6-2019

An Epidemiological Model with Simultaneous Recoveries

Ariel B. Farber

University of Maine, ariel.farber@maine.edu

AN EPIDEMIOLOGICAL MODEL WITH SIMULTANEOUS RECOVERIES

By

Ariel Briana Farber

B.A. Humboldt State University, 2014

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

(in Mathematics)

The Graduate School

The University of Maine

August 2019

Advisory Committee:

David Hiebeler, Ph.D., Professor of Mathematics, Advisor

Jaehong Jeong, Ph.D., Assistant Professor of Statistics

Peter Stechlinski, Ph.D., Assistant Professor of Mathematics

AN EPIDEMIOLOGICAL MODEL WITH SIMULTANEOUS RECOVERIES

By Ariel Briana Farber

Thesis Advisor: David Hiebeler, Ph.D.

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Arts
(in Mathematics)
August 2019

Epidemiological models are an essential tool in understanding how infection spreads throughout a population. Exploring the effects of varying parameters provides insight into the driving forces of an outbreak. In this thesis, an SIS (susceptible-infectious-susceptible) model is built partnering simulation methods, differential equations, and transition matrices with the intent to describe how simultaneous recoveries influence the spread of a disease in a well-mixed population. Individuals in the model transition between only two states; an individual is either susceptible — able to be infected, or infectious — able to infect others.

Events in this model (infections and recoveries) occur by way of a Poisson process. In a well-mixed population, individuals in either state interact at a constant rate where interactions have the potential to infect a susceptible individual (infection event). Recovery events, during which infectives transition from infectious to susceptible, occur at a constant rate for each infected individual. SIS models mimic the behavior of diseases that do not confer immunity to those previously infected. Examples of such diseases are the common cold, head lice, and many STIs [2]. This model describes the effects the scale of recovery events have on an outbreak. Thus, for each recovery event, k number of infectives recover. The rate at which recoveries occur is inversely proportionate to k in order to maintain the average per-capita rate of recovery.

A system of ordinary differential equations (ODEs) is derived and supported by simulated data to describe the first and second moments (used to describe mean and variance) of the probability density function defining the number of infectious individuals in the population. Additionally, a Markov chain describes the process via transition matrices, which provide insight on extinctions caused by large-scale recoveries and their effect on the mean.

The research shows as the values of k increase, there is a statistically significant decline in the average infection level and an increase in the standard deviation. The most extreme changes in the average infection level are observed under conditions that increase the probability of extinction. Even in small populations where the decreased infection level is not biologically significant, the results are beneficial. Because large-scale recovery events have no negative impact on average infection levels, treatment methods that may reduce costs and increase accessibility could be adopted.

Healthcare professionals utilize epidemiological models to understand the severity of an outbreak and the effectiveness of treatment methods. A key feature of mathematically modeling real-world processes is the level of abstraction it offers, thus making the models applicable to many fields of study. For instance, those interested in agricultural development use these models to treat crops efficiently and optimize yield, cybersecurity experts use them to investigate computer viruses and worms, and ecologists implement them when studying seed dispersal.

DEDICATION

To my mom, who always listens.

To my sisters, who never fail to believe.

To my dad, who would be proud of us all.

I love you.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. David Hiebeler for his patience and guidance as my advisor and mentor through this program and research. Regardless of his bustling schedule, he was always available to help troubleshoot code and find the right words when I was at a loss. I would also like to thank my colleague and friend, Camden Bock, for hours both saved and spent at our "accountability meetings" to keep each other chugging along. A special thanks to Betsy Graves Rose for her kindness and chats that meant more than she knows, and her loving pup, Tony, who could always brighten my day. Finally, a heartfelt thank you to my best friend and partner, Leif Korth, who has been by my side, unwavering, every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LISTINGS

CHAPTER 1
INTRODUCTION

Epidemiological models are an essential tool in understanding how infections, either in the literal sense or symbolic e.g., ideas, seeds, etc., spread throughout a population. Therefore, studying these models intrigues not only those with purely mathematical interests but also professionals in the healthcare industry, environmental sciences, and cybersecurity to name a few. Analyzing the effects of different parameters provides insight to healthcare professionals as to how, when, and where to deploy treatment methods thus optimizing the treatment's effect. In addition, data generated by modeling potential outbreaks helps determine the need for preventative action and serves as a resource in lobbying for needed funding, technology, and policies to stay ahead of threats.

In this thesis, a SIS (susceptible-infectious-susceptible) model is built by partnering simulation methods, differential equations, and transition matrices. The intention of this research is to investigate and have the model describe how simultaneous recoveries affect the endemic behavior of a disease spreading in a well-mixed population. Individuals in SIS models can be in only one of two states; either susceptible or infectious. Susceptible individuals are vulnerable to infection, while infectious individuals are those capable of infecting others. In general, SIS models illustrate the spread of diseases that confer no temporary resistance or immunity. Examples of such diseases are the common cold, head lice, and many STIs [2]. In a well-mixed population such as this, individuals interact with one another at a constant rate, with said interactions having the potential of fostering new infectives. We chose to examine what effect simultaneous recoveries has on the overall expected proportion of infected individuals in a well-mixed population.

A paper written by David Hiebeler examined the effect of block-recovery modeled on a lattice structure. The results showed that recovering large blocks of the population, less often, drove the average contamination down [5]. We examine whether or not simultaneous

recoveries will yield a similar result in a well-mixed network model. The model is an adaptation of the classic SIS model described in work by Kermack and McKendrick [8]. The adjustment includes a new parameter that represents how many simultaneous recoveries will happen during a recovery event.

Ordinary differential equations (ODEs) are derived to describe the behavior of the model over time, the solutions to which are the moments of the model's probability distribution. The differential equations are derived in such a way as to implicitly capture stochastic variation seen in realizations of the process. This method results in a system of ODEs in which the equation describing any given moment includes the moment of the next-higher order. That is, the n^{th} moment depends on the $(n+1)^{th}$ moment, resulting in an open system of equations. Therefore, moment-closure approximations are used to truncate the model and close the system. Support for the approximations is obtained in data generated by simulations of the process.

Phenomena observed in realizations of the simulation that were not described by the differential equations gave the motivation to model the process as a Markov chain. Therefore, transition matrices are developed to provide explanation and further insight into the dynamical effects of simultaneous recoveries on the epidemiological process. Aspects of this model, which are novel compared to previous studies, include the use of simultaneous recoveries. Although tools such as moment-closure approximations and Markov chain models are often utilized in applied mathematics, their application in this study is new. Furthermore, the said application of these tools in the model studied here is explored primarily independently rather than following the path of previous study.

CHAPTER 2
BACKGROUND THEORY

2.1 Poisson Processes

The events in the model (additional infections and the recoveries of individuals) occur by way of a Poisson process, meaning the time between any two consecutive events are exponentially distributed. In general, when mathematically modeling real-world occurrences an exponential distribution for certain random variables, e.g., the time between events, is often assumed. Both the ease of working with an exponential distribution and the veracity of the model to the actual distribution of time between events motivate the assumption [11]. The aforementioned ease is in part due to the memoryless property of the exponential distribution, i.e., for any exponential random variable X

$$P\{X > t + \Delta t | X > t\} = P\{X > \Delta t\} \text{ for all } t, \Delta t \geq 0. \tag{2.1}$$

This carries through into the Poisson process resulting in another useful property. Given disjoint time intervals, the numbers of events that occur within each are independent and the process is said to have independent increments.

To clarify, consider for a moment the number of emails a person is likely to receive throughout the day, and partition the day by the hour. Unless that person has used up their storage, it makes sense to believe that the probability of receiving a new email during a given hour is not affected by the number of emails that have accumulated prior. Now, reduce the intervals to those with the length of a half-hour. Again, it is believable that the probability of receiving an email within the time increment is unaffected by how many have come before it. Continue to reduce the intervals from minutes to seconds to mere milliseconds and still further, pushing the process to resemble continuous time. In any case, the probability of receiving a new email depends not on how many have already flooded the inbox but rather the interval of time for which the opportunity is available and the rate

of the arrival of emails. Thus, the occurrence of new events is independent of events that have happened previously.

In addition to independent increments and the memoryless properties, Poisson processes are often implemented when modeling discrete events in continuous time because they possess other accessible properties that preserve the accuracy of the model to the actual behavior of events. The properties in Table 2.1 lend themselves to both writing mathematical expressions for the probabilities of events occurring during short intervals of time, as well as simulating dynamical systems.

Given a Poisson process with rate λ per unit of time, let $N(t) = n$ for $n = 0, 1, 2, ...$ be the number of events that occur in the time interval $[0, t]$.

(1) The process is memoryless (defined in text by Eqn. 2.1)

(2) The process has independent increments.

(3) The inter-arrival time of events are exponentially distributed with mean $\frac{1}{\lambda}$.

(4) The probabilities for the number of events occurring in an interval of time are given by:

  (a) $P\{N(t + \Delta t) - N(t) = 1\} = \lambda \Delta t + o(\Delta t)$

  (b) $P\{N(t + \Delta t) - N(t) \geq 2\} = o(\Delta t)$

(5) Suppose $N_1(t), N_2(t)$ are Poisson processes with rates λ_1 and λ_2.
Then $N(t) = N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$

(6) If the Poisson process $N(t)$ has probability, $P\{\text{an event is counted}\} = p$, then let $\tilde{N}(t)$ be another Poisson process for the kept events with a rate of λp.

