (Basic reproduction number)**.** *Basic reproduction number R_0 of infection is the expectation of number of infections generated by one infected individual.*

In case of R-F the R_0 is taken as $R_0 = p \cdot S_0$

5.2 Epidemics dynamics

Unfortunately, the there is not exact closed formulas of distributions of infected and susceptible individuals for an arbitrary timestep. However, one can try to approximate the dynamics of Reed-Frost modeled epidemics with the system of differential equations under some additional assumptions.

Proposition 9. *Assuming the number of exposures to infected individuals grows linearly within the unit time interval, as the size of the population N grows, the dynamics of the epidemics can be approximated by the system of the differential equations*

$$\begin{cases} \frac{ds}{dt} = R_0 si \\ \frac{di}{dt} = R_0 si - i \\ \frac{dr}{dt} = i \end{cases}$$

where $s(t)$, $i(t)$, $r(t)$ are the susceptible, infected and recovered proportions of the population at time moment t.

Proof. Recall that $I_{t+1} \sim Binom(S_t, (1-(1-p)^{I_t})$. To move to continuous model, we need to make an additional assumption. Initially, we assumed that the probability for a susceptible individual to avoid infection between timesteps t and $t+1$ is $(1-q^{I_t})$. Now we also assume that if the time interval between timestep is a unit time interval 1, then for time interval $\Delta t < 1$, the probability for a susceptible individual to avoid infection between timesteps t and $t + \Delta t$ is $1 - q^{I_t \Delta t}$. The logic behind this assumption is that the number of exposures to infected individuals grows linearly with time interval. Applying this assumption, we can transform our formula to:

$$I_{t+\Delta t} \sim Binom(S_t, (1-(1-p)^{I_t \Delta t})$$

Substitying $R_0 = pN$, we get:

$$(1-p)^{I_t \Delta t} = \left(1 - \frac{R_0}{N}\right)^{I_t \Delta t} = \left(1 - \frac{R_0}{N}\right)^{N i_t \Delta t}$$

Assuming N is large, we can use well known limit:

$$\lim_{N\to\infty}\left(1-\frac{R_0}{N}\right)^{Ni_t\Delta t}=\left(\left(1-\frac{R_0}{N}\right)^{N}\right)^{i_t\Delta t}=e^{-R_0 i_t\Delta t}$$

So $I_{t+\Delta t}\sim Binom(S_t, 1-e^{-R_0 i_t\Delta t})$. $I_{t+\Delta t}$ can be treated as the sum of S_t independent Bernoulli trials with probability of success $1-e^{-Ri_t\Delta t}$. Therefore, as N tends to infinity, the S_t also tends to infinity, and by the law of large numbers,

$$\frac{i_{t+\Delta t}}{s_t}=\frac{I_{t+\Delta t}}{S_t}\approx\mathbb{E}(Bernoulli(1-e^{-R_0 i_t\Delta t}))=1-e^{-R_0 i_t\Delta t} \tag{2}$$

$$i_{t+\Delta t}=s_t(1-e^{-R_0 i_t\Delta t}) \tag{3}$$

$$\tag{4}$$

If Δt is sufficiently small, we can take $e^{-R_0 i_t\Delta t}\approx 1-R_0 i_t\Delta t$:

$$i_{t+\Delta t}=s_t(1-(1-R_0 i_t\Delta t))=R_0 s_t i_t\Delta t \tag{5}$$

$$\tag{6}$$

Next, initially we assumed that all I_t individuals become recovered between timesteps t and $t+1$. Similarly as for the probability of avoiding infection between t and $t+\Delta t$, we assume that $I_t\Delta t$ recover between t and $t+\Delta t$. Therefore we can derive the formula of change of infected individuals:

$$i_{t+\Delta t}-i_t=R_0 s_t i_t\Delta t-i_t\Delta t \tag{7}$$

$$\frac{i_{t+\Delta t}-i_t}{\Delta t}=R_0 s_t i_t-i_t \tag{8}$$

$$\tag{9}$$

Considering the limit case, we finally get:

$$\frac{di}{dt}=R_0 si-i$$

Now, for $s_{t+\Delta t}$ we have $s_{t+\Delta t}=s_t-i_{t+\Delta t}$, so:

