Stockholms universitet

Stochastic models for epidemics with and without inclusion of superspreaders

Felicia Benedetti Yu

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2022:1
Matematisk statistik
Januari 2022

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2022:1**
http://www.math.su.se

Stochastic models for epidemics with and without inclusion of superspreaders

Felicia Benedetti Yu*

January 2022

Abstract

In this thesis we investigate the effects of inclusion of superspreaders in a SIR epidemic model. We create a standard SIR epidemic model and a two-type SIR epidemic model with the same basic reproductive number R_0, to describe an epidemic in a homogeneous population and in a heterogeneous population respectively. Then, with the help of branching process approximation of these models we calculate the probability that the epidemic in the two cases stays small, and conclude that it is higher in the case of superspreaders. We then investigate the expected size, when observing an emerging epidemic at some future generation, k in the homogeneous and heterogeneous case respectively. We could then conclude that we will expect the size of the epidemic in generation k to be bigger if the population contain superspreaders.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: felicia.benedetti98@gmail.com. Supervisor: Pieter Trapman.

Acknowledgements

I want to start by thanking my supervisor Pieter Trapman at the department of mathematics at Stockholm university for his meaningful thoughts and guidance throughout the process of this bachelor thesis.

Contents

1 Introduction

This thesis focuses on infectious diseases that spreads from person to person in a population. Example of such diseases are Covid-19, the common cold, measles, influenza etc. We do not consider sexually transmitted diseases since the assumptions needed for these kind of diseases differ from the non-sexual transmitted ones. Our analysis is first and foremost be based on the SIR epidemic model which is described in Section 2. In the simplest of SIR epidemic models, every individual in the population spreads the infectious disease to others with the same rate. The population in this case is called homogeneous. This is a helpful model but it does not represent a real life scenario all that well. In the real world, the rate in which a infectious individual transmits the disease to others differ from person to person. That leads us to so called superspreaders. A superspreader is a person that is responsible for more of the spreading of the disease than the average individual in the population. This can for example be a person that attends parties with a lot of people while being infectious. In an article by (Lloyd-Smith et al. 2005), the authors provides an analysis about superspreaders in the SARS-epidemic in Singapore, in which it is concluded that the majority of infected individuals were barely infectious while a small fraction of the infected were highly infectious. Since a model with superspreaders seems to represent the real world better than the standard SIR model, a further investigation of the impact of superspreaders in a model is worth to look into, which is the purpose of this thesis.

One can think of superspreaders in two different ways. The first way is that all people, before getting infected have the same properties, but by randomness, some of the infected will be more infectious for some reason like for example attending a so called superspreading event. These individuals will therefore be seen as superspreaders. This is the way that (Lloyd-Smith et al. 2005) chooses to think of superspreaders in their analysis but in this thesis we think of this in the second way; we assume that all people have different properties from the beginning that determine whether or not a person is a superspreader. This results in the population being split into two groups; non-superspreaders and superspreaders. Note that these two approaches are just two different ways of thinking regarding the same thing, meaning they are equivalent in terms of modelling.

By analyzing the impact on the epidemic of these superspreading events, one can get a better understanding on how an epidemic is going to play out in different situations and from that knowledge being able to set up necessary restrictions in case of more serious diseases.

In Section 5 we analyze the difference between a homogeneous SIR epidemic model and a SIR epidemic model with two groups of individuals with different infectious rates. In the sections previous to Section 5 our two models will be defined.

2 Definition of model

In this section we will define the standard SIR epidemic model and the two-type SIR epidemic model, which are the ones we later compare to each other. The majority of the theory in this section is retrieved from the book written by (Andersson and Britton 2000).

First we want to introduce an important number when analyzing epidemics; the basic reproduction number, R_0. This is the expected number of secondary infections caused by one typical infected individual when the whole population is assumed to be susceptible. A large outbreak of the disease is possible if and only if $R_0 > 1$ (Barratt, Kirwan, and Shantikumar 2018).

2.1 The standard SIR epidemic model

The letters S, I and R in the standard SIR epidemic model stands for the states *susceptible*, *infectious* and *removed* respectively. A susceptible individual is one that is not infected and therefore susceptible for the disease. An infectious individual is one that is infected and have the possibility to spread the disease further to those still susceptible. Eventually after some time, an infected individual either dies or recovers and is therefore immune. This individual is then in the removed state and plays no further roll in the epidemic.

In this model we assume that the population in question is closed and, like mentioned in the introduction assumed to be homogeneous. We also assume that it is homogeneously mixing, meaning that we do not take different social groups and structures into account, i.e every individual is equally likely to come in contact with any of the other individuals in the population.