Table 2.1. Summary of key properties of Poisson processes.
[11]

Explanation for items 1-3 of Table 2.1 are already given. In reference to the probabilities defined by item 4 of the list, notice that these probabilities depend only on the rate of the process and the length of the time interval, just as was discussed in the email example. To further understand the probabilities it is necessary to recall what it means for

a function $f(\Delta t)$ to be $o(\Delta t)$, the latter referring to the terminology "little-oh" of Δt.

Definition: A function $f(\Delta t) = o(\Delta t)$ as $\Delta t \to 0$ if $\lim_{\Delta t \to\ 0} \dfrac{f(\Delta t)}{\Delta t} = 0$. [13]

Thus, the $o(\Delta t)$ terms in a function refer to quantities that are negligible when taking the limit as $\Delta t \to 0$. For this reason, the $o(\Delta t)$ terms are omitted when deriving the differential equations for the model in Section 3.3. The remaining properties in Table 2.1, items 5 and 6, summarize the combining of different types of Poisson processes and the filtering of a certain process that is part of a combination, respectively.

CHAPTER 3
THE MODEL

3.1 Model Specification

In this thesis, an SIS epidemiological model for a well-mixed population with a fixed population size N is developed. Any individual in the population is either susceptible or infectious at a given time t. In the model, infectious individuals contact other individuals (referred to as targets) at a constant rate with the potential to infect said targets. Therefore, as time passes, allowing the infection to spread, an individual's state may change. As the name suggests, a well-mixed population implies that every target has the same probability of being contacted by an infected individual.

Given a total of N individuals, let $I(t)$ represent the number of individuals that are infectious, i.e., hosts, at time t. Likewise, let $S(t) = N - I(t)$ represent the number of individuals susceptible at time t. Each host randomly contacts targets at the rate of ϕ. If a target is contacted and susceptible to infection, then that target becomes infected. In the event that a target is already a host, it is of no consequence and the potential to infect a new individual lost.

Infectives in the population each recover at rate μ. Once an individual has recovered from the infection, they immediately return to the susceptible state with no resistance, meaning the likelihood that a recovered individual will be re-infected is equal to that of any other susceptible individual becoming infected for the first time. This aspect of SIS models mimics the behavior of diseases that do not confer immunity to those who have been previously infected such as the common cold, head lice, and many STIs. The intention of this model is to describe the effect that the scale of recoveries has on the behavior of an outbreak. To accomplish this, an adjustment is made to the model that allows for k individuals to recover simultaneously during a recovery event, with the rate at

which these events occur inversely proportional to k, thus maintaining μ as the average per-capita rate of recovery for the entire population.

3.2 Discrete Simulation Methods

Simulations allow for numerous realizations of a process in a relatively short amount of time. Analysis of the data from these experiments provide valuable information on the behavior of an outbreak that would otherwise take years to collect if these phenomena were only observed in nature. Thus, simulations are a useful tool in modeling the spread of disease because they are a formal re-creation of a process that can be implemented as many times as desired. To begin, a simulation of the discrete process is developed and the observed values (data) are used to calculate the sample mean and variance for the number of infectives at each time-step. These quantities are then compared to those produced by the simulations of the ODEs derived in Section 3.3. To simplify, the discrete process simulation will be referred to as the simulation(s), whereas the solutions and values given by simulations of the ODEs will exclude the term.

In this model each infected individual contacts others at the rate of ϕ. Given that a contact only results in a new infection if the target is in the susceptible state, successful infections for the entire population occur at the rate

$$\Phi = \underbrace{\phi I(t)}_{\substack{\text{population's} \\ \text{contact rate}}} \underbrace{\left(1 - \frac{I(t)}{N}\right)}_{\substack{\text{vulnerable} \\ \text{proportion} \\ \text{of the population}}} \tag{3.1}$$

Likewise, given the per-capita rate of recovery for hosts is μ, the total rate of recovery events is

$$M = \frac{\mu I(t)}{k}. \tag{3.2}$$

Note that the population's rate of recovery is given by $\mu I(t)$ and each recovery event affects k individuals. Thus, these events occur $\frac{1}{k}$ less often to preserve the average per-capita rate of recovery.

Therefore the total rate, λ, at which events occur in the model is

$$\lambda = \text{infection rate} + \text{recovery rate} = \phi I(t)\left(1 - \frac{I(t)}{N}\right) + \frac{\mu I(t)}{k} = \Phi + M$$

Where property 5 in Table 2.1 applies to the sum (the result of the combination of multiple Poisson processes is itself a Poisson process with the rate equal to the sum of the rates of the combined processes.)

The first step in simulating the process is determining at what time the next event occurs and updating the time in the simulation. Because the inter-arrival time between events is exponential (property 3, Table 2.1), this is as simple as generating a random number, Δt, from an exponential distribution with a mean $\frac{1}{\lambda}$ and adding it to the current recorded time in the simulation. The next step is to determine which kind of event took place. This is similar to flipping a weighted coin of which one side is an infection event and the other is a recovery event. Let P_Φ be the probability that the event is a new infection and P_M the probability that there is a recovery event. These probabilities are given by

$$P_\Phi = \frac{\text{infection rate}}{\text{total rate}} = \frac{\Phi}{\lambda}$$

$$P_M = \frac{\text{recovery rate}}{\text{total rate}} = \frac{M}{\lambda}.$$

The "flipping" of the coin is done by sampling a random value from a uniform distribution with bounds $(0, 1)$. If the value is less than P_Φ, then the event is an infection and the event is processed by setting $I(t + \Delta t) = I(t) + 1$. Otherwise, it is a recovery event and processed by $I(t + \Delta t) = I(t) - min(k, I(t))$, where the minimum between $I(t)$ and k is taken from the infected population to ensure a negative population isn't produced. Then the process is repeated until the desired amount of time for the simulation has been reached.

3.3 Derivation of the ODEs

Along with simulations, differential equations are a common tool in developing continuous-time models such as the one being explored here. Differential equations describe the behavior of unknown functions in terms of the function itself and its derivatives. In this model, ordinary differential equations are derived to describe the change in moments of the unknown density function describing the infection level with respect to time. Modeling processes using differential equations provides a general, or abstract, description that can be applied to many different processes in nature. The benefit of abstraction allows the application of mathematical theorems and methods to analyze the quantitative and qualitative behavior of a system that can then be interpreted in terms of the study, e.g., the spread of an infection through a population.

The methods used to derive ODEs are similar to those in Hiebeler's paper studying the effect of other parameters in SIS models [7]. The following theorem will be useful in construction of the ODEs.

Theorem 3.3.1

Given any two random variables X and Y

$$E[X] = E\left[E[X|Y]\right].$$

Thus, if Y is a discrete random variable then

$$E[X] = \sum_{y} E\left[X|Y = y\right] \cdot P\{Y = y\}. \quad [4]$$

To begin deriving differential equations describing this system, let $X(t)$ be the random variable that represents the actual number of infected individuals at time t and let $I(t) := E[X(t)]$. Given that the number of infected individuals at time t is $X(t)$, then the

number of infected individuals at the time $(t + \Delta t)$ for some small Δt will be:

$$X(t+\Delta t) = \begin{cases} X(t)+1 & \text{with probability } \phi X(t)\Delta t\left(1-\frac{X(t)}{N}\right) \\ X(t)-k & \text{with probability } \frac{\mu X(t)\Delta t}{k} \\ X(t) & \text{with probability } 1-\frac{\mu X(t)\Delta t}{k}-\phi X(t)\Delta t\left(1-\frac{X(t)}{N}\right) \end{cases} \tag{3.3}$$

where the expressions $o(\Delta t)$ in the probabilities seen in Table 2.1, item 4, have been omitted, as they would vanish in a subsequent step.

Therefore, the expected value of the above quantity is

$$\begin{aligned} E[X(t+\Delta t)|X(t)] &= (X(t)+1)\left(\phi X(t)\Delta t\left(1-\frac{X(t)}{N}\right)\right) \\ &\quad +(X(t)-k)\left(\frac{\mu X(t)\Delta t}{k}\right) \\ &\quad +X(t)\left(1-\frac{\mu X(t)\Delta t}{k}-\phi X(t)\Delta t\left(1-\frac{X(t)}{N}\right)\right) \\ &= \phi X^2(t)\Delta t+\phi X(t)\Delta t-\phi\Delta t\frac{X^3(t)}{N}-\phi\Delta t\frac{X^2(t)}{N} \\ &\quad +\frac{\mu X^2(t)\Delta t}{k}-\mu\Delta t X(t) \\ &\quad +X(t)+\phi\Delta t\frac{X^3(t)}{N}-\phi X^2(t)\Delta t-\frac{\mu X^2(t)\Delta t}{k} \\ &= \phi X(t)\Delta t-\mu X(t)\Delta t+X(t)-\phi\Delta t\frac{X^2(t)}{N}. \end{aligned}$$

Now by theorem 3.3.1 the expected value of $E\left[X(t+\Delta t)\right]$ is

$$\begin{aligned} E[X(t+\Delta t)] &= E[E[X(t+\Delta t)|X(t)]] \\ &= E\left[\phi X(t)\Delta t-\mu X(t)\Delta t+X(t)-\phi\Delta t\frac{X^2(t)}{N}\right] \\ &= \phi\Delta t E[X(t)]-\mu\Delta t E[X(t)]+E[X(t)]-\phi\Delta t\frac{E\left[X^2(t)\right]}{N} \\ &= \phi\Delta t I(t)-\mu\Delta t I(t)+I(t)-\phi\Delta t\frac{E\left[X^2(t)\right]}{N}, \end{aligned}$$

and,

$$I(t+\Delta t) = E[X(t+\Delta t)] = \phi\Delta t I(t)-\mu\Delta t I(t)+I(t)-\phi\Delta t\frac{E\left[X^2(t)\right]}{N}.$$

