# The Effect of Movement on the Early Phase of an Epidemic

by

Jason Rose

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Mathematics
University of Manitoba
Winnipeg

**Abstract**

A Markov chain model for the early stochastic phase of the transmission of an infectious pathogen is studied, investigating its properties in the case of an isolated population and of two coupled populations with explicit movement of infectious individuals. Travel was found to play a role in the early development and spread of an infectious disease, particularly in the case of differing basic reproduction numbers in the connected locations.

**Acknowledgements**

I would first like to thank my thesis supervisor Dr. Julien Arino. I will be forever grateful for his patience and expertise throughout my studies. I also wish to thank the Department of Mathematics for financial support throughout my studies, and in particular Professor Steve Kirkland for the opportunities granted to me over the duration of my degree.

I would also like to thank my parents, Terry and Breda Rose, for their endless support even after I moved across the world to pursue my studies. Most of all, I wish to thank my partner Jane for her understanding and help throughout my time studying. I could not have achieved what I have without her.

# Contents

**Introduction** **1**

Motivation . . . . . 1

Compartmental framework for disease modelling . . . . . 4

Spatial aspects of disease modelling . . . . . 5

The early phase of an epidemic . . . . . 7

Outline of thesis . . . . . 8

**1 Mathematical Background** **10**

1.1 Stochastic processes . . . . . 10

1.2 Finite Markov chains . . . . . 12

1.2.1 Introduction to Markov chains . . . . . 12

1.2.2 Some additional terminology from matrix theory . . . . . 16

1.2.3 Directed graph representation of a Markov chain . . . . . 16

1.2.4 Limiting behaviour of an ergodic Markov chain . . . . . 21

1.2.5 Limiting behaviour of a non-ergodic Markov chain . . . . . 22

1.2.6 Absorption probability . . . . . 30

1.2.7 Individual realizations of the Markov chain . . . . . 32

**2 Epidemiological Background** **34**

2.1 Compartmental models . . . . . 34

2.2 Compartmental models in epidemiology . . . . . 35

2.2.1 Movement between compartments . . . . . 36
2.2.2 Some common assumptions . . . . . 38
2.3 Spatial aspects in epidemiology . . . . . 39
2.3.1 Bartlett, 1956 . . . . . 39
2.3.2 Baroyan and Rvachev, 1969, 1971 . . . . . 40
2.3.3 Kermack-McKendrick style metapopulation models . . . . . 40
2.4 The early phase of an epidemic . . . . . 41
2.5 Stochastic models in epidemiology . . . . . 42
2.5.1 Chain binomial models [2] . . . . . 42
2.6 Methods of measuring the expected effects of a disease outbreak . . . 43
2.6.1 Basic reproduction number . . . . . 44
2.6.2 Attack rate . . . . . 46
2.6.3 Final size of an epidemic . . . . . 47

**3 A Markov chain model for an SIS-type disease in an isolated location 48**

3.1 The model . . . . . 48
3.2 Mathematical analysis . . . . . 50
3.2.1 Underlying deterministic model . . . . . 50
3.2.2 Analysis of the Markov chain . . . . . 53
3.3 Numerical simulations . . . . . 55
3.3.1 Parameters . . . . . 55
3.3.2 Effect of $N$, $P$ and $\Delta t$ . . . . . 55
3.3.3 Investigations into the probability of disease extinction . . . . 60
3.3.4 Final size of the epidemic . . . . . 62

**4 Two interconnected locations 65**

4.1 The model . . . . . 65

4.2 Transition probabilities . . . . . 68
4.3 States with a given prevalence . . . . . 70
4.4 Transition matrix . . . . . 73
4.4.1 Infection submatrices . . . . . 74
4.4.2 Recovery (and death) submatrices . . . . . 77
4.4.3 Movement (and absence of transition) . . . . . 80
4.5 Mathematical analysis . . . . . 82
4.6 Numerical simulations . . . . . 84
4.6.1 Algorithm for two-location numerics . . . . . 84
4.6.2 Terminology . . . . . 84
4.6.3 Travel rates . . . . . 85
4.6.4 Time step for connected populations, $\Delta t$ . . . . . 85
4.6.5 Investigating the impact of travel . . . . . 86
4.6.6 Direct comparison to the single location case . . . . . 89
4.6.7 Final size of the epidemic . . . . . 92
4.6.8 Further investigations into the final size of the epidemic for the two locations case . . . . . 93

**5 Conclusion** **97**

**Bibliography** **99**

# List of Figures


1 Number of enplaning and deplaning passengers in Canada from 1997 to 2014 [12] . . . . . 3

2 Estimated prevalence of influenza in two neighbouring regions of France for the 2011-2012 epidemic season. . . . . 6

3 Positive test results for influenza in the Canadian Prairies from week ending 27-07-2013 to week ending 26-07-2014. . . . . 7

1.2 Three graphs representing (a) an absorbing Markov chain, (b) an ergodic chain, and (c) a regular Markov chain. . . . . 18

1.3 Weighted directed graph representing the transition matrix (1.12) . . 29

2.1 Flow diagram for an SIS model. . . . . 37

3.1 Random walk used in isolated locations. . . . . 49

3.2 Value of the upper bound for $\Delta t$ given by (3.9) as a function of the absorption threshold $N$, when $\mathcal{R}_0 = 1$. . . . . 57

3.3 Probability of a disease outbreak as a function of the number of cases required for an outbreak ($N$) and the initial number of infectious cases. 59

3.4 Probability of disease extinction due to variations of $\mathcal{R}_0$ and the number of infectious individuals in a population. . . . . 61

3.5 Probability of disease extinction due to variations of $\mathcal{R}_0$ and the number of infectious individuals in a population. . . . . 61

3.6 Average final size of the epidemic for $0.5 \leq \mathcal{R}_0 \leq 2$. . . . . 63
3.7 Average duration of the stochastic phase when the initial number of infected varies from 1 to 25 and $\mathcal{R}_0$ varies from 0.3 to 3. . . . . 64

4.1 The first few states of the chain corresponding to a maximum prevalence of 3 in the coupled two-locations system, when $N^A > 3$ and $N^B > 3$. The doubly circled state is absorbing. . . . . 67
4.2 States directly connected to a state with $i$ infectious individuals in location $A$ and $j$ individuals in location $B$. . . . . 68
4.3 States of the chain when $N^A = 2$ and $N^B = 3$. Doubly circled states are absorbing. . . . . 70
4.4 Probability of disease extinction as a function of $\mathcal{R}_0$ in the absence (dashed curve) and presence (continuous curve) of travel between locations. . . . . 87
4.5 Difference in Figure 4.4 (travel minus no travel) between the probabilities of disease extinction. . . . . 87
4.6 Effect of variation of travel on the probability of disease extinction when $\mathcal{R}_0(A) = 1$ and $\mathcal{R}_0(B) = 2$. . . . . 88
4.7 Change in probability of disease extinction due to variation of $\mathcal{R}_0(B)$ with $\mathcal{R}_0(A) = 1$ and travel rates fixed. . . . . 89
4.8 Probability of disease extinction with varying values of $\mathcal{R}_0$ and initial number of infected individuals in location A. The number of initially infected individuals in location B is assumed to be zero. . . . . 90
4.9 Effect of travel on the probability of extinction when $\mathcal{R}_0$ is allowed to vary in each location. . . . . 91
4.10 Probability of overall disease extinction given that $\mathcal{R}_0(A)$ is 1, with varying $\mathcal{R}_0(B)$ and travel rates. . . . . 92

4.11 Final size of the epidemic with $\mathcal{R}_0$ varying from 0.5 to 2 starting with one infected individual in location A with the corresponding result for the one location case overlaid. . . . . 94
4.12 Final size of the epidemic with $\mathcal{R}_0$ varying from 0.5 to 2 starting with five infected individuals in location A with the corresponding result for the one location case overlaid. . . . . 95
4.13 Average duration of the stochastic phase for two locations (continuous curve) compared with one location (dashed curve) . . . . . 96

# List of Tables

4.1 Meaning of the symbols used in Figures 4.1 and 4.2. . . . . 67

4.2 Transitions and corresponding probabilities in the model for two locations A and B, assuming that the states and transitions are allowable. See the text for details. . . . . 68

# Introduction

## Motivation

The movement of infectious diseases has been occurring and studied for millennia. Rasmussen *et al.* in 2015 [34] found traces of Yersinia Pestis (the bacterium which causes plague) in the teeth of seven skeletons - the oldest of which dating over 5000 years old. In [39], Thucydides, a historian and political philosopher (460 - 400 BC) described a "plague" descending upon Athens. In slightly more recent times (1347-1351 AD) the Black Death, one of the most devastating plagues ever to hit humanity, arrived in Europe following trade routes from the far east and was responsible for the deaths of an estimated 30-50% of European populations [16].

The $20^{\text{th}}$ century saw the emergence of a number of newer diseases including Lyme disease (1975), Legionnaire's disease (1976), the human immunodeficiency virus (HIV), hepatitis C (1989), hepatitis E (1990), and hantavirus (1993) [20]. In 2003, Severe Acute Respiratory Syndrome (SARS) surfaced for the first time in China, becoming the first pandemic threat of the $21^{\text{st}}$ century. Through measures invoked by the World Health Organization which included heightening public awareness, screening of international travellers, isolation of infected individuals and quarantining of close contacts with infected individuals, the spread of the disease was successfully arrested [36].

Aside from the threat to human life, many infectious diseases carry a severe

monetary cost as well. For example, in 2010, Thompson *et al* showed that the monetary cost of eradication of poliomyelitis (polio) is smaller in the long run than control measures (i.e., less intense vaccination programs) [38]. In US dollars, they estimate the worldwide net benefits of global eradication of polio to be \$40-\$50 billion between 1988 and 2035. Between 1967 and 1980, a global effort was made to achieve the eradication of smallpox. The World Health Organisation (WHO) approximates the cost of the scheme to be \$300 million US dollars, but since then, the annual worldwide benefit of the eradication of smallpox is estimated to be over \$1 billion [33].

With the constant emergence and reemergence of infectious diseases along with increasing interconnectivity of virtually every region of the world due to air travel, it has become more and more important to study the dynamics of disease transmission and to build frameworks for studying disease transmission which can be adapted to different diseases. Consider Figure 1, generated from the data in [12]. From the figure, we can clearly see a steady upward trend in the number of enplaning and deplaning passengers in Canada since 2003, except for the small slump in 2009 coinciding with the worldwide economic downturn that started at that time.

With these rising figures in air travel comes an increased danger in worldwide pandemics. Aside from the increase in air travel, global interconnectedness is increasing in other ways as well. For example, in Canada, the number of vehicles per 1000 members of the population rose from 292 to 581 from 1960 to 2002 [15]. With increased number of vehicles comes increased interconnectivity between people living in cities and those in the rural communities or satellite towns. Every additional passenger travelling out of a disease-stricken region brings with him or her an additional risk of the movement of an infectious disease, and with these increased dangers comes an increased need for disease modelling capable of accounting for these new dangers.

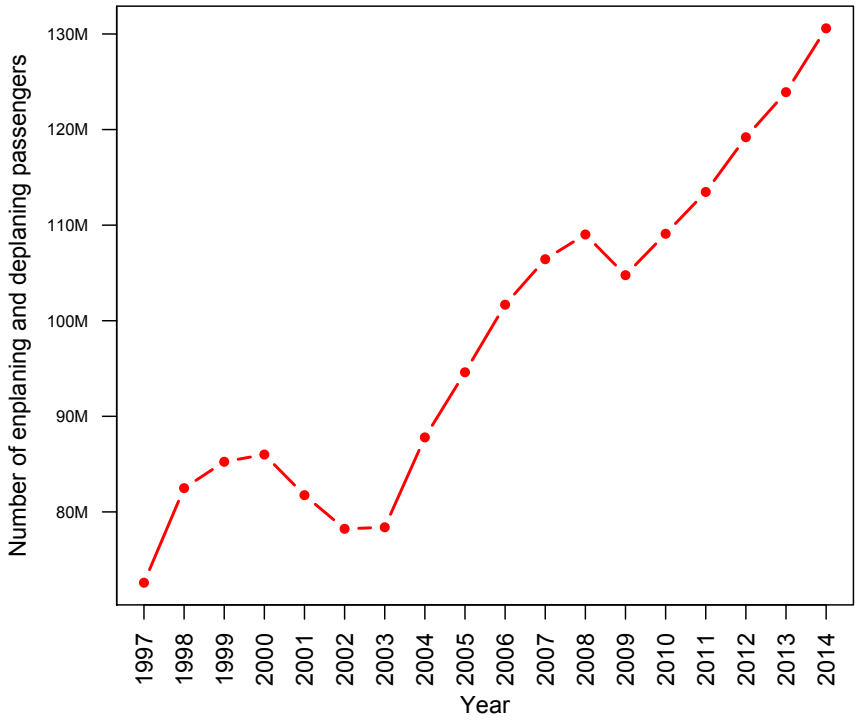



Figure 1: Number of enplaning and deplaning passengers in Canada from 1997 to 2014 [12]

.

# Compartmental framework for disease modelling

In order to model disease movement, it is convenient to divide the at-risk population into distinct *compartments* representing the affected population's status with respect to the disease in question. For example, when a disease is active in a population, some members of the population have not yet contracted the disease, so those individuals would be classified as *susceptible*. Also present in the population are individuals who have already contracted the disease and are currently *infectious*, and are classified as such. Depending on the disease, there can be other compartments; for example, some diseases leave individuals with immunity to the disease after the infectious period has ceased. For those particular diseases another compartment is required for some of the affected individuals; a *recovered* compartment. We would call a disease an SIR disease if it causes an individual to undergo the transitions described above (susceptible $\rightarrow$ infected $\rightarrow$ recovered). If there is no recovered class (i.e., the disease does not grant immunity to future infection), then the individuals who are infected return to the susceptible class for the possibility of future infection, making it an SIS-type disease.

Compartmental models may seem simple, but this style of modelling allows us to find a number of quantities which are useful in assessing the level of threat of an infectious disease. The use of compartmental modelling involves parameters describing rates of movements between compartments. With the help of these parameters and others, some very useful information about the expected progression of the disease in the population can be obtained.

One of these quantities is known as the *basic reproduction number* and denoted $\mathcal{R}_0$. The basic reproduction number is one of the more popular indicators of the level of threat of a disease, and is defined as the number of secondary infections caused by one infectious individual in a completely susceptible population. If $\mathcal{R}_0$ is greater than one, then (in general) it can be expected that the disease outbreak

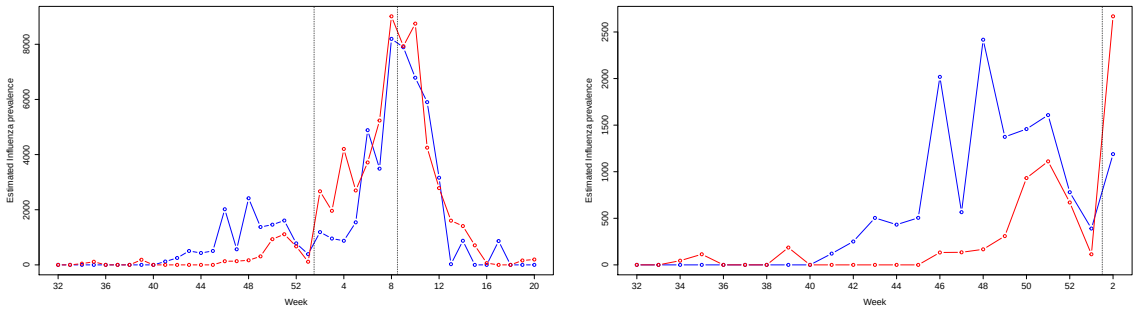
will be large, while if it is less than one (supposing that the possible disease event is an epidemic rather than endemic situation) we can conclude that the disease outbreak will not be a major one (if the disease may become endemic, then in general, $\mathcal{R}_0 < 1$ leads to the disease becoming extinct). The information required to calculate the basic reproduction number includes the probability that a contact between an infectious and a susceptible person results in a new infection, the contact rate between susceptible and infectious individuals and the average duration of the infection. In many cases, this information may be difficult to find but for certain disease types, there are other ways of formulating the basic reproduction number, as we will see in Chapter 2.

## Spatial aspects of disease modelling

For the above description of compartmental models we need to assume homogeneity of the population which in many cases is an unrealistic assumption. Consider a population in which large groups of people are clustered together in isolated locations – Canada for instance. It is unrealistic to assume individuals mix homogeneously in this setting. Two individuals both living in Vancouver are (in general) more likely to come into contact than one individual living in Vancouver and another in Toronto. Clearly, the physical contact required for the transmission of an infectious disease can only take place when individuals are in the same location. Aside from the probabilities of contact, a given disease is more likely to spread faster in a densely populated region than in a sparsely populated one [37], requiring different parameters to be used in the modelling; so different locations with differing population densities further complicates the situation.

A workaround for this issue might be to model the disease in each city or isolated location separately. This would allow us to adjust the values of the parameters

according to each individual location's needs, but it would not account for travel between the locations.



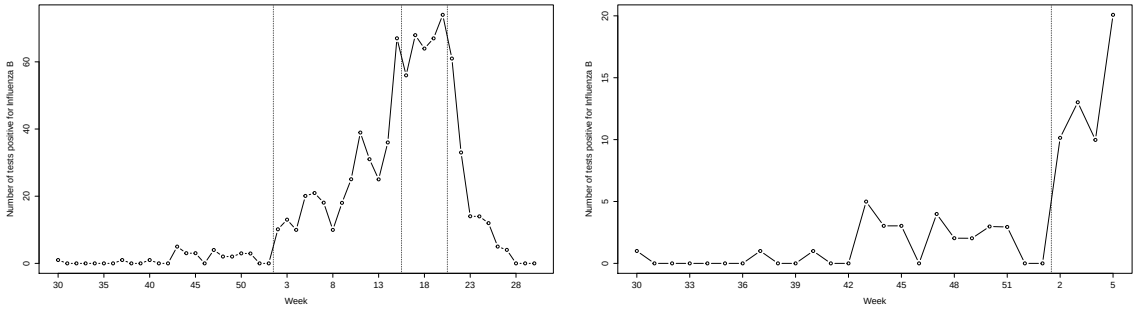

(a) One year. (b) Zoom on first phase.

Figure 2: Estimated prevalence of influenza in two neighbouring regions of France for the 2011-2012 epidemic season.

In order to better understand the role of space, we now consider the data in Figure 2, which shows the estimated prevalence (i.e., the number of infected individuals per 100,000 inhabitants) in two regions of France during the 2011-2012 winter influenza season. The data is obtained from Réseau Sentinelles. The regions under consideration are in the south west of the country: Aquitaine (blue/darker curve) and Midi-Pyrénées (red/lighter curve). Here, the epidemic can be decomposed in roughly three phases shown by the vertical lines in Figure 2a. Focusing on the initial phase of the outbreak, we observe that there are times when the disease is absent in one region and present in the other, and vice-versa. Also, during the initial phase there are periods of time when the prevalence in Aquitaine undergoes wide variations while that in Midi-Pyrénées increases steadily. Both regions then undergo a sharp drop in prevalence before actually going into the later phases in which we see a much larger prevalence. Clearly, if there were no importations of cases into these regions, the epidemic should have died out in both. More details are given on the work done in accounting for the movement of disease across separate locations in Chapter 2.

# The early phase of an epidemic

As can be seen in Figure 2, before infectious diseases become full-blown epidemics, they go through an initial phase during which the number of infected individuals rises and falls in a manner subject to probabilistic effects. We will refer to this period as the "stochastic phase" or "initial phase" of the outbreak. For instance, consider the weekly count of positive influenza tests for the three provinces of the Canadian Prairies (Alberta, Saskatchewan and Manitoba) as given by the Public Health Agency of Canada's Respiratory Virus Detection Surveillance System [32]. Clearly, the epidemic in Figure 3 can be decomposed into several phases. During an




(a) One year. (b) Zoom on first phase.

Figure 3: Positive test results for influenza in the Canadian Prairies from week ending 27-07-2013 to week ending 26-07-2014.

initial phase lasting up to about week 1, the number of laboratory confirmed cases is very low, with a maximum value of 5. The second phase then begins and although there are variations, it is clear that the number of cases is increasing sharply until week 15, at which point the rate of apparition of new cases slows and then starts diminishing. Figure 3b provides a closer look at the initial phase of the outbreak.

Of course, the data in Figure 3 should be used with caution: these are laboratory confirmed cases, which represent but a fraction of the actual cases. Also of importance is the spatial resolution of the data: the Prairies extend for almost 1,400 kilometres along the border with the USA and there are almost 2,000 kilometres from the northwestern-most point in Alberta to the southeastern-most point in Manitoba.

## Outline of thesis

In this thesis, I will focus exclusively on the stochastic phase of an epidemic. In order to do so, it is assumed that there is a prevalence level above which one does not need to consider stochastic effects anymore: if prevalence exceeds this threshold, the disease becomes epidemic. Many questions will be addressed regarding the ramifications of movement between two locations affected by a disease outbreak. For example, does the potential movement of infectious individuals between two coupled locations render a full-blown epidemic more likely? Does it have an effect on the expected duration of the stochastic phase? If the two locations in question have different basic reproduction numbers, does this make an epidemic more likely and if so, how much more likely? Must travel be halted completely to stop the spread of disease or can its spread be arrested by reducing travel by some amount? If so, by how much must it be reduced?

To answer these questions, we set up a stochastic model. Deterministic models describe what happens "on average" in a population and are therefore more suited to analysing situations with larger populations [41]. In deterministic models, the number of individuals in compartments can take non-integer values, and this makes less sense when dealing with a small number of individuals. Furthermore, stochastic models allow the number of people who move from susceptible to infected to vary through chance, rather than at a sort of "predetermined" rate as in a deterministic model. In examining the flutters mentioned above, it is more practical to allow this element of randomness into the model, as it is more reflective of the real life situation.

In [3], three models are discussed; an SIS model with constant population size, variable population size, and an SIR model with constant population size. For each, the deterministic system is given, followed by a corresponding stochastic model. We are interested in building on the stochastic SIS model of constant population. First, we rework the one-location SIS model in [3], then consider the situation of two

isolated locations connected by transport. In the one-location stochastic SIS model in [3], the model is first set up as a Markov chain where transition probabilities follow a Poisson process and the time between transitions is exponentially distributed. Next, a closer relationship is shown between the deterministic and stochastic models when another probability distribution called the quasi-stationary distribution is used. The mean number of infected individuals is then calculated for both probability distribution systems and compared with the deterministic system, followed by the relation of the system to a random walk (when the total number in the population is high), calculations regarding the expected duration of the epidemic and further numerical results.

Few works have considered discrete-time metapopulation models. When they have, a reaction-diffusion process is regularly used to model the inter-location movement [13, 14, 29, 31]. Here, we first use a Markov chain (which will be defined in the next chapter) to re-formulate the model in an isolated location (with no transport in or out), then consider distinct locations with their Markov processes coupled to form one new Markov process.

# 1

# Mathematical Background

The following material is a summary of the mathematics used in this work. A more detailed account on any of these topics can be found in [10, 11, 17, 24, 25, 35].

## 1.1 Stochastic processes

In order to properly define a stochastic process, it is necessary to use the following definitions.

**Definition 1.1.** Let $\Omega \neq \emptyset$ be a set, and let $2^\Omega$ represent the power set of $\Omega$ (the set of all subsets of $\Omega$). Then $\mathcal{A} \subset 2^\Omega$ is a **$\sigma$-algebra** if:

1. $\Omega \in \mathcal{A}$.
2. $\mathcal{A}$ is closed under complements, i.e., $A^C := \Omega \backslash A \in \mathcal{A}$ for any $A \in \mathcal{A}$.
3. $\mathcal{A}$ is closed under countable unions, i.e., $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ for any choice of countably many sets $A_1, A_2, \ldots \in \mathcal{A}$.

**Definition 1.2.** A pair $(\Omega, \mathcal{A})$, with $\Omega$ a nonempty set and $\mathcal{A} \subset 2^\Omega$ a $\sigma$-algebra is a ***measurable space*** with $A \in \mathcal{A}$ ***measurable sets***.

**Definition 1.3.** Let $\mu : \mathcal{A} \to \mathbb{R}$. $\mu$ is a ***measure*** if it satisfies

1. For all $A \in \mathcal{A}$, $\mu(A) \geq 0$.

2. $\mu(\emptyset) = 0$.