Let $S(t)$, $I(t)$ and $R(t)$ denote the number of susceptible, infectious and removed in the population respectively, at time t after the start of the epidemic. Then we assume that $S(0) = n$, $I(0) = m$ and $R(0) = 0$.

Every infected individual is infectious for a period with distribution according to a random variable D, where all of the individual infectious periods are independent of each other. We are going to assume that D is exponential distributed with mean μ. While infectious, an individual makes infectious contact with others at time points following a homogeneous Poisson process with a rate λ_H, where H stands for homogeneous. These Poisson processes are independent of each other as well as of the infectious periods. If an infectious individual makes contact with a susceptible individual it is assumed that the susceptible immediately gets infected and thereby becomes infectious as well. We denote this process as (Andersson and Britton 2000), by $E_{n,m}(\lambda_H, D)$.

According to (Andersson and Britton 2000), the basic reproductive number, for this model is given by $R_0 = \lambda_H \mu$ since an infected person is on aver-

age infectious for a period of length μ and during this time infects initially susceptible individuals with rate λ_H.

2.2 The two-type SIR epidemic model

In reality it is not that likely that all individuals have the same properties. Therefore we introduce the two-type SIR epidemic models that assumes that the population is split into two groups with different properties. This is a special case of the multi-type SIR epidemic model. We assume that individuals of type 1 is non-superspreaders and those of type 2 is superspreaders.

The population in this case is still considered to be closed, but are no longer homogeneous. We do not take into account for social structures in this case either and the rate at which the infected makes infectious contact with a susceptible is only dependent on the type of that infected individual.

Let us denote the number of susceptible, infectious and removed in the whole population as above and assume again that $S(0) = n$, $I(0) = m$ and $R(0) = 0$. Then we let $S_i(t)$, $I_i(t)$ and $R_i(t)$ be the number of susceptible, infectious and removed respectively of type i at time t, $i = 1, 2$. Assume then that there are initially n_i susceptible and m_i infectious of type i, so that

$$n = n_1 + n_2, \qquad m = m_1 + m_2.$$

Let the fraction of superspreaders in the population be $\pi_2 = \pi = \frac{n_2}{n}$, and the fraction of non-superspreaders be $\pi_1 = 1 - \pi = \frac{n_1}{n}$.

Every infected individual, regardless of which group they belong to is infectious for a period with distribution according to the same random variable D introduced in the standard model. Under this period an infectious individual of type i makes infectious contact with others at time points following a homogeneous Poisson process with rate λ_i. All infectious periods and Poisson processes are assumed to be independent of each other. Here we have assumed that all individuals in the population has the same possibility to get in contact with an infectious individual. Let $\lambda_1 = \lambda_M$, where M stands for multi-type. The superspreaders in this model are assumed to be x, $x > 1$ times more infectious than the non-superspreaders, giving superspreaders the infection rate $\lambda_2 = x\lambda_M$.

Let $\mathbf{n} = (n_1, n_2)$ and $\mathbf{m} = (m_1, m_2)$, and denote this process by $E_{\mathbf{n},\mathbf{m}}(\lambda_M, D, x)$.

3 Branching Process

In this section we are going to discuss two types of branching processes; the single-type branching process and the two-type branching process, where the latter is a special case of multi-type branching processes. The reason for this is that these branching processes can represent an approximation of the two

SIR models described in the previous section. More of this approximation will be explained in Section 4.

The following theory is mainly retrieved from (Allen 2015).

3.1 Single-type branching process

A single-type branching process is a process in which the element of the branching process, let us call them individuals, give birth to other individuals independently and identically distributed. Let Y be the so called offspring random variable, representing the number of children of an individual. Let us denote the branching process $\{X(t); t \in [0, \infty)\}$, which consists of a set of discrete random variables with non-negative integers as values. We then say that $X(t)$ represent the number of individuals in the branching process at time t. In our case we also assume that the process have the Markov property, thus being a Markov chain in continuous time and the time between events being exponentially distributed. From this assumption it follows that the individuals in the branching process have a independently, equally and exponentially distributed life-spans. During their life-spans they give birth to others according to a Poisson process. We also assume that the process is homogeneous in time.

We let the transition probability for the process to get to state j in the time period Δt when in state i be denoted by

$$p_{ij}(\Delta t) = P(X(t + \Delta t) = j; X(t) = i).$$

According to (Allen 2015, p. 2), the transition probabilities satisfy the following.

$$\sum_{j=0}^{\infty} p_{ij}(t)s^j = \left(\sum_{j=0}^{\infty} p_{1j}(t)s^j \right)^i, \qquad s \in [0, 1].$$

Implying that a branching process that begins with i individuals, i.e $X(0) = i$, is equal to the sum of i processes where $X(0) = 1$.