Subtracting $I(t)$ from both sides of equation 3.4 and dividing by Δt yields

$$I(t+\Delta t) - I(t) = \phi \Delta t I(t) - \mu \Delta t I(t) - \phi \Delta t \frac{E\left[X^2(t)\right]}{N},$$
$$\frac{I(t+\Delta t) - I(t)}{\Delta t} = \phi I(t) - \mu I(t) - \phi \frac{E\left[X^2(t)\right]}{N}. \tag{3.4}$$

Finally, by the definition of the derivative and taking the limit as $\Delta t \to 0$ of equation 3.4 results in the differential equation

$$\frac{dI(t)}{dt} = (\phi - \mu) I(t) - \phi \frac{E\left[X^2(t)\right]}{N}.$$

Note that the $o(\Delta t)$ terms omitted in equation 3.3 would vanish in the above step as $\Delta t \to 0$, and notice the above equation contains the term, $E[X^2(t)]$. Furthermore, if zero variance were assumed i.e., $E[X^2(t)] = E[X^2(t)]$ and $N = 1$, the classic SIS model

$$I' = \phi SI - \mu I$$

would be recovered. However, to capture the stochastic variation seen in realizations of the process, zero variance is not assumed ($E[X^2(t)]$ is not necessarily equal to $E[X(t)]$) and the second moment is a free variable. Therefore, the equation is not closed and has infinite solutions. With the introduction of this new variable the differential equation for the second moment is derived in a similar fashion. Starting with the below values and taking the expected value of $E[X^2(t+\Delta t)]$ conditioned on $X(t)$,

$$X^2(t+\Delta t) = \begin{cases} (X(t)+1)^2 & \text{with probability } \phi X(t)\Delta t(1-\frac{X(t)}{N}) \\ (X(t)-k)^2 & \text{with probability } \frac{\mu X(t)\Delta t}{k} \\ X^2(t) & \text{with probability } 1-\frac{\mu X(t)\Delta t}{k}-\phi X(t)\Delta t(1-\frac{X(t)}{N}) \end{cases}$$

$$\begin{aligned} E[X^2(t+\Delta t)|X(t)] &= (X(t)+1)^2(\phi X(t)\Delta t(1-\tfrac{X(t)}{N})) \\ &\quad +(X(t)-k)^2(\tfrac{\mu X(t)\Delta t}{k}) \\ &\quad +X^2(t)(1-\tfrac{\mu X(t)\Delta t}{k}-\phi X(t)\Delta t(1-\tfrac{X(t)}{N})) \\ &= \phi\Delta t X^3(t)+2\phi\Delta t X^2(t)+\phi X(t)\Delta t \\ &\quad -\phi\Delta t\tfrac{X^4(t)}{N}-2\phi\Delta t\tfrac{X^3(t)}{N}-\phi\Delta t\tfrac{X^2(t)}{N} \\ &\quad +\tfrac{\mu\Delta t X^3(t)}{k}-2\mu\Delta t X^2(t)+k\mu X(t)\Delta t \\ &\quad +X^2(t)-\phi\Delta t X^3(t)+\phi\Delta t\tfrac{X^4(t)}{N}-\tfrac{\mu\Delta t X^3(t)}{k} \\ &= X^2(t)+\phi X(t)\Delta t+2\phi\Delta t X^2(t)+\tfrac{\phi X^2(t)\Delta t}{N}-\tfrac{2\phi X^3(t)\Delta t}{N} \\ &\quad +k\mu X(t)\Delta t-2\mu\Delta t X^2(t). \end{aligned} \tag{3.5}$$

Once more, with a bit of algebraic manipulation the expected value of equation 3.5 results in a difference quotient as shown,

$$\begin{aligned} E[X^2(t+\Delta t)] &= E[E[X^2(t+\Delta t)|X(t)]] \\ &= E[X^2(t)]+\phi\Delta t E[X(t)]+2\phi\Delta t E[X^2(t)]-\frac{\phi\Delta t E\left[X^2(t)\right]}{N} \\ &\quad -2\frac{\phi\Delta t E[X^3(t)]}{N}+k\mu\Delta t E[X(t)]-2\mu\Delta t E[X^2(t)] \end{aligned}$$

$$\begin{aligned} \frac{E[X^2(t+\Delta t)]-E[X^2(t)]}{\Delta t} &= \phi\left(I(t)+2E[X^2(t)]-\frac{E[X^2(t)]+2E[X^3(t)]}{N}\right) \\ &\quad +\mu(kI(t)-2E[X^2(t)]). \end{aligned} \tag{3.6}$$

Taking the limit of equation 3.6 as $\Delta t \to 0$ the ODE for the second moment is

$$\frac{dE\left[X^2(t)\right]}{dt} = \phi\left(I(t) + 2E[X^2(t)] - \frac{E[X^2(t)] + 2E[X^3(t)]}{N}\right) + \mu(kI(t) - 2E[X^2(t)]).$$

Again, the equation contains the following moment, $E[X^3(t)]$, resulting in an open system of equations. Additional ODEs could be derived for successive moments in the same manner and a general equation for the n^{th} moment could be explored, as was done in a paper by Hiebeler in 2006 [6]. In any case the system would eventually have to be closed by some means for evaluation. Thus the open system will be truncated to a system of the two differential equations

$$\begin{aligned} \frac{dI(t)}{dt} &= (\phi - \mu)I(t) - \phi\frac{E\left[X^2(t)\right]}{N} && (3.7) \\ \frac{dE\left[X^2(t)\right]}{dt} &= \phi\left(I(t) + 2E[X^2(t)] - \frac{E[X^2(t)] + 2E[X^3(t)]}{N}\right) \\ &\quad +\mu\left(kI(t) - 2E[X^2(t)]\right). && (3.8) \end{aligned}$$

and closed using moment-closure approximations.

3.3.1 Moment-Closure Approximations

Approximating $E[X^3(t)]$ in terms of $E[X^2(t)]$ and $I(t) = E[X(t)]$ allows the system of ODEs (equations 3.7 and 3.8) to be closed. According to a paper by Singh and Hespanha [12] in 2007, most moment closure techniques begin by assuming specific distributions and using the knowledge of such distributions to express higher-order moments in terms of those with lower order, the mean and variance for example. Then, such approximations are justified through empirical and analytical methods [12].

Four moment-closure approximations for $E[X^3(t)]$ are considered to truncate the system, and are listed below. The first was investigated by Hiebeler [7] and analyzed in Singh and Hespanha's [12] paper. The following three were also implemented in Hiebeler [7] and the final two were further explored and analyzed by Brown and Bolker [3] and Murrell [9].

$$\text{Normal:} \quad E[X^3(t)] \approx 3E[X^2(t)]I(t) - 2I(t)^3 \tag{3.9}$$

$$\text{Conditional:} \quad E[X^3(t)] \approx \frac{E[X^2(t)]^2}{I(t)} \tag{3.10}$$

$$\text{Power-2:} \quad E[X^3(t)] \approx \frac{1}{2}\left(\frac{3E[X^2(t)]^2}{I(t)} - I(t)\right)^3 \tag{3.11}$$

$$\text{Power-3:} \quad E[X^3(t)] \approx \left(\frac{E[X^2(t)]}{I(t)}\right)^3 \tag{3.12}$$

Though the number of infectives is a discrete random variable, the normal distribution is assumed to approximate these values. Though more rigorously analyzed by Singh and Hespanha [12], the normal approximation for $E[X^3(t)]$ can be heuristically derived using the moment generating function for the normal distribution. It is important to note that for this model, the normal moment-closure approximation exhibits serious deficiencies under certain circumstances that will be addressed in Section 5.2.1. The conditional approximation was originally developed by Dr. Hiebeler and assumes, given three individuals are sampled from the population, independence between two individuals being infectious when conditioned on a third. The remaining approximations have commonly been used in applied mathematics and are included with the normal and conditional approximations as tools for developing the model. The values obtained by these approximations are compared to simulation results to determine their validity as reasonable assumptions.

CHAPTER 4
RESULTS

4.1 Rates of Change

Respectively, Figures 4.1 and 4.2 show the rates of change $\frac{dI(t)}{dt}$ and $\frac{dE[X^2(t)]}{dt}$ along with the average rates of change measured from the simulation data over different values of k. The quantities for the rates defined by equations 3.7 and 3.8 do not come from numerical solutions. Instead they are obtained by substituting simulation measurements for $I(t)$ and $E[X^2(t)]$ into the equations and implementing the different moment-closure approximations introduced in section 3.3.1.

The images in Figures 4.1 and 4.2 confirm that the approximated values of the rates, $\frac{dI(t)}{dt}$ and $\frac{dE[X^2(t)]}{dt}$, are representative of the average rates produced in the simulations and maintain consistency as k varies. However, as k gets larger, the errors between the curves (those of the ODEs and the simulation) increase. Notice that only $\frac{dE[X^2(t)]}{dt}$ is dependent on k and $E[X^3(t)]$. While $\frac{dI(t)}{dt}$ is not influenced by k directly, the simulation for the first moment is. Thus, the values of $\frac{dI(t)}{dt}$ are indirectly affected by k and the ODE is still compared to simulation results over k. However, since neither the simulation of the mean nor $\frac{dI(t)}{dt}$ depend on $E[X^3(t)]$, the moment-closure approximations do not apply and it is not useful to compare their effect.

