Mathematical Statistics
Stockholm University

Effects of random distributions on infectious time periods in epidemic modeling

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control

Research Report 2005:13

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2005:13**,
http://www.math.su.se/matstat

Effects of random distributions on infectious time periods in epidemic modeling

Tommi Asikainen,
Stockholm University and
The Swedish Institute for Infectious Disease Control*

December 2005

Abstract

This paper aims at making sensitivity analysis in epidemic models. We will concentrate on the effect of different random distributions on the infectious time period. One issue is how many persons will be totally infected during an outbreak when no major outbreaks are possible. With a major outbreak we mean the equivalence with branching processes, when a population will explode in size. We find out that the expected value is low, but the variance can be considerably high. This causes problems since often having more than one case makes it to an epidemic among medically trained individuals.

The second problem considered is how long time will it take for a major epidemic to occur. Given that the population is in a state, that major outbreaks can occur.

KEY WORDS: Herd immunity, epidemic modeling, time until outbreak, basic reproductive distribution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91Stockholm, Sweden. E-mail: tommi.asikainen@smi.ki.se. Financial support from the Swedish Institute for Infectious Disease Control, Karolinska Institute, DG Sanco project MODELREL and Martin Rinds foundation is gratefully acknowledged.

Contents

List of Figures

1. Introduction

The aim is to investigate the sensitivity in epidemic models. The emphasis will be on the effect of different random distributions on infectious time periods.

But first we will have to define the term epidemic (or outbreak). There is no universal definition for an epidemic. (Heymann (2004)) describes an epidemic as "The occurrence, in a defined community or region, of cases of illness (or an outbreak) with a frequency clearly in excess of normal expectancy", this is for a disease which could occur seasonally. In the case of new diseases it is described as "A single case of a communicable disease long absent from a population or the first invasion by a disease not previously recognized in that area requires immediate reporting and full field epidemiological investigation; 2 cases of such a disease associated in time and place are sufficient evidence of transmission to be considered an epidemic". These definitions are useful for medical people and epidemiologists. Statistically an outbreak could be divided into a minor or major outbreak A minor outbreak could be when the size of a population in a branching process approximation will not explode, a major outbreak would mean that the population explodes in size. It is important to be aware of the differences between the medical and statistical view of outbreaks.

First we study the total number of infected in a community where the probability for a major outbreak (epidemic) is zero. Secondly we study the time until a major outbreak (epidemic) appears of an infectious disease in a community who are subject to infection, with a positive probability for major outbreak.

The infection process will be started by imported cases, that is an infectious person enters the population.

One example is the measles is Sweden, for which the vaccination program started 1982. Before that it was in a state that a major outbreak occurred with 3 - 4 year intervals (Ström (1964)). After the introduction of the vaccine the number of resistant individuals increased. These were enough to put the population in a state, where the probability for a major outbreak is zero. Measles is a disease which in many countries with childhood vaccination programs against measles can cause minor outbreaks by introduction of imported cases. The number of infected in Sweden, started by one case have varied between 1 - 80 since 1997. 80 sounds like a major outbreak from a medical viewpoint, but calculating prediction intervals will not make it to a major outbreak from a statistical viewpoint.

Due to some articles claiming that the vaccine could cause autism (Wakefield et al. (1998), Pounder et al. (1995)), the proportion vaccinated decreased.

If this would have continued, the population could again be in a state where major outbreaks would be possible (Asikainen et al. (2003)). This shows that the question of how long time it will take before a major outbreak will occur is also of importance.

These things are interesting for both known (present and past) and new "emerging" diseases. Known diseases could be measles, rubella or smallpox. New ones pandemic influenza caused by avian flu or some other strain.

Depending on how infectious a disease is and how many in the population are immune against this, through vaccination or having encountered the disease earlier in life, there is a a probability for a major outbreak. The quantity describing infectiousness is called basic reproductive number. We will discuss some problems in interpreting this quantity.