(Allen 2015, p. 3) also states an assumption regarding the offspring which is that for the probability, p_0 of extinction at each generation in the process the following is true.

$$0 < p_0 < 1.$$

That is, the probability for the process to reach the zero-state is positive for $t > 0$. If $X(t) = 0$ for some $t > 0$ there are no individuals left to give birth to more individuals, thus the zero-state is an absorbing state. The goal from the application of branching processes in this thesis is to find the probability of ultimate extinction, q for the approximating branching process, as it equals the probability for the epidemic to stay small. To calculate q we need to know what the so called probability generating function of the offspring

random variable Y is. The following definition is stated by (Gut 2009, p. 59).

Definition 1
Let X be a non-negative integer valued random variable. The probability generating function of X is

$$g_X(t) = E(t^X) = \sum_{n=0}^{\infty} t^n \cdot P(X = n).$$

From this definition we get that the probability generating function of the offspring random variable Y is

$$g_Y(s) = \sum_{j=0}^{\infty} s^k \cdot p_k, \qquad s \in [0, 1].$$

The probability generating function, g_Y is well defined and continuously differentiable on $[0, 1]$ with the properties

$$g_Y(0) = p_0,$$
$$g_Y(1) = 1,$$
$$g'_Y(1) = \sum_{k=0}^{\infty} k p_k.$$

Then the probability of ultimate extinction is given by the minimal solution to the equation

$$g_Y(s) = s, \qquad s \in (0, 1].$$

(Allen 2015, p. 4) states this in the following theorem, while also giving a proof of this.

Theorem 1
The probability of ultimate extinction of $X(t)$ as $X(0) = 1$ is given by the smallest fixed point, q^* of the offspring probability generating function g_Y on $(0, 1]$.

The assumptions stated above results in that the probability of ultimate extinction for a branching process that begins with m individuals, $X(0) = m$, is equal to $q = (q^*)^m$ where q^* is the probability of ultimate extinction for a branching process where $X(0) = 1$.

It can also be shown that a fixed point smaller than 1 of g_Y exists if $g'_Y(1) > 1$. If $g'_Y(1) \leq 1$ then the only fixed point is 1 and therefore $q^* = 1$. In that case, ultimate extinction is certain, and if $g'_Y(1) < 1$ then the branching process is said to be *subcritical* and if $g'_Y(1) = 1$ it is called *critical*. But in the case if $g'_Y(1) > 1$ then the branching process is not certain to die out and in that case it is said to be *supercritical*.

3.2 Two-type branching process

Let $\mathbf{X}(t)$ be a vector of two stochastic processes, $\{X_1(t); t \in [0,\infty)\}$ and $\{X_2(t); t \in [0,\infty)\}$. Then $\{\mathbf{X}(t); t \in [0,\infty)\}$ is a multi-type branching process of two dimensions, i.e a two-type branching process.

We denote the transition probability for this process to get to state $\mathbf{j} = (j_1, j_2)$ from state $\mathbf{i} = (i_1, i_2)$ in a time period Δt by

$$p_{\mathbf{ij}}(\Delta t) = P(\mathbf{X}(t + \Delta t) = \mathbf{j}; \mathbf{X}(t) = \mathbf{i}).$$

Now let Y_{ij} be a discrete, non-negative random variable corresponding to the number of offspring of type j from an individual of type i. Assume that every individual of type i gives birth independently of each other with the same probability at all times. Let the probability that an individual of type i, $i = 1, 2$ gives birth to k_1 offspring of type one and k_2 offspring of type two be

$$p_{i,\mathbf{k}} = P(Y_{i1} = k_1, Y_{i2} = k_2), \qquad \mathbf{k} = (k_1, k_2).$$

Then the probability generation function of the offspring random variable $Y_i = (Y_{ii}, Y_{ij})$ is given by

$$g_{Y_i}(\mathbf{s}) = \sum_{\mathbf{k}} \mathbf{s}^{\mathbf{k}} \cdot p_{i,\mathbf{k}} = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} s_1^{k_1} \cdot s_2^{k_2} \cdot p_{i,\mathbf{k}}.$$

Let g_{Y_1} and g_{Y_2} be the probability generating functions of the offspring from an individual from type 1 and 2 respectively. From this we can get the probability of ultimate extinction vector $\mathbf{q}^* = (q_1^*, q_2^*)$ by the following theorem, which is a modification of theorem 1.2 in (Allen 2015, p. 9).