4.2 Equilibrium

Figures 4.3, 4.4, 4.5 all illustrate the average proportion of the population (infection level) that is infected, $\frac{I(t)}{N}$ $(N = 1,000)$, at time t, over varying values of k. Each individual figure demonstrates some nuances within the model. For instance, Figure 4.3 shows the results given an initial infection level of 0.5%, while the results shown in Figure 4.5 were produced with an initial infection level of 10%. Figure 4.4 has the same initial conditions

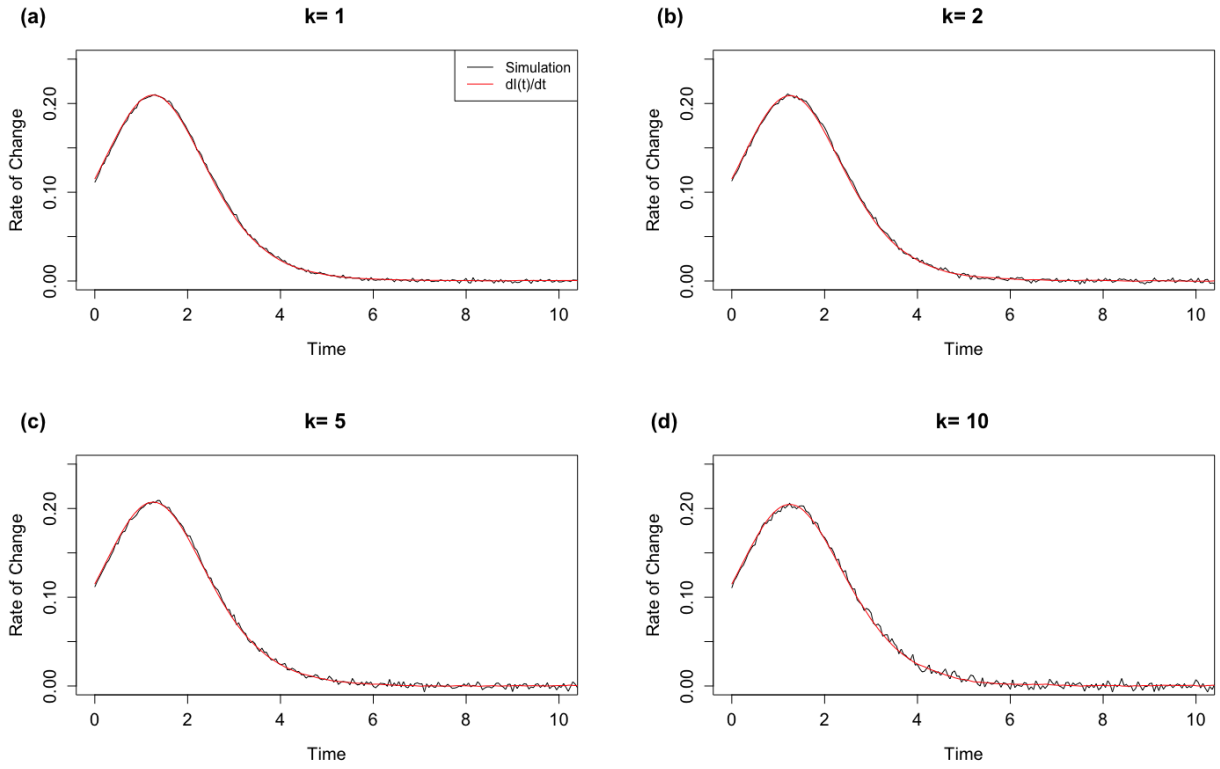

Figure 4.1. $\frac{dI(t)}{dt}$ vs Time.
The instantaneous rate of change, $\frac{dI(t)}{dt}$ against time, t for various values of k with parameters $\phi = 2$, $\mu = .7$, and $N = 1000$. The values for the simulations were obtained from 10,000 realizations of the process and approximated by the average rates of change over small time intervals i.e., $\frac{I(t+\Delta t)-I(t)}{\Delta t}$ with $\Delta t = .05$. The values for the rates were computed by substituting simulated values for $I(t)$ and $E[X^2(t)]$ into Eqn. 3.7.

as Figure 4.3, but the simulation does not allow for the infection to go extinct. This is achieved by setting $I(t) = 1$ whenever the simulation resulted in $I(t) = 0$ in an effort to see the effect of k assuming nonextinction with small initial level of infection.

Together in each image are the average infection level generated by simulation data along with error-bars denoting ± 1 standard deviations, and ODE solutions with the different moment closure approximations. Solutions for the ODEs were found numerically using the lsoda function from the deSolve package – available in the R software [10].

The simulations and solutions to the ODEs in Figures 4.3, 4.4, and 4.5 reveal a logistic growth process with a positive equilibrium at approximately 0.65, however, this is not in

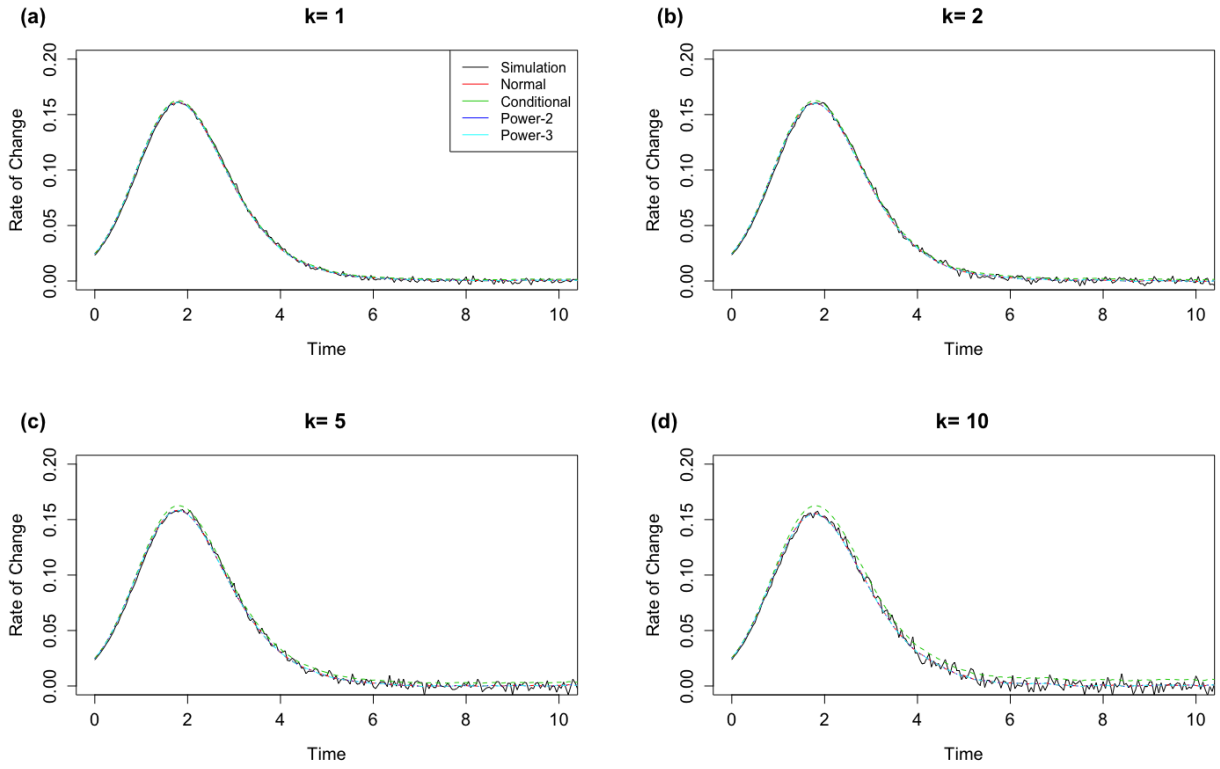

Figure 4.2. $\frac{dE[X^2(t)]}{dt}$ vs Time.
The instantaneous rate of change, $\frac{dE[X^2(t)]}{dt}$ against time, t, for various values of k and moment-closure approximations with parameters $\phi = 2$, $\mu = .7$, and $N = 1000$. The simulation values were approximated using the same method as in Figure 4.1. The values for the rates are found using the moment-closure approximations given by Eqs. 3.9-3.12 and making the same substitutions for $I(t)$ and $E[X^2(t)]$ used in Fig. 4.1.

fact a true equilibrium for the model. In this model, extinction is an absorbing state, meaning that once extinction is reached the infection level remains at zero as there are no new infections introduced to the system. That is, for any t such that $I(t) = 0$ then $I(\hat{t}) = 0$ for all $\hat{t} > t$. Furthermore, recall that the population, N, is fixed and finite in the model. It can be shown for such processes that extinction is certain. In other words, suppose $P_0(t)$ is the probability the infection goes extinct. Then for $N < \infty$ and extinction the only absorbing state, $\lim_{t\to\infty} P_0(t) = 1$. Therefore, the only true equilibrium for $I(t)$ in this model is zero. Prior to extinction, however, the infection level appears stationary and positive for long period (of time) as is shown in the results. Such behavior is referred to as