If the proportion immune is high enough, the probability for a major outbreak is zero. This is linked with basic reproductive number.

An other example is the time until a major outbreak might occur. This has been an interesting question concerning the recent outbreaks or fear of it, of SARS and avian flu. Both could be capable of making a major outbreak, but have been kept in control by massive counter measures by the authorities.

The most modeling done in the field do not take the stochasticity in the start of the epidemic into account. The interesting part is how long time from breaking the barrier when major outbreaks are possible, it will take until the major outbreak (epidemic) will take place. It is also a believe among medical people that this will take place instantly. We try to show that this should be viewed as a stochastic process.

2. Models for spread of infection

2.1 *Disease and infection*

The development of disease can be described as shown below and in Figure 1 (Giesecke (2002)):

1. Healthy The person has not yet experienced the disease and is not protected against it.
2. Incubation period A susceptible person gets infected by an infectious individual. It takes a certain amount of time until the disease develops in the body and the first signs of disease appear (these can be sub clinical).
3. Clinical disease Signs of the disease are appearing

[!bp]

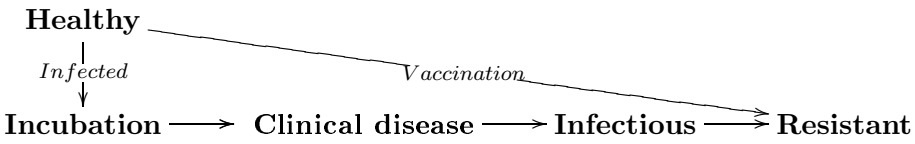

Figure 1. Progress of disease

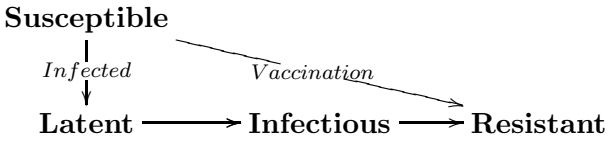

Figure 2. Progress of infection

4. Resistant The person has recovered from the disease, or has been transferred from susceptible through a successful vaccination.

How infection is spread, differs from the above. This is because an individual can be infectious before or even after the stage of clinical disease. This process consists of the following states and is shown in Figure 2 (Giesecke (2002)):

1. Susceptible The person has not yet experienced the disease and is not protected against it.
2. Latent period A susceptible person gets infected by an infectious individual. The virus is fighting with the body's defense system.
3. Infectious period The person is now infectious and can transmit the disease to others.
4. Resistant The person has recovered from the disease, or has been transferred from susceptible through a successful vaccination.

The number of individuals in the different states are usually of interest. Although in many cases, each individual has to be followed separately during the epidemic. These are the two important factors which will decide how the epidemic is progressing:

- The time spent at each state, and its random distribution
- How many susceptible are infected per time unit

When an individual has moved to Resistant state, there will be two possibilities. For individuals who have been infected with measles they will stay there for the rest of their life. For persons vaccinated it is not known how long time the protection holds. The other possibility is with diseases like influenza which come in different forms every year and this is why people can be susceptible even after encountering the disease earlier.

2.2 *Basic assumptions in epidemic models and some terms*

Infectious contacts made by an infectious person in the population are assumed to occur according to a Poisson process $\boldsymbol{K}$ with rate λ. This mean that time periods between contacts are exponentially distributed. The number of contacts during a time period is Poisson distributed.

If a proportion of the population are immune, due to earlier encounter with the disease or through vaccination, the rate λ will reduce in relation to this.

Different kind of mixing is possible in the population, through families, work etc. In this report the population is homogeneously mixing. That is there is no special division of the population into different subgroups.

An index case is the first infective in an outbreak. The others infected by this index case are called secondary cases. Tertiary cases are those infected by the secondary cases. In reality this ordering is very hard to follow due to the uncertainty in who has infected whom. It is also possible to have several index cases, people exposed to a sudden release of an agent for example. The index case will be assumed to be equally infectious as other individuals during the epidemic. Often it might happen that the index case is infectious before entering the population, making him less infectious that the others. This is not investigated in this paper.