Theorem 2
The probability of ultimate extinction of the continuous-time branching process $\mathbf{X}(t)$ when $X_i(0) = 1$ and $X_j(0) = 0$, $i \neq j$, is the fixed point $\mathbf{q}^* = (q_1^*, q_2^*)$ of the system consisting of g_{Y_1} and g_{Y_2} where $q_1^*, q_2^* \in [0, 1]$. The probability of ultimate extinction of $\mathbf{X}(t)$ when $X_1(0) = m_1$ and $X_2(0) = m_2$ is given by $q = (q_1^*)^{m_1} \cdot (q_2^*)^{m_2}$.

4 Branching process approximation

The two SIR epidemic models described in Section 2 can be approximated with a single-type- and two-type branching process respectively. Where the state of a branching process corresponds to the number of infected in the epidemic. The birth of an individual in the branching process equals the event that a new individual has been infected and the death of an individual equals that individual to be removed from the epidemic. Ultimate extinction of the approximating branching processes implies that the epidemic stays small.

4.1 The homogeneous case

Firstly we assume that the branching process $X(t)$ starts with m individuals, $X(0) = m$. Then we make the assumption that each individual have a life-span distributed as the random variable D described in Section 2.1. During this time, the individuals give birth at time points from a Poisson process with intensity λ_H. Let us denote this process as (Andersson and Britton 2000), by $E_m(\lambda_H, D)$.

According to (Andersson and Britton 2000, p. 32), $I(t_0)$ in the epidemic process $E_{n,m}(\lambda_H, D)$ in Section 2.1 converges to $X(t_0)$ for each fixed t_0 almost surely. Thus we approximate our standard SIR epidemic model, $E_{n,m}(\lambda_H, D)$ by the single-type branching process, $E_m(\lambda_H, D)$.

4.2 The two-type case

In the two-type case we assume that the branching process $\mathbf{X}(t)$ starts with m_1 and m_2 individuals of type 1 and type 2 respectively and let $\mathbf{m} = (m_1, m_2)$. Assume then that each individual in the branching process has a life-span with distribution according to the random variable D, and during this time, an individual of type i gives birth to another individual of type j at time points from a Poisson process with intensity $\lambda_i \pi_j$, where $\lambda_1 = \lambda_M$ and $\lambda_2 = x\lambda_M$ and $\pi_1 = 1 - \pi$ is the fraction on non-superspreaders in the population and $\pi_2 = \pi$ is the fraction of superspreaders. We then denote this branching process by $E_{\mathbf{m}}(\lambda_M, D, x)$.

Let $\mathbf{M}$ be a matrix with element m_{ij} corresponding to the average secondary cases of type j generated by an individual of type i.

$$\mathbf{M} = \begin{pmatrix} \mu\lambda_1\pi_1 & \mu\lambda_1\pi_2 \\ \mu\lambda_2\pi_1 & \mu\lambda_2\pi_2 \end{pmatrix} = \begin{pmatrix} \mu\lambda_M(1-\pi) & \mu\lambda_M\pi \\ \mu x\lambda_M(1-\pi) & \mu x\lambda_M\pi \end{pmatrix}.$$

According to (Andersson and Britton 2000, p. 61) we can approximate the process $E_{\mathbf{n},\mathbf{m}}(\lambda_M, D, x)$ in Section 2.2 with the branching process $E_{\mathbf{m}}(\lambda_M, D, x)$ when n is large, which we assume. Then the basic reproduction number, R_0 is given by the largest eigenvalue of the matrix $\mathbf{M}$, which turns out to be $\mu\lambda_M(1 + \pi(x - 1))$.

5 Results

5.1 Distribution of the offspring random variable

In this section, we want to get the distribution of the offspring variables Y and Y_i in the branching approximations of the standard SIR epidemic model and the two-type SIR epidemic model respectively. To do this we seek the probabilities p_k and $p_{i,\mathbf{k}}$.