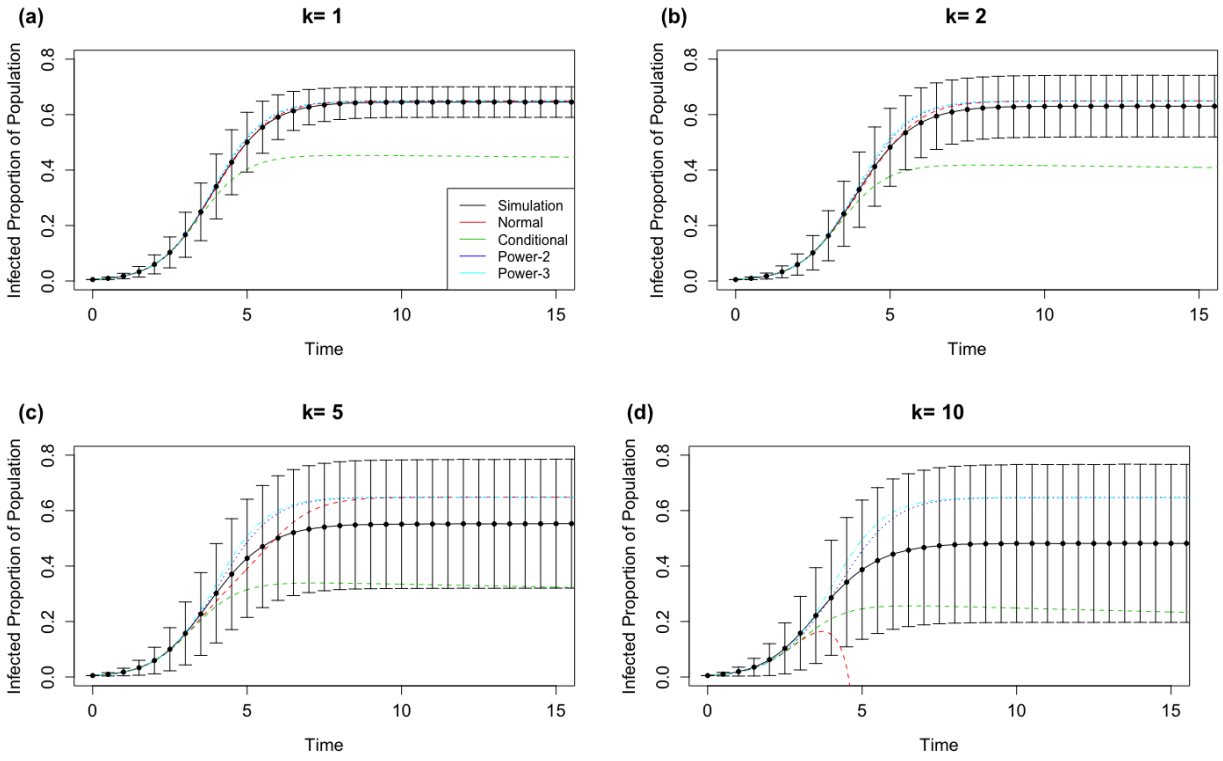

Figure 4.3. Average Infection Level with Small Initial Value.
Average level of infection against time for various values of k with parameters $\phi = 2$, $\mu = .7$, $N = 1000$, and $I(0) = 5$. The ODEs were numerically solved for $I(t)$ for each moment-closure approximation. The simulated means were generated by averaging values from 10,000 realizations of the process. The error-bars represent ± 1 standard deviation of the simulated means. Notice that the normal approximation behaves badly when $k > I(0)$.

quasistationary or a *quasiequilibrium* [1]. The quasiequilibrium depicts the expected infection level conditioning on nonextinction, and is compared to the ODE equilibria. For this reason the quasistationary state will often be referred to as the equilibrium henceforth, whereas extinction is attributed to the true equilibrium of $I(t) = 0$.

Figure 4.3 demonstrates that increasing the value of k with an initial infection level of 0.5% drastically decreases the simulated equilibrium for the expected infection level, $I(t)$. However, when extinction is not allowed as in Figure 4.4, this radical change does not occur, and the ODEs accurately represent the equilibrium seen in the simulations. Likewise, when allowing extinction with the higher initial infection level of 10% Figure 4.5 shows the ODEs fitting the simulation data.

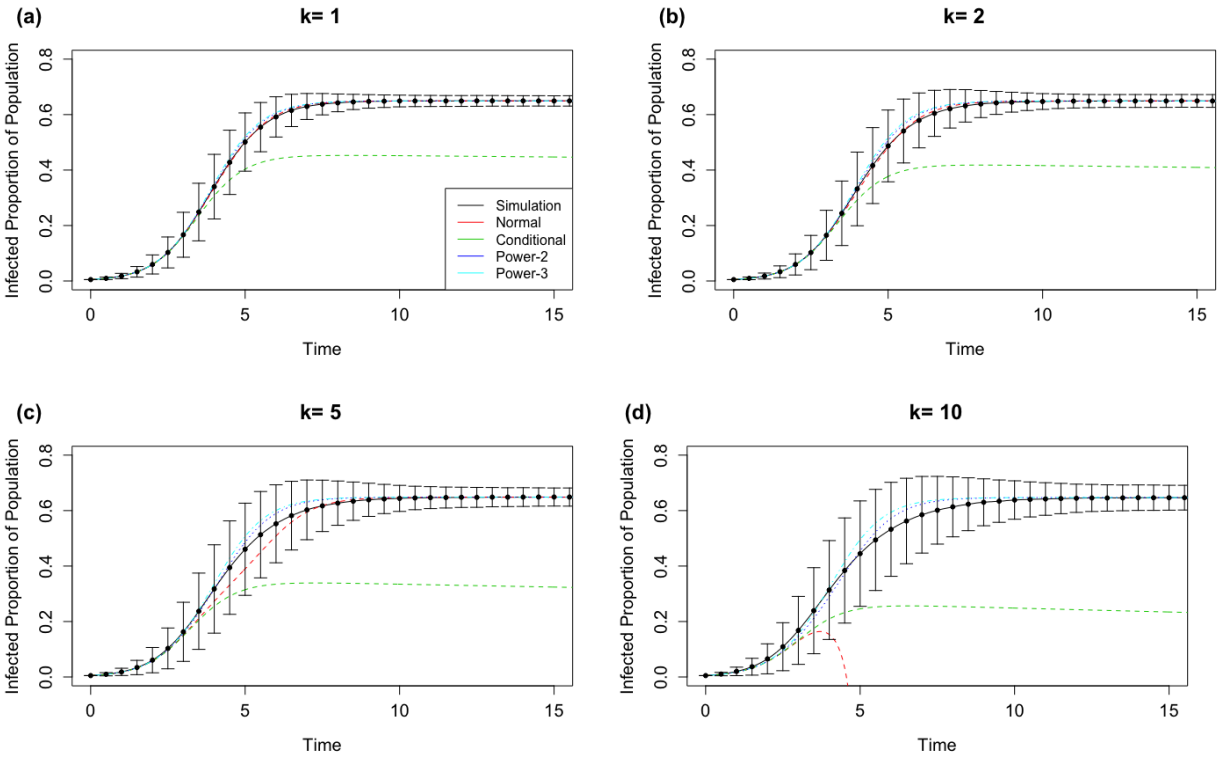

Figure 4.4. Average Infection Level with No Extinction.
Average level of infection, $I(t)$, barring extinction vs. time, t, for various values of k with parameters and methods for obtaining values the same as in Fig. 4.3

In each case, the conditional moment-closure approximation underestimates the equilibrium. The normal moment-closure approximation performs slightly better than the remaining two approximations. However, for small initial values the approximation quickly loses its accuracy and ultimately behaves very badly as k increases. Specifically, when $k > I(0)$ the approximation produces values of $I(t)$ beyond what would be allowed given the population size, i.e, $I(t) < 0$ or $I(t) > N$.

The remaining approximations (power-2 and power-3) maintain congruence with the simulation curve the best with all values of k. In Figure 4.4, most of the error between the three curves is seen during the transient state with a maximum absolute error for the infection level of 0.062. At first, both approximations underestimate the growth, then as the growth begins to slow down they surpass the simulation curve and underestimate time

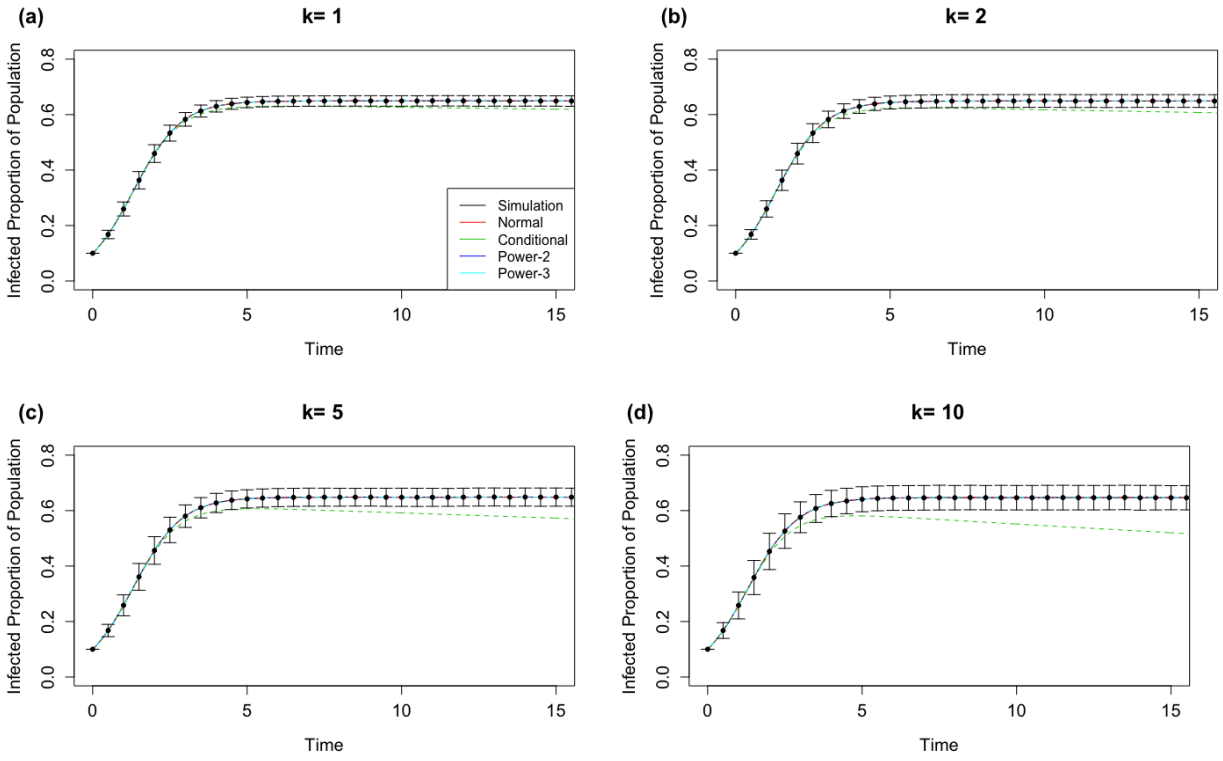

Figure 4.5. Average Infection Level with Large Initial Value.
Average level of infection vs. time for various values of k with parameters ϕ, μ, and N, same as in Fig. 4.3 and initial condition $I(0) = 100$. The methods used to obtain values again, are the same as in Figs. 4.3 and 4.4. Notice, here the normal approximation behaves well unlike in Fig. 4.3.