In order to start an epidemic the index cases are assumed to appear according to Poisson process $\boldsymbol{K_i}$ with rate λ_i. Index case be a person who gets infected abroad and is still infectious or becomes infectious when returning home.

2.3 *Time spent in different states*

An infectious individual will remain infectious for a stochastic period of time $\boldsymbol{I}$, according to a random distribution.

When trying to model the transitions between different states the time spent in different states can have different distributions. The two most common distributions are delta distribution (constant time period) and exponential distribution.

Constant time periods are useful in certain situations. The case with exponential distribution gives some stochasticity. But very few infections, if any,

follows an exponential distribution in the different states.

In our paper we try a number of different distributions and do comparisons between these. We will use the gamma distribution, since it can be viewed as a sum of exponential distributions. This way we can keep the nice Markovian properties and add more reality to our model.

We will also investigate the effect of time, when the critical level for having major outbreaks is achieved. The time for a major outbreak, can be depending on the assumptions concerning the distribution of time being infectious. The disease might also have different states, where individuals are transmitting different amounts of infection.

2.4 *Basic reproductive distribution*

Assume we have knowledge of the number of infectious contacts per time unit, Chapter 2.2. The random distribution of the infectious time period $\boldsymbol{I}$ is also known, Chapter 2.3.

The number of new cases made by an infectious person will be a stochastic process. We will propose the term "Basic reproductive distribution", $\boldsymbol{RD}$, for the number of new cases. Depending on the random distributions on the input parameters, $\boldsymbol{RD}$ will follow a random distribution. In some situations it will be a very simple expression.

Using $\boldsymbol{RD}$ as a tool, there will be a lot of flexibility. Assume that the infectious time period is constant and that the rate of making infectious contacts is exponentially distributed. This is equivalent for having exponentially distributed infectious time period and a constant rate of making infectious contacts.

2.5 *Basic reproductive number*

An often used parameter in epidemic modeling, is the so called "basic reproductive number", denoted $\boldsymbol{R_0}$. In its easiest form it can be expressed as the expected number of new infections generated by one infectious individual, during the time he is infectious in a population where everyone is susceptible. It can be seen as the mean of the basic reproductive distribution.

This is although a very "misused" parameter since it is often estimated from previous outbreaks. Usually how the population structure is set can give major deviations in the estimation. (Asikainen (2005), Gani and Leach (2001))

If $\boldsymbol{R_0} < 1$ no major outbreaks (epidemics) will occur, only small outbreaks are possible. Although a small can be of size even exceeding 100. If $\boldsymbol{R_0} > 1$ there will be a positive probability for a major outbreak, meaning that a considerable proportion of the population will get infected.

Basic reproductive number is related to the term, herd immunity, which means that $\boldsymbol{R_0} < 1$, giving only small outbreaks. It means that there are so few susceptible in the population that even if an infectious individual enters the population, he will not start an epidemic. This term should be used very carefully. For example 1997 a person infected with measles entered a discotheque in a Stockholm, which due to mass vaccination is in a state of herd immunity. There he infected a number of others and the total epidemic consisted of about 80 cases before it was stopped.

If the population possesses some resistance to the disease, not all infectious contacts will cause new cases. In this case $\boldsymbol{R_0}$ will be scaled to suit the case, often denoted $\boldsymbol{R_a}$, $\boldsymbol{R_*}$ or $\boldsymbol{R_e}$. We will write $\boldsymbol{R_0}$ even in this case.

3. Epidemic models

In Chapter 2 we mentioned the basic assumptions and some common terms. In this chapter we mention a number of models used.

In describing the models all will have the same value on $\boldsymbol{R_0}$, namely λ/γ.

The notation for the number of susceptible, latent, infectious and removed at time point t is $\mathcal{S}(t)$, $\mathcal{E}(t)$, $\mathcal{I}(t)$ and $\mathcal{R}(t)$.