Like we mentioned before, in the model of an epidemic with one type, every infected individual is infectious for a period $D \sim Exp(\mu)$ with probability density function $f_D(x) = \frac{1}{\mu} e^{-\frac{1}{\mu}x}$. During this time they infect others according to a Poisson process with rate λ_H, implying that under a period of length t an infectious individual infects k other individuals with probability $\frac{(\lambda_H t)^k}{k!} e^{-\lambda_H t}$. Using the law of total expectation, the probability for an individual to infect k other individuals is given by

$$\begin{aligned} p_k &= \int_0^\infty \frac{(\lambda_H t)^k}{k!} e^{-\lambda_H t} \frac{1}{\mu} e^{-\mu t} dt \\ &= \frac{1}{\mu\lambda_H + 1} \left(1 - \frac{1}{\mu\lambda_H + 1}\right)^k, \end{aligned}$$

yielding that the offspring random variable Y is given by a geometric distribution with parameter $\frac{1}{\mu\lambda_H + 1}$.
This results in the probability generating function of Y being given by

$$g_Y(s) = \sum_{k=0}^{\infty} s^k \frac{1}{\mu\lambda_H + 1} \left(1 - \frac{1}{\mu\lambda_H + 1}\right)^k.$$

In the model of an epidemic with two types, every individual is infectious for a period distributed as in the single-type case. During this time an individual of type i infect $(k_1 + k_2)$ other individuals with probability

$$\begin{aligned} &\frac{((\lambda_i\pi_1 + \lambda_i\pi_2)t)^{k_1+k_2}}{(k_1+k_2)!} e^{-(\lambda_i\pi_1 + \lambda_i\pi_2)t} \\ &= \frac{(\lambda_i t)^{k_1+k_2}}{(k_1+k_2)!} e^{-\lambda_i t} \end{aligned}$$

Where k_1 of them are of type 1 and k_2 of them are of type 2 and $\lambda_i\pi_j$ is the rate that an individual of type i infects an individual of type j. This results in the probability for this to happen to be

$$\begin{aligned} p_{i,\mathbf{k}} &= p_{i,(k_1,k_2)} \\ &= \frac{1}{\mu\lambda_i + 1} \left(1 - \frac{1}{\mu\lambda_i + 1}\right)^{k_1+k_2} \cdot \binom{k_1 + k_2}{k_1} (1-\pi)^{k_1} \pi^{k_2}. \end{aligned}$$

This is the probability density function of $Y_i = (Y_{i1}, Y_{i2})$. We then get that the probability generation function of Y_i is given by

$$\begin{aligned} g_{Y_i}(\mathbf{s}) = \sum_{k_1=0}^{\infty} \sum_{k_1=0}^{\infty} s_1^{k_1} \cdot s_2^{k_2} \cdot \frac{1}{\mu\lambda_i + 1} \left(1 - \frac{1}{\mu\lambda_i + 1}\right)^{k_1+k_2} \cdot \\ \binom{k_1 + k_2}{k_1} (1-\pi)^{k_1} \pi^{k_2}. \end{aligned}$$

5.2 Calculation of the probability of a small epidemic

Let us now calculate the probability that the epidemics in our two cases stays small, with the help of calculating the probability of extinction of the approximated branching processes. We do this by solving

$$g_Y(s) = s \qquad s \in [0,1] \tag{1}$$

for the standard SIR model and

$$\begin{cases} g_{Y_1}(\mathbf{s}) = s_1 \\ g_{Y_2}(\mathbf{s}) = s_2 \end{cases} \qquad \mathbf{s} \in [0,1] \tag{2}$$

for the two-type SIR model. We get that the smallest positive solution to (1) is $q^* = \frac{1}{\lambda_H \mu} = \frac{1}{R_0}$, thus the probability of the epidemic to stay small for the simple SIR model is $q_H = \frac{1}{R_0}^m$.

For (2) the solution is a bit more complicated and is given by

$$\begin{cases} q_1^* = \dfrac{1 - x(1+\mu\lambda_M) + \sqrt{(x-1)^2 + 2(2\pi-1)(x-1)x\mu\lambda_M + (\mu\lambda_M)^2}}{2(\pi-1)(x-1)\mu\lambda_M} \\ q_2^* = \dfrac{x(1-\mu\lambda_M) - 1 + \sqrt{(x-1)^2 + 2(2\pi-1)(x-1)x\mu\lambda_M + (\mu\lambda_M)^2}}{2(\pi-1)(x-1)x\mu\lambda_M} \end{cases}$$

This results in the probability of a small epidemic for the two-type model to be

$$q_M = (q_1^*)^{m_1} \cdot (q_2^*)^{m_2} = ((q_1^*)^{1-\pi} \cdot (q_2^*)^{\pi})^m.$$

5.3 How does the probability of ultimate extinction depend on the fraction of superspeaders?

In this section we explore the probability that the epidemic, in the case of superspreaders stays small. We do this by looking into the probability of ultimate extinction of the approximated branching process $E_{\mathbf{m}}(\lambda_M, D, x)$ as a function of the fraction, π of superspreaders in the population. Of course for $\pi = 0$ and $\pi = 1$ the model represents an epidemic with a homogeneous population and would thus be a standard SIR epidemic model and thus the probability of ultimate extinction of the branching process is given by $\frac{1}{R_0}$ in that case.