taken to reach the same equilibrium as the simulation. With the larger initial condition, $I(0) = 100$ or 10% of the population, the maximum absolute error decreases to 0.0018.

The same methods that were used to obtain values for $I(t)$ were also used to evaluate the second moment, $E[X^2(t)]$, and generate Figure 4.6 which represents the standard deviations for $I(t)$ with the corresponding k values as in Figure 4.5 (initial infection level of 10%). As k increases from zero to ten the standard deviation increases by more than a factor of two. Aside from the conditional closure approximation, which overestimates the standard deviation, the approximations tend to underestimate the variance as k increases during the transient states.

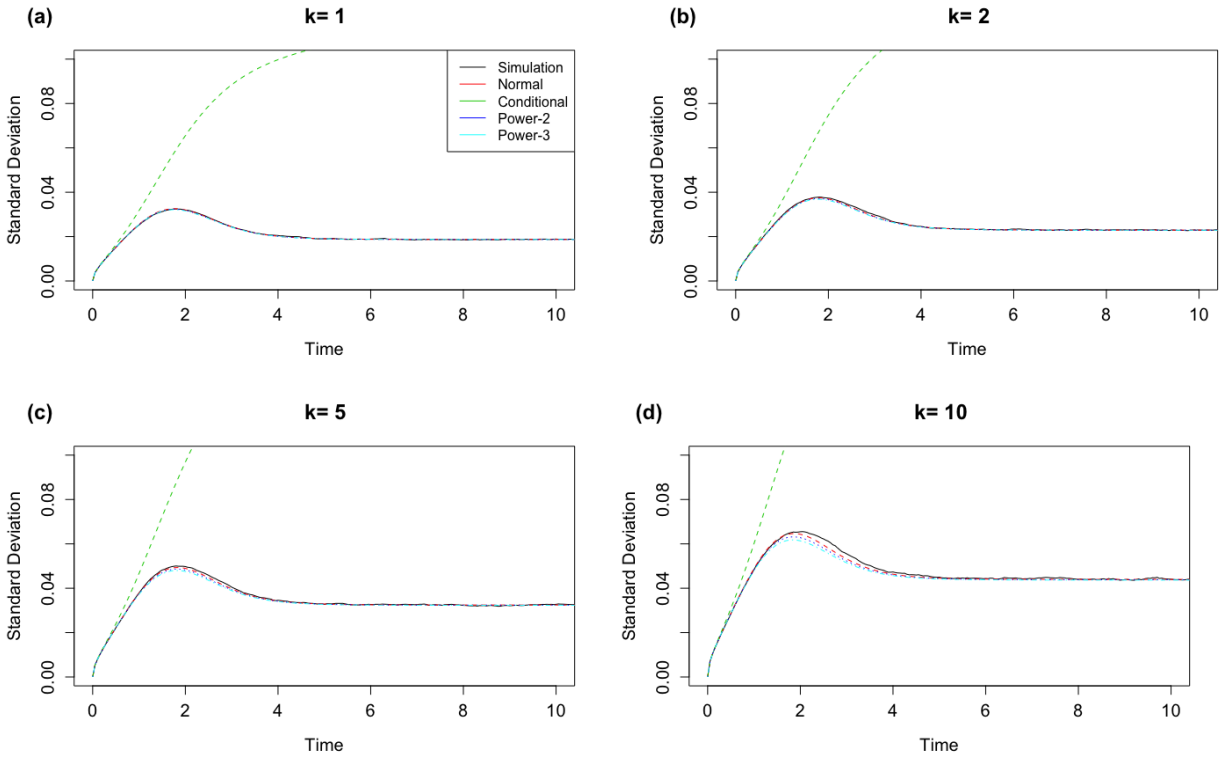

Figure 4.6. Standard Deviation of Infection Level.
The standard deviation of the infection level against time for various values of k with parameters $\phi = 2$, $\mu = .7$, $N = 1000$, and $I(0) = 100$ (same as in Fig. 4.5). Values were obtained in the same fashion for $E[X^2(t)]$ as they were for $I(t)$ and used to calculate the variances and standard deviations.

Figure 4.7 shows the difference in the equilibrium of the infection level when $k = 1$ as k increases from 1 to 500. The first two values of $k = 1, 5$ from which point k increases to 500 in increments of 5. The dots represent simulation means during the quasistationary state and were observed over 1,000 realizations of the process with parameters $\phi = 2$, $\mu = .7$, $N = 10,000$, and $I(0) = 1000$. A linear model was fit to the data and the regression line included in the Figure (the solid black line). The fitted values estimate a decrease of 0.02 in the infection level when $k = 500$ from the infection level when $k = 1$. A summary of the model confirms the decline in equilibrium as k increases, with a p-value of 2.2×10^{-16}. This means that, though the decrease in equilibrium is small, it is still statistically significant.

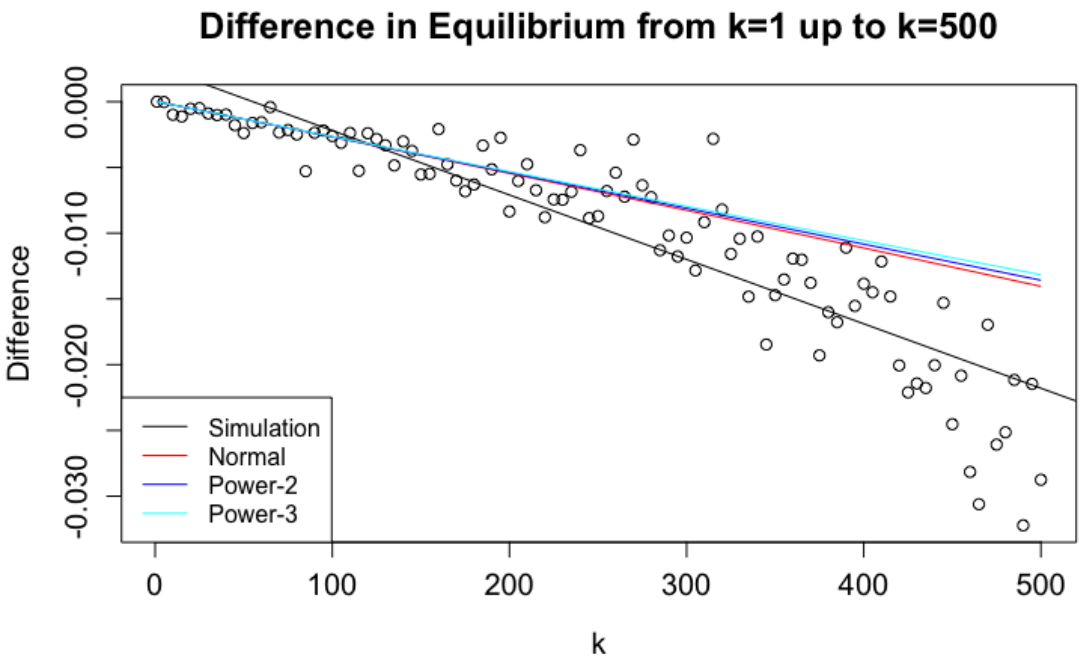

Figure 4.7. The Difference in Equilibrium for Scaled Recovery Events.
The difference in average infection level at equilibrium from 1,000 realizations with parameters $\phi = 2$, $\mu = .7$, $N = 10,000$, and $I(0) = 1000$ as k increases from $1 - 500$. Included is the linear regression of the data (represented by the black line) as well as the differences in the approximations at equilibria of the relevant ODE approximations Eqs. 3.9, 3.11, and 3.12.

The same values were calculated for the ODE solutions for all approximations, except the conditional-closure, and predict a slower but similar decline in infection levels.

CHAPTER 5
DISCUSSION

5.1 Rates of Change

Often, SIS models include ODEs that directly describe both the number of infectious and susceptible individuals in the system [8]. In this model the state variable, $S(t)$, which represents the latter, is written in terms of the number of infectives, $I(t)$, and the population, N. This eliminates a variable in the system of equations and is done to focus on the number of infectious individuals in the system at time t, and derive the ODEs that describe the first and second moments of $I(t)$'s probability distribution (Eqs. 3.7 and 3.8). The ODE for the third moment, $\frac{dE[X^3(t)]}{dt}$, was also derived using the same methods, though it was ultimately set aside. The first and second moments of the distribution provide information regarding the mean and variance of the distribution, and the decision to exclude the third ODE kept the toolbox of established moment-closure approximations at hand.