Assume that a person spends in average $1/\beta$ in the latent phase.

3.1 *Constant value on the infectious time period*

Also known as the Reed-Frost epidemic model (Abbey (1952)), assumes that the infectious time period $\boldsymbol{I}$ is constant, with a value of $1/\gamma$.

$$\begin{aligned} \boldsymbol{I} &\sim \delta\left(\frac{1}{\gamma}\right) \\ \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \boldsymbol{K}(\boldsymbol{I}) &\sim \boldsymbol{RD} \\ \boldsymbol{R_0} &= \mathrm{E}[\boldsymbol{K}(\boldsymbol{I})] = \frac{\lambda}{\gamma} \end{aligned}$$

A model can be made for both discrete and continuous time.

Assume that infections occur at discrete time (and that the time between generations is of length 1 for simplicity). The number of infections at different

time points will be binomially distributed. Let N denote the total size of the population.

$$\begin{aligned}
&\text{P(A certain individual not infected by 1 infective)} = e^{-\frac{\lambda}{N\gamma}} \\
&\text{P(A certain individual not infected by } \mathcal{I}(t) \text{ infective)} = e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}} \\
&\text{P(A certain individual infected by } \mathcal{I}(t) \text{ infective)} = 1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}
\end{aligned}$$

This gives a way to calculate the number of infected at different generations:

$$\begin{aligned}
\mathcal{I}(t+1) &\sim \text{Bin}\Big(\mathcal{S}(t), 1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}\Big) \\
\text{E}[\mathcal{I}(t+1)] &= \mathcal{S}(t)\Big(1 - e^{-\frac{\lambda \mathcal{I}(t)}{N\gamma}}\Big) \approx \frac{\lambda}{\gamma n}\mathcal{S}(t)\mathcal{I}(t)
\end{aligned}$$

In continuous time, the infection is a Poisson process with rate $\frac{\lambda}{\gamma}$. The time between new infections will be exponentially distributed with rate λ, with expected value $\frac{1}{\lambda}$. Let $\boldsymbol{\xi}$ = number of new cases made by $\mathcal{I}(t)$ infective. $\boldsymbol{\xi} \sim \boldsymbol{RD}$.

$$\begin{aligned}
\boldsymbol{\xi} &\sim \text{Po}\Big(\frac{\lambda}{\gamma}\Big) \\
\text{Proportion susceptible} &= \frac{\mathcal{S}}{N} \\
&\Downarrow \\
\boldsymbol{\xi} &\sim \text{Po}\left(\frac{\lambda \mathcal{S}}{\gamma N}\mathcal{I}\right) \\
\text{E}[\boldsymbol{\xi}] &= \frac{\lambda \mathcal{S}}{\gamma N}\mathcal{I}
\end{aligned}$$

The expected value of the processes is the same. The difference is that in continuous time new infections do not have to happen at certain generations.

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why $\frac{\mathcal{S}}{N}$, proportion of susceptible has been used.

The Basic reproductive distribution, $\boldsymbol{RD}$ with constant value on the infectious time period will be a Poisson distribution.

3.2 *Exponentially distributed infectious time period*

Also called the general epidemic model (Kermack and McKendrick (1927)), the infectious time $\boldsymbol{I}$ is assumed to be exponentially distributed with rate γ.

$$\begin{aligned} \boldsymbol{I} &\sim \text{Exp}(\gamma) \\ \text{E}[\boldsymbol{I}] &= 1/\gamma \\ \text{Var}(\boldsymbol{I}) &= \frac{1}{\gamma^2} \\ \boldsymbol{\xi} &\sim \boldsymbol{RD} \\ \boldsymbol{R_0} &= \text{E}[\boldsymbol{\xi}] = \frac{\lambda}{\gamma} \end{aligned}$$

This model is widely used, specially for simulating epidemics. Since the times are