We investigate this for the Delta variant of the ongoing epidemic Covid-19. We therefore set R_0 to 5.08 which is the average basic reproductive number for Delta Covid-19 according to (Liu and Rocklöv 2021). We let $\mu = 7$ based on the recommended days by (Folkhälsomyndigheten.se 2021) for infected individuals to stay home after the first symptoms. We then assume that superspreaders are eight times more infectious than non-superspreaders, yielding $x = 8$.

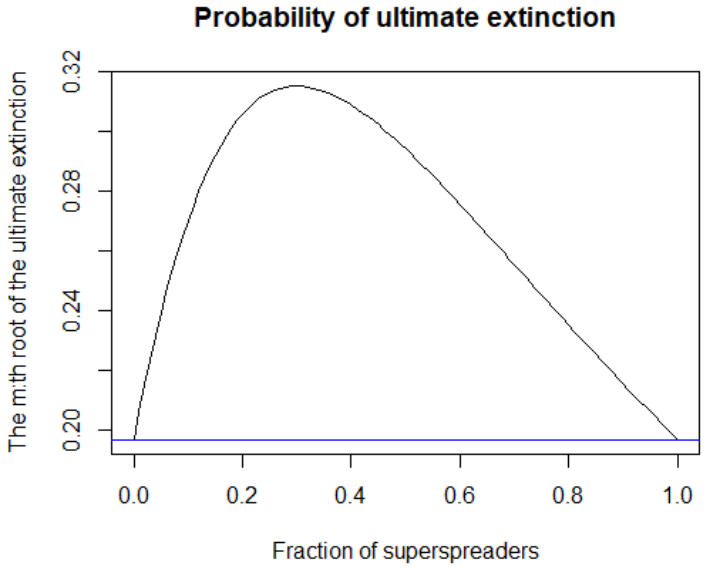

Figure 1: The m:th root probability of ultimate extinction, $\sqrt[m]{q}$ as a function of π for the two-type SIR epidemic model with parameter values $R_0 = 5.08, \mu = 7, x = 8$ and $\lambda_M = \frac{5.08}{7(1+7\pi)}$ and m initially infected individuals. The blue line represents q in a standard SIR epidemic model.

In Figure 1 the probability of ultimate extinction starts of at $\frac{1}{5.08}$ at $\pi = 0$ to then grow as the fraction of superspreaders increases. After passing a certain fraction the probability of ultimate extinction then starts to decrease again to eventually land at the value $\frac{1}{5.08}$ yet again as $\pi = 1$. The explanation for this behavior has to do with the fact that the basic reproductive number is fixed. When $\pi = 0$ the infection rate for every individual is given by $\frac{R_0}{\mu}$. But as π increases and we therefore have the situation of a population with two types of individuals, where the infectious rate for the non-superspreaders in the population is given by $\frac{R_0}{\mu(1+\pi(x-1))} < \frac{R_0}{\mu}$ and the infection rate for the superspreaders is given by $\frac{xR_0}{\mu(1+\pi(x-1))} > \frac{R_0}{\mu}$. When the fraction of superspreaders in the population is low, the vast majority of the infected individuals have an infection rate being smaller than in the individuals in a homogeneous model. While the superspreaders in the population have a higher infection rate than $\frac{R_0}{\mu}$, they make up for a very small fraction of the transmission of the disease since there is not many infected individuals of this type, if even any. Although if the fraction of superspreaders is big enough the superspreaders would "take over" the transmissions of the disease and thus reduce the probability of the epidemic to stay small.

5.4 Observing an emerging epidemic

Say that we observe an emerging epidemic and want to make some conclusions regarding the epidemics behavior in the future. But when one observes a ongoing epidemic it is going to be biased because when observing we inevitably make assumptions regarding the survival of the epidemic. Below we show the existence of such bias when calculating the expected number of infected in a homogeneous population, with the help of the real time branching process $\{X(t); t \in [0, \infty)\}$ defined in Section 3.1.