The ODEs were validated through the comparison of the simulation for large and small values for both N and k. The results shown in Figure 4.1 were chosen to demonstrate that they maintain their accuracy as k varies, as the effects of the scale of k are the focus of the study, and one can see from the figures that they hold. Quantitative analysis confirms that the error between the ODEs and simulated rates of change does increase along with k. The explanation, in part, can also be seen in the figure by noticing a variability or "jaggedness" of the simulation results. The dramatic spikes are a result of approximating the instantaneous rates of change using discrete methods.

5.2 Equilibrium: Fit of ODEs to Simulation Data

After confirming the ODEs fit the simulation data, the system was numerically solved with an initial infection level of 0.5% over many values of k. The results in Figure 4.3 show a dramatic decrease in the simulation's equilibrium, thus increasing the error between its value and the values of the ODEs. Small values for both $I(0)$ and k are attributed to the reduced equilibrium seen in Figure 4.3. These conditions lead to recovery events that occur while the infection level is small enough to be eradicated during the initial transient state. For example, when $I(0) \leq k$ it would take at most two recovery events in the first $2 \times I(0)$ total events to cause extinction. Furthermore, when k is also small, the probability of a recovery event is higher than that of large k. In conjunction, $I(0) \leq k$ and small values of k increase probability of the infection going extinct during the initial transitive state. This phenomena will be referred to as an *early extinction* henceforth.

When the simulations were run under conditions that nullified or decreased the probability of early extinction, as shown in Figures 4.4 and 4.5, the simulation is much less volatile and behaves per the ODEs description, thus reinforcing that early extinction is the cause for discrepancy. Of the two explorations, values produced using the larger initial condition resulted in less variability. Because of this, the larger initial infection level of 10% was chosen to further study the effect of k on the equilibrium conditioning on nonextinction. The results in Figure 4.7 show a statistically significant decrease in the equilibrium as k increases. Even in smaller populations where a 2% decline may not be biologically significant, the results suggest that the large-scale treatment less frequently is still an efficient practice as it could lower provider costs and increase accessibility with no negative impact on the expected infection level. For instance, employing fewer providers at temporary clinics in areas with little access would be less expensive than the infrastructure and overhead for numerous permanent operations scattered throughout a region.

The standard deviation was most affected by increases in k. Figure 4.6d shows when $k = 10$, 1% of the population size $N = 1,000$, the variance has already increased by greater

than a factor of two. As k increases, the impact of a recovery event on the infection level is amplified. Large recovery events cause disturbances about the equilibrium with greater magnitude than those caused by new infections and smaller recovery events. In addition, though the probability of a recovery event decreases as k grows, the potential for the event to cause extinction increases along with it. Together, both the disturbances and extinctions contribute to an increase in variability.

5.2.1 Moment-Closure Approximations

Though the conditional moment-closure approximation suited the simulation for the rate of change, it fails to accurately describe the behavior of the spread of infection for all k by underestimating the expected infection level. The normal approximation performs well in some cases but is volatile with small values of $I(0)$ that are close to k, thus is limited in application. It can be shown that when substituting the normal approximation for $E[X^3(t)]$, the set of values bounded by $I(t)$ and $E[X^2(t)]$ is not positive invariant, meaning there exists some $I(0) \in (0, N)$ such that $I(t) \notin (0, N)$ for $t < \infty$, which yields unbounded values for $I(t)$ and causes the bad behavior seen in Figure 4.3. The values of $I(0)$ for which this behavior occurs depends on k.

Barring such circumstances, the normal approximation performs quite well and is most accurate to the simulation throughout the transient state with a small initial infection level. Though the power-2 and power-3 approximations stray from the simulated values during the transient state in Figures 4.3 and 4.4, they are the most dependable approximations in the sense that they behave well over varying values of $I(0)$ and k. As $I(0)$ increases, these two approximations' accuracy throughout the transient state improves to the point at which they perform as well as the normal approximation. Therefore, which approximation would be the *best* method is dependent on the parameters k and $I(0)$, and the behavior of which state (transient or [quasi]stationary) is being described.

CHAPTER 6
MARKOV CHAIN AND TRANSITIONS MATRICES

6.1 Introduction

Thus far, the focus of this research has been concerned with the effects of k on the conditional equilibrium of the model, i.e., the quasiequilibrium assuming nonextinction. However, the model as it stands does little to predict the frequency or consequences of early extinction phenomena discussed at the beginning of section 5.2. Under such conditions that lead to early extinction, the ODEs fail to represent the infection level. To account for this, transition matrices are constructed to analyze the process as a Markov chain. A Markov chain is a stochastic process that takes on finite values referred to as *states* and transition from one state to another with a fixed probability that is only dependent on the current state [11]. A transition matrix is a probability matrix with values representing the probability of the process transitioning from one state to another in a given number of steps.

Let i be the number of infectious individuals in the system with $i = 0, 1, 2, ...N$, and $I_n = i$ mean the process is in state i after n events. The value p_{ij} is the probability that during the next event, the process will transition from state i into state j. Together, the probabilities for each combination of i and j make up the following one-step transition matrix, $\mathbf{P}$, for the process.

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0j} & \cdots & p_{0N} \\ p_{10} & p_{11} & \cdots & p_{1j} & \cdots & p_{1N} \\ \vdots & \vdots & \ddots & & & \vdots \\ p_{i0} & p_{i1} & \cdots & p_{ij} & & p_{iN} \\ \vdots & \vdots & & & \ddots & \vdots \\ p_{N0} & p_{N1} & \cdots & p_{Nj} & \cdots & p_{NN} \end{bmatrix} \tag{6.1}$$

Equation 6.1 is an example of a general transition matrix with N states. Note the values of i indicate the current state and row, whereas j values represent the column and next state.

6.2 Equilibrium and Early Extinction

In this model, the term "step" refers to the number of events that have occurred i.e., a one-step transition matrix represents the probabilities of transitions after the first event, two-step for the second event, and so on. Transition matrices of steps greater than one are represented by $\mathbf{P}^n$ for $n = 1, 2, ...$, and as mentioned, the values of which represent the n-step probability, $p_{ij}^{(n)}$. Similarly, define the state probability $P\{I_n = i\} =: P_i^{(n)}$. Throughout the remaining text assume a transition matrix be of one-step when referenced unless otherwise specified.

For $0 < i < N$, when an event occurs, the state of the system can either increase by one or decrease by k, and when $i = N$ only the latter can occur. Extinction (state zero) is an absorbing state, therefore $p_{00} = 1$. If everyone in the population is infected then $p_{N,N+1} = 0$ and $p_{N,N-k} = 1$. Therefore the probabilities for each one-step transition are as follows:

$$p_{ij} = \begin{cases} P_\Phi(i) & \text{for } 0 < i < N;\ j = i + 1 \\ P_M(i) & \text{for } 0 < i < N;\ j = \max(0, i - k) \\ 1 & \text{for } i, j = 0 \\ 1 & \text{for } i = N;\ j = N - k \\ 0 & \text{else} \end{cases} \tag{6.2}$$

where $P_\Phi(i) = P_\Phi$ and $P_M(i) = P_M$ from section 3.2 evaluated for $I(t) = i$. Note that when $i < k$, a recovery event would lead to extinction, and thus, under such conditions the state of the process becomes zero rather than $i - k$. For all $j \neq i + 1$ or $j \neq i - k$, $p_{ij} = 0$ resulting in a transition matrix that is composed of mostly zeros. Let $\mathbf{P}_{[k]}$ be the transition matrix for the process allowing k recoveries, e.g.,

$$\mathbf{P}_{[1]} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ P_{M(1)} & 0 & P_{\Phi(1)} & \ddots & & & & \vdots \\ 0 & P_{M(2)} & 0 & P_{\Phi(2)} & \ddots & & & \vdots \\ \vdots & \ddots & P_{M(3)} & 0 & P_{\Phi(3)} & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & P_{M(N-2)} & 0 & P_{\Phi(N-2)} & 0 \\ \vdots & & & & \ddots & P_{M(N-1)} & 0 & P_{\Phi(N-1)} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{P}_{[2]} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ P_{M(1)} & 0 & P_{\Phi(1)} & \ddots & & & & \vdots \\ P_{M(2)} & 0 & 0 & P_{\Phi(2)} & \ddots & & & \vdots \\ 0 & P_{M(3)} & 0 & 0 & P_{\Phi(3)} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & P_{M(N-2)} & 0 & 0 & P_{\Phi(N-2)} & 0 \\ \vdots & & & \ddots & P_{M(N-1)} & 0 & 0 & P_{\Phi(N-1)} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{P}_{[k]} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ P_{M(1)} & 0 & P_{\Phi(1)} & \ddots & & & & & & \vdots \\ P_{M(2)} & 0 & 0 & P_{\Phi(2)} & \ddots & & & & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & & & \vdots \\ P_{M(k)} & 0 & & & \ddots & P_{\Phi(k)} & \ddots & & & \vdots \\ 0 & P_{M(k+1)} & \ddots & & & \ddots & P_{\Phi(k+1)} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & P_{M(N-2)} & \ddots & & & \ddots & P_{M(N-2)} & 0 \\ \vdots & & & \ddots & P_{M(N-1)} & \ddots & & & \ddots & P_{M(N-1)} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

Above are transition matrices for $k = 1, 2$ followed by a general matrix for any $0 < k < N$. In general, the patterns on the main diagonal and superdiagonal (diagonal branching from $p_{0,1}$) are the same for all values of k. The main diagonal begins with $p_{00} = 1$ then consists of only zeros, which is clearly expected, and the superdiagonal is composed of values representing the probability of infection, $P_\Phi(i)$. Notice the diagonal pattern for $P_M(i)$ emerges from the $-k^{th}$ diagonal (diagonal branching from $p_{k,0}$), and that the column representing the probability of extinction has positive values for all $i \leq k$.