$$s_{t+\Delta t}-s_t=-i_{t+\Delta t}=-R_0 s_t i_t\Delta t \tag{10}$$

$$\frac{s_{t+\Delta t}-s_t}{\Delta t}=-R_0 s_t i_t \tag{11}$$

$$\frac{ds}{dt}=-R_0 si \tag{12}$$

Finally, by assumption $I_t\Delta t$ are recovered at time interval Δt $r_{t+\Delta t}=r_t+i_t\Delta t$, so

$$r_{t+\Delta t}-r_t=i_t\Delta t \tag{13}$$

$$\frac{r_{t+\Delta t}-r_t}{\Delta t}=i_t \tag{14}$$

$$\frac{dr}{dt}=i \tag{15}$$

Combining, we get

$$\begin{cases} \frac{ds}{dt} = -R_0 si \\ \frac{di}{dt} = R_0 si - i \\ \frac{dr}{dt} = i \end{cases}$$

□

5.3 Final size of the outbreak

This approximation gives us an opportunity to estimate the final size of the outbreak. The answer depends on R_0. If $R_0 < 1$, that means each infected individual produces on average $R_0 < 1$ infected individuals, which means that epidemics will die without touching the significant proportion of the population. However, if $R_0 > 1$, the outbreak will likely explode. In both cases, we can estimate the final size of the epidemics.

Proposition 10. *The final size of the outbreak can be approximated as the solution of the equation*

$$x = 1 - e^{-R_0 x}$$

Proof. Since $\frac{ds}{dt} = -R_0 si$ and $\frac{dr}{dt} = i$, we have

$$\frac{ds}{dt} = -R_0 s \frac{dr}{dt}$$

From the theory of differential equations, we can find that:

$$s(t) = s(0)e^{-R_0(r(t)-r(0))}$$

Assuming the initial number of infected people is small ($i(0) = 0$), we have $s(0) = 1$, and taking into account that $r(0) = 0$, we have:

$$s(t) = e^{-R_0 r(t)}$$

The final size of the outbreak would be $r(\infty) = \lim_{t\to\infty} r(t)$. Note that since at the end of the outbreak, each individual is either susceptible or recovered, we have $s(\infty) + r(\infty) = 1$. Consequently,

$$s(\infty) = e^{-R_0 r(\infty)} \tag{16}$$

$$1 - r(\infty) = e^{-R_0 r(\infty)} \tag{17}$$

$$r(\infty) = 1 - e^{-R_0 r(\infty)} \tag{18}$$

□

While this solution of this equation doesn't have a simple closed form, it can be solved numerically.

Now, looking closer at this equation, we can see that it has trivial solution $x = 0$. That corresponds to the outcome when epidemics extincts without touching significant proportion of the population.

Taking derivative of $f(x) = x - 1 + e^{-R_0 x}$, we get: $f'(x) = 1 - R_0 e^{-R_0 x}$. It becomes clear that if $R_0 < 1$, $f'(x) > 0\ \forall x > 0$, so the only solution is $x = 0$. However, if $R_0 > 1$, $f'(0) < 0$. That means that $1 - e^{-R_0 x}$ grows faster that x for small x, however, $1 > 1 - e^{-R_0}$, which means that plots of these function intersect for some $x > 0$ implying that there is a non-zero solution. So if $R_0 > 1$ the equation has two solutions, zero and non-zero.

6 Simulation of Real World Scenarios

6.1 Reed-Frost Model

The Reed-Frost Model assumes a constant population size N, and a constant transmission probability p between any two members of the population, at every timestep. We will choose $N = 100000$ and $p = 0.00002$. We begin the simulation with a 3 infected individuals.

At each step we find p_0 the probability of being not being infected by anybody, $p_0 = (1-p)^{I_{t-1}}$. Then we find the number of new infections $I_t \sim \text{Binom}(S_t, 1-p_0)$, where S_t is the susceptible population at timestep t. Plotting the healthy and newly infected populations according to this simulation yields the following graph:

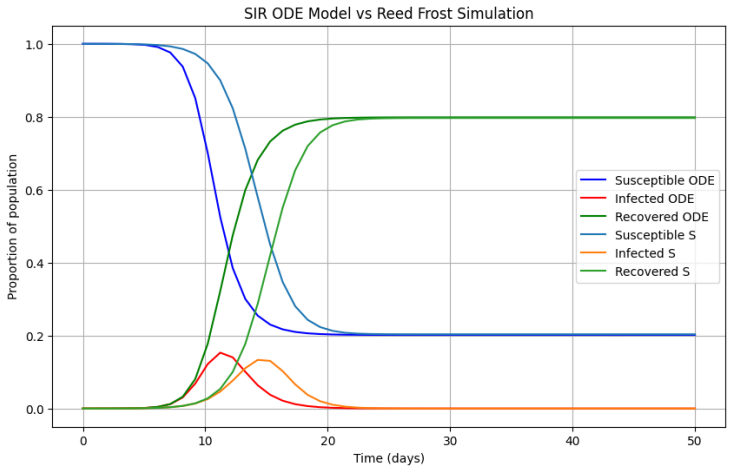

Figure 1: Epidemic dynamic predicted by solving system of ODE and Reed-Frost epidemic simulation showing susceptible, infected and recovered proporions of populations over time.

In this simulation, our reproduction number R_0 is $p \cdot N = 2$, this means that an infected person is expected to cause, on average, 2 infections among the healthy population. The inflection point of the epidemic curve occurs when $1 - \frac{1}{R_0}$ of the susceptible population has been infected; in this case 50% of our starting population, which aligns with our results. As can be seen on Figure 1, the plots of simulation and ODE system solution do not match, and we can observe the shift of 7 days between the same values of corresponding curves. However, it can be seen that recovery proportion in both simulation and analytical solution both tend to one value, 0.8, which means that the system of ODE described in previous section can be succesfully used to predict the final size of the outbreak.

6.2 Modifying the Reed-Frost Model

We can introduce a natural addition to our model by having infected people recover, after a time of being infectious. A recovered person is then once again.

Using the same conditions as above, we introduce a recovery time of 5 timesteps.

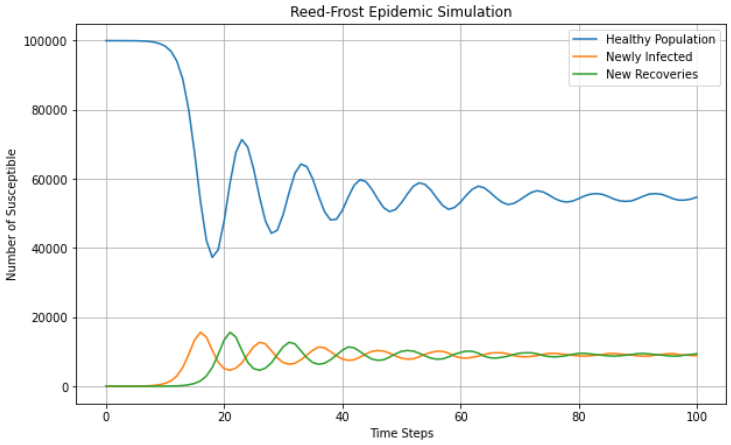

Figure 2: Reed-Frost epidemic simulation showing susceptible, newly infected and newly recovered populations over time.

This introduces some oscillatory behaviour, which can reach an equilibrium when the newly infected and newly recovered populations are approximately equal at each timestep. Increasing the recovery time removes this periodic behaviour, since the epidemic wave will run its course before any infected become susceptible again. Decreasing the recovery time makes the model reach an equilibrium much faster.

6.3 Student Movement on Campus

We simulate the movement of student on campus using the following transition matrix at each timestep:

$$P = \begin{array}{r|ccccc} & \text{AS} & \text{BM} & \text{SC} & \text{LCC} & \text{MD} \\ \hline \text{Alexandra Square} & 0.80 & 0.04 & 0.07 & 0.08 & 0.01 \\ \text{Bowland Main} & 0.30 & 0.70 & 0.00 & 0.00 & 0.00 \\ \text{South Campus} & 0.00 & 0.00 & 0.90 & 0.05 & 0.05 \\ \text{Lancaster City Centre} & 0.10 & 0.00 & 0.00 & 0.90 & 0.00 \\ \text{Mathematics Department} & 0.15 & 0.00 & 0.05 & 0.00 & 0.80 \end{array}$$

Each row represents the pmf of a student's next location given their current location. For example, a student currently at Alexandra Square has an 80% probability of remaining there, a 4% probability of moving to Bowland Main, and so on.

We begin the simulation with 10000 students in Alexandra Square and 5000 in the city centre, and simulate 25 timesteps:

Let $\mathbf{x}^{(n)}$ be the vector of students at different locations on campus at timestep n. For example $\mathbf{x}^{(0)} = [10000, 0, 0, 5000, 0]$. At each timestep:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} P$$

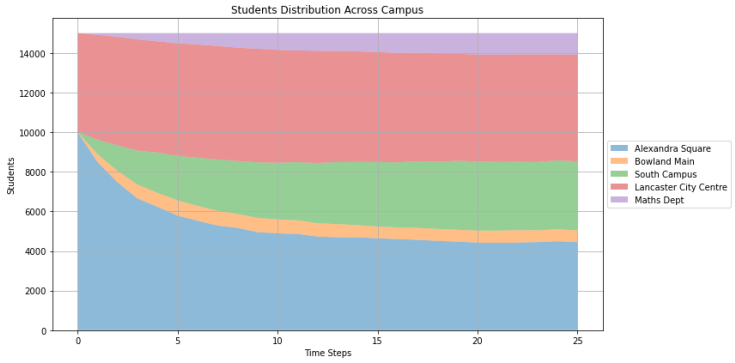

Figure 3: Simulation of student movement on campus.

so

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)} P^n$$

Since our Markov chain is irreducible and aperiodic, by the Perron-Frobenius theorem for stochastic matrices, we have:

$$\lim_{n\to\infty} P^n = \mathbf{1}\pi^T$$

Where π is the unique stationary distribution and $\mathbf{1}$ is a column vector only containing ones. So the system will converge to the stationary distribution π.

The stationary distribution π is a probability vector that satisfies:

$$\pi P = \pi$$

In this case:

$$P^{1000} = \begin{bmatrix} 0.29126214 & 0.03883495 & 0.24133148 & 0.35367545 & 0.07489598 \\ 0.29126214 & 0.03883495 & 0.24133148 & 0.35367545 & 0.07489598 \\ 0.29126214 & 0.03883495 & 0.24133148 & 0.35367545 & 0.07489598 \\ 0.29126214 & 0.03883495 & 0.24133148 & 0.35367545 & 0.07489598 \\ 0.29126214 & 0.03883495 & 0.24133148 & 0.35367545 & 0.07489598 \end{bmatrix}$$

so our limiting distribution is (2.d.p) $[0.29, 0.04, 0.24, 0.35, 0.08]$, which matches the results of our simulation.

References

[1] Pishro-Nik, H. (2014) Introduction to probability, statistics, and random processes. Available at: `https://www.probabilitycourse.com, KappaResearchLLC,2014` (Accessed: March 2025).

[2] ABBEY, H. (1952) 'AN EXAMINATION OF THE REED-FROST THEORY OF EPIDEMICS', Human biology, 24(3), pp. 201–233.

[3] Leveque, O. (no date) *Lecture notes on Markov chains*. Available at: `https://ipgold.epfl.ch/~leveque/Lecture_Notes/markov.pdf` (Accessed: March 2025).

[4] Ludwig, D. (1975) 'Final size distribution for epidemics', Mathematical biosciences, 23(1), pp. 33–46. Available at: https://doi.org/10.1016/0025-5564(75)90119-4.

[5] Lefevre, C. and Picard, P. (1990) 'A non-standard family of polynomials and the final size distribution of Reed-Frost epidemic processes', Advances in applied probability, 22(1), pp. 25–48. Available at: https://doi.org/10.1017/S0001867800019315.

[6] Miller, J.C. (2012) 'A Note on the Derivation of Epidemic Final Sizes', Bulletin of mathematical biology, 74(9), pp. 2125–2141. Available at: https://doi.org/10.1007/s11538-012-9749-6.

[7] Halloran, M.E., Longini, I.M. and Struchiner, C.J. (2009) Design and Analysis of Vaccine Studies. New York, NY: Springer New York.