Let us assume that the epidemic we want to observe starts with one initially infected person, $X(0) = 1$. We assume that the epidemic will grow exponentially through time t, as it is common (Trapman, Ball, et al. 2016), with an epidemic rate $r = \lambda_H - \frac{1}{\mu}$. Thus one would expect that the expected number of individuals in the branching process, i.e the expected number of infected at time t to be given by

$$E[X(t)] = e^{rt}.$$

But when analyzing an emerging epidemic we need to assume the fact that the epidemic is still spreading at time t since otherwise, $E[X(t)]$ would of course be 0. Thus, what we actually observe is the expected number of infected at time t conditioned on the fact that the epidemic is not yet extinct. Let Q_t be the event that the epidemic is extinct at time t and let Q_t^c be the event that it is still spreading. Then the probability of Q_t is given

by

$$p_0 = p_{10}(t) = \frac{e^{rt} - 1}{R_0 e^{rt} - 1}.$$

The derivation of this is given by (Trapman, Meester, and Heesterbeek 2004).

We can now calculate the expectation of the size at time t through the following

$$\begin{aligned} E[X(t)|Q_t^c] &= \frac{E[X(t), \mathbb{1}_{Q_t^c}]}{1 - p_0} \\ &= \frac{E[X(t)]}{1 - \frac{e^{rt}-1}{R_0 e^{rt}-1}} = \frac{R_0 e^{rt} - 1}{R_0 - 1} \\ &= \frac{R_0 e^{rt}}{R_0 - 1} - \frac{1}{R_0 - 1}. \end{aligned}$$

Since an epidemic with a reproductive number less than 1 will not lead to a large outbreak and due to this is not of great interest for analysis, we can assume that $R_0 > 1$. This means that the term $-\frac{1}{R_0 - 1}$ is relatively small, thus we approximate the expectation of the size at time t by

$$E[X(t)|Q_t^c] \approx \frac{R_0 e^{rt}}{R_0 - 1},$$

which is larger than $E[X(t)]$. So in conclusion, when observing an emerging epidemic, what we calculate the expected size of the epidemic to be at time t is larger than in reality.

5.4.1 Observation-bias for the two models

In this section we want to analyze the difference in observing an epidemic where the population is homogeneous and where the population consists of two types of people; non-superspreaders and superspreaders. What we now want to do is to examine what the expected size of the epidemic is in the two-type case, like we did for the single-type case above. But for the multi-type case, this becomes rather complicated when working with branching processes in real time, like the branching process introduced in Section 3.2. Because of this, we choose to work with branching processes in discrete time in both cases. With this approach we can work with generations in the branching processes instead of time since the start of the epidemic. This means that the initially individuals in the branching process, i.e the initially susceptible constitutes for the first generation. All secondary cases generated by the first generation belongs to the second generation and so on. Our aim is then reformed to giving an expression for the expected size of generation k. The basic reproductive number, R_0 is the expected number

of susceptible individuals that a typical infected individual will infect. Then intuitively the expected number of infected in generation k, $E[Z_k]$ would be $R_0 \cdot E[Z_{k-1}]$ and let us then assume that there initially is one infected, then ultimately the expected number of infected in generation k should be given by

$$E[Z_k] = R_0^k, \qquad k > 1.$$

By the law of total expectation, the following applies

$$E[Z_k] = E[Z_k|Q_k] \cdot P(Q_k) + E[Z_k|Q_k^c] \cdot P(Q_k^c) = E[Z_k|Q_k] \cdot P(Q_k).$$

Where Q_k is the event that the epidemic is still spreading up to generation k and Q_k^c is the event that the epidemic is extinct by generation k.

This applies to both our models. But like stated above, when observing an emerging epidemic, we inevitably make the assumption that the epidemic is spreading up to generation k, so what we will observe is in fact $E[Z_k|Q_k]$. To make matters easier we assume that k is large since we interested in how the epidemic potentially will play out in the far future. Because k is large we can then approximate Q_k with ultimate survival of the epidemic, so we approximately observe $E[Z_k|survival]$.

In conclusion, we have the following relation

$$E[Z_k] \approx E[Z_k|survival] \cdot (1-q)$$

$$\Longleftrightarrow$$

$$E[Z_k|survival] \approx \frac{R_0^k}{1-q}.$$

Where q, is the probability of ultimate extinction of the approximated branching process for the model in question.

We know from Section 5.3 that the probability of ultimate extinction of an approximated branching process for an epidemic with superspreaders is bigger than the same epidemic without superspreaders, provided that the fraction of superspreaders is not large which it most likely is not. Given this fact we see that

$$E[Z_k|survival]_H < E[Z_k|survival]_M.$$

Where H stands for homogeneous and M for multi-type.

In other words, we observe that an epidemic in a homogeneous population will grow with a smaller rate than the epidemic in a population with superspreaders.