Therefore the values, $p_{ij}^{(n)}$ of $\mathbf{P}_{[k]}^n$, represent the probability that there are j infectious individuals in the system given $I(0) = i$ after n events have occurred. Recall that $P_0 = 1$ is the probability that the infection will eventually go extinct. Therefore, it comes as no surprise that for all k, the eigenvector corresponding to the eigenvalue 1 of matrix $\mathbf{P}_{[k]}$, has 1 as its first element and is otherwise empty e.g.,

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Consider the eigenvalues of $\mathbf{P}_5$ (chosen to examine phenomena seen in Figure 4.3). Eigenanalysis of $\mathbf{P}_5$ results in the expected eigenvalue 1, along with a handful of eigenvalues very close to 1 (within the tolerance of 1×10^{-4}). To investigate the quasiequilibrium, eigenvalue 1 and the corresponding eigenvector are disregarded. Analysis of the eigenvectors corresponding to the remaining eigenvalues within tolerance reveals the average infection level conditioning on nonextinction of 0.65.

Let $\hat{P}_0$ be the probability of early extinction, meaning for $I(0) = i$

$$\hat{P}_0 = p_{i0}^{(n)} \text{ for } n \text{ large .} \tag{6.3}$$

Evaluation of $\mathbf{P}^n_{[5]}$ for large values of n (values observed having enough number of events to allow the process to reach equilibrium) produces an approximation for $\hat{P}_0$ with $I(0) = 5$ given by $\hat{P}_0 \approx p^n_{50} = 0.15$. Together these quantities (the equilibrium and $\hat{P}_0$) result in an expected infection level of

$$\frac{E[I(t)]}{N} = 0.65(1 - .15) = 0.55 \tag{6.4}$$

Likewise, when $k = 10$ the n-step transition matrices for large enough n, $\mathbf{P}^n_{[10]}$ produces the value $\hat{P}_0 \approx p^n_{50} = 0.26$ and the average infection level of 0.65 and

$$\frac{E[I(t)]}{N} = 0.65(1 - 0.26) = 0.48 \tag{6.5}$$

These quantities align with the results seen in Figure 4.3, which shows the behavior predicted by the above results. The simulated equilibrium is seen to decrease from 0.65 when $k = 1$ to 0.55 when $k = I(0) = 5$, and further still to 0.48 when $k = 10$. Variation in these respective results occur when the quantities are extended to further decimal places; however, the discrepancies are insignificant for the population size.

CHAPTER 7
CONCLUSION

The SIS epidemiological model in this thesis was developed in order to investigate the effects of simultaneous recoveries on the expected proportion of infected individuals in a well-mixed population. Initially, the model consisted of differential equations which describe the moments of the model's probability distribution. The ODEs were developed to organically include the stochastic variation that is lost by or later added to deterministic models. This method of development results in an open system of equations, and thus, moment-closure approximations are used to truncate the system after the second moment.

The approximations of the differential equations fit the rates of change observed in simulation data, and allow for the system to be numerically solved. Comparing the approximations to data simulated for various initial conditions, values of k, and population sizes, makes it clear that which approximation performs best depends on said parameters. Aside from the conditional closure, equation 3.10, investigating the equilibrium conditioned on nonextinction using moment-closure approximations provide reliable insight. The approximations depict the qualitative behavior observed in simulations and produce quantitative results agreeable to those obtained with simulation data. Additionally, the solutions preserve the variability of the mean between multiple realizations of the process seen in Figure 4.6.

It is when parameters that result in early extinction are used, that the approximations for the differential equations fail to describe the results by simulations. Particularly, small values of k and $I(0) \leq k$ result in the largest errors between the approximations and the simulated values. Under these conditions, two of the differential equations with solutions that otherwise agree with simulated values, overestimate the equilibrium, while the other plummets towards negative infinity. To explore this phenomena a description of the process as a Markov chain and transition matrices are added to the model. This approach presents

the opportunity to analyze the system by the means of eigenanalysis to obtain the conditional equilibrium, while additionally providing a tool with which the conditional probability of early extinction due to k for any given $I(0)$ can be measured. The expected infection level calculated using these quantities accurately depicts the behavior seen in the simulations and accounts for the errors generated by the approximations.

More often than not, the approximations for equations 3.7 and 3.8 model the epidemiological process with accuracy and capture the nuances in the system. Analysis of the transition matrices results in quasiequilibria similar to those obtained by approximated solutions of the ODEs, thus supporting the results. Finally, calculating the expected level of infection given the conditional probability of early extinction demonstrates phenomena seen in the simulation and validates the model. Future study would build upon the elements of this model to include stochastic variation to recovery events. This would be achieved by defining a hypergeometric random variable representing the number of successful recoveries in k attempts.

REFERENCES

[1] L. J. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2003.

[2] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, no. 40 in Texts in applied mathematics, Springer, New York, 2001.

[3] D. H. Brown and B. M. Bolker, *The effects of disease dispersal and host clustering on the epidemic threshold in plants*, Bulletin of Mathematical Biology, 66 (2004), pp. 341–371.

[4] G. Casella and R. L. Berger, *Statistical Inference*, Brooks/Cole Cengage Learning, Belmont, CA, 2017.

[5] D. Hiebeler, *Spatially correlated disturbances in a locally dispersing population model*, Journal of Theoretical Biology, 232 (2005), pp. 143–149.

[6] ——, *Moment Equations and Dynamics of a Household SIS Epidemiological Model*, Bulletin of Mathematical Biology, 68 (2006), pp. 1315–1333.

[7] D. E. Hiebeler, R. M. Rier, J. Audibert, P. J. LeClair, and A. Webber, *Variability in a Community-Structured SIS Epidemiological Model*, Bulletin of Mathematical Biology, 77 (2015), pp. 698–712.

[8] W. Kermack and A. McKendrick, *A contribution to the mathematical theory of epidemics*, Proceedings of the Royal Society of London A, 115 (1927), pp. 700–721.

[9] D. J. Murrell, U. Dieckmann, and R. Law, *On moment closures for population dynamics in continuous space*, Journal of Theoretical Biology, 229 (2004), pp. 421–432.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.

[11] S. M. Ross, *Introduction to Probability Models*, Academic Press, 2007.

[12] A. Singh and J. P. Hespanha, *A Derivative Matching Approach to Moment Closure for the Stochastic Logistic Model*, Bulletin of Mathematical Biology, 69 (2007), pp. 1909–1925.

[13] R. S. Strichartz, *The way of analysis*, World Publishing Corporation, Beijing, 2010.

CHAPTER 8
APPENDIX

```
## Simulation of an SIS model with initial conditions v0, infection rate phi, recovery mu,
## simultaneous recoveries k, time units allotted finalT,
##record number of infectives recordDeltaT,
## ext = 1 or 0, 0 allows for model to go extinct while 1 does not

KSIS1 = function(v0=c(995,5), phi=2, mu=0.7, k=2, finalT=20, recordDeltaT=0.05, ext = 0) {
  S0 = v0[1]
  I0 = v0[2]
  N = S0 + I0
  totalVecLength = ceiling(finalT/recordDeltaT) + 1
  I.v = rep(NA, totalVecLength)
  currentT = 0
  I.v[1] = I0
  t.v = seq(0, (totalVecLength-1)*recordDeltaT, len=totalVecLength)
  lastRecordedT = currentT
  index = 1
  eventCount=0
  currentI = I.v[1]
  while (index <= totalVecLength-1) {
    if(currentI == 0) {
      I.v[(index+1): totalVecLength] = 0
      break
    }

    ## number of events

    eventCount = eventCount +1

    ## contactRate = phi*currentI
    successfulContactRate = phi*currentI*(1-currentI/N)
    recoveryRate = mu*currentI/k
    totalRate = successfulContactRate + recoveryRate
    currentT = currentT + rexp(1, totalRate)


    ## Record stuff as necessary
    spotsLeft = totalVecLength - index
    numTimesToRecord = min(floor((currentT - lastRecordedT) / recordDeltaT), spotsLeft)
    indices = index + seq_len(numTimesToRecord)
    I.v[indices] = currentI
    index = index + numTimesToRecord
    lastRecordedT = t.v[index]
    ## End of recording stuff


    ## Model stuff
    probEventIsInfection = successfulContactRate / totalRate
    if (runif(1) < probEventIsInfection) { # successful infection
      currentI = currentI + 1
    } else {  # recovery
      currentI = currentI - min(currentI-ext,k)
    }
  }
  return(list(t.v=t.v, I.v=I.v, eventCount=eventCount))
}
```

Listing 8.1. Discrete Process Simulation

BIOGRAPHY OF THE AUTHOR

Ariel Farber was born in Boise, Idaho and graduated from Centennial High School in 2010. She attended the Humboldt State University in Arcata, California from 2010-2013 and completed her degree from afar at Boise State University in the spring of 2014, with a Bachelor if Arts in Mathematics. Ariel plans to attend medical school and pursue a career in medicine. Ariel Briana Farber is a candidate for the Master of Arts degree in Mathematics from The University of Maine in August 2019.