3. For all countable collections of pairwise disjoint sets $\{A_i\} \in \mathcal{A}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

**Definition 1.4.** For $\Omega$, $\mathcal{A}$ and $\mu$ defined as above, the triple $(\Omega, \mathcal{A}, \mu)$ defines a ***measure space***. If, in addition, $\mu(\Omega) = 1$, then $(\Omega, \mathcal{A}, \mu)$ defines a ***probability space*** and the sets $A \in \mathcal{A}$ are ***events***. $\mu$ is then denoted by $\mathbb{P}$ (for probability).

**Definition 1.5.** Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, and $(S, \xi)$ be a measurable space. A ***random variable*** is a function $X : \Omega \to S$. $S$ is often referred to as the ***state space***. If $S$ is countable, then the random variable can be called a ***discrete*** random variable.

**Definition 1.6.** Given a probability space $(\Omega, \mathcal{A}, \mu)$ and a measurable space $(S, \xi)$, a ***stochastic process*** is a collection of random variables indexed by a totally ordered set $T$, usually representing time.

**Definition 1.7.** If $X$ is a finite, discrete random variable which can take the values $x_1, \ldots, x_k$ with probability $p_1, \ldots, p_k$ respectively, then the ***expected value*** or ***expectation*** of $X$ is

$$E(X) = \sum_{i=1}^{k} p_i x_i.$$

**Definition 1.8.** If $X$ is a finite, discrete random variable which can take the values $x_1, \ldots, x_k$ with probability $p_1, \ldots, p_k$ respectively, then the ***variance*** is defined as

$$Var(X) = \sum_{i=1}^{k} p_i \left(x_i - E(X)\right)^2.$$

## 1.2 Finite Markov chains

### 1.2.1 Introduction to Markov chains

In general, for discrete-time stochastic processes, the value of the random variable $X_n$ will depend upon the values of the earlier random variables $X_{n-1}, X_{n-2}, \ldots, X_0$. Therefore, we are often interested in the conditional probability $\mathbb{P}[X_{n+1} = s_{n+1} | X_n = s_n, X_{n-1} = s_{n-1}, \ldots, X_0 = s_0]$, the probability the random variable $X_n$ takes the value $s_n$ given all the values of the previous random variables.

The possible values of the random variables $X_i$ form a countable set $S = \{s_1, s_2, \ldots\}$ which is called the ***state space***.

**Definition 1.9.** A stochastic process is called a ***Markov chain*** if it satisfies the property that

$$\mathbb{P}[X_{n+1} = s_{n+1} | X_n = s_n, \ldots, X_0 = s_0] = \mathbb{P}[X_{n+1} = s_{n+1} | X_n = s_n].$$

The full definition of a Markov chain is much more involved than that given above, but for the purpose of this thesis, the above definition is sufficient. For the remainder of this chapter, assume all Markov chains mentioned are finite Markov chains. i.e., there are a countable and finite set of states for the chain to occupy. For a Markov chain $X = \{X_n | n = 0, 1, \ldots\}$, the probability

$$p_{ij} = \mathbb{P}[X_k = s_j | X_{k-1} = s_i].$$

is the transition probability from state $i$ to state $j$.

Given a finite state space $\{s_1, \ldots, s_n\}$ and transition probabilities $p_{ij}$, we form

the one-step transition matrix of the Markov chain

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & & \ddots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}. \tag{1.1}$$

This matrix is ***stochastic*** (all rows sum to one) since

$$\begin{aligned} \sum_{j=1}^{n} p_{ij} &= \sum_{j=1}^{n} \mathbb{P}[X_1 = s_j | X_0 = s_i] \\ &= \mathbb{P}[X_1 \in S | X_0 = s_i] \\ &= 1, \end{aligned}$$

because $S$ contains $s_i$ and all other states to which the system can move from $s_i$, and that transition must take place.

Suppose we want to find the probability $p_{ij}^{(2)}$ that the system is in state $j$ after two transitions given it started in state $i$. It is easy to see that the transition from state $i$ to state $j$ in two steps can be achieved by the system moving from state $i$, to any state $k$ in the first step and then from that state $k$ to state $j$ in the second, with probability $p_{ik}p_{kj}$. Therefore $p_{ij}^{(2)}$ can be obtained by summing $p_{ik}p_{kj}$ for all $k$, so that

$$p_{ij}^{(2)} = \sum_{k=1}^{n} p_{ik}p_{kj} \tag{1.2}$$

Notice that this computation can be described by multiplying the $i^{th}$ row of the matrix $P$ by the $j^{th}$ column of $P$. This will be true for all $i$ and $j$, so we can see that the ***two-step transition matrix***, and inductively, the $n$***-step transition***

***matrix*** are given by $P^2$ and $P^n$ respectively so that

$$P^n = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \cdots & p_{1n}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \cdots & p_{2n}^{(n)} \\ \vdots & & \ddots & \\ p_{n1}^{(n)} & p_{n2}^{(n)} & \cdots & p_{nn}^{(n)} \end{bmatrix}. \tag{1.3}$$

It is easy to see that the matrix in (1.3) is stochastic from a similar argument to that in (1.2). In fact, the product of any two stochastic matrices is also stochastic: Let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ stochastic matrices. The $(i,j)^{th}$ entry of $AB$ is $\sum_{k=1}^{n} a_{ik}b_{kj}$. Summing over $j$, we get that the sum of the $i^{th}$ row is

$$\begin{aligned} \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ik}b_{kj} &= (a_{i1}b_{11} + \cdots + a_{in}b_{n1}) + \cdots + (a_{i1}b_{1n} + \cdots + a_{in}b_{nn}) \\ &= a_{i1}(b_{11} + \cdots + b_{1n}) + \cdots + a_{in}(b_{n1} + \cdots + b_{nn}) \\ &= a_{i1} + \cdots + a_{in} \\ &= 1, \end{aligned}$$

which will be true for every row $i$ of the matrix $AB$.

**Definition 1.10.** The following are useful definitions regarding the states of a Markov chain.

- We say that a state $s_j$ is ***accessible from*** a state $s_i$, and write $s_i \to s_j$, if there exists a $t$, such that $p_{ij}^{(t)} \neq 0$. That is, $s_i \to s_j$ if it is possible to get from state $i$ to state $j$ in a finite number of steps.
- $s_i$ and $s_j$ are said to ***communicate*** if $s_i \to s_j$ and $s_j \to s_i$.
- A state $s_i$ is called ***essential*** if for all $j$ such that $s_i \to s_j$, it is also true that $s_j \to s_i$.

- A state that is not essential is called ***inessential***, i.e., $s_i$ is inessential if there exists a $j$ such that $s_i \to s_j$ but $s_j \nrightarrow s_i$. Inessential states are sometimes called ***transient***.

- All essential states can be grouped into ***essential classes*** in which all states in the class communicate with each other. In each class, there are no transitions leading out of the class. The same can be done for the inessential states although there may be transitions from inessential states leading out of the class.

- An essential state which forms an essential class on its own is called ***absorbing***.

Using the above definitions regarding the relationships between the states of a Markov chain, we now give some results and definitions in relation to the Markov chain as a whole.

**Definition 1.11.**

- A Markov chain is said to be ***absorbing*** if

  1. There is at least one absorbing state in the Markov chain.

  2. The set of absorbing states is accessible from any non-absorbing state.

- A Markov chain is called ***ergodic*** if it is possible get from any state $i$ to any state $j$ in some number of steps.

- A matrix $P$ representing the transition matrix of a Markov chain, is called ***regular*** if there exists some number $t$ such that all entries of $P^t$ are positive.

- A Markov chain is called ***periodic*** if every state of the Markov chain is periodic, i.e., every state $i$ has the property that every return to state $i$ must occur in multiples of some integer $d > 1$.

### 1.2.2 Some additional terminology from matrix theory

The following results apply only to non-negative matrices. A non-negative matrix is a matrix containing only non-negative entries. If a matrix $M$ is non-negative, we write $M \geq 0$. Note also that if $M$ is an $n \times n$ (square) matrix, we write $M \in \mathcal{M}_n$.

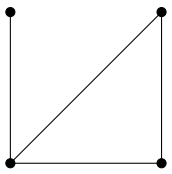
**Definition 1.12.** Let $M \in \mathcal{M}_n$ and $M \geq 0$.

- $M$ is said to be ***irreducible*** if for all $i$ and $j$, there exists some number $t$ such that the entry $m_{ij}^{(t)}$ is positive; i.e., every state is accessible from every other state. It is ***reducible*** if it is not irreducible. (A characterization of the latter is that with the necessary ordering of states, the matrix can be put in block lower triangular form.)

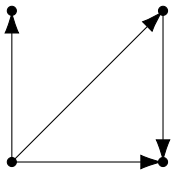
- $M$ is said to be ***primitive*** if there exists some number $t$ such that all entries of $M^t$ are positive; i.e., there exists some number $t$ such that every state $s_j$ is accessible from every other state $s_i$ in exactly $t$ steps.

### 1.2.3 Directed graph representation of a Markov chain

**Definition 1.13.** A ***graph*** is an ordered pair $G = (V, E)$, where $V$ is a set of vertices and $E$ is a set of edges. A ***directed graph (digraph)*** is a graph where each edge has a direction. In this case, edges are often called ***arcs***.



(a) A graph



(b) A directed graph

Before we examine the relationship between graph theory and Markov chains, let us first go through some basic definitions and concepts in graph theory. For the

purpose of this discussion, all graphs will be finite (graphs with a finite number of edges and vertices).

**Definition 1.14.** Let $G$ be a graph and $D$ be a digraph. Then

- We say two vertices in $G$ are ***adjacent*** if there is an edge or arc joining them.
- Suppose there are $n$ vertices in a directed graph $D$. The ***adjacency matrix*** $A$ of the digraph $D$ is the matrix of zeros and ones, such that the entry $(i, j)$ of $A$ is 1 if there is an arc from vertex $i$ to vertex $j$, and 0 if there is no arc connecting vertex $i$ to vertex $j$.
- If a vertex $v$ is an initial point or endpoint of an edge or arc $e$, we say $v$ is ***incident*** to $e$.
- A ***walk*** is an alternating sequence of vertices and edges or arcs, starting and ending in a vertex, where each vertex in the sequence is adjacent to the next vertex in the sequence. If the vertices are distinct, we call the walk a ***path***. If in addition, the edges or arcs are distinct we call the walk a ***trail***. If we have a path starting and ending on the same vertex, we call it a ***cycle***.
- A digraph is called ***strongly connected*** if there is a path between each pair of vertices in $D$ in each direction.
- A ***strongly connected component*** of $D$ is a maximal set of vertices in $D$ such that for every pair of vertices $a$ and $b$ in the set, there is a path from $a$ to $b$. Every digraph can be broken down into its strongly connected components.

**Definition 1.15.** Let $M \in \mathcal{M}_n$. The digraph $D(M)$ representing $M$ is the directed graph on vertices labelled $1, \ldots n$, with an arc from vertex $i$ to vertex $j$ if and only if $m_{ij} \neq 0$.

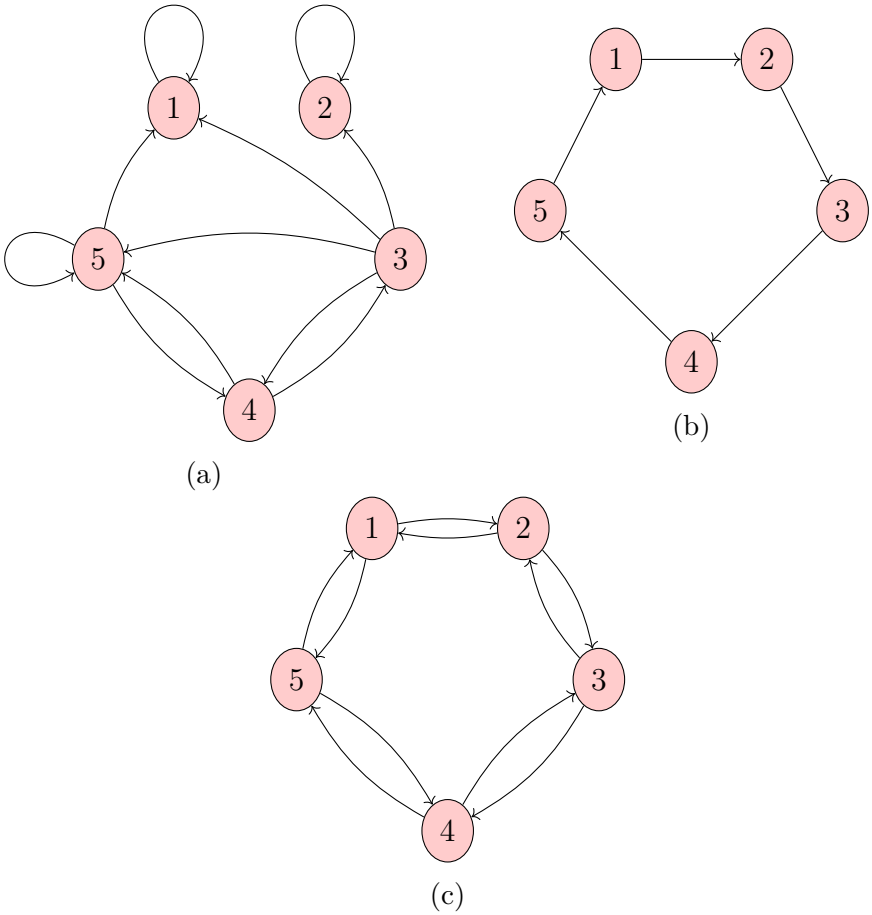



Figure 1.2: Three graphs representing (a) an absorbing Markov chain, (b) an ergodic chain, and (c) a regular Markov chain.

It is easy to visualize a Markov chain using the digraph associated to the transition matrix representing the Markov chain and there is much we can deduce about the Markov chain using techniques from graph theory. Consider Figure 1.2.

1. Looking at Figure 1.2a, we can see that states 1 and 2 are absorbing, states 3, 4 and 5 are transient, and there is a path from each transient state to an absorbing state. Hence, this graph represents an absorbing Markov chain.

2. A glance at Figure 1.2b tells us that this graph represents an ergodic Markov chain. Using the transitions $1 \to 2 \to 3 \to 4 \to 5 \to 1$, it is clearly possible to reach any state from any state. However, although the Markov chain is ergodic, it is not regular. To see this, consider the adjacency matrix $A$ associated to the graph in Figure 1.2b.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

   A quick check using for instance MATLAB reveals that none of $A^2, A^3, A^4$ or $A^5$ have all entries positive. Furthermore, $A^6 = A$, so the process repeats itself, never giving all-positive entries. In fact, it is easy to see that the graph is periodic, with each state having period 5.

3. The digraph in Figure 1.2c represents a regular Markov chain. This is a little more difficult to see, but is quickly checked using MATLAB. Consider the

adjacency matrix $A$ associated to the graph in Figure 1.2c. We find that

$$A^4 = \begin{bmatrix} 6 & 1 & 4 & 4 & 1 \\ 1 & 6 & 1 & 4 & 4 \\ 4 & 1 & 6 & 1 & 4 \\ 4 & 4 & 1 & 6 & 1 \\ 1 & 4 & 4 & 1 & 6 \end{bmatrix}.$$

Since each entry of $A^4$ is positive, $A$ is primitive. Note that when we search for a power $k$ of $A$ such $A^k > 0$, we need not search through every natural number in the hopes of finding the appropriate $k$. From [21, Corollary 8.5.3], we have that if $A \in \mathcal{M}_n$ is nonnegative, then $A$ is primitive if and only if

$$A^{n^2-2n+2} > 0.$$

Since all the relevant background in Markov chains, matrix theory and graph theory has been given, we now go through the links between each.

**Theorem 1.16.** The following are equivalent:

1. The digraph representing a Markov chain is strongly connected.
2. The Markov chain is ergodic.
3. The transition matrix representing a Markov is irreducible.

**Theorem 1.17.** The following are equivalent:

1. The directed graph representing a Markov chain is strongly connected and the greatest common divisor of the lengths of the cycles is one.
2. The Markov chain is regular.

3. The transition matrix representing the Markov chain is primitive.

Further to the links between graph theory and Markov chain theory is the connection between essential classes and strongly connected components. The states of the Markov chain that form an essential class are represented in the digraph as a strongly connected component. The essential classes can be seen in the matrix representing the Markov chain as absorbing blocks, which will be discussed later in this chapter; see Equation (1.4).

### 1.2.4 Limiting behaviour of an ergodic Markov chain

There are two different cases of limiting behaviour of ergodic Markov chains we must examine, as the behaviour is different for each. It is proven in [11, Lemma 3.4.1] that an ergodic Markov chain can only be either periodic or regular so we look at both of these cases now.

Let $\boldsymbol{\pi}(0) = [\pi_1\, \pi_2\, \ldots\, \pi_n]$ denote the ***initial distribution*** of the Markov chain, where $\pi_i$ is the probability that the Markov chain is initially in state $i$. $\sum_i \pi_i = 1$ since $\boldsymbol{\pi}$ is a probability vector. Then $\boldsymbol{\pi}(1) = \boldsymbol{\pi}(0)P$ gives the probabilities the Markov chain is in each state after one time step. Similarly, $\boldsymbol{\pi}(2) = \boldsymbol{\pi}(1)P = \boldsymbol{\pi}(0)P^2$ gives the probabilities after two time steps, and inductively, $\boldsymbol{\pi}(0)P^k$ gives the probability distribution after $k$ steps of the chain.

**Definition 1.18.** The ***stationary distribution vector*** of a finite, ergodic Markov chain with transition matrix $P$ is the probability distribution vector $\boldsymbol{\pi}$, such that

$$\boldsymbol{\pi}P = \boldsymbol{\pi}.$$

Thus, viewed in terms of dynamical systems, $\boldsymbol{\pi}$ is a fixed point of the chain. Note also that the fixed point equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ implies that $\boldsymbol{\pi}$ is a left eigenvector of $P$

associated to the eigenvalue 1, which we know to be an eigenvalue of $P$ since $P\mathbf{1} = \mathbf{1}$ where $\mathbf{1} = (1, \ldots, 1)^T$.

The stationary vector has a different interpretation depending on whether the Markov chain is regular or periodic. In the case that the ergodic Markov chain is regular, we know that

$$\lim_{k\to\infty} \boldsymbol{\pi}(0)P^k = \boldsymbol{\pi},$$

i.e., regardless of the initial distribution $\boldsymbol{\pi}(0)$, the $i^{th}$ entry of the vector $\boldsymbol{\pi}$ represents the long term probability that the system is in state $i$. If the ergodic Markov chain is periodic, a weaker result holds [28]:

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \boldsymbol{\pi}(0)P^j = \boldsymbol{\pi},$$

so that we still gain some information about the long-term behaviour of the system. Informally, the left side of the above equation can be interpreted as a vector containing the average proportions of time the chain spends in each state, which is another interpretation of $\boldsymbol{\pi}$. Note that in both cases, $\boldsymbol{\pi}$ is unique.

### 1.2.5 Limiting behaviour of a non-ergodic Markov chain

When the Markov chain has essential classes, the associated digraph has strong components, one for each essential class, which are proper subsets of the set of states of the chain. In other words, the transition matrix is reducible. The ***canonical form*** of the transition matrix of a Markov chain with essential classes is obtained by rearranging the states (using simultaneous row and column permutations) so that the essential classes are first grouped together, followed by the inessential classes.

The resulting $n \times n$ matrix will be of the form

$$P = \left[\begin{array}{ccc|c} P_1 & & O & \\ & P_2 & & O \\ & \ddots & & \\ O & & P_k & \\ \hline & R & & Q \end{array}\right] = \left[\begin{array}{c|c} \bigoplus_{i=1}^{k} P_i & O \\ \hline R & Q \end{array}\right], \tag{1.4}$$

where $P_i$ are the essential classes, $Q$ is the block matrix describing the transitions between inessential states and $R$ contains the transitions from inessential states to essential states. Each $P_i$ is an $n_i \times n_i$ (square) matrix, $Q$ is an $m \times m$ square matrix, meaning that $R$ is an $m \times (n-m)$ rectangular matrix. Here, $\sum_i n_i = n - m$.

This rearranging can also be done using the digraph representing the Markov chain. We first identify the strongly connected components of the digraph. Next we relabel the vertices, enumerating them so that the strongly connected components are together, taking care to list the essential classes first, and the inessential classes after. When the corresponding transition matrix is written, the matrix will be in the form of (1.4). In fact, we use this method in the algorithm for the numerics in Chapters 3 and 4.

For the purpose of this thesis, we need not examine the general case of essential classes, but rather absorbing states (i.e., the case where each essential class consists of only one state), so from now on, we can simplify our considerations from essential classes to absorbing states. Thus, (1.4) will become

$$P = \left[\begin{array}{ccc|c} 1 & & O & \\ & 1 & & O \\ & \ddots & & \\ O & & 1 & \\ \hline & R & & Q \end{array}\right] = \left[\begin{array}{c|c} \mathbb{I} & O \\ \hline R & Q \end{array}\right]. \tag{1.5}$$

**Definition 1.19.** Let $P$, the transition matrix of an absorbing Markov chain be in canonical form as above. The ***fundamental matrix*** of the chain is the matrix

$$N = (\mathbb{I} - Q)^{-1}, \tag{1.6}$$

where $\mathbb{I}$ represents the identity matrix of the appropriate size, i.e., if $Q$ is an $m \times m$ matrix, then $\mathbb{I}$ is also an $m \times m$ matrix.

The fundamental matrix is a very useful tool in analysing a Markov chain. We now show that the inverse given in Definition 1.19 always exists. Once this is established, we know we can use the fundamental matrix for any absorbing Markov chain.

In the proof of the proposition to follow, we will be using the fact that in any finite, absorbing Markov chain, the probability that the process is in an essential class after $t$ steps tends to 1 as $t \to \infty$ [24, Theorem 3.1.1].

**Proposition 1.20.** Let $Q$ be the submatrix containing the transient state transitions of the absorbing Markov chain as written in (1.4). Then $(\mathbb{I} - Q)$ is invertible.

*Proof.* Assume that the $n \times n$ transition matrix $P$ is in the form (1.4), with $Q$ an $m \times m$ matrix, with $0 < m < n$. Partition the state vector $\boldsymbol{\pi}(t)$ as $\boldsymbol{\pi}(t) = [\boldsymbol{y}(t)\, \boldsymbol{z}(t)]$, where $\boldsymbol{y}(t)$ and $\boldsymbol{z}(t)$ are $(n-m)$- and $m$-vectors representing the essential and transient states respectively. Then, the evolution of the chain

$$\boldsymbol{\pi}(t+1) = \boldsymbol{\pi}(t)P$$

takes the form

$$[\boldsymbol{y}(t+1)\, \boldsymbol{z}(t+1)] = [\boldsymbol{y}(t)\, \boldsymbol{z}(t)]P,$$

or using (1.4),

$$[\boldsymbol{y}(t+1)\,\boldsymbol{z}(t+1)] = [\boldsymbol{y}(t)\,\boldsymbol{z}(t)] \left[\begin{array}{c|c} \bigoplus_{i=1}^{k} P_i & O \\ \hline R & Q \end{array}\right]$$

Rewriting this,

$$\begin{aligned} \boldsymbol{y}(t+1) &= \boldsymbol{y}(t) \bigoplus_{i=1}^{k} P_i + \boldsymbol{z}(t)R(t) && (1.7) \\ \boldsymbol{z}(t+1) &= \boldsymbol{z}(t)Q. && (1.8) \end{aligned}$$

By the theorem mentioned before Proposition 1.20, for any initial distribution $\boldsymbol{\pi}(0)$, the process is in an essential class with probability 1 as $t \to \infty$. Thus, $\boldsymbol{z}(t) \to 0$ as $t \to \infty$ for all possible $\boldsymbol{z}(0)$.

In other words, for all $\boldsymbol{z}(0)$,

$$0 = \lim_{t\to\infty} \boldsymbol{z}(t) = \lim_{t\to\infty} \boldsymbol{z}(0)Q^t = \boldsymbol{z}(0) \lim_{t\to\infty} Q^t. \tag{1.9}$$

As a consequence,

$$\lim_{t\to\infty} Q^t = 0.$$

Assume now that there exists $\boldsymbol{x}$ such that $(\mathbb{I} - Q)\boldsymbol{x} = 0$. We have

$$\begin{aligned} (\mathbb{I} - Q)\boldsymbol{x} = 0 &\iff x = Qx \\ &\iff \boldsymbol{x} = Q^2\boldsymbol{x} \\ &\iff \dots \\ &\iff \boldsymbol{x} = Q^t\boldsymbol{x}, \forall t \geq 0. \end{aligned}$$

Using (1.9), for all $\boldsymbol{x}$, it follows that

$$\boldsymbol{x} = \lim_{t\to\infty} Q^t \boldsymbol{x} = \left(\lim_{t\to\infty} Q^t\right) \boldsymbol{x} = \mathbf{0},$$

so $\boldsymbol{x} = \mathbf{0}$, i.e., $(\mathbb{I} - Q)\boldsymbol{x} = 0 \iff \boldsymbol{x} = 0$, so $(\mathbb{I} - Q)$ is invertible.

□

Continuing on, notice that

$$\begin{aligned}(\mathbb{I} - Q)(I + Q + Q^2 + \cdots + Q^t) &= \mathbb{I} + Q + Q^2 + \cdots + Q^t \\ &\quad -Q - Q^2 - \cdots - Q^t - Q^{t+1} \\ &= \mathbb{I} - Q^{t+1}.\end{aligned}$$

Since $\mathbb{I} - Q$ is invertible,

$$\mathbb{I} + Q + Q^2 + \cdots + Q^t = (\mathbb{I} - Q)^{-1}(\mathbb{I} - Q^{t+1}),$$

and when $t \to \infty$, as established in the proof of Proposition 1.20, $Q^{t+1} \to 0$, so we get

$$(\mathbb{I} - Q)^{-1} = \mathbb{I} + Q + Q^2 + Q^3 + \cdots = \sum_{t=0}^{\infty} Q^t.$$

Recall that

$$Q^n = \begin{bmatrix} q_{11}^{(n)} & q_{12}^{(n)} & \cdots & q_{1n}^{(n)} \\ q_{21}^{(n)} & q_{22}^{(n)} & \cdots & q_{2n}^{(n)} \\ \vdots & & \ddots & \\ q_{n1}^{(n)} & q_{n2}^{(n)} & \cdots & q_{nn}^{(n)} \end{bmatrix}, \tag{1.10}$$

for all $n$, and $q_{ij}^{(n)}$ is the probability that the system is in transient state $j$ after $n$ steps, given it started in transient state $i$. With this in mind, it is now easy to see the following properties:

1. The $(i,j)^{th}$ entry of $(\mathbb{I}-Q)^{-1}$ gives the expected number of visits to the $j^{th}$ transient state before absorption given that the system started in the $i^{th}$ transient state.

2. The $i^{th}$ row sum of $(\mathbb{I}-Q)^{-1}$ gives the expected number of steps until absorption into an essential class, given that the system started in the $i^{th}$ transient state.

Now consider the matrix $(\mathbb{I}-Q)^{-1}R$. In order to understand the meaning of the entries of this matrix, consider for instance a Markov chain with two absorbing states and three transient states, so that

$$(\mathbb{I}-Q)^{-1} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} \qquad \text{and} \qquad R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \\ r_{31} & r_{32} \end{bmatrix}. \tag{1.11}$$

In this case, $(\mathbb{I}-Q)^{-1}R$ is a $2\times 3$ matrix. By multiplying the first row of $(\mathbb{I}-Q)^{-1}$ by the first column of $R$ we find that the $(1,1)$ entry of $(\mathbb{I}-Q)^{-1}R$ is $\rho_{11}r_{11}+\rho_{12}r_{21}+\rho_{13}r_{31}$. Recall, for example, that $\rho_{12}$ represents the expected number of visits to transient state 2 given the system started in transient state 1, so multiplying it by $r_{21}$ gives us the probability that given the system started in transient state 1, it ends up being absorbed in absorbing state 1 directly from transient state 2. The same idea will hold for each $\rho_{ij}$ and $r_{ji}$, so that

3. The $(i,j)^{th}$ entry of $(\mathbb{I}-Q)^{-1}R$ gives us the probability of the system being absorbed in absorbing state $j$ given it started in transient state $i$.

We also have some other useful results coming from the matrix $N=(\mathbb{I}-Q)^{-1}$. Let $N_{dg}$ be the matrix obtained by setting all the off-diagonal components of $N$ to zero, and $N_{sq}$ be the matrix obtained by squaring each entry of $N$ [24]. Then

4. The variance of the number of visits to transient state $j$ before absorption

given the system started in transient state $i$ is the $(i,j)^{th}$ entry of the matrix $N_2 = N(2N_{dg} - \mathbb{I}) - N_{sq}$.

5. Let $s = N\mathbf{1}$, where $\mathbf{1}$ denotes all-ones column vector. Given the system started in transient state $i$, the variance of the number of steps before absorption is the $i^{th}$ entry of the vector $(2N - \mathbb{I})s - s_{sq}$.

**Example:**

Suppose we have a Markov chain with the transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0.1 & 0.12 & 0 & 0.4 & 0.38 \\ 0 & 0 & 0.45 & 0.3 & 0.25 \\ 0.15 & 0 & 0 & 0.6 & 0.25 \end{bmatrix}. \tag{1.12}$$

Notice that the digraph representing the Markov chain is the same as that given in Figure 1.2a, except that the edges are now given weights representing the transition probabilities in (1.12).

From both the graph and the transition matrix, we can see that the absorbing states are the states $\{1, 2\}$, and the transient states are $\{3, 4, 5\}$. Note that states 3, 4 and 5 will sometimes be referred to as transient states 1, 2 and 3 respectively in the forthcoming discussion.

Since the transition matrix is already in canonical form, we have

$$Q = \begin{bmatrix} 0 & 0.4 & 0.38 \\ 0.45 & 0.3 & 0.25 \\ 0 & 0.6 & 0.25 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 0.1 & 0.12 \\ 0 & 0 \\ 0.15 & 0 \end{bmatrix}. \tag{1.13}$$

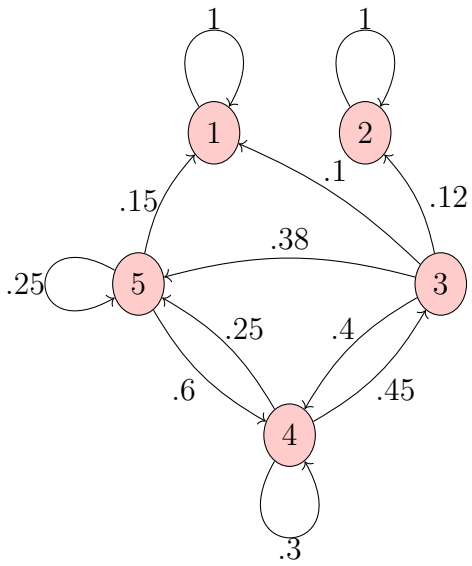



Figure 1.3: Weighted directed graph representing the transition matrix (1.12)

.

$N$ is calculated to be

$$N = (\mathbb{I} - Q)^{-1} = \begin{bmatrix} 2.7293 & 3.8428 & 2.6638 \\ 2.4563 & 5.4585 & 3.0640 \\ 1.9651 & 4.3668 & 3.7846 \end{bmatrix} \tag{1.14}$$

and

$$NR = \begin{bmatrix} 0.6725 & 0.3275 \\ 0.7052 & 0.2948 \\ 0.7642 & 0.2358 \end{bmatrix}. \tag{1.15}$$

From (1.14) and (1.15), we have gained new information about the system. For example,

- The expected number of visits to transient state 2 (state 4), given the system started in transient state 3 (state 5) is 4.3668.
- Given the system started in transient state 2 (state 4), the expected number of steps until it becomes absorbed in one of the absorbing states is 2.4563 +

$5.4585 + 3.0640 = 10.9788$.

- Given the system started in transient state 3 (state 5), the probability of absorption into state 1 is 0.7642 and absorption into state 2 is 0.2358.

Then, we find that

$$N_2 = N(2N_{dg} - \mathbb{I}) - N_{sq} = \begin{bmatrix} 4.7196 & 23.3420 & 10.4030 \\ 4.9180 & 24.3369 & 10.7398 \\ 4.8998 & 24.2368 & 10.5384 \end{bmatrix} \tag{1.16}$$

and

$$(2N - \mathbb{I})s - s_{sq} = \begin{bmatrix} 94.1526 \\ 95.7089 \\ 96.2973 \end{bmatrix}. \tag{1.17}$$

From (1.16) and (1.17), we gain some helpful information regarding the standard deviation (square root of the variance) in the system. For example,

- The standard deviation of the number of visits to transient state 3 (state 5) given the system started in transient state 1 (state 3) is 3.225.

- Given the system started in transient state 2 (state 4), the standard deviation of the number of steps until absorption is 9.783.

### 1.2.6 Absorption probability

Further to the fundamental matrix, there is another way to analyse the limiting behaviour of a non-ergodic Markov chain which achieves the same results as in the previous section. Suppose there is more than one essential class in a Markov chain and we want to find the probability of absorption into each essential class given we started in some transient state. The ***absorption probability*** describes the

probability of reaching and permanently staying in a particular essential class of the Markov chain.

Suppose that starting in state $i$, we want to calculate the long term probability of the Markov chain ultimately ending up in state $j$. We will denote this $\zeta_{ij}$. Note that $X_\infty$ will be used to represent a large enough amount of time so that the system is absorbed in state $j$. We have

$$\begin{aligned}\zeta_{ij} &= \mathbb{P}[X_\infty = j | X_0 = i] \\ &= \sum_{k=1}^{n} \mathbb{P}[X_\infty = j | X_1 = k, X_0 = i] p_{ik} \\ &= \sum_{k=1}^{n} \mathbb{P}[X_\infty = j | X_1 = k] p_{ik},\end{aligned}$$

by the Markov property. So we get

$$\zeta_{ij} = \sum_{k=1}^{n} p_{ik} \zeta_{kj}.$$

This must be true for all $i$, so we can rewrite this in matrix form as

$$\begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \\ \vdots \\ \zeta_{nj} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \\ \vdots \\ \zeta_{nj} \end{bmatrix}.$$

Notice that this is the equation for the right eigenvectors of the matrix P corresponding to the eigenvalue one. Although the fundamental matrix of the Markov chain already gives us this result, this is nonetheless another way to calculate these particular absorption probabilities.

For example, going back to (1.12) we get two right eigenvectors associated to the

eigenvalue 1:

$$v_1 = \begin{bmatrix} 0.6282 \\ 0 \\ 0.4225 \\ 0.4431 \\ 0.4801 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} 0 \\ 0.8945 \\ 0.2930 \\ 0.2637 \\ 0.2109 \end{bmatrix}.$$

At first glance these vectors do not appear to represent probabilities, but recall that MATLAB does not automatically scale an eigenvector, so we do so now. The first entry of $v_1$ represents the probability that given the system begins in absorbing state 1, it ends up absorbed in state 1. Obviously this number should be 1, so if we scale our eigenvector $v_1$ by dividing every entry by 0.6282 and by the same logic divide $v_2$ by 0.8945, we get the altogether more realistic vectors $v_1$ and $v_2$ to be

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0.6725 \\ 0.7052 \\ 0.7642 \end{bmatrix} \quad \text{and} \quad v_2 = \begin{bmatrix} 0 \\ 1 \\ 0.3275 \\ 0.2948 \\ 0.2358 \end{bmatrix},$$

and as we can see, this gives us the same information as $NR$ from (1.15).

### 1.2.7 Individual realizations of the Markov chain

For the forthcoming work in Chapters 3 and 4 it will be necessary to implement individual realizations of the Markov chain, i.e., generate a sequence of states the Markov chain successively occupies based on the probabilities of transition from state to state. The following is an algorithm describing how the realizations will be computed:

1. Pick an initial distribution $\boldsymbol{\pi}(0)$ so that the probability of being in a specified

initial state is 1. For instance, to generate a realization where the initial state is state 2, one would use $\boldsymbol{\pi}(0) = (0, 1, 0, \ldots, 0)$.

2. Apply the transition matrix of the Markov chain to the distribution $\boldsymbol{\pi}(0)$. This will give the probability distribution of transitions to the different states accessible from the initial state $\boldsymbol{\pi}(1) = \boldsymbol{\pi}(0)P$.

3. The nonzero entries in $\boldsymbol{\pi}(1)$ are the states to which transition from the initial state is possible. The values in $\boldsymbol{\pi}(1)$ are the probabilities of such transitions. A random number between 0 and 1 is generated and used to determine which transition occurs.

4. Based on the outcome of Step 3, create a new distribution vector with a 1 in the appropriate position (the position representing which state was chosen) and 0 everywhere else.

5. Repeat Steps 2 - 4 until the chain reaches an absorbing state (or if the chain is not absorbing, a determined number of steps have been computed) using the newly computed $\boldsymbol{\pi}(1)$ to play the role of $\boldsymbol{\pi}(0)$ in Step 1 and record the resulting sequence of the $\boldsymbol{\pi}(i)$.

# 2

# Epidemiological Background

In this chapter, I will give an outline of the background in mathematical epidemiology required in this thesis. The following summary is heavily influenced by the papers in [2].

## 2.1 Compartmental models

Compartmental models are used in many areas of science to model processes. Although the nature and interpretation of the compartments involved in the model can vary, they all have one common defining facet: a ***compartment*** is a collection of kinetically homogenous material or content. This implies that a new component added to the compartment immediately mixes with all of the original contents of the compartment [23].

In medicine, compartmental models are used to study the flow of chemicals such as nutrients and hormones in the body - in this case the compartments would be the organs. In pharmacokinetics, the effect of drugs administered in living organism is modelled. In this case, the compartments are not the organs in the body, but rather the stages the drug undergoes as it takes effect in the body – absorption, distribution, metabolization and excretion.

## 2.2 Compartmental models in epidemiology

In this thesis, we will be using compartmental models, but in a different way to those mentioned above. Most epidemiological models involve grouping the population into compartments in which the individuals' status with respect to the disease in question are assumed to be identical. For example, most compartmental models start by assuming every member of the population automatically begins in the susceptible compartment **S**, and many then assume that once a person is infectious, they enter a new compartment, **I**, for infectious. Below is a list of commonly used letters and terms for different compartments.

- **S** : Susceptible. Individuals who have no immunity to the disease, and are at risk of contracting the disease.
- **I** : Infectious. Individuals currently infected with the disease and can cause infection to others.
- **R** : Recovered. Individuals who have recovered from contracting the disease and are no longer infectious. These individuals sometimes gain immunity from the disease for a period of time and eventually go back to the susceptible class.
- **E** : Exposed. Individuals who have been infected by the disease but who are not yet showing symptoms and are assumed to be unaware they have contracted the disease. The use of **E** is historical; this compartment should in fact be labelled **L** for "latently infected".

This is not a comprehensive list, but details the most regularly used compartments for most diseases. Some diseases require a specific set of compartments. For example, in [5], there were many more compartments required: $\mathbf{S_T}, \mathbf{L}, \mathbf{L_T}, \mathbf{I_T}, \mathbf{A}, \mathbf{A_T}$ representing treated susceptible, latent, treated latent, treated infective, asymptomatic and treated asymptotic respectively. Note the use of the word "infectious" rather than

"infected". An infected individual is someone who has the disease while an infectious individual is someone who has disease and can transmit it to others. An infectious compartment is a subset of the infected compartment – one of many possible infected compartments. If there is only one infected compartment then infectious and infected mean the same thing and are therefore interchangeable.

Different diseases cause individuals to go through compartments in unique ways. For example, a disease such as gonorrhea might cause an individual to move from the susceptible compartment, **S**, to the infectious compartment, **I**, and then back to **S**. Hence, in order to model gonorrhea, we would use an SIS model. Other diseases such as influenza may cause an individual to undergo a different process through the compartments: individuals begin as susceptible (**S**), may spend some time infected with the disease but without showing symptoms (**E**), then becoming infectious (**I**) for a period of time until moving into the recovered class (**R**) in which they are no longer susceptible to that particular strain of influenza. Hence, an SEIR model is used in order to study the dynamics of the disease. For the remainder of this section, we will deal with an SIS model, as this will be the type of model under investigation in this thesis. i.e., the disease in question will cause members of the population to move from the susceptible class **S**, to the infected class **I**, and rather than recover from the disease and go to a recovered class with immunity, the individual goes directly back to the susceptible class.

### 2.2.1 Movement between compartments

Before we proceed, note the distinction between **S** and $S$. The former is to denote the compartment **S**, while $S_t = S(t)$ will be used to represent the number of susceptible individuals in that compartment at time $t$. Obviously the same applies with **I** and $I_t = I(t)$. In order to discuss the rate of movement of individuals between compartments, then for the SIS model we need to introduce a few more terms:

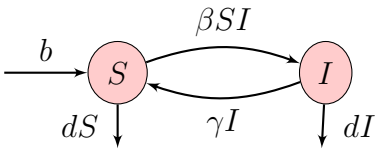



Figure 2.1: Flow diagram for an SIS model.

- $\beta$: The disease transmission coefficient. This is the rate at which infected individuals transmit the disease to susceptibles when a contact occurs.

- $\gamma$: The recovery rate. This is the rate at which individuals recover from the disease, i.e., the rate of movement from compartment **I** to compartment **S**. Note that since $\gamma$ is constant per capita and that this is an ODE compartmental model, $\frac{1}{\gamma}$ gives the average duration of the infection.

Suppose an infectious individual is introduced into a population of $N$ individuals. If everyone is initially in the susceptible compartment, i.e. $S_0 = N$, the total population size, then the rate of infection is $\beta N$. As the disease progresses, and not all of the population is susceptible anymore, the probability that a contact by the infected individual is with a susceptible person at time $t$ is $\frac{S_t}{N}$, so the infection rate becomes $(\beta N)\left(\frac{S_t}{N}\right) = \beta S_t$. If we suppose now that there are $I_t$ infected individuals at time $t$, this becomes $\beta S_t I_t$. This type of formulation of the infection rate is known as ***mass action incidence*** and assumes that members of the population are homogeneously mixed and all individuals are equally likely to come into contact. Infected individuals leave the compartment **I** at a per capita rate of $\gamma I_t$. For the modelling in this thesis, I will be assuming mass-action incidence.

There are also individuals leaving both compartments due to natural death at a per capita rate of $d$ and entering the susceptible class due to birth at a rate of $b$; see Figure 2.1.

### 2.2.2 Some common assumptions

When formulating a mathematical model to describe disease transmission dynamics, we often need to make many assumptions – some to more accurately represent the nature of the disease in question and some to avoid our model being overly complex and unusable. What follows is a list of commonly-made assumptions and a brief description of what that assumption means.

- *No deaths due to the disease*: Many diseases do not have a high mortality rate – for example, although influenza may be fatal to an elderly or frail individual, these deaths usually do not have a high impact on the overall dynamics of the system and can therefore be ignored.

- *No recruitment*: The population in question is closed, meaning there are no individuals entering or leaving the population. This is an allowable assumption for instance if the period of time considered is small.

- *No vertical transmission*: Individuals are not born into the infected compartment. True for some diseases, and untrue for others.

- *Constant population*: Assuming the total population remains constant often makes it much easier to analyze the dynamics of a disease, especially in stochastic models. This can be done by assuming there are no births or deaths at all, or by assuming births and deaths occur at an equal, per capita rate, or using a formulation such that the population is eventually constant.

There are many more assumptions that can be made which may be necessary for different situations.

## 2.3 Spatial aspects in epidemiology

In recent years it has become increasingly important to examine spatial aspects in epidemiology. Interconnectivity between virtually all areas of the world has rapidly grown and as a result, contacts between different groups of people has skyrocketed. Our analysis of the spread of infectious diseases must evolve to meet the demands of the current situation. In fact, in 2013, Khan *et al* found that during the earlier stages of the H1N1 pandemic, most public health benefits could have been attained by screening at just 8 airports [27]. Results like this indicate that research into the geographical spread of disease is becoming more and more necessary in the fight against infectious diseases.

What follows is a short review of some of the work done in order to account for the geographical spread of disease via movement of individuals between separate patches. The following review was helped and influenced substantially by [4].

### 2.3.1 Bartlett, 1956

In 1956, Bartlett [9] gave a model accounting for movement of both susceptible and infected individuals across two distinct patches for a disease fitting an SI model.

$$\begin{aligned} S_1' &= -(\beta_1 I_1 + \beta_2 I_2)S_1 + b + m_S(S_2 - S_1) \\ I_1' &= (\beta_1 I_1 + \beta_2 + I_2)S_1 - (d+\rho)I_1 + m_I(I_2 - I_1) \\ S_2' &= -(\beta_1 I_1 + \beta_2 I_2)S_2 + b + m_S(S_1 - S_2) \\ I_2' &= (\beta_1 I_1 + \beta_2 + I_2)S_2 - (d+\rho)I_2 + m_I(I_1 - I_2). \end{aligned}$$

This model incorporates both movement of infectious and susceptible individuals, and allows for different values of the disease transmission coefficient $\beta$ for each location. Note that the indices specify the location. In this model, $\beta_1$ and $\beta_2$ represent

the disease transmission coefficient in location 1 and 2, $d + \rho$ together is the rate at which individuals leave the infected compartment, whether by natural death ($d$) or death due to the disease ($\rho$) and $b$ is the birth rate. $m_S$ and $m_I$ are the rates of travel of susceptible and infectious individuals respectively. Note that this is quite a naïve model; infection occurs both within and between locations.

### 2.3.2 Baroyan and Rvachev, 1969, 1971

In 1969, Baroyan and Rvachev studied the spatial spread of influenza across cities in the Soviet Union [7, 8]. The model was formulated by splitting the Soviet Union into smaller sub-regions representing the cities. Transport between each of the cities is considered and within each city, influenza was modelled deterministically using an SIR compartmental model. This framework was used in later studies into the global spread of influenza.

### 2.3.3 Kermack-McKendrick style metapopulation models

Consider the Kermack-McKendrick SIR model [26]:

$$\begin{aligned} S' &= -\beta SI \\ I' &= \beta SI - \gamma I \\ R' &= \gamma I. \end{aligned}$$

This is an SIR model without demography (notice that natural birth and death rates are not considered). There has been some work on incorporating metapopulations into the Kermack-McKendrick model above. For example, Faddy in 1986 [18]

introduced an SI model without demography:

$$\begin{aligned} S_i' &= -S_i \sum_{j=1}^{n} \beta_{ji} I_j \\ I_i' &= S_i \sum_{j=1}^{n} \beta_{ji} I_j - \gamma_i I_i + \sum_{j \neq i}^{n} m_{ij} I_j - \sum_{j \neq i} m_{ji} I_i, \end{aligned}$$

where $\gamma_i$ represents the sum of all removals from each infected class $I_i$. Notice that this model incorporates both movement ($m_{ij}$) of infected individuals and "long range" infection ($\beta_{ji}$) allowing infected individuals from location $j$ to infect individuals in location $i$. Formulating the model this way allows for possible transmission of infection through contact between individuals coming into contact on the border of neighbouring locations as well as those moving from location to location for example by aeroplane. This is one example of quite a few metapopulation models which ignore demography.

## 2.4 The early phase of an epidemic

One of the many benefits of using a stochastic approach to model the spread of disease is its application in the early stages of an outbreak. In 1980, Longini [30] showed that deterministic models are inadequate for describing the infection process in groups of less than 35 individuals - the deterministic model is incapable of recognizing the stochastic fluctuations when the population is small enough. In this thesis, the same idea is applied to the early stages of an outbreak, but without imposing limitations on the size of the population. In the early stages of a new disease outbreak, we frequently observe "stochastic" effects on the spread of the disease - a stage of initial "flutters" where the disease has not fully taken hold. See Figure 3 and the corresponding discussion in the introduction section for more on this.

In this thesis, I will study and make use of this phenomenon to model the early

stages of a disease outbreak. Using an appropriate (relatively small) number $x$ of individuals as an upper bound, we will say that once there are more than $x$ infected individuals in the system, then it is extremely likely that the disease will grow into an epidemic. In fact, we will set $x$ high enough so that if the system reaches $x$ individuals infected, we consider an epidemic a certainty. The work of Whittle [42] (which will be detailed later) will be applied to determine the appropriate value of $x$.

## 2.5 Stochastic models in epidemiology

### 2.5.1 Chain binomial models [2]

Two well-known discrete-time Markov chain models for the spread of disease are the Greenwood model and the Reed-Frost model. Both are examples of chain binomial models and are referred to as such because a binomial distribution is used to determine the number of new infectious individuals after a time step. The argument behind using a binomial distribution in mathematical epidemiology is as follows: Suppose we take a time step small enough so that only one contact sufficient for transmission of an infection may take place during one time step. Let $p$ be the probability of the successful contact occurring in a time interval, and assume the probabilities of contact in in each time interval are independent. Then the distribution for the total number of contacts over all time is the binomial distribution.

As before, $S_t$ and $I_t$ represent the number of individuals in the susceptible and infected compartments at time $t$ respectively. Here however, they are random variables, taking values $s_t, i_t \in \{0, 1, \ldots, N\}$ where $N$ is the size of the population. The models assume that $I_0 \geq 1$ initially. At time $t$, the infected individuals are in contact with the susceptible individuals, and the susceptible individuals do not become infectious until time $t+1$. At time $t+1$, the individuals who were infectious at time

$t$, are no longer infectious.

The models differ in the assumption regarding the probability of infection. Let $p_i$ be the probability that a susceptible person is not infected, given there are $i$ infected individuals in the system. The Greenwood model assumes that the probability of infection $p_i = p$ is constant, while the Reed-Frost model accounts for dependence on the number of infected individuals, i.e., $p_i = p^i$.

**Greenwood model [19]**

In the Greenwood model, the random variable $S_{t+1}$ is a binomial random variable that depends on $S_t$ and $p$, the probability a susceptible individual does not get infected; i.e., $S_{t+1} \sim b(S_t, p)$. The probability of a transition from $(s_t, i_t)$ to $(s_{t+1}, i_{t+1})$ is given by

$$p_{s_{t+1},s_t} = \begin{pmatrix} s_t \\ s_{t+1} \end{pmatrix} p^{s_{t+1}}(1-p)^{s_t - s_{t+1}}.$$

**Reed-Frost model [1]**

In the Reed-Frost model, the random variable $S_{t+1}$ is binomially distributed and satisfies $S_{t+1} \sim b(S_t, p^{I_t})$. The probability of a transition from $(s_t, i_t)$ to $(s_{t+1}, i_{t+1})$ is given by

$$p_{(s,i)_{t+1},(s,i)_t} = \begin{pmatrix} s_t \\ s_{t+1} \end{pmatrix} (p^{i_t})^{s_{t+1}}(1-p^{i_t})^{s_t - s_{t+1}}.$$

## 2.6 Methods of measuring the expected effects of a disease outbreak

In epidemiology, there are many methods employed to "measure" the effects of a disease outbreak. In this section we detail some of the major quantities used for this purpose.

### 2.6.1 Basic reproduction number

The ***basic reproduction number*** $\mathcal{R}_0$ is defined as the number of secondary infections caused by a single infectious individual in a completely susceptible population. $\mathcal{R}_0$ is extremely useful in the analysis of the spread of an infectious disease because in general,

- If $\mathcal{R}_0 < 1$, then the disease will die out;
- If $\mathcal{R}_0 > 1$, then the disease outbreak will be a major one (in an epidemic case) or will become established in the population (in an endemic case).

The case of $\mathcal{R}_0$ exactly equal to 1 would be extremely rare, but in theory, if this happened then the prevalence level of the disease should remain the same. Note that $\mathcal{R}_0$ is a dimensionless number and not a rate, so it does not have units in time or any other quantity. This is why we call it the basic reproduction *number* and not basic reproductive *rate* as it is sometimes incorrectly labelled.

Intuitively, we can calculate $\mathcal{R}_0$ by the formula

$$\mathcal{R}_0 = \rho c k,$$

where $\rho$ is the probability that a contact between an infectious and susceptible individual results in an infection, $c$ is the contact rate between infectious and susceptible individuals and $k$ is the duration of an infection. We can even see from this formulation of $\mathcal{R}_0$ that it is a dimensionless quantity; the units of each quantity cancel, leaving $\mathcal{R}_0$ unit-less:

$$\mathcal{R}_0 = \left(\frac{\text{infection}}{\text{contact}}\right) \cdot \left(\frac{\text{contact}}{\text{time}}\right) \cdot \left(\frac{\text{time}}{\text{infection}}\right).$$

Provided one has all the relevant contact information for the disease, it is easy to calculate $\mathcal{R}_0$. However, most of the time not all of this information is readily

available. Given a mathematical model for disease spread, how does one compute $\mathcal{R}_0$ in practice? We illustrate this for an ODE SIS model deduced from Figure 2.1. From the diagram,

$$S' = b - dS - \beta SI + \gamma I \tag{2.1a}$$

$$I' = \beta SI - \gamma I - dI. \tag{2.1b}$$

Here we have assumed that the birth rate is constant, rather than proportional to the number of people in the population. This is a reasonable assumption when the population under consideration is large. It is easy to see that the disease-free equilibrium is $(S^\star, I^\star) = (\frac{b}{d}, 0)$. Using this and the method in Chapter 6 of [40] we can calculate $\mathcal{R}_0$. Suppose there are $n$ disease compartments and $m$ non-disease compartments. We denote $\mathcal{F}_i$ the rate at which infections increase the $i^{th}$ disease compartment and by $\mathcal{V}_i$ the rate disease progression, death and recovery decrease the $i^{th}$ compartment. The model can now be written in the form

$$\begin{aligned} x_i' &= \mathcal{F}_i(x,y) - \mathcal{V}_i(x,y), \qquad i = 1, \ldots, n \\ y_j' &= g_j(x,y) \qquad j = 1, \ldots, m. \end{aligned}$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that $x_i$ and $y_j$ represent the net rate of transitions in and out of the $i^{th}$ disease and $j^{th}$ non-disease compartments respectively. $g_j(x,y)$ is the transitions into non-disease compartment $j$ minus transitions out of non-disease compartment $j$.

The square matrices $\mathbf{F}$ and $\mathbf{V}$ are the Jacobian matrices of $\mathcal{F}$ and $\mathcal{V}$ evaluated at the DFE.

$$\mathbf{F} = \left[\frac{\partial \mathcal{F}_i}{\partial x_j}\left(\frac{b}{d}, 0\right)\right] \qquad \text{and} \qquad \mathbf{V} = \left[\frac{\partial \mathcal{V}_i}{\partial x_j}\left(\frac{b}{d}, 0\right)\right].$$

The matrix $\mathbf{K} = \mathbf{F}\mathbf{V}^{-1}$ is known as the ***next generation matrix***, and the basic reproduction number $\mathcal{R}_0$ is obtained by calculating the spectral radius (the eigenvalue of largest modulus) of $\mathbf{K}$.

For the SIS model (2.1), we only have one disease compartment and one non-disease compartment, so $\mathbf{F}$ and $\mathbf{V}$ are scalars. In our case,

$$\begin{aligned} \mathcal{F} = \beta SI \quad &\Longrightarrow \quad \mathbf{F} = \beta S^\star \\ &\Longrightarrow \quad \mathbf{F} = \beta \frac{b}{d} \end{aligned}$$

and

$$\mathcal{V} = dI + \gamma I \implies \mathbf{V} = d + \gamma.$$

Since $\mathbf{F}$ and $\mathbf{V}$ are scalars, $\mathcal{R}_0$, the spectral radius of $\mathbf{K} = \mathbf{F}\mathbf{V}^{-1}$, is

$$\mathcal{R}_0 = \frac{b}{d}\frac{\beta}{d+\gamma} = \frac{\beta}{d+\gamma}N^\star, \tag{2.2}$$

where $N^\star = S^\star + I^\star = b/d$ is the equilibrium population.

### 2.6.2 Attack rate

Another worthwhile quantity in the analysis of a disease outbreak is the ***attack rate***. The attack rate is obtained by taking the number of new cases of the infection in the population in question and dividing it by the number of susceptible individuals in the population, i.e.,

$$\text{Attack rate} = \frac{\text{number of new cases in the population}}{\text{number of susceptible individuals}}.$$

### 2.6.3 Final size of an epidemic

The ***final size of an epidemic*** can be informally defined as the number of people experiencing infection during an outbreak [22]. In order to model such a quantity, we need to take a slightly different approach to the Markov chain model as in Chapter 1 – the fundamental matrix is very useful in determining long-term probabilities, but will not give information about the total number of infections experienced during an outbreak.

To calculate the final size of an epidemic, we will be simulating individual realizations of the Markov chain and counting the number of times an infection takes place. Note that since we only examine the early stage of the outbreak, this data can only be gathered for cases in which the disease becomes extinct. If the Markov chain reaches an absorbing state which implies an epidemic is forthcoming, the simulation stops, preventing us from continuing to count the number of infections. Thus, the analysis of the final size of an outbreak will be conditioned on disease extinction in this thesis.

The final size of an epidemic is calculated using the algorithm from Section 1.2.7, but with some minor adjustments:

- In Step 3, when the algorithm chooses the next state for the Markov chain, if the chosen state represents an infection event, then the final size is increased by 1.

- In Step 5, if the absorbing state the Markov chain reaches represents an epidemic, then this particular realization of the Markov chain is discarded. If the realization has not been discarded, the number of infection events from Step 3 are counted. The sum of the infection events is the final size of the epidemic.

This process is repeated a large number of times so that an average final size of the epidemic can be determined.

# 3

# A Markov chain model for an SIS-type disease in an isolated location

In the next chapter, a Markov chain model will be developed to describe the movements of infected individuals across two populations connected by transport. Before establishing the model for two interconnected locations in Chapter 4, we study the model for a disease in an isolated location. The model here is heavily influenced by the work of Allen and Burgin [3].

## 3.1 The model

Consider an isolated location. In this location, the model is a variation on the SIS system in [3] and other simple stochastic epidemic models. First, note that we assume the total population to be asymptotically constant and equal to $P$; see Section 3.2.1. Individuals can be in one of two states: susceptible to the disease or infected (and infectious) with the disease. The numbers of individuals in each compartment at time $t$ are denoted $S(t)$ and $I(t)$, respectively. Upon infection, susceptible individuals

transition to the infectious class. They remain there for some time until they recover. Upon recovery, infectious individuals immediately are susceptible again to infection.

Because the total population is asymptotically constant, the model can focus on the number of infectious individuals in the population, with the number of susceptible individuals given by $S(t) = P - I(t)$. As in [3], we formulate a discrete-time Markov chain for the number of individuals infected (and infectious) with the disease. The main variation from [3] comes from the introduction of a threshold $N$, $0 < N \leq P$, above which the epidemic is assumed to go into an exponential growth phase, at which point it becomes irrelevant to the present study. Thus the states are $I = \{0, \ldots, N\}$, with both 0 and $N$ absorbing.

Assume that the time step $\Delta t$ is sufficiently small that only one change in state is possible per unit time. Transition probabilities are then given, for $i = 1, \ldots, N-1$, by

$$\mathbb{P}\{I(t+\Delta t) = i+1 | I(t) = i\} = \beta(P-i)i\ \Delta t \tag{3.1a}$$

$$\mathbb{P}\{I(t+\Delta t) = i-1 | I(t) = i\} = (\gamma + d)i\ \Delta t \tag{3.1b}$$

$$\mathbb{P}\{I(t+\Delta t) = i | I(t) = i\} = 1 - (\beta(P-i)i + (\gamma+d)i)\ \Delta t. \tag{3.1c}$$

All other transitions have probability 0. To simplify notation, we denote $\Pi_i = \beta(P-i)i\Delta t$, $\Gamma_i = (\gamma+d)i\Delta t$ and $p_i = 1 - (\Pi_i + \Gamma_i)$, representing an infection, recovery or death and no change in the number of infected, respectively. Note that

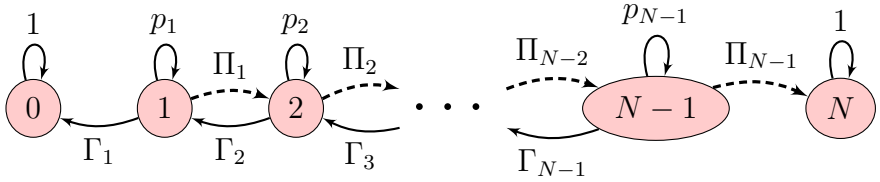



Figure 3.1: Random walk used in isolated locations.

we have used mass action incidence in (3.1a), whereas [3] considered several different

incidence functions.

System (3.1) is a random walk on the number $I(t)$ of infectious individuals in the population. The transition matrix of the Markov chain takes the form

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & & & 0 \\ \Gamma_1 & p_1 & \Pi_1 & 0 & \dots & & & 0 \\ 0 & \Gamma_2 & p_2 & \Pi_2 & \dots & & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & & \Gamma_{N-1} & p_{N-1} & \Pi_{N-1} \\ 0 & 0 & 0 & & \dots & 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

Note that we are using $\mathbf{T} = [t_{ij}]$ with the convention that $t_{ij}$ represents the probability of transition from state $i$ to state $j$.

## 3.2 Mathematical analysis

### 3.2.1 Underlying deterministic model

As in [3], the underlying discrete-time deterministic model describing the evolution of average numbers in each compartment is first studied. Here, it takes the form:

$$S(t + \Delta t) = S(t) + \{b + \gamma I(t) - \beta S(t)I(t) - dS(t)\}\,\Delta t \tag{3.3a}$$

$$I(t + \Delta t) = I(t) + \{\beta S(t)I(t) - (d + \gamma)I(t)\}\,\Delta t, \tag{3.3b}$$

where $\Delta t$ is the (fixed) time step, $b$ is the birth rate, $d$ is the *per capita* death rate, $\beta$ is the infection parameter and $\gamma$ is the *per capita* rate of recovery. Initial conditions are assumed to be $S(0) + I(0) = P > 0$. We now show that the total population $S(t) + I(t)$ is asymptotically constant. The total population follows the difference

equation

$$P(t+\Delta t) = P(t) + \{b - dP(t)\}\,\Delta t.$$

Let $P^\star$ be such that $P^\star(t+\Delta t) = P^\star(t)$ (i.e., $P^\star$ is a fixed point). Thus,

$$P^\star = P^\star + \{b - dP^\star\}\,\Delta t \implies P^\star = \frac{b}{d}.$$

Let

$$f(P) = P + \{b - dP\}\Delta t.$$

Then

$$f'(P^\star) = 1 - d\Delta t.$$

Since both $d$ and $\Delta t$ are less than one, $|f(P^\star)| < 1$, so the fixed point $P^\star$ is attracting. Therefore, the total population approaches $b/d$.

The major difference between the model we have set up and that in [3] is that in this model, we have two absorbing states; at $I(t) = 0$ and $I(t) = N$. In [3], there is only one absorbing state, $I(t) = 0$. We also have the following result.

**Theorem 3.1.** Let

$$\mathcal{R}_0 = \frac{\beta}{d+\gamma}\frac{b}{d} = \frac{\beta}{d+\gamma}P^\star. \tag{3.4}$$

If $\mathcal{R}_0 < 1$, then (3.3) has only the disease-free fixed point (DFFP)

$$(S_{DF}, I_{DF}) = \left(\frac{b}{d}, 0\right), \tag{3.5}$$

which is locally attractive. If $\mathcal{R}_0 > 1$, then the DFFP (3.5) is unstable and there is an endemic fixed point

$$(S_{EFP}, I_{EFP}) = \left(\frac{d+\gamma}{\beta}, \frac{b}{d} - \frac{d+\gamma}{\beta}\right) = \left(\frac{P^\star}{\mathcal{R}_0}, \left(1 - \frac{1}{\mathcal{R}_0}\right)P^\star\right), \tag{3.6}$$

which is locally attractive.

*Proof.* The values of the fixed points (3.5) and (3.6) are easily found by solving the fixed point problem. At an arbitrary point $(S, I)$, the Jacobian matrix of (3.3) is

$$J(S,I) = \begin{bmatrix} 1-(\beta I + d)\Delta t & (\gamma - \beta S)\Delta t \\ \beta I \Delta t & 1-(\gamma + d - \beta S)\Delta t \end{bmatrix}. \tag{3.7}$$

Evaluating 3.7 at the disease free fixed point gives

$$J(S_{DF}, I_{DF}) = \begin{bmatrix} 1-d\Delta t & (\gamma - \beta P^\star)\Delta t \\ 0 & 1-(\gamma + d - \beta P^\star)\Delta t \end{bmatrix}, \tag{3.8}$$

and thus, the eigenvalues of $J(S_{DF}, I_{DF})$ are $1 - d\Delta t$ and $1 - (\gamma + d - \beta P^\star)\Delta t$. We have $1 - d\Delta t < -1 \iff \Delta t > 2/d$, which in practice would be an extremely large and unrealistic time step, so we assume the eigenvalue is in the unit circle (see remark after proof). Therefore, the local stability of the DFFP is governed by the eigenvalue $1 - (\gamma + d - \beta P^\star)\Delta t$. As for $1 - d\Delta t$, this value becoming less than -1 is unlikely, so we focus on conditions such that

$$1 - (\gamma + d - \beta P^\star)\Delta t > 1.$$

This is equivalent to $\gamma + d - \beta P^\star < 0$, i.e., $\mathcal{R}_0 > 1$. Hence the result for the DFFP.

Evaluating $J$ at the EFP gives the eigenvalues $1 - d\Delta t$ and $1 + (d + \gamma - \beta P^\star)\Delta t$. Clearly the latter is in the unit disk when $1 + (d + \gamma - \beta P^\star)\Delta t$ is outside it, and vice versa. So we have the result. □

**Remark:**

If an average lifetime of 70 years is used and the time units used are days, this would mean that in order for the eigenvalue $1 - d\Delta t$ to be outside the unit circle,

$\Delta t$ would have to be greater than

$$\frac{2}{d} = \frac{2}{\frac{1}{70\times 365.25}} = 140\,\text{years}.$$

Clearly the model here is a model for an endemic disease, so it would not describe a situation such as that seen in Figure 3.1. However, we will work with values of $I_{EFP}$ much larger than the absorbing threshold $N$, so this is not an issue.

### 3.2.2 Analysis of the Markov chain

We now write the transition matrix $\mathbf{T}$ in canonical form as explained in Chapter 1, reordering states as $\{0, N, 1, \ldots, N-1\}$. In this form, the matrix is block lower triangular,

$$\mathbf{T} = \begin{bmatrix} \mathbb{I}_2 & 0 \\ \mathbf{R} & \mathbf{Q} \end{bmatrix},$$

where $\mathbb{I}_2$ is the $2\times 2$ identity matrix, $\mathbf{R}$ is a $(N-2)\times 2$-matrix giving the probability of making a transition into the absorbing states from transient states,

$$\mathbf{R} = \begin{bmatrix} \Gamma_1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \Pi_{N-1} \end{bmatrix}$$

and $Q$ is the $(N-2)\times(N-2)$-matrix comprising the transition probabilities within the transient states,

$$\mathbf{Q} = \begin{bmatrix} p_1 & \Pi_1 & 0 & \dots & & 0 \\ \Gamma_2 & p_2 & \Pi_2 & \dots & & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & & & \Gamma_{N-2} & p_{N-2} & \Pi_{N-2} \\ 0 & & & & \Gamma_{N-1} & p_{N-1} \end{bmatrix}.$$

The fundamental matrix is then defined as $\mathbf{N} = (\mathbb{I}_{N-2} - \mathbf{Q})^{-1}$. For $i = 1, \dots, N-1$, we have $p_i = 1 - (\Gamma_i + \Pi_i)$, so the matrix $\mathbb{I}_{N-2} - \mathbf{Q}$ takes the form

$$\mathbb{I}_{N-2} - \mathbf{Q} = \Delta t \begin{bmatrix} \Gamma_1 + \Pi_1 & -\Pi_1 & 0 & \dots & 0 \\ -\Gamma_2 & \Gamma_2 + \Pi_2 & -\Pi_2 & \dots & 0 \\ & \ddots & \ddots & \ddots & \\ & & & -\Gamma_{N-1} & \Gamma_{N-1} + \Pi_{N-1} \end{bmatrix}.$$

Finding an explicit form for $\mathbf{N}$ is possible as a formula exists for the inverse of tridiagonal matrices such as $\mathbb{I}_{N-2} - \mathbf{Q}$ [43]. However, the computation is very involved and not worth the effort here: the fundamental matrix will be used only in numerical investigations of the properties of the system.

## 3.3 Numerical simulations

### 3.3.1 Parameters

The model has few parameters. They are chosen here to loosely match those of influenza, although it is clear that the model is not descriptive of influenza because it is a little too simple. The average duration $1/\gamma$ of the infection is taken to be 7 days. $\beta$ is chosen according to the values of $\mathcal{R}_0$ under investigation according to Equation (2.2). The death rate $d$ is chosen under the assumption that the average life expectancy in the population is 70 years. The birth rate $b$ is chosen so that the population will approach the desired fixed point $b/d$ (although $b$ does not present explicitly in the model, as we only model the number of infected individuals, while births only happen in the susceptible compartment).

### 3.3.2 Effect of $N$, $P$ and $\Delta t$

Before we begin a short exploration of the numerical properties of the model in the single location case, let us consider the influence of *structural* parameters, namely $N$, $P$ and $\Delta t$.

#### The total population, $P$

The total population $P$ is typically chosen to represent a small (1,000), medium (50,000) or large (1,000,000) community. Since the model is set up so that the Markov chain terminates when the number of infected people reaches either 0 or $N$, the only part $P$ plays here is in the choosing of $\beta$, via the rearranged formula for the basic reproduction number

$$\beta = \frac{\mathcal{R}_0(d+\gamma)}{P}.$$

**The role of the time step, $\Delta t$**

The parameter $\Delta t$ plays an important role. As usual, the probabilistic argument is used in that $\Delta t$ is chosen small enough so that only one event takes place at each time step. However, this is of little use in practice and it is better here to consider the role of $\Delta t$ in (3.2). Consider the row sums of off-diagonal entries there. For $i = 1, \ldots, N-1$, that sum is

$$\Gamma_i + \Pi_i = (\beta i(P-i) + (\gamma + d)i)\,\Delta t = \left(-\beta i^2 + (\beta P + \gamma + d)i\right)\Delta t.$$

As (3.2) must be a stochastic matrix, this quantity cannot be larger than 1, for $i = 1, \ldots, N-1$. The polynomial $Q(i) = -\beta i^2 + (\beta P + \gamma + d)i$ is 0 when $i = 0$ and increases until it reaches a maximum at $i = P + \frac{\gamma+d}{\beta}$, which may or may not be an integer but is larger than $i = P$. As $N \leq P$, this implies that, for the model, $Q(i)$ is maximum at $i = N-1$. It follows that $\Delta t$ is constrained by

$$\Delta t \leq \frac{1}{(\beta P + \gamma + d)(N-1) - \beta(N-1)^2}. \tag{3.9}$$

Note that this is a technical condition: in practice, $\Delta t$ is chosen smaller than this bound in order to satisfy the assumption that only one event can take place during each time step.

We now check the range of values taken by the upper bound of the time step given by (3.9) as a function of $N$ to make sure that this does not lead to unrealistically small values. For P = 1,000,000, and $\beta$ chosen so that $\mathcal{R}_0 = 1$, the results are as shown in Figure 3.2. Although it is not visible in the graph, the lowest value of $\Delta t$ (0.07142054) is achieved when $N$ reaches 50. In practice, for whatever values of $\mathcal{R}_0$, $N$ and $P$ in question, we will choose $\Delta t$ to be half the lowest value we find in the corresponding simulation.

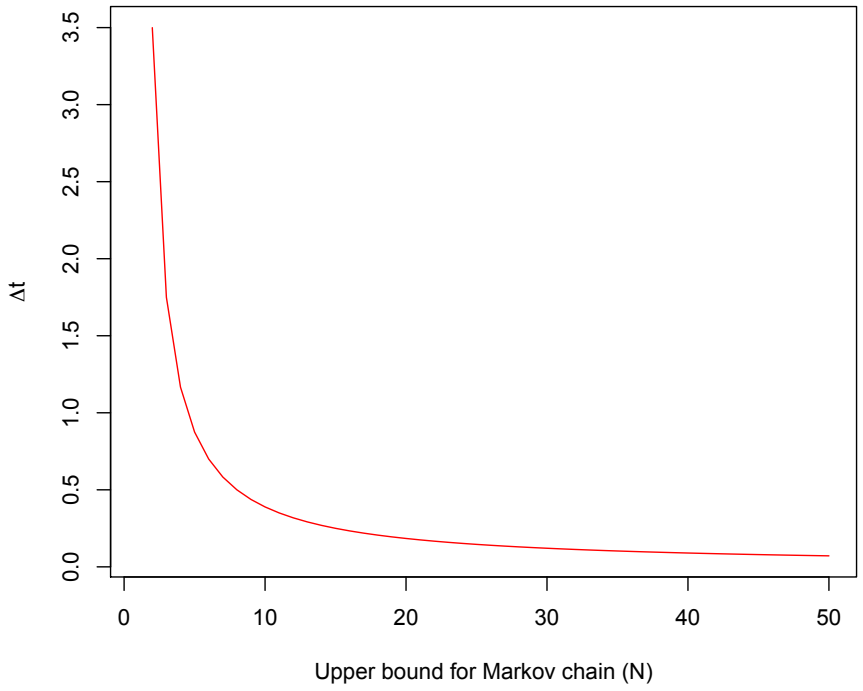



Figure 3.2: Value of the upper bound for $\Delta t$ given by (3.9) as a function of the absorption threshold $N$, when $\mathcal{R}_0 = 1$.

### The upper threshold for the Markov chain, $N$

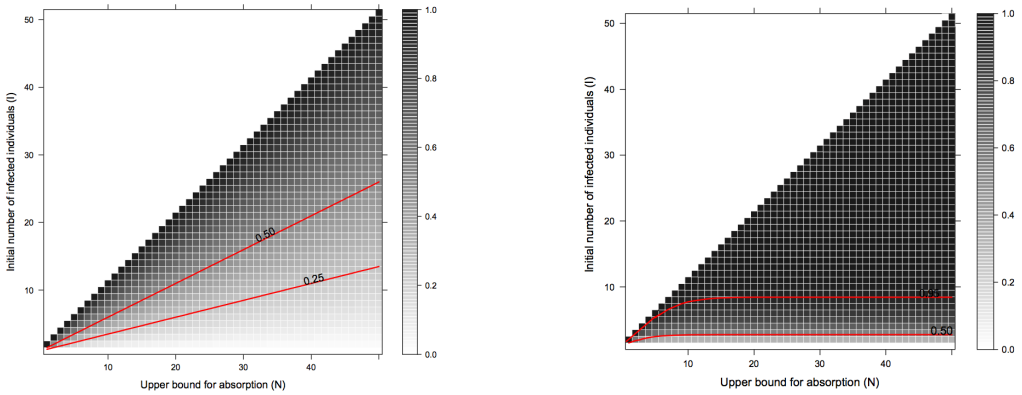
Before starting the investigations into the probability of disease extinction in earnest, we now explore the choice of the parameter $N$. For our first example, consider Figure 3.3.

In this figure, we are investigating the likelihood of a major outbreak as a function of the the absorption threshold $N$ of our Markov chain. When $N$ is chosen relatively small, the probability of a major outbreak changes drastically with the initial number of infected individuals. Thus, a very small $N$ is not ideal and probably not representative of reality. However, when $N$ is relatively large, the probability of a major outbreak is not so sensitively dependent on the initial number of infected individuals. These graphs help give us a sense for the threshold number to use in the Markov chain; i.e., the number of individuals $N$, for which once the number of infected $I$ exceeds $N$, an epidemic is likely. Choosing $N$ too small could cause us to incorrectly predict a major epidemic when that may not be the case, but a quick look at the graphs show us that we need not take $N$ excessively large to be confident in our predictions.

From this analysis, we can conclude that after a certain number of individuals in a population become infected an epidemic is extremely likely, and we will use this threshold number in later simulations. Hence, rather than running computationally expensive simulations with hundreds of thousands or even millions of states for the Markov chain to occupy, we can set the Markov chain to terminate after the number of infected individuals reaches this threshold quantity with satisfactory certainty that the disease will indeed become an epidemic.

Consider Figure 3.3b. The upper curve indicates which values of $N$ and $I$ give rise to a 0.95 probability of a major outbreak, with everything above the curve representing a probability greater than 0.95. As we can see, there is some value of $N$ for which the curve noticeably begins to level out. The number of initially

infected individuals needed so that there is a probablity of 0.95 for an epidemic is almost the same at $N = 12$ as at $N = 40$ (judging from the diagram, about $I = 7$). This levelling out of the curve can be seen for any value of $\mathcal{R}_0 > 1$. The closer we take $\mathcal{R}_0$ to 1, the longer it takes for the curve to level out, but it does so eventually nonetheless. As we can see from Figure 3.3a, when $\mathcal{R}_0 = 1$, the relationship is linear for each probability we choose.




(a) When $R_0 = 1$.

(b) When $R_0 = 1.5$.

Figure 3.3: Probability of a disease outbreak as a function of the number of cases required for an outbreak ($N$) and the initial number of infectious cases.

Figure 3.3b indicates that for $\mathcal{R}_0 > 1$, there is some number of infected individuals $I$ which indicates a high likelihood of an epidemic regardless of the threshold quantity $N$ for the Markov chain (we can see this from the levelling of the curve). However, we need not rely solely on numerics to reach this conclusion. In 1955, Whittle [42] gave a relationship between the number of infected individuals $i_0$ in a population for a continuous time analog of the present model, the basic reproduction number $\mathcal{R}_0$ and the probability $p$ of the disease quickly dying out (so that $1-p$ is the probability of a major epidemic). That relationship comes in the form of the equation

$$1 - p = \left(\frac{1}{\mathcal{R}_0}\right)^{i_0}, \tag{3.10}$$

for $\mathcal{R}_0 > 1$. This is a useful formula for setting up the Markov chain model. In

order to choose the upper bound for high likelihood of a major outbreak, we need only choose an appropriately high probability $p$ and set $\mathcal{R}_0$ according to the disease in question. For example, setting $1 - p = 0.99$ and $\mathcal{R}_0 = 1.5$, we get from (3.10) that $i_0 = 11.36$, so we use 12 infected individuals as the threshold quantity $N$, above which we assume the disease outbreak becomes an epidemic.

To run a simulation in which $\mathcal{R}_0$ is made to vary between two bounds greater than 1, we need to slightly modify our use of Whittle's formula (3.10) in choosing our upper bound. If we calculate the threshold quantity $i_0$ from the lowest value over which $\mathcal{R}_0$ ranges, then it is easy to see that any larger values of $\mathcal{R}_0$ in the formula will only result in lower probabilities of the disease dying out.

### 3.3.3 Investigations into the probability of disease extinction

In the simulation shown in Figure 3.4, we vary $1.1 \leq \mathcal{R}_0 \leq 1.5$. Using $\mathcal{R}_0 = 1.1$ in Whittle's formula, we find the threshold quantity of infectious individuals to be 49 (for this simulation we use 50 for convenience). Hence, the Markov chain has two absorbing states at 0 and 50 infected individuals. The graph then ranges for $1.1 \leq \mathcal{R}_0 \leq 1.5$ and the initial number of infectious individuals from 1 to 49. The contour on the graph highlights which values of $\mathcal{R}_0$ and $I$ give us a probability of 0.1 of the disease dying out without first becoming an epidemic. As we can see, any pair of $\mathcal{R}_0$ and $I$ to the right of the line represent a probability of less than 0.1 of the disease dying out, so the likelihood of a major epidemic is high.

In order to perform simulations where $\mathcal{R}_0$ is allowed to vary below 1, we must now abandon Whittle's formula. In this case, we pick a reasonably high number for the threshold value to use in the simulation. For example, choosing 100 as the upper bound, and varying $\mathcal{R}_0$ between 0.5 and 2, the situation is as shown in Figure 3.5. In this case, the curve highlights all the pairs of $I$ and $\mathcal{R}_0$ for which the disease has probability of 0.5 of both dying out and of becoming an epidemic.

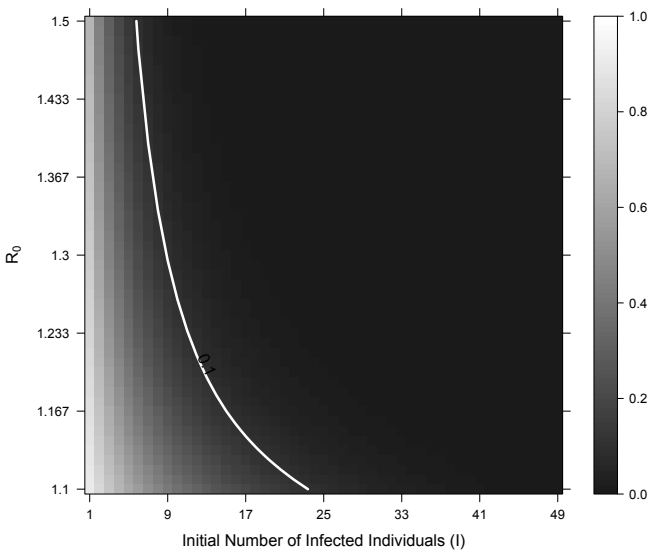



Figure 3.4: Probability of disease extinction due to variations of $\mathcal{R}_0$ and the number of infectious individuals in a population.

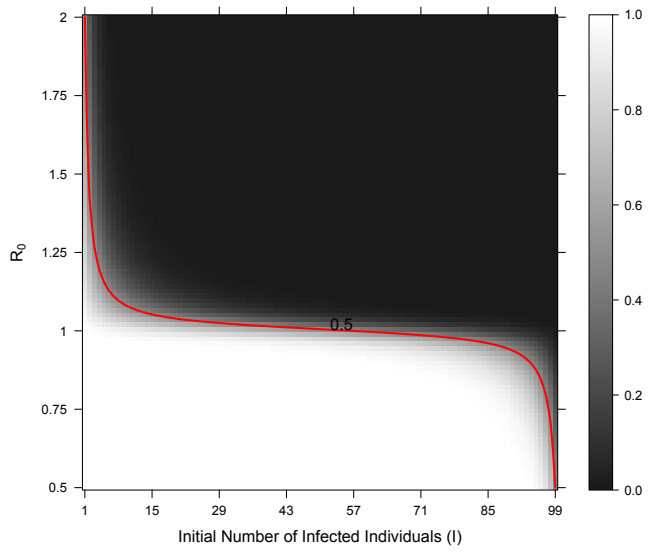



Figure 3.5: Probability of disease extinction due to variations of $\mathcal{R}_0$ and the number of infectious individuals in a population.

### 3.3.4 Final size of the epidemic

Suppose now that for a disease with some particular $\mathcal{R}_0$, we want to calculate the final size of the epidemic. Recall from Section 2.6.3 that the final size of an epidemic is defined to be the total number of individuals who experience infection during an outbreak. The method of calculating these numbers follows that given there.

Recall also that since the model only investigates the early phase of an outbreak, we must first condition on the disease becoming extinct.

Suppose the disease in question has basic reproduction number $\mathcal{R}_0 = 1.2$. From Whittle's formula (3.10), we find that for this choice of $\mathcal{R}_0$, the threshold $N$ must be at least 26 (i.e., to be at least 99% sure an epidemic occurs, the system must reach 26 infectious individuals). For these parameters, and starting the simulation with one infected individual, after 1000 realizations of the Markov chain, we find the average final size of the epidemic to be 4.6. Of the 1000 realizations, 820 resulted in disease extinction while 144 resulted in an epidemic with 36 realizations still not absorbed after the initial 90 days. These numbers reflect the predictions of the matrix $NR$ reasonably well; with parameters as above, the matrix $NR$ gives a probability of 0.83 of disease extinction given one initially infected individual while 85% of the absorbed realizations that ended in disease extinction.

Suppose now that we allow $\mathcal{R}_0$ to vary between 0.5 and 2. As before, Whittle's formula no longer applies in choosing the upper bound for the Markov chain when $\mathcal{R}_0$ falls below one, so we let $N = 50$. The results are as shown in Figure 3.6.

Note that as $\mathcal{R}_0$ grows larger and larger, the results here become less meaningful. When $\mathcal{R}_0 > 1$ is increasing, we typically expect an outbreak to become more and more likely. Therefore, conditioning on the disease becoming extinct is essentially conditioning on unexpected behaviour, and because the vast majority of our simulations will be discarded (the ones that end up absorbed in the upper threshold), the data we gain from this simulation is representative of fewer realizations.

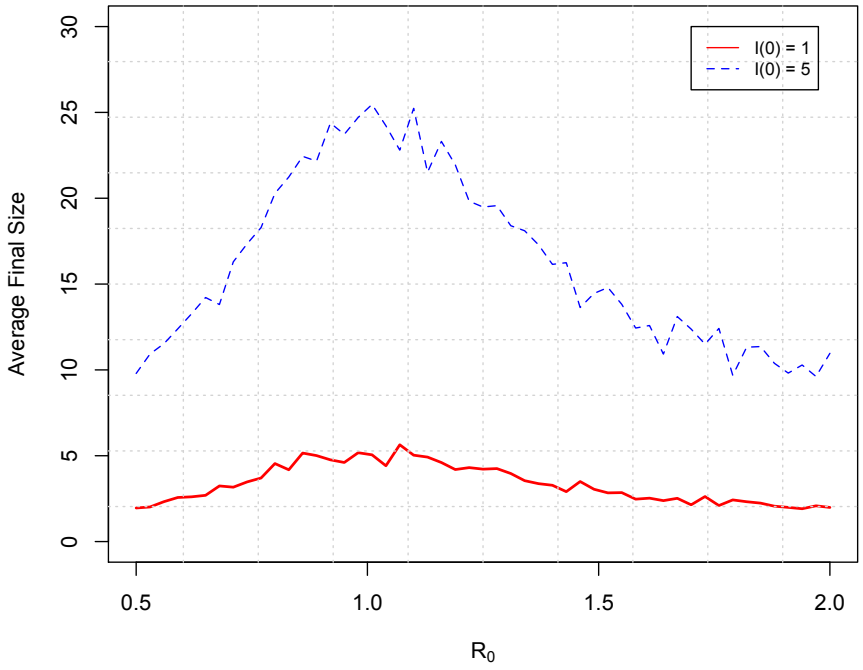



Figure 3.6: Average final size of the epidemic for $0.5 \leq \mathcal{R}_0 \leq 2$.

The blue (dashed) curve represents the average final size of the outbreak when there are initially 5 people infected, while the red (continuous) curve is the average final size when there is just one person initially infected. In both curves we see a slight increase in final size when $\mathcal{R}_0$ is around 1.

We conclude the investigation into the final size of the epidemic by examining the expected duration of the stochastic phase of the outbreak, i.e., the time until the chain is absorbed in either of the absorbing states. For this simulation, we allow $\mathcal{R}_0$ to vary from 0.3 to 3, and the initial number of infected from 1 to 49 with $N = 50$.

Figure 3.7 shows that the the average duration of the stochastic phase is longest when $\mathcal{R}_0$ is around 1 and the initial number of infected people is between 6 and 17 with the largest length of time until absorption reaching up to just over 8 weeks.

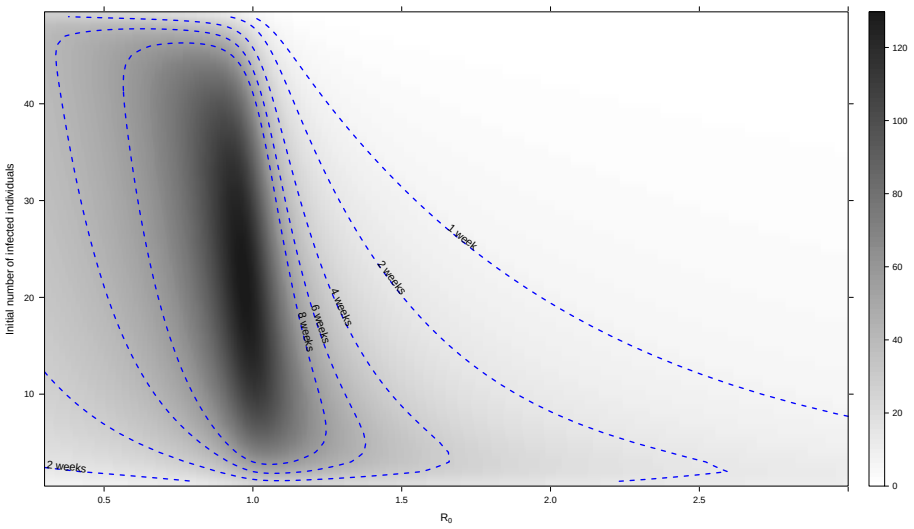



Figure 3.7: Average duration of the stochastic phase when the initial number of infected varies from 1 to 25 and $\mathcal{R}_0$ varies from 0.3 to 3.

# 4

# Two interconnected locations

We now couple two systems of the form defined in Chapter 3 in order to describe the movement of infectious individuals between two distinct locations and the effect this has on a nascent epidemic.

## 4.1 The model

Let $A$ and $B$ be the two locations, $S^A(t)$, $S^B(t)$, $I^A(t)$ and $I^B(t)$ be random variables representing the number of susceptible individuals in locations $A$ and $B$ and infectious individuals in locations $A$ and $B$, respectively. Let $P^A$ and $P^B$ be the total population in $A$ and $B$, respectively. In principle, one should consider these numbers as variables, as the total population in a location changes as individuals move in and out of it. However, as we are considering the early phase of an epidemic, when $S^A \simeq P^A$ and $S^B \simeq P^B$, we consider $P^A$ and $P^B$ to be fixed. As a consequence, we need only track $I^A(t)$ and $I^B(t)$. For $X \in \{A, B\}$, denote $I_i^X(t) = \{I^X(t) = i\}$, i.e., the situation where there are $i$ infectious individuals in location $X$ at time $t$. We omit the dependence on $t$ if that does not lead to confusions.

In order to formulate a Markov chain to describe the evolution of both $I^A$ and $I^B$, we must consider the states of the Markov chains in the two locations as pairs

and thus consider states of the form

$$I_i^A I_j^B,$$

meaning that there are $i$ infectious people in location A and $j$ infectious people in location B.

As in the one population case of Chapters 3, we assume that there is a threshold number of infected individuals above which an epidemic takes place. We assume that this threshold can be different in the two locations and denote the thresholds $N^A$ in location A and $N^B$ in location B. When either location reaches its threshold, it is assumed that a major outbreak occurs, i.e., the system has reached an absorbing state. Thus,

$$I_i^A I_j^B, \qquad i = 0, \ldots, N^A, \quad j = 0, \ldots, N^B, \quad 0 \leq i + j < N^A + N^B \tag{4.1}$$

are all the possible states in the chain. We call such states *admissible*. A state that is not admissible is called *unreachable*. Note that the right inequality is strict: the state $I_{N^A}^A I_{N^B}^B$ is indeed unreachable, since prior to reaching this state, the chain would already have been absorbed in $I_{N^A}^A$ or $I_{N^B}^B$.

We call *prevalence* the total number $k$ of infectious individuals in the system and denote

$$\mathcal{I}_k = \{I_A^i I_B^j \mid i + j = k, \quad i = 0, \ldots, N^A, \quad j = 0, \ldots, N^B\}$$

the set of admissible states in the chain achieving prevalence $k$. From the remark above, prevalence is such that $0 \leq k < N^A + N^B$.

As in Chapter 3, we assume that the time step $\Delta t$ is small enough that events can only affect a single individual. As a consequence, only a small number of transitions are possible, as represented in Figure 4.1, which shows the transition graph of the

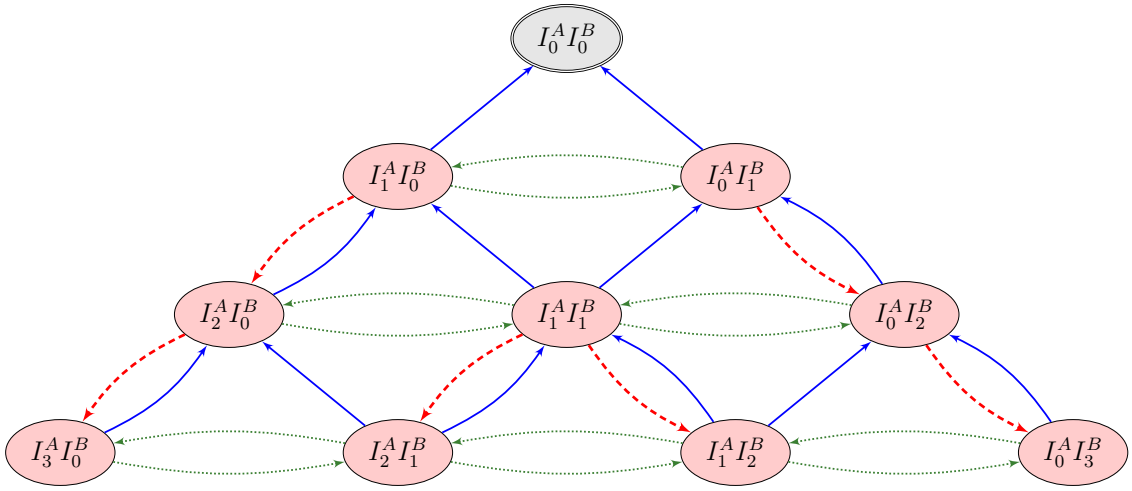



Figure 4.1: The first few states of the chain corresponding to a maximum prevalence of 3 in the coupled two-locations system, when $N^A > 3$ and $N^B > 3$. The doubly circled state is absorbing.

system for a prevalence equal to 3 or less. To simplify the graph, we have omitted a self loop on each state $I_i^A I_j^B$. Each row in this representation shows the states corresponding to a given prevalence, from 0 at the top to 3 at the bottom. The

| | |
|---|---|
| ⇢ | Movement between locations. |
| → | Recovery of an infectious individual. |
| ⇢ | New infection. |

Table 4.1: Meaning of the symbols used in Figures 4.1 and 4.2.

signification of the arrows used in Figure 4.1 is explained in Table 4.1.

Now consider an arbitrary admissible state $I_i^A I_j^B$, assuming a prevalence $k = i + j \geq 2$. First, assume that the state is "interior" to the digraph, i.e., $0 < i < N^A$ and $0 < j < N^B$. Then the situation is as represented in Figure 4.2, which shows all the possible transitions to and from any state $I_i^A I_j^B$ provided they exist (for example for a state such as $I_1^A I_0^B$, we cannot have recovery of an infectious individual in location $B$ so there is no transition to $I_i^A I_{0-1}^B$).

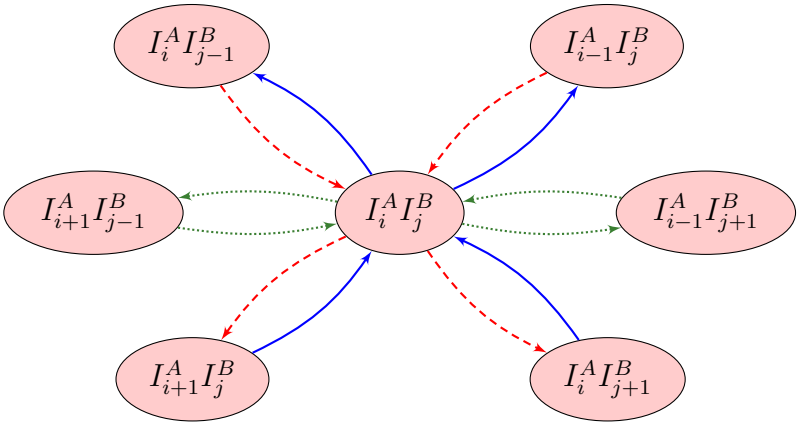



Figure 4.2: States directly connected to a state with $i$ infectious individuals in location $A$ and $j$ individuals in location $B$.

## 4.2 Transition probabilities

| Transition | Probability | Meaning |
|---|---|---|
| $I_i^A I_j^B \to I_{i-1}^A I_j^B$ | $\Gamma_i^A$ | Recovery or death in $A$ |
| $I_i^A I_j^B \to I_{i+1}^A I_j^B$ | $\Pi_i^A$ | Infection in $A$ |
| $I_i^A I_j^B \to I_i^A I_{j-1}^B$ | $\Gamma_j^B$ | Recovery or death in $B$ |
| $I_i^A I_j^B \to I_i^A I_{j+1}^B$ | $\Pi_j^B$ | Infection in $B$ |
| $I_i^A I_j^B \to I_{i+1}^A I_{j-1}^B$ | $M_{ij}^{AB}$ | Movement of infectious from $A$ to $B$ |
| $I_i^A I_j^B \to I_{i-1}^A I_{j+1}^B$ | $M_{ij}^{AB}$ | Movement of infectious from $B$ to $A$ |
| $I_i^A I_j^B \to I_i^A I_j^B$ | $p^{ij}$ | Nothing happens |

Table 4.2: Transitions and corresponding probabilities in the model for two locations A and B, assuming that the states and transitions are allowable. See the text for details.

As in Chapter 3, we assume that the probability of a new infection in location $X \in \{A, B\}$ is

$$\Pi_i^X = \beta^X (P^X - i^X) i^X \, \Delta t, \tag{4.2}$$

respectively, where $s^X = P^X - i^A$ and $i^X$ are the numbers of susceptible and infectious individuals in location $X \in \{A, B\}$, respectively. Similarly, the probability of

recovery or death in $X \in \{A, B\}$ is given by

$$\Gamma_i^X = (\gamma^X + d) i^X \, \Delta t. \tag{4.3}$$

Note that these processes are "localized": it is required for individuals to be physically present in their location for the event to take place. In addition to local processes, we introduce transport events, through which infectious individuals in a location travel to the other location. We define the probability of movement of an infectious individual from location A to location B and vice versa in the time interval $\Delta t$ as

$$M_{ij}^{AB} = m^{AB} i^A \, \Delta t \tag{4.4}$$

$$M_{ij}^{BA} = m^{BA} i^B \, \Delta t \tag{4.5}$$

where $m^{AB}$ and $m^{BA}$ are the probabilities that a member of population A travels to population B and vice versa in a time step.

Adding a second location and taking into account that both locations have absorbing states greatly complicates the situation compared to what it is in Chapter 3. The discussion on values of probabilities that precedes assumes that all states involved are admissible; the situation is in fact a little more complicated. Take for instance (4.4). While correct in general, it should be noted that the precise form of the equation is as follows:

$$M_{ij}^{AB} = m^{AB} i^A \, \mathbf{1}_{\{1,\ldots,N_A-1\}}(i^A) \mathbf{1}_{\{1,\ldots,N_B-1\}}(i^B) \, \Delta t,$$

where $\mathbf{1}_A(x)$ is the indicator function of set $A$. Indeed, if, for example, $i^B$ has been absorbed into $N^B$, then the chain has reached an absorbing state and no further transition is possible. See Table 4.2 for a summary of transitions in the model.

## 4.3 States with a given prevalence

Before detailing the transition matrix, it is useful to consider the sets of states achieving a given prevalence, as they play a major role in determining factors such as the size, shape and structure (zero/nonzero pattern) of the transition matrix. Let us illustrate the role of prevalence by considering the situation for $N^A = 2$ and $N^B = 3$, shown in Figure 4.3. Consider prevalences $0 < k < 2 = \min(N^A, N^B)$. Then as in Figure 4.2, the system can transition from any state with prevalence $k$ to any other state with prevalence $k + 1$ via an infection in either location. When $k = \min(N^A, N^B) = 2$, the possible transitions depend on whether all infectious are in A (in which case, the chain has reached an absorbing state and no further transitions are possible) or if $i^A < N^A$ (in which case, infections are possible in both locations).

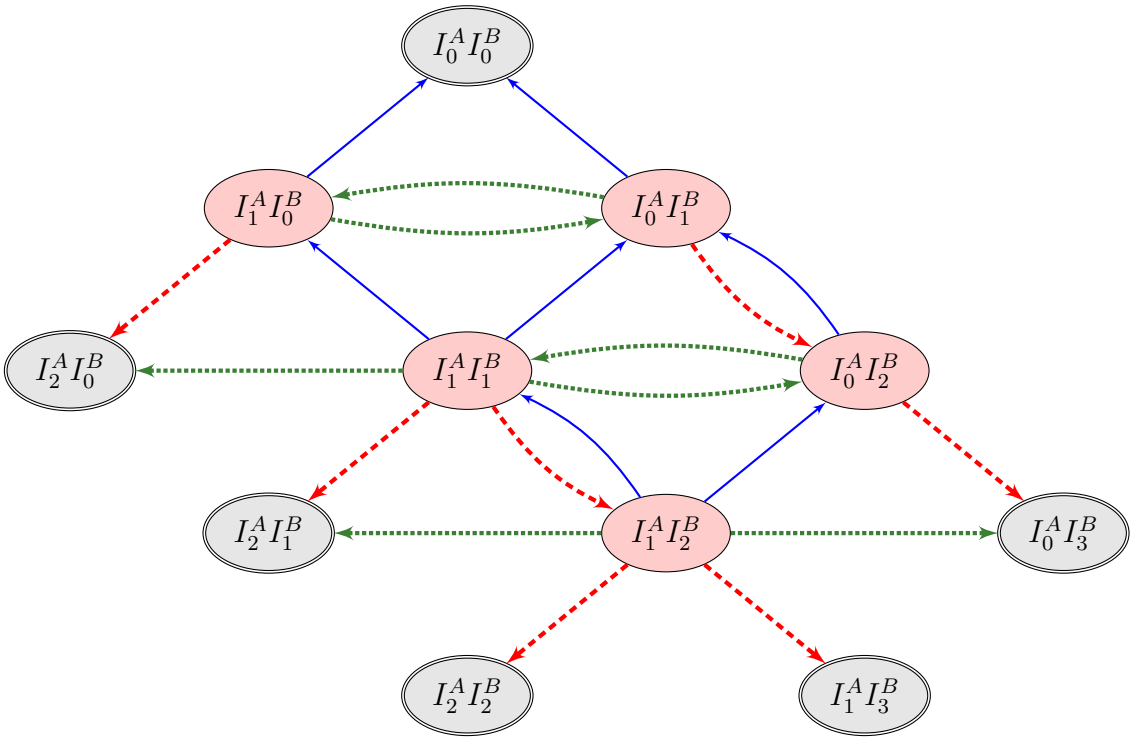



Figure 4.3: States of the chain when $N^A = 2$ and $N^B = 3$. Doubly circled states are absorbing.

Let us now detail the number of states associated to each prevalence level.

1. First, suppose that the prevalence $k = 0$. Then there is one admissible state, $I_0^A I_0^B$ which is absorbing.

2. Second, let prevalence $k$ be such that $0 < k < \min(N^A, N^B)$. Then there are $k+1$ admissible states, which are ordered as

$$\mathcal{I}_k = \{I_k^A I_0^B, I_{k-1}^A I_1^B, \ldots, I_0^A I_k^B\}. \tag{4.6a}$$

   All states in (4.6a) are transient.

3. Third, suppose that prevalence $k$ is such that $\min(N^A, N^B) \leq k < \max(N^A, N^B)$. Note that if $N^A = N^B$, this case does not arise. If $N^A \neq N^B$, assume without loss of generality that $N^A < N^B$. Keeping the same ordering for states of prevalence $k$ as in (4.6a), it is clear that states $I_{k-\ell}^A I_\ell^B$, for $\ell = 0, \ldots, k - N^A$, are either unreachable (when $\ell < k - N^A$) or absorbing (when $\ell = k - N^A$). The former are not considered, so the ordered list of states takes the form

$$\mathcal{I}_k = \{I_{N^A}^A I_{k-N^A}^B, I_{N^A-1}^A I_{k-N^A+1}^B, \ldots, I_0^A I_k^B\}. \tag{4.6b}$$

   For each $k$ in this range, $\mathcal{I}_k$ has $N^A + 1$ states regardless of the value of $k$. The state $I_{N^A}^A I_{k-N^A}^B$ is absorbing while all other states are transient.

4. Fourth, suppose that prevalence $k$ is such that $\max(N^A, N^B) \leq k < N^A + N^B$. Assuming again that $N^A < N^B$, the ordered list of states of prevalence $k$ is

$$\mathcal{I}_k = \{I_{N^A}^A I_{k-N^A}^B, I_{N^A-1}^A I_{k-N^A+1}^B, \ldots, I_{k-N^B}^A I_{N^B}^B\}. \tag{4.6c}$$

   There are $N^A + N^B + 1 - k$ states in (4.6c). Among these states, both $I_{N^A}^A I_{k-N^A}^B$ and $I_{k-N^B}^A I_{N^B}^B$ are absorbing; the remainder are transient.

5. Finally, as remarked earlier, it is impossible to reach state $I_{N^A}^A I_{N^B}^B$. Indeed, to

reach it, the process would have to leave one of the absorbing states $I^A_{N^A-1}I^B_{N^B}$ or $I^A_{N^A}I^B_{N^B-1}$.

**Lemma 4.1.** If $N^A < N^B$, then the Markov chain has $(N^A+1)(N^B+2)-1$ admissible states. If $N^A = N^B$, then the Markov chain has $(N^A+1)^2-1$ admissible states. In both cases, $N^A+N^B+1$ of these states are absorbing.

*Proof.* In order to obtain the result, we examine the cases separately based on the different prevalence classes detailed above. First, assume that $N^A < N^B$.

1. For a given prevalence $k \leq N^A - 1$, there are $k+1$ admissible states, giving $N^A(N^A+1)/2$ admissible states with prevalence $k \leq N^A - 1$.

2. Now consider prevalence in the range $N^A \leq k \leq N^B - 1$. There are $N^A + 1$ admissible states for each $k$ in this range giving a total of $((N^B-1)-N^A)(N^A+1)$ admissible states.

3. Finally, for prevalences from $N^B \leq k \leq N^A + N^B - 1$, the number of states decreases from $N^A+1$ to 2, giving $(N^A+1)(N^A+2)/2-1$ admissible states.

Taking the sum of the results from the three cases above, the number of states is

$$\begin{aligned}\frac{N^A(N^A+1)}{2} + (N^B-N^A-1)(N^A+1) + \frac{(N^A+1)(N^A+2)}{2} - 1 &= (N^A+1)\left(N^A + (N^B-N^A+1)+1\right) - 1 \\ &= (N^A+1)(N^B+2) - 1.\end{aligned}$$

In the case that $N^A = N^B$, we do not include the states for $N^A \leq k \leq N^B - 1$, so we get that the number of states is

$$\begin{aligned}\frac{N^A(N^A+1)}{2} + \frac{(N^A+1)(N^A+2)}{2} - 1 &= (N^A+1)\left(\frac{N^A}{2} + \frac{N^A+2}{2}\right) - 1 \\ &= (N^A+1)(N^A+1) - 1 \\ &= (N^A+1)^2 - 1.\end{aligned}$$

Let us now count the number of absorbing states. Consider the same prevalence ranges as those in the analysis of admissible states above. If $N^A < N^B$, then each

prevalence level $\mathcal{I}_k$ for $N^A \leq k \leq N^B - 1$ has one absorbing state and there are $N^B - N^A$ such states. Prevalences from $k = N^B$ to $k = N^A + N^B - 1$ have two absorbing states, giving $2N^A$ such states. Adding the zero absorbing state gives the result.

In the case $N^A = N^B$, intermediate prevalences do not come into play and there are thus there are a total of $2N^A + 1 = N^A + N^B + 1$ absorbing states. □

## 4.4 Transition matrix

In the transition matrix associated to this system, states are grouped first by prevalence. When doing so, the transition matrix is a block tridiagonal matrix,

$$\mathbf{T} = \begin{bmatrix} 1 & \mathbf{I}_0 & & & & & & \\ \mathbf{R}_1 & \mathbf{M}_1 & \mathbf{I}_1 & & & & & \\ & \mathbf{R}_2 & \mathbf{M}_2 & \mathbf{I}_2 & & & & \\ & & & & & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & & & \mathbf{I}_{N^A+N^B-2} & \\ & & & & & \mathbf{R}_{N^A+N^B-1} & \mathbf{M}_{N^A+N^B-1} & \end{bmatrix}. \tag{4.7}$$

Blocks in $\mathbf{T}$ have varying sizes and structures depending on the prevalence $k$ and represent the following transitions:

- $\mathbf{I}_k$ represent transitions from prevalence class $\mathcal{I}_k$ to prevalence class $\mathcal{I}_{k+1}$, i.e., new infections,
- $\mathbf{R}_k$ describe transitions from prevalence class $\mathcal{I}_{k+1}$ to prevalence class $\mathcal{I}_k$, i.e., recoveries and deaths,
- $\mathbf{M}_k$ describe transitions within the prevalence class $\mathcal{I}_k$, i.e., movement between

locations as well as absence of transition. These are square matrices that sit along the main diagonal of blocks.

Prevalence level plays an important role in the size, shape and structure of the submatrices making up the transition matrix $\mathbf{T}$. The specific role of prevalence differs depending on the type of transition encoded by the matrices, so we now investigate each matrix type separately. From now on, without loss of generality we assume that $N^A \leq N^B$, so that $\min(N^A, N^B) = N^A$ and $\max(N^A, N^B) = N^B$. For context on how any of the upcoming submatrices fit into the overall transition matrix, see the sample matrix in (4.18) with $N^A = 3$ and $N_B = 5$.

### 4.4.1 Infection submatrices

Infection takes the system from states in set $\mathcal{I}_k$ to states in set $\mathcal{I}_{k+1}$. Thus, the shape (and size) of $\mathbf{I}_k$ depends on the prevalence of the disease. The situation is as follows:

1. $\underline{\mathcal{I}_\mathbf{0} \text{ and } \mathcal{I}_\mathbf{1}}$

   If $k = 0$, the infection matrix is the $1 \times 2$ zero vector, i.e.,

   $$\mathbf{I}_0 = [0 \quad 0]\,,$$

   while if $k = 1$, it becomes

   $$\mathbf{I}_1 = \begin{bmatrix} \Pi_1^A & 0 & 0 \\ 0 & 0 & \Pi_1^B \end{bmatrix}.$$

2. $\underline{\mathcal{I}_\mathbf{k} \text{ for } \mathbf{2 \leq k < N^A}}$

   If $2 \leq k < N^A$, the infection matrices take the form of the $(k+1) \times (k+2)$-

matrix

$$\mathbf{I}_k = \begin{array}{c} \\ I_k^A I_0^B \\ I_{k-1}^A I_1^B \\ \vdots \\ I_1^A I_{k-1}^B \\ I_0^A I_k^B \end{array} \begin{array}{c} \begin{array}{ccccccc} I_{k+1}^A I_0^B & I_k^A I_1^B & & \cdots & & I_1^A I_k^B & I_0^A I_{k+1}^B \end{array} \\ \begin{bmatrix} \Pi_k^A & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \Pi_{k-1}^A & \Pi_1^B & & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & & \Pi_1^A & \Pi_{k-1}^B & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \Pi_k^B \end{bmatrix} \end{array}. \tag{4.8}$$

Note that for convenience, the states from which the system starts (row) and the state to which the system proceeds (column) have been indicated on the borders of $I_k$. This will be done throughout this section wherever necessary.

3. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^A \leq k < N^B}}$

When the overall prevalence reaches $N^A$, we see a change in both the shape and the structure of the infection matrices. When the prevalence $N^A \leq k < N^B$, the general form of infection matrices is similar to (4.8), except that the top row of $\mathbf{I}_k$ is zero, since it corresponds to a transition from state $I_{N^A}^A I_{k-N^A}^B$, which is impossible since it is absorbing. For each $N^A \leq k < N^B$, the infection

matrices $\mathbf{I}_k$ are square and of constant size $(N^A+1)\times(N^A+1)$.

$$\mathbf{I}_k = \begin{array}{c} \\ I^A_{N^A} I^B_{k-N^A} \\ I^A_{N^A-1} I^B_{k-N^A+1} \\ \vdots \\ I^A_1 I^B_{k-1} \\ I^A_0 I^B_k \end{array} \begin{array}{c} \begin{array}{cccccc} I^A_{N^A} I^B_{k-N^A+1} & I^A_{N^A-1} I^B_{k-N^A+2} & & \cdots & I^A_1 I^B_k & I^A_0 I^B_{k+1} \end{array} \\ \begin{bmatrix} 0 & 0 & \cdots & & 0 \\ \Pi^A_{N^A-1} & \Pi^B_{k-N^A+1} & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \Pi^A_1 \; \Pi^B_{k-1} & 0 \\ 0 & \cdots & & 0 & \Pi^B_k \end{bmatrix} \end{array}. \tag{4.9}$$

Recall that if $N^A = N^B$, these matrices do not occur.

4. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^B \le k < N^A + N^B - 1}}$

Consider finally sets $\mathcal{I}_k$ for prevalence $N^B \le k < N^A + N^B - 1$. The general form of the infection matrices is again similar to (4.8). In addition to the row of zeros as seen in (4.9), we also have a row of zeros at the bottom of the matrix, corresponding to an impossible transition from the absorbing state $I^A_{k-N^B} I^B_{N^B}$. For prevalence in this range, the matrices are no longer square, but have size

$(N^A + N^B + 1 - k) \times (N^A + N^B - k)$ and take the form

$$\mathbf{I}_k = \begin{array}{c} \\ I^A_{N^A} I^B_{k-N^A} \\ I^A_{N^A-1} I^B_{k-N^A+1} \\ \vdots \\ I^A_{k-N^B+1} I^B_{N^B-1} \\ I^A_{k-N^B} I^B_{N^B} \end{array} \begin{array}{c} \begin{array}{cccccc} I^A_{N^A+1} I^B_{k-N^A} & I^A_{N^A} I^B_{k-N^A+1} & \cdots & I^A_{k-N^B} I^B_{N^B-1} & I^A_{k-N^B+1} I^B_{N^B} \end{array} \\ \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \Pi^A_{N^A-1} & \Pi^B_{k-N^A+1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \Pi^A_1 & \Pi^B_{k-1} \\ 0 & \cdots & 0 & 0 \end{bmatrix} \end{array}. \tag{4.10}$$

### 4.4.2 Recovery (and death) submatrices

Before detailing the specifics of the recovery and matrices, some terminology is needed. We use the word *shape* of a matrix, to specify whether the matrix is square or rectangular. The word *structure* refers to the zero-nonzero pattern of entries within the matrix. Note that even though these submatrices include transitions due to recovery and death, we refer to them as "recovery matrices" from now on.

When a recovery occurs, the system transitions from prevalence states $\mathcal{I}_k$ to prevalence states $\mathcal{I}_{k-1}$. Similar to infection matrices, prevalence plays a role in the size and structure of the recovery matrices.

1. $\underline{\mathcal{I}_\mathbf{0}}$

   When the prevalence is zero there is no recovery matrix $\mathbf{R}_0$.

2. $\underline{\mathcal{I}_\mathbf{k} \textbf{ for } \mathbf{1 \leq k < N^A}}$

For $1 \leq k < N^A$, $\mathbf{R}$ is a $(k+1) \times k$-matrix and takes the form

$$\mathbf{R}_k = \begin{array}{c} \\ I_A^k I_B^0 \\ I_A^{k-1} I_B^1 \\ \\ \\ \\ I_A^1 I_B^{k-1} \\ I_A^0 I_B^k \end{array} \begin{array}{c} \begin{array}{cccc} I_A^{k-1} I_B^0 & I_A^{k-1} I_B^1 & & I_A^1 I_B^{k-2} \quad I_A^0 I_B^{k-1} \end{array} \\ \begin{bmatrix} \Gamma_k^A & 0 & 0 & 0 \\ \Gamma_1^B & \Gamma_{k-1}^A & 0 & 0 \\ & & \ddots & \\ 0 & 0 & \Gamma_B^{k-1} & \Gamma_A^1 \\ 0 & 0 & 0 & \Gamma_B^k \end{bmatrix} \end{array}. \tag{4.11}$$

3. $\underline{\mathcal{I}_{\mathbf{N^A}}}$

When the prevalence reaches $N^A$, we see a change in the structure of the matrix. The recovery matrix is similar to that in (4.11) except that all entries of the first row become zero because that row now represents transitions from the absorbing state $I_A^{N^A} I_B^0$, which are impossible. When the system reaches this prevalence level, the recovery matrix is an $(N^A + 1) \times N^A$-matrix.

$$\mathbf{R}_{N^A} = \begin{array}{c} \\ I_A^{N^A} I_B^0 \\ I_A^{N^A-1} I_B^1 \\ \vdots \\ \\ \\ I_A^1 I_B^{N^A-1} \\ I_A^0 I_B^{N^A} \end{array} \begin{array}{c} \begin{array}{cccc} I_A^{N^A-1} I_B^0 & I_A^{N^A-2} I_B^1 & \cdots & I_A^1 I_B^{N^A-2} \quad I_A^0 I_B^{N^A-1} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \Gamma_1^B & \Gamma_{N^A-1}^A & 0 & 0 \\ & & \ddots & \\ 0 & 0 & \Gamma_B^{N^A-1} & \Gamma_A^1 \\ 0 & 0 & 0 & \Gamma_B^{N^A} \end{bmatrix} \end{array}. \tag{4.12}$$

4. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^A < k < N^B}}$

Suppose now that $N^A < k < N^B$. Then recovery matrices $\mathbf{R}_k$ are similar to those in (4.12) except that the dimension is $(N^A+1)\times(N^A+1)$. The transition matrix looks the same as that in (4.12) but with a column of zeros attached on the left.

$$\mathbf{R}_k = \begin{array}{c} \\ I^A_{N^A} I^B_{k-N^A} \\ I^A_{N^A-1} I^B_{k-N^A+1} \\ \vdots \\ I^A_1 I^B_{k-1} \\ I^A_0 I^B_k \end{array} \begin{array}{c} \begin{array}{cccccc} I^A_{N^A} I^B_{k-N^A-1} & I^A_{N^A-1} I^B_{k-N^A} & & \cdots & & I^A_0 I^B_{k-1} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \Gamma^B_{N^A-1} & \Gamma^A_{k-N^A+1} & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & & & \Gamma^{k-1}_B & \Gamma^1_A \\ 0 & 0 & & \cdots & 0 & \Gamma^k_B \end{bmatrix} \end{array}. \quad (4.13)$$

5. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^B \le k \le N^A + N^B - 1}}$

When the prevalence reaches $N^B$, the size of the matrix remains the same as in (4.13), but the last row is now also a row of zeros. This is because in this prevalence range, the last row of $\mathbf{R}_k$ contains transitions from $I^A_{k-N^A} I^B_{N^B}$, which is an absorbing state.

$$\mathbf{R}_k = \begin{array}{c} \\ I^A_{N^A} I^B_{k-N^A} \\ I^A_{N^A-1} I^B_{k-N^A+1} \\ \vdots \\ I^A_{k-N_A+1} I^B_{N^B-1} \\ I^A_{k-N^A} I^B_{N^B} \end{array} \begin{array}{c} \begin{array}{cccccc} I^A_{N^A} I^B_{k-N^A-1} & I^A_{N^A-1} I^B_{k-N^A} & & \cdots & & I^A_{k-N^A} I^B_{N^B-1} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \Gamma^B_{N^A-1} & \Gamma^A_{N^B-N^A+1} & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & & & \Gamma^{N^B-1}_B & \Gamma^1_A \\ 0 & 0 & & \cdots & 0 & 0 \end{bmatrix} \end{array}. \quad (4.14)$$

6. $\underline{\mathcal{I}_{\mathbf{N^A+N^B-1}}}$

For the prevalence level $N^A + N^B - 1$, the recovery matrix takes the form of the $2 \times 3$ zero matrix

$$\mathbf{R}_{N^A+N^B-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

### 4.4.3 Movement (and absence of transition)

Movement and absence of transition does not change prevalence; movement only affects the current location of infectious individuals. As with the infection and recovery matrices, although the movement matrices are always square, the prevalence determines the size and structure of the matrices.

1. $\underline{\mathcal{I}_{\mathbf{0}}}$

   When the prevalence is zero, we do not have a movement and absence of transition matrix because we have just one (absorbing) state. This is represented in our overall transition matrix as a 1 in the $(1, 1)$ position.

2. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{1 \leq k < N^A}}$

   When prevalence is in this range, all movement transitions are possible – since prevalence is between 1 and $N^A$, there are no absorbing states. The movement

matrices are $k \times k$ matrices of the form

$$\mathbf{M}_k = \begin{array}{c} \\ I_k^A I_0^B \\ I_{k-1}^A I_1^B \\ I_{k-2}^A I_2^B \\ \vdots \\ I_2^A I_{k-2}^B \\ I_1^A I_{k-1}^B \\ I_0^A I_k^B \end{array} \begin{array}{c} \begin{array}{ccccccc} I_k^A I_0^B & I_{k-1}^A I_1^B & I_{k-2}^A I_2^B & \cdots & I_2^A I_{k-2}^B & I_1^A I_{k-1}^B & I_0^A I_k^B \end{array} \\ \begin{bmatrix} p_{k,0} & M_{k,0}^{AB} & 0 & \cdots & 0 & 0 & 0 \\ M_{k-1,1}^{BA} & p_{k-1,1} & M_{k-1,1}^{AB} & & 0 & 0 & 0 \\ 0 & M_{k-2,2}^{BA} & p_{k-2,2} & & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & & p_{2,k-2} & M_{2,k-2}^{AB} & 0 \\ 0 & 0 & 0 & & M_{1,k-1}^{BA} & p_{1,k-1} & M_{1,k-1}^{AB} \\ 0 & 0 & 0 & \cdots & 0 & M_{0,k}^{BA} & p_{0,k} \end{bmatrix} \end{array}. \quad (4.15)$$

3. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^A \leq k < N^B}}$

Suppose now that $N^A \leq k < N^B$. In this range, the first row of the movement matrices represent transitions from the absorbing state $I_{N^A}^A I_{k-N^A}^B$, so the transition matrix is similar to that in (4.15), except that the first row is adjusted to account for this absorbing state:

$$\mathbf{M}_k = \begin{array}{c} \\ I_{N^A}^A I_{k-N^A}^B \\ I_{N^A-1}^A I_{k-N^A+1}^B \\ I_{N^A-2}^A I_{k-N^A+2}^B \\ \vdots \\ I_2^A I_{k-2}^B \\ I_1^A I_{k-1}^B \\ I_0^A I_k^B \end{array} \begin{array}{c} \begin{array}{ccccccc} I_{N^A}^A I_{k-N^A}^B & I_{N^A-1}^A I_{k-N^A+1}^B & I_{N^A-2}^A I_{k-N^A+2}^B & \cdots & I_2^A I_{k-2}^B & I_1^A I_{k-1}^B & I_0^A I_k^B \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ M_{N^A-1,k-N^A+1}^{BA} & p_{N^A-1,k-N^A+1} & M_{N^A-1,k-N^A+1}^{AB} & & 0 & 0 & 0 \\ 0 & M_{N^A-2,k-N^A+2}^{BA} & p_{N^A-2,k-N^A+2} & & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & & p_{2,k-2} & M_{2,k-2}^{AB} & 0 \\ 0 & 0 & 0 & & M_{1,k-1}^{BA} & p_{1,k-1} & M_{1,k-1}^{AB} \\ 0 & 0 & 0 & \cdots & 0 & M_{0,k}^{BA} & p_{0,k} \end{bmatrix} \end{array}. \quad (4.16)$$

4. $\underline{\mathcal{I}_{\mathbf{k}} \textbf{ for } \mathbf{N^B \leq k < N^A + N^B - 1}}$ Similar to how (4.16) was adjusted to account for the absorbing state in the first row, the final row of the movement matrices now represents a transition from another absorbing state $I^A_{k-N^B} I^B_{N^B}$, so movement matrices change as shown below.

$$\mathbf{M}_k = \begin{array}{c} \\ I^A_{N^A} I^B_{k-N^A} \\ I^A_{N^A-1} I^B_{k-N^A+1} \\ I^A_{N^A-2} I^B_{k-N^A+2} \\ \vdots \\ I^A_{k-N^B+2} I^B_{N^B-2} \\ I^A_{k-N^B+1} I^B_{N^B-1} \\ I^A_{k-N^B} I^B_{N^B} \end{array} \begin{array}{c} \begin{array}{ccccccc} I^A_{N^A} I^B_{k-N^A} & I^A_{N^A-1} I^B_{k-N^A+1} & I^A_{N^A-2} I^B_{k-N^A+2} & \cdots & I^A_{k-N^B+2} I^B_{N^B-2} & I^A_{k-N^B+1} I^B_{N^B-1} & I^A_{k-N^B} I^B_{N^B} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ M^{BA}_{N^A-1,k-N^A+1} & p_{N^A-1,k-N^A+1} & M^{AB}_{N^A-1,k-N^A+1} & & 0 & 0 & 0 \\ 0 & M^{BA}_{N^A-2,k-N^A+2} & p_{N^A-2,k-N^A+2} & & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & & p_{2,k-2} & M^{AB}_{2,k-2} & 0 \\ 0 & 0 & 0 & & M^{BA}_{k-N^B+1,N^B-1} & p_{k-N^B+1,N^B-1} & M^{AB}_{k-N^B+1,N^B-1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \end{array}. \quad (4.17)$$

5. $\underline{\mathcal{I}_{\mathbf{N^A+N^B-1}}}$

   Finally, when prevalence reaches $N^A + N^B - 1$, there are no movements possible since each of the remaining states $I^A_{N^A} I^B_{N^B-1}$ and $I^A_{N^A-1} I^B_{N^B}$ are absorbing. Thus, we are left with the $2 \times 2$ identity matrix

$$\mathbf{M}_{N^A+N^B-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

## 4.5 Mathematical analysis

An analysis similar to the one carried out in Chapter 3 could be performed, but this is outside of the scope of this thesis. In order to keep the focus of this work on the stochastic phase of disease outbreaks, we instead move straight to the numerical simulations.

| | $I_0^A I_0^B$ | $I_1^A I_0^B$ | $I_0^A I_1^B$ | $I_2^A I_0^B$ | $I_1^A I_1^B$ | $I_0^A I_2^B$ | $I_3^A I_0^B$ | $I_2^A I_1^B$ | $I_1^A I_2^B$ | $I_0^A I_3^B$ | $I_3^A I_1^B$ | $I_2^A I_2^B$ | $I_1^A I_3^B$ | $I_0^A I_4^B$ | $I_3^A I_2^B$ | $I_2^A I_3^B$ | $I_1^A I_4^B$ | $I_0^A I_5^B$ | $I_3^A I_3^B$ | $I_2^A I_4^B$ | $I_1^A I_5^B$ | $I_3^A I_4^B$ | $I_2^A I_5^B$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $I_0^A I_0^B$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_1^A I_0^B$ | $\Gamma_1^A$ | $p_{1,0}$ | $M_{1,0}^{AB}$ | $\Pi_1^A$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_0^A I_1^B$ | $\Gamma_1^B$ | $M_{0,1}^{BA}$ | $p_{0,1}$ | 0 | 0 | $\Pi_1^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_2^A I_0^B$ | 0 | $\Gamma_2^A$ | 0 | $p_{2,0}$ | $M_{2,0}^{AB}$ | 0 | $\Pi_2^A$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_1^A I_1^B$ | 0 | $\Gamma_1^B$ | $\Gamma_1^A$ | $M_{1,1}^{BA}$ | $p_{1,1}$ | $M_{1,1}^{AB}$ | 0 | $\Pi_1^A$ | $\Pi_1^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_0^A I_2^B$ | 0 | 0 | $\Gamma_2^B$ | 0 | $M_{0,2}^{BA}$ | $p_{0,2}$ | 0 | 0 | 0 | $\Pi_2^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_3^A I_0^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_2^A I_1^B$ | 0 | 0 | 0 | $\Gamma_1^B$ | $\Gamma_2^A$ | 0 | $M_{2,1}^{BA}$ | $p_{2,1}$ | $M_{2,1}^{AB}$ | 0 | $\Pi_2^A$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_1^A I_2^B$ | 0 | 0 | 0 | 0 | $\Gamma_2^B$ | $\Gamma_1^A$ | 0 | $M_{1,2}^{BA}$ | $p_{1,2}$ | $M_{1,2}^{AB}$ | 0 | $\Pi_1^A$ | $\Pi_2^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_0^A I_3^B$ | 0 | 0 | 0 | 0 | 0 | $\Gamma_3^B$ | 0 | 0 | $M_{0,3}^{BA}$ | $p_{0,3}$ | 0 | 0 | 0 | $\Pi_3^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_3^A I_1^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_2^A I_2^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_2^B$ | $\Gamma_2^A$ | 0 | $M_{2,2}^{BA}$ | $p_{2,2}$ | $M_{2,2}^{AB}$ | 0 | $\Pi_2^A$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_1^A I_3^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_3^B$ | $\Gamma_1^A$ | 0 | $M_{1,3}^{BA}$ | $p_{1,3}$ | $M_{1,3}^{AB}$ | 0 | $\Pi_1^A$ | $\Pi_3^B$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_0^A I_4^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_4^B$ | 0 | 0 | $M_{0,4}^{BA}$ | $p_{0,4}$ | 0 | 0 | 0 | $\Pi_4^B$ | 0 | 0 | 0 | 0 | 0 |
| $I_3^A I_2^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $I_2^A I_3^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_3^B$ | $\Gamma_2^A$ | 0 | $M_{2,3}^{BA}$ | $p_{2,3}$ | $M_{2,3}^{AB}$ | 0 | $\Pi_2^A$ | $\Pi_3^B$ | 0 | 0 | 0 |
| $I_1^A I_4^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_4^B$ | $\Gamma_1^A$ | 0 | 0 | $M_{1,4}^{BA}$ | $p_{1,4}$ | $M_{1,4}^{AB}$ | 0 | $\Pi_1^A$ | $\Pi_4^B$ | 0 | 0 |
| $I_0^A I_5^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $I_3^A I_3^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $I_2^A I_4^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\Gamma_4^B$ | $\Gamma_2^A$ | 0 | $M_{2,4}^{BA}$ | $p_{2,4}$ | $M_{2,4}^{AB}$ | $\Pi_2^A$ | $\Pi_4^B$ |
| $I_1^A I_5^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| $I_3^A I_4^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| $I_2^A I_5^B$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

(4.18)

Transition matrix for $N^A = 3$ and $N^B = 5$.

## 4.6 Numerical simulations

### 4.6.1 Algorithm for two-location numerics

The following is an outline of the process that was used in creating the transition matrix and its decomposition into canonical form as in Chapter 1, page 23.

1. In a first pass, given $N^A$ and $N^B$, an ordered list is created containing all admissible states. The states are ordered as in the sample matrix (4.18).

2. A graph of states is then created with arcs describing transitions as strings.

3. Given parameter values, the strings are evaluated, giving numerical weights to the arcs. These weights correspond to entries in the transition matrix. For example, a weight of 0.01 from vertex $i$ to vertex $j$ represents a transition probability of 0.01 from state $i$ to state $j$ of the Markov chain.

4. A graph decomposition algorithm is then run to identify the transient states and strong components with one vertex (absorbing states). This allows us to easily compute the canonical form of the matrix.

### 4.6.2 Terminology

Before proceeding with the simulations, we need to introduce some new terminology and recall some of the terminology from earlier in this chapter. In the case that we need to consider different basic reproduction numbers in each location, we will refer to the basic reproduction number in locations A and B as $\mathcal{R}_0(A)$ and $\mathcal{R}_0(B)$, respectively.

Recall that we assume there is a threshold number of infected individuals above which an epidemic takes place and that these thresholds are denoted $N^A$ and $N^B$ in locations A and B, respectively. For the numerics in Chapter 3, we used Whittle's

formula (3.10) to deduce these threshold values. Where possible (i.e., when $\mathcal{R}_0(A)$ and $\mathcal{R}_0(B)$ are greater than 1), we will use this formula for each location. If we require $\mathcal{R}_0(A)$ or $\mathcal{R}_0(B)$ less than 1 for a simulation, we will then set $N^A$ and $N^B$ to 50.

### 4.6.3 Travel rates

In order to choose accurate travel rates, we used data from Bluedot [6], taking the example of Thompson and Winnipeg in Manitoba. Counting the number of travellers to Thompson (location A) from Winnipeg (location B) in 2013 (and vice versa) and dividing by 365, we found out the average number of travellers per day between Thompson and Winnipeg. Dividing these numbers by the population of Thompson gives

$$\begin{aligned} m^{AB} &= 0.006748909, \\ m^{BA} &= 0.007425633, \end{aligned} \tag{4.19}$$

which we will use as sample travel rates for the numerics. These will be the travel rates used throughout unless otherwise stated.

### 4.6.4 Time step for connected populations, $\Delta t$

As in Chapter 3, $\Delta t$ must be chosen small enough so that only one event happens per time step. However, because of the extra terms in the transition matrix for two locations, the upper bound for $\Delta t$ must must be recalculated. This time, the row sums of the off-diagonal entries in the transition matrix are given by

$$\Gamma_j^A + \Gamma_i^A + M_{i,j}^{BA} + M_{i,j}^{AB} + \Pi_i^A + \Pi_j^B.$$

As before, we require the sums of the off-diagonal entries in each row to be less than one, so using (4.2), (4.3) and (4.4), we get

$$\left(\gamma^A i^A + \gamma^B i^B + m^{BA} i^B + m^{AB} i^A + \beta^A (P^A - i^A) i_A + \beta^B (P^B - i^B) i^B\right) \Delta t \leq 1,$$

so

$$\Delta t \leq \frac{1}{\gamma^A i^A + \gamma^B i^B + m^{BA} i^B + m^{AB} i^A + \beta^A (P^A - i^A) i_A + \beta^B (P^B - i^B) i^B}.$$

### 4.6.5 Investigating the impact of travel

Consider Figure 4.4, which shows the probability of extinction of the disease with $\mathcal{R}_0(A) = \mathcal{R}_0(B)$ varying together in the case where travel is included (continuous line) and no travel (dashed line). We assume there is initially one infected person in location A. For this simulation, the threshold numbers were $N^A = N^B = 50$. As we would expect, the graph indicates that travel does not cause significant unexpected behaviour in the system as the basic reproduction number varies. However, from a very careful inspection, it would appear that as $\mathcal{R}_0$ approaches 1, the probability of disease extinction begins to reduce a little sooner in the case with no travel than when travel is included. Figure 4.5 confirms this suspicion, though the difference is minor.

It is often possible for the same disease to have different basic reproduction numbers in different locations. This can be caused by a number of factors such as the differing population densities (causing $\beta$ to be different [37]), slightly different death rates and life expectancy, or even factors such as different climates. It is therefore natural to investigate the probability of overall disease extinction with respect to different or varying reproductions numbers in each location when the model does and does not incorporate travel. One would expect that imposing different basic

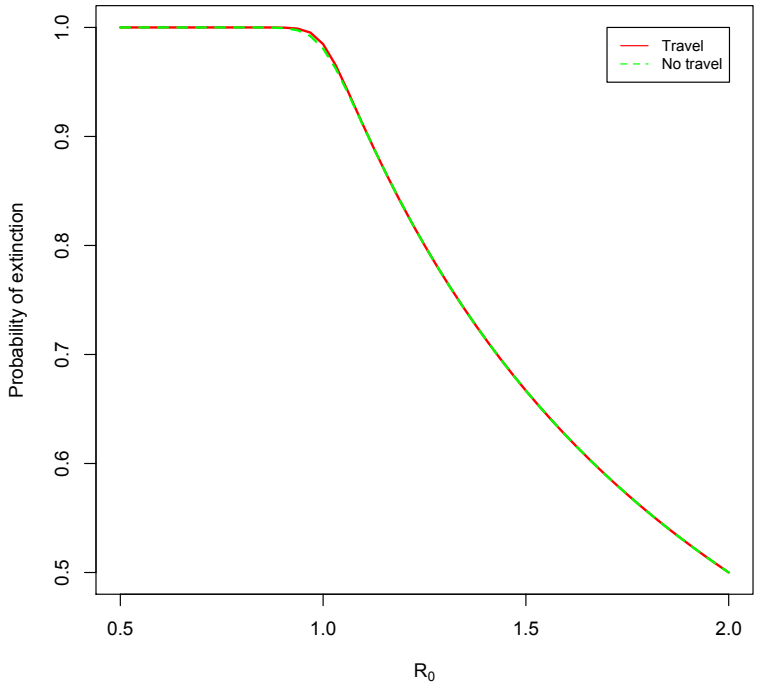



Figure 4.4: Probability of disease extinction as a function of $\mathcal{R}_0$ in the absence (dashed curve) and presence (continuous curve) of travel between locations.

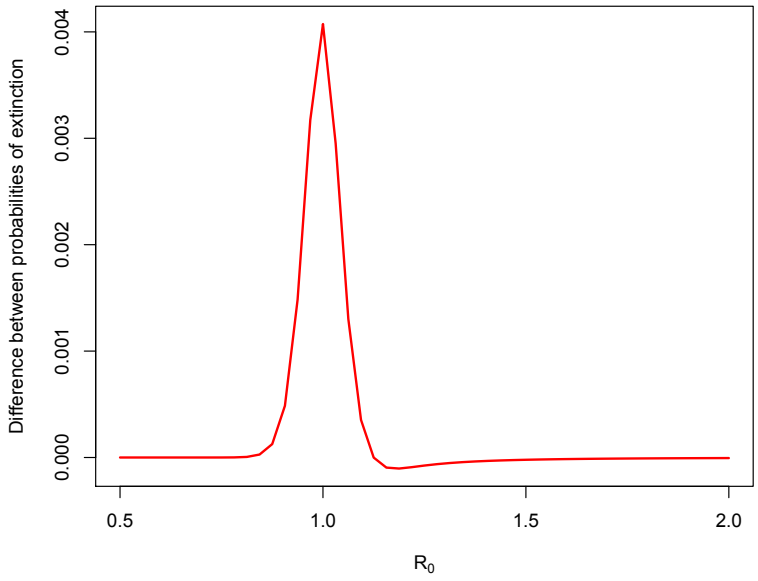



Figure 4.5: Difference in Figure 4.4 (travel minus no travel) between the probabilities of disease extinction.

reproduction numbers on each location should amplify the effects of travel, so for the upcoming simulation, we will allow the disease to have different basic reproduction numbers in each location.

We now examine the effect of travel more directly on the system by varying the travel rates and plotting the probability of disease extinction. For this simulation, we vary the travel rates from 0 to $m_{AB}$ and $m_{BA}$. Suppose again that $N^A = N^B = 50$, and the system starts in state $I^A_{\frac{1}{2}N^A} I^B_0$; i.e., the number of infected individuals in the system is 25 in location A and 0 in location B. Setting $\mathcal{R}_0(A) = 1$ and $\mathcal{R}_0(B) = 2$, we can see from Figure 4.6 that the role of travel in this model in certainly non-negligible. The figure clearly shows that as travel to and from the population with the higher reproduction number increases, the likelihood of disease extinction is reduced.

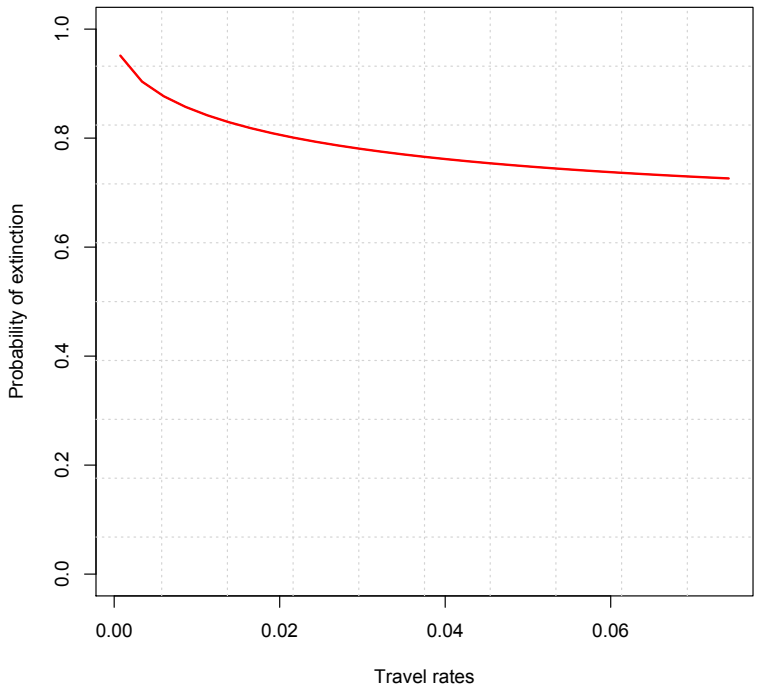



Figure 4.6: Effect of variation of travel on the probability of disease extinction when $\mathcal{R}_0(A) = 1$ and $\mathcal{R}_0(B) = 2$.

From Figure 4.6, we can see that travel plays a role in the outcome of an infectious

disease when the reproduction numbers are different in each location. Obviously the effect of travel is likely to be augmented when there is a large disparity in reproduction numbers, so we now investigate this. Suppose again that $\mathcal{R}_0(A) = 1$, travel is fixed at the usual rates in (4.19). We now vary the values of $\mathcal{R}_0(B)$ and observe the probability of disease extinction. Again $N^A = N^B = 50$ and the system starts in state $I_{25}^A I_0^B$. The results are show in Figure 4.7.

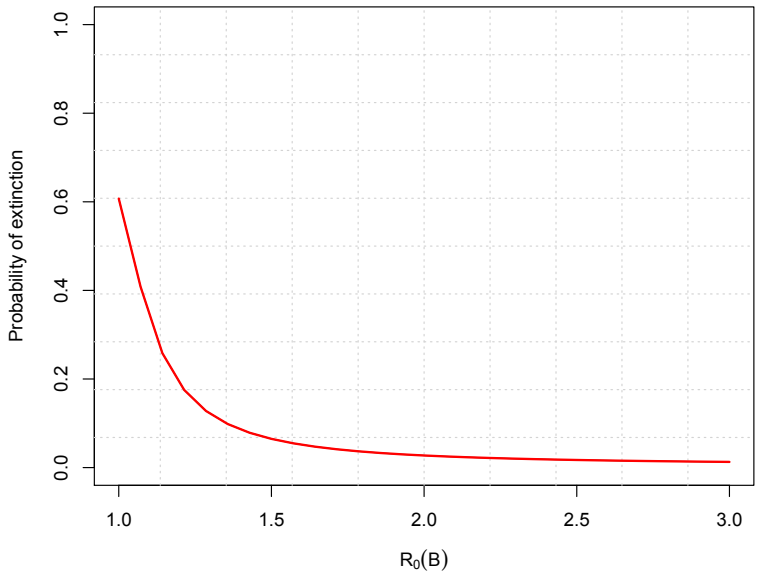



Figure 4.7: Change in probability of disease extinction due to variation of $\mathcal{R}_0(B)$ with $\mathcal{R}_0(A) = 1$ and travel rates fixed.

As we can see, the probability of disease extinction decreases sharply as $\mathcal{R}_0(B)$ increases. When $\mathcal{R}_0(B)$ is around 1, the probability of extinction is greater than 0.5, but as $\mathcal{R}_0(B)$ increases to around 1.1, the probability of extinction has already decreased to below 0.4.

### 4.6.6 Direct comparison to the single location case

Let us now compare the system with transport more directly to the system for isolated locations in Chapter 3. First, assume that there are zero individuals infected

in location B. As we vary $\mathcal{R}_0$ and the initial number of infected in location A, Figure 4.8 shows similar behaviour to Figure 3.5 in Chapter 3. With travel taken into account, we see roughly the same probabilities of disease extinction for $0.5 \leq \mathcal{R}_0 \leq 2$ and the number of individuals initially infected with the disease in location A ranging from 1 to 49 (i.e., $N^A = 50$).

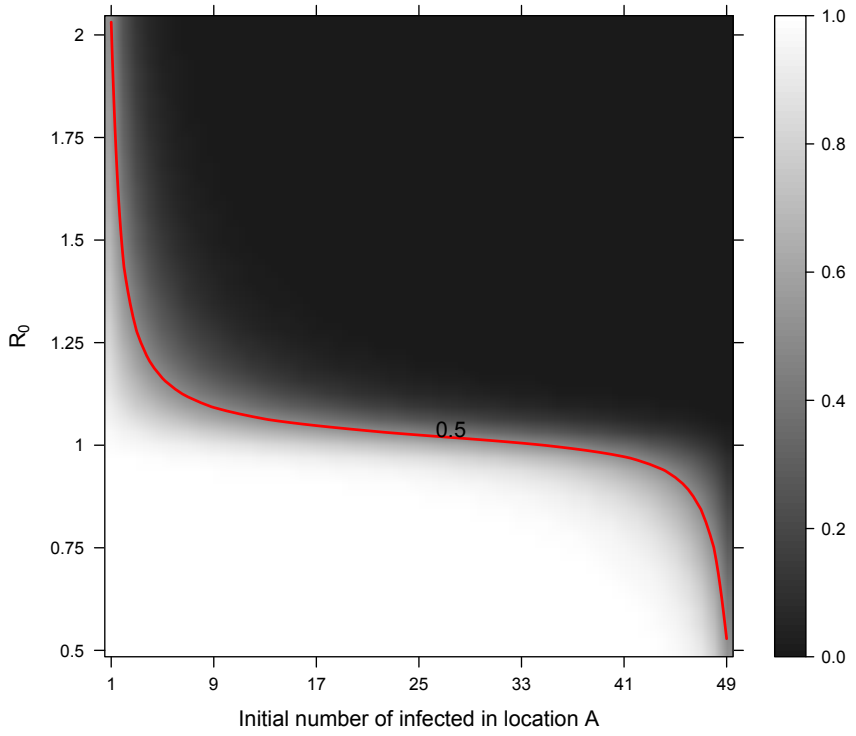
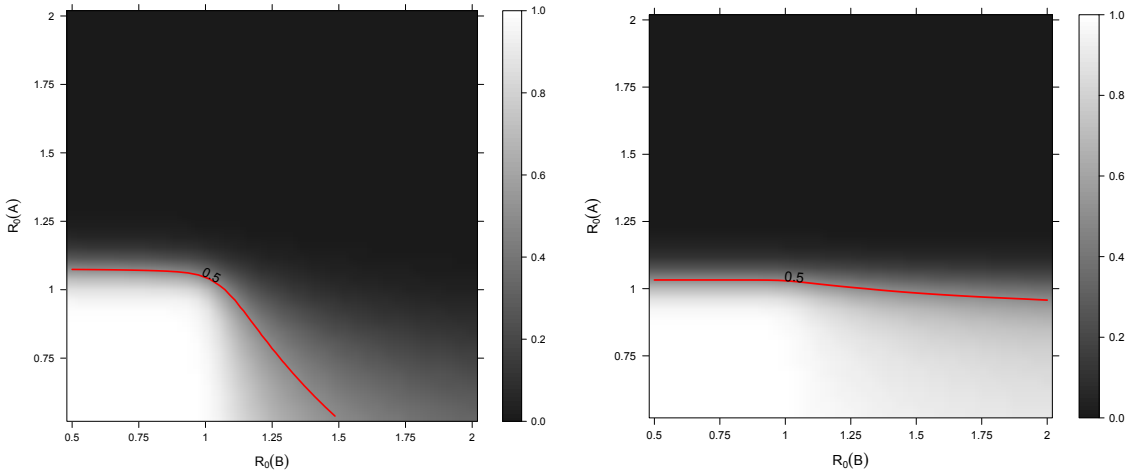



Figure 4.8: Probability of disease extinction with varying values of $\mathcal{R}_0$ and initial number of infected individuals in location A. The number of initially infected individuals in location B is assumed to be zero.

Suppose again that $N^A = N^B = 50$, and the system starts off in the state $I^A_{\frac{1}{2}N^A} I^B_0$. Figure 4.9 shows how the probability of overall disease extinction changes with the basic reproduction number varying in each location. As we can see, transport plays an important role when the reproduction numbers are allowed to vary. From Figure 4.9a, we can see that a relatively low $\mathcal{R}_0(A)$ does not ensure that the disease will die out even when the all infected are initially based in location A. If $\mathcal{R}_0(B) > 1$, it becomes more likely that the disease will not die out, but rather evolve into an

epidemic. On the other hand, if travel is reduced tenfold, provided the reproduction number is lower than 1 in location A (the location with the outbreak), then the threat of a major epidemic is significantly reduced.




(a) Normal travel rates. (b) Travel rates reduced tenfold.

Figure 4.9: Effect of travel on the probability of extinction when $\mathcal{R}_0$ is allowed to vary in each location.

Figure 4.9 shows that travel can play a major role in the development of an infectious disease when the basic reproduction numbers are allowed to vary from location to location. In particular, when there is an outbreak in location A, travel plays a significant role when the reproduction number in location A is approximately one.

Let us explore the role of a travel a little further in this aspect. Consider Figure 4.10. For this plot, we assume that $\mathcal{R}_0(A) = 1$, $\mathcal{R}_0(B)$ varies and travel rates vary from ten times lower than the standard travel rates to ten times higher. In this plot we can see the role of travel very clearly. When travel is reduced, we do not see much change as $\mathcal{R}_0(B)$ varies from 0.5 to 2, but as travel is increased, the change is much more pronounced. From the graph, we can see that any combination of $\mathcal{R}_0(B)$ and travel rates above the rightmost line indicate a probability of less than 0.2 of disease extinction, meaning that for this range of values an epidemic is the more

likely outcome. In the same way, any combination of $\mathcal{R}_0(B)$ and travel rates above the leftmost line imply a probability of disease extinction is greater than 0.8, so in this range of $\mathcal{R}_0(B)$ and travel rates, extinction becomes the more likely outcome. The region between the curves represents the values of travel rates and $\mathcal{R}_0(B)$ for which the outcome of the outbreak is uncertain; each pair of values in this region gives a probability of disease extinction between 0.2 and 0.8.

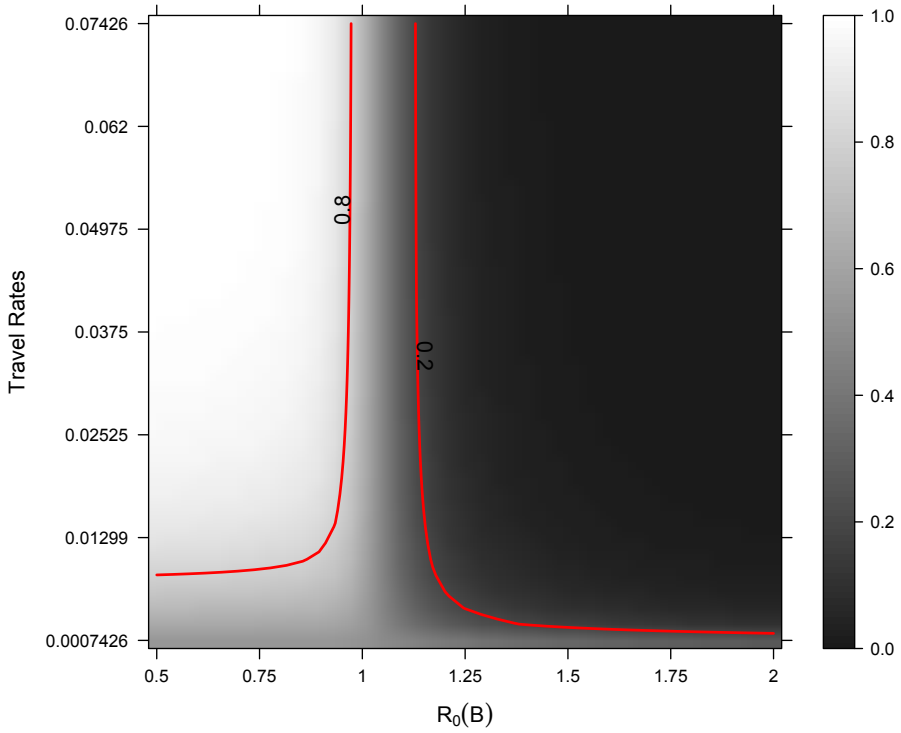



Figure 4.10: Probability of overall disease extinction given that $\mathcal{R}_0(A)$ is 1, with varying $\mathcal{R}_0(B)$ and travel rates.

### 4.6.7 Final size of the epidemic

As in Section 3.3.4, we investigate the final size of the epidemic. For the same reasons as before, we condition on disease extinction.

First, in order to compare directly with the findings in Section 3.3.4, consider the case where $\mathcal{R}_0 = 1.2$ in each location. Using (3.10), we find the threshold value for a

full-blown epidemic to be 26 for each location. Starting with one infected individual in location A and zero in location B (i.e., state $I_1^A I_0^B$), after 1000 realizations of the Markov chain, we find the average final size of the epidemic to be 4.7. Recall that this number was 4.6 when the corresponding simulation was run in Chapter 3.

In keeping with the investigations in Chapter 3, suppose now that $\mathcal{R}_0$ is varied from 0.5 to 2. As before, we cannot use Whittle's formula (3.10) in choosing $N^A$ and $N^B$ when $\mathcal{R}_0$ varies below 1, so let $N^A = N^B = 50$. Figure 4.11 shows how the final size of the epidemic changes with $\mathcal{R}_0$ for the travel rates as in (4.19) when the initial number of infected individuals in location A is 1, and in location B is zero. The red (continuous) curve represents the result for the two location case, while the blue (dashed) is the corresponding result from Figure 3.6 in Chapter 3 overlaid. Figure 4.12 is the same simulation for 5 initially infected individuals in location A and zero in location B, again with red (continuous) curve representing the two location case with the blue (dashed) curve from Figure 3.6 in Chapter 3.

As we can see, travel does not play a major role in the average final size of the epidemic. In both simulations for one and five initially infected individuals in location A, the numbers for the two location case closely follow those of the one location case for every $\mathcal{R}_0$ in the given range.

### 4.6.8 Further investigations into the final size of the epidemic for the two locations case

Finally, we look at the average duration of the stochastic phase for two locations. Figure 4.13 shows the expected time until the chain is absorbed for $0.3 \leq \mathcal{R}_0(A) = \mathcal{R}_0(B) \leq 3$ and the initial number of infected ranging from 1 to 49. As before, $N^A = N^B = 50$. The red (continuous) lines show the different lengths of time in two locations while dashed lines are the results from one location superimposed onto the graph. As we can see, the addition of travel causes the average duration to increase

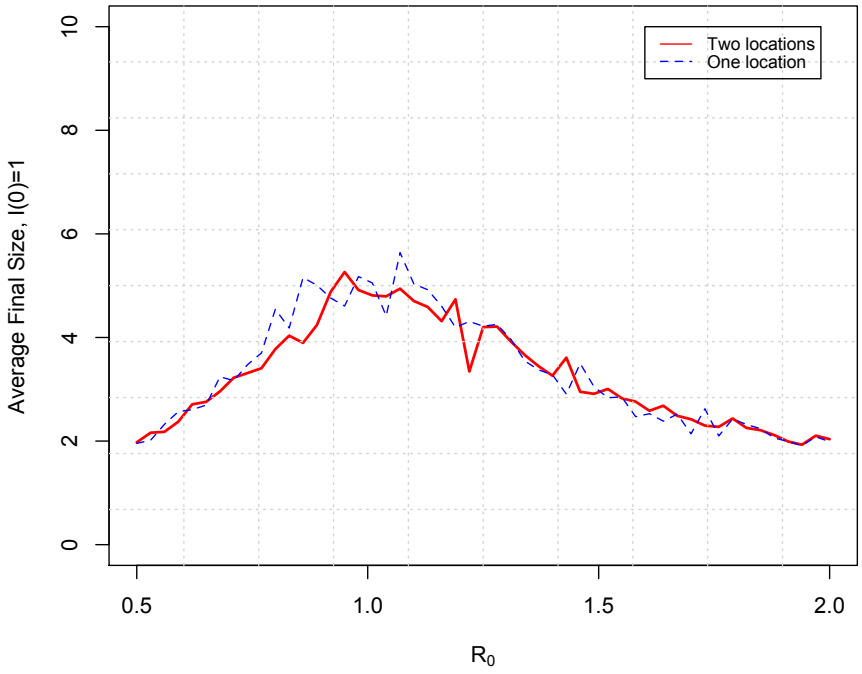



Figure 4.11: Final size of the epidemic with $\mathcal{R}_0$ varying from 0.5 to 2 starting with one infected individual in location A with the corresponding result for the one location case overlaid.

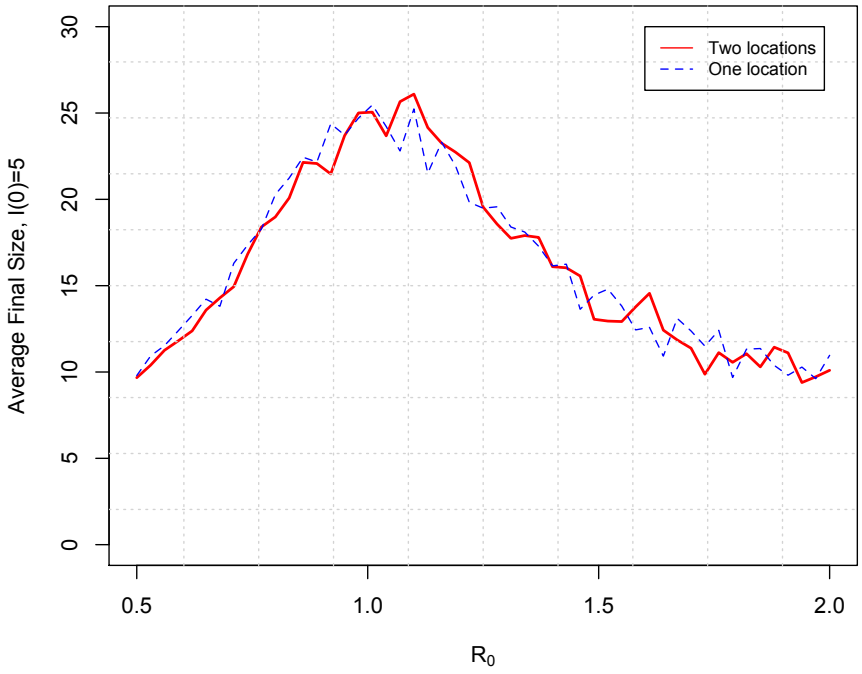



Figure 4.12: Final size of the epidemic with $\mathcal{R}_0$ varying from 0.5 to 2 starting with five infected individuals in location A with the corresponding result for the one location case overlaid.

very slightly.

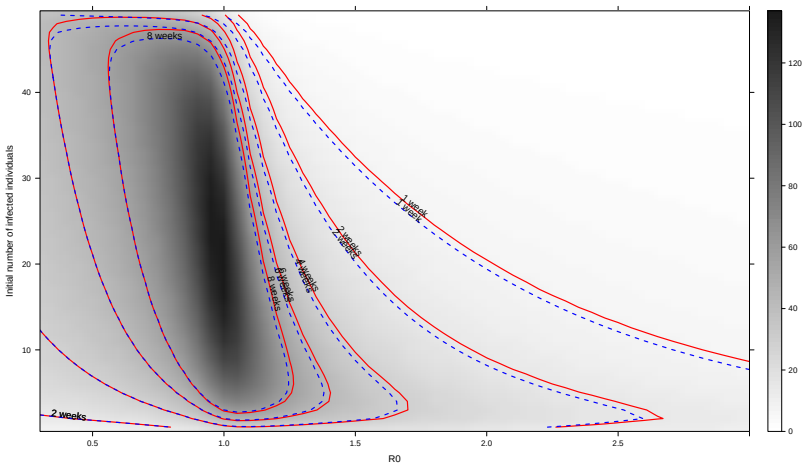



Figure 4.13: Average duration of the stochastic phase for two locations (continuous curve) compared with one location (dashed curve)

# 5

# Conclusion

The primary goal of this work was to investigate the effect of travel on the characteristics of the stochastic phase of disease spread in populations. After building a Markov chain model for both one isolated location and two distinct but connected locations, many numerical simulations were run.

First, the probability of disease extinction was measured for varying values of $\mathcal{R}_0$ and numbers of infected individuals in the population. The addition of travel to and from another location did not appear to have a major effect on the simulations in this regard – in each case, probability of extinction was heavily dependent on $\mathcal{R}_0$ and the number of individuals initially infected in the population.

Some of the simulations did reveal more dependence on travel rates. For example, Figure 4.9 investigates the situation of an outbreak in progress in location A (where there are 25 currently infected) and location B disease free. The figure shows the change in the probability of disease extinction due to variations of $\mathcal{R}_0(A)$ and $\mathcal{R}_0(B)$. Figure 4.9a shows the behaviour for normal travel rates, while Figure 4.9b shows the situation for travel rates reduced significantly (tenfold). The results imply that if the disease has different reproduction numbers in each location, then travel between the locations may result in the overall likelihood of disease extinction reducing quite

a lot, while if travel rates are reduced, the probability of disease extinction remains in line with $\mathcal{R}_0(A)$. This implies that if a disease is prevalent in one location but not another, and it may have different basic reproduction numbers in each location, it is best to limit travel in order to lower the probability of a major outbreak.

The versatility of a Markov chain model makes it an attractive method for modelling disease movement. With a Markov chain model we can calculate probabilities of extinction, absorption into various absorbing states and durations of time spent in individual or all transient states. Although this model revealed a lot of information about the dynamics of the disease, there are still quite a few ways it could be furthered and improved upon.

First, an obvious extension to the model would be to add a third connected location, and then to generalize it to $n$ connected locations. The difficulty here is that the size of the Markov chain would soon become enormous, as the number of possible states for the Markov chain to access increases (almost) exponentially. For example, in one location, an absorbing threshold of 50 infected individuals means 51 states, while in two locations, by Lemma 4.1, absorbing thresholds of 50 in each location leads to 2600 states (101 of which are absorbing). For 3 and $n$ locations, the transition matrix representing the Markov chain has dimension $51^3 \times 51^3$ and $51^n \times 51^n$, respectively, making numerics increasingly difficult.

Another extension would be to apply a Markov chain model to a more complex disease type. In this work, the model was applied to a disease that follows an SIS progression. Because we had only two compartments and a constant population, we could model the disease solely based on the number of infected individuals at a given time. With more compartments this would become more difficult but still possible – perhaps using more complex functions describing the entries in the Markov chain to account for the changes in overall structure of the population. It may not be possible, but is nonetheless worthwhile investigating.

# Bibliography


[1] H. Abbey. An examination of the Reed-Frost theory of epidemics. *Human biology*, 24(3):201, 1952.

[2] L.J.S. Allen, F. Brauer, P. van den Driessche, and J. Wu, editors. *Mathematical Epidemiology*. Springer, 2008.

[3] L.J.S. Allen and A.M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163:1–33, 2000.

[4] J. Arino. Diseases in metapopulations. *Modeling and dynamics of infectious diseases*, 11:65–123, 2009.

[5] J. Arino, F. Brauer, P. van den Driessche, J. Watmough, and J. Wu. A model for influenza with vaccination and antiviral treatment. *Journal of Theoretical Biology*, 253(1):118–130, 2008.

[6] J. Arino and K. Khan. *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, chapter Using mathematical modelling to integrate disease surveillance and global air transportation data. Wiley, 2014. To appear.

[7] V.O. Baroyan and L.A. Rvachev. Deterministic epidemic models for a territory with a transport network. *Kibernetica*, 3:67–73, 1967.

[8] V.O. Baroyan, L.A. Rvachev, U.V. Basilevsky, V.V. Ezmakov, K.D. Frank, M.A. Rvachev, and V.A. Shaskov. Computer modeling of influenza epidemics for the whole country (USSR). *Adv. App. Prob.*, 3:224–226, 1971.

[9] M.S. Bartlett. Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume IV, pages 81–109. University of California Press, 1956.

[10] J.A. Bondy and U.S.R. Murty. *Graph Theory With Applications*. Macmillan London, 1976.

[11] R.A. Brualdi and H.J. Ryser. *Combinatorial Matrix Theory*. Cambridge University Press, 1991.

[12] Stats Canada. Table for enplaning and deplaning passengers in canada, 1997 - 2014, 2014. http://www.statcan.gc.ca/pub/51-203-x/2014000/longdesc-ct001-eng.htm, [Online; accessed 29-March-2016].

[13] V. Colizza and A. Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical Review Letters*, 99(14):148701, 2007.

[14] V. Colizza and A. Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251(3):450–467, 2008.

[15] J. Dargay, D. Gately, and M. Sommer. Vehicle ownership and income growth, worldwide: 1960-2030. *The Energy Journal*, pages 143–170, 2007.

[16] S.N. DeWitte and J.W. Wood. Selectivity of black death mortality with respect to preexisting health. *Proceedings of the National Academy of Sciences*, 105(5):1436–1441, 2008.

[17] J.L. Doob. *Stochastic Processes*, volume 101. New York Wiley, 1953.

[18] M.J. Faddy. A note on the behavior of deterministic spatial epidemics. *Mathematical Biosciences*, 80:19–22, 1986.

[19] M. Greenwood. On the statistical measure of infectiousness. *Journal of Hygiene*, 31(03):336–351, 1931.

[20] H.W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[21] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[22] T. House, J.V. Ross, and D. Sirl. How big is an outbreak likely to be? Methods for epidemic final-size calculation. In *Proc. R. Soc. A*, volume 469, page 20120436. The Royal Society, 2013.

[23] J.A. Jacquez and C.P. Simon. Qualitative theory of compartmental systems. *Siam Review*, 35(1):43–79, 1993.

[24] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*, volume 356. van Nostrand Princeton, NJ, 1960.

[25] J.G. Kemeny, J.L. Snell, and G.L. Thompson. *Introduction to Finite Mathematics*. Prentice-Hall Englewood Cliffs, NJ, 1957.

[26] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. London, Ser. A*, 115:700–721, 1927.

[27] K. Khan, R. Eckhardt, J.S. Brownstein, R. Naqvi, W. Hu, D. Kossowsky, D. Scales, J. Arino, M. MacDonald, J. Wang, J. Sears, and M.S. Cetron. Entry and exit screening of airline travellers during the A(H1N1) 2009 pandemic: a retrospective evaluation. *Bulletin World Health Organisation*, 91(5):368–376, 2013.

[28] S. Kirkland. On the Kemeny constant and stationary distribution vector for a Markov chain. *Electronic Journal of Linear Algebra*, 27(1):23, 2014.

[29] S.Y. Liu, A. Baronchelli, and N. Perra. Contagion dynamics in time-varying metapopulation networks. *Physical Review E*, 87(3):032805, 2013.

[30] I.M. Longini. A chain binomial model of endemicity. *Mathematical Biosciences*, 50(1):85–93, 1980.

[31] S. Merler, M. Ajelli, A. Pugliese, and N.M. Ferguson. Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in europe: implications for real-time modelling. *PLoS Comput Biol*, 7(9):e1002205, 2011.

[32] Public Health Agency of Canada. Influenza statistics, 2014. http://www.phac-aspc.gc.ca/bid-bmi/dsd-dsm/rvdi-divr/index-eng.php, [Online; accessed 29-March-2016].

[33] World Health Organisation. Approximation for cost of smallpox from 1988 to 2035, 2014. http://www.who.int/mediacentre/news/notes/2010/smallpox_20100517/en/, [Online; accessed 29-March-2016].

[34] S. Rasmussen, M.E. Allentoft, K. Nielsen, L. Orlando, M. Sikora, K-G. Sjögren, A.G. Pedersen, M. Schubert, A. Van Dam, C.M.O. Kapel, H.B. Nelson, S. Brunak, A. Pavel, A. Epimakhov, M.V. Khalyapin, A. Gnuni, A. Kriiska, I. Lasak, M. Metspalu, V. Moiseyev, A. Gromov, D. Pokutta, L. Saag, L. Varul, L. Yepiskoposyan, T. Sicheritz-Ponten, R.A. Foley, M.M. Lahr, R. Nielsen, Kristiansen K., and E. Willerslev. Early divergent strains of Yersinia Pestis in Eurasia 5,000 years ago. *Cell*, 163(3):571–582, 2015.

[35] A. Rényi. *Probability Theory*. North-Holland series in applied mathematics and mechanics. Elsevier, 1970.

[36] D.M. Skowronski, C. Astell, R.C. Brunham, D.E. Low, M. Petric, R.L. Roper, P. J. Talbot, T. Tam, and L. Babiuk. Severe Acute Respiratory Syndrome (SARS): a year in review. *Annu. Rev. Med.*, 56:357–381, 2005.

[37] P.M. Tarwater and C.F. Martin. Effects of population density on the spread of disease. *Complexity*, 6(6):29–36, 2001.

[38] R.J.D. Tebbens, M.A. Pallansch, S.L. Cochi, S.G.F. Wassilak, J. Linkins, R. W. Sutter, R.B. Aylward, and K.M. Thompson. Economic analysis of the global polio eradication initiative. *Vaccine*, 29(2):334–343, 2010.

[39] Thucydides. *The History of the Peloponnesian War*. The Internet Classics Archive, Circa 430 B.C. http://classics.mit.edu/Thucydides/pelopwar.html.

[40] P. van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.*, 180:29–48, 2002.

[41] E. Vynnycky and R. White. *An Introduction to Infectious Disease Modelling*. Oxford University Press, 2010.

[42] P. Whittle. The outcome of a stochastic epidemic. *Biometrika*, 42(1-2):116–122, 1955.

[43] C.Y. Wu, T.Z. Huang, L. Li, and X.G. Lv. Inverses of block tridiagonal matrices and rounding errors. *Bulletin of the Malaysian Mathematical Sciences Society*, 34(2), 2011.