6 Summary and discussion

In this thesis we have compered two types of SIR epidemic models; the standard SIR epidemic model where the population is assumed to be homogeneous, and the two-type SIR epidemic model where there are two types of individuals in the population; superspreaders and non-superspreaders. In both models the population is assumed to be closed and homogeneous mixing. The comparisons are made when assuming that the epidemic have the same basic reproductive number, R_0 in the two cases. Further assumptions made are that the infected individuals are infectious for a exponential period of time and during this time infect others according to a Poisson distribution.

The two SIR models are then approximated with a single-type branching process and a two-type branching process respectively. In order to make this approximation we made the assumptions that the branching processes had the Markov property and that the initially susceptible in the population is large.

What we saw was that the expression for the probability for an epidemic described by the standard SIR epidemic model to stay small where given by $\frac{1}{R_0}$, but for an epidemic described by the two-type model the expression was very complicated and it was hard to make any conclusions from that. Because of that we needed to make some assumptions about the values of some parameters in order to compare the two models. When doing that, we could conclude that the fraction of superspreaders in an epidemic plays a big roll in the probability for a small outbreak. We saw that the higher the fraction of superspreaders, the higher the probability of a small outbreak was up to a point. When the fraction of superspreaders where big enough the probability started to go down. But the fraction of superspreaders is not likely to be high and thus the inclusion of superspreaders in a model to describe an epidemic shows a more positive future regarding the possibility of the epidemic to die out.

We then analyzed the bias that emerge when calculating the size of an emerging epidemic in the future. When approximating the epidemic in the homogeneous case by a branching process in real time, we saw that we observe that the epidemic will grow more rapidly than in reality. When approximating the standard and the two-type SIR epidemic model with branching processes in discrete time, we saw that the in an emerging epidemic with superspreaders we observe that the epidemic grows even more rapidly than in the homogeneous case.

So although an epidemic with superspreaders are more likely to go extinct than in the homogeneous case, if the epidemic does not go extinct, we can expect it to grow faster than if there where no superspreaders.

7 Improvements

In further work it would be interesting to analyze the case where the population in question is not closed, since that is not usually the case in reality. It is also more likely that the population is not homogeneous mixing and it would be interesting to investigate how the two-type model would look if λ_{ii} was not necessarily equal to λ_{ij}. This would of course make calculations more complicated and thus was not considered in this bachelor thesis.

In this thesis we also assumed that a fraction, π of superspreaders in a population implicates that the fraction of superspreaders in the initially infected is also π. This is not necessarily the case.

To get a better overall view of the epidemic in the homogeneous and the two-type case respectively, it would also be interesting to see some simulations of the epidemic through time.

References

Allen, Linda JS (2015). "Stochastic population and epidemic models". In: *Mathematical biosciences lecture series, stochastics in biological systems*.

Andersson, Hakan and Tom Britton (2000). *Stochastic epidemic models and their statistical analysis*. Vol. 151. Springer Science & Business Media.

Barratt, H, M Kirwan, and S Shantikumar (2018). "Epidemic theory (effective & basic reproduction numbers, epidemic thresholds) & techniques for analysis of infectious disease data (construction & use of epidemic curves, generation numbers, exceptional reporting & identification of significant clusters)". In: *Health Knowledge*.

Folkhälsomyndigheten.se (2021). *Hur Länge Bör Jag Stanna Hemma? - folkhälsomyndigheten*. URL: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/skydda-dig-och-andra/hur-lange-bor-jag-stanna-hemma/ (visited on 12/21/2021).

Gut, Allan (2009). *Intermediate course in probability [electronic resource]*. Springer.

Liu, Ying and Joacim Rocklöv (Aug. 2021). "The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus". In: *Journal of Travel Medicine* 28.7. taab124. ISSN: 1708-8305. DOI: 10.1093/jtm/taab124. eprint: https://academic.oup.com/jtm/article-pdf/28/7/taab124/41825935/taab124.pdf. URL: https://doi.org/10.1093/jtm/taab124.

Lloyd-Smith, James O et al. (2005). "Superspreading and the effect of individual variation on disease emergence". In: *Nature* 438.7066, pp. 355–359.

Trapman, Pieter, Frank Ball, et al. (2016). "Inferring R 0 in emerging epidemicsâthe effect of common population structure is small". In: *Journal of The Royal Society Interface* 13.121, p. 20160288.

Trapman, Pieter, Ronald Meester, and Hans Heesterbeek (2004). "A branching model for the spread of infectious animal diseases in varying environments". In: *Journal of mathematical biology* 49.6, pp. 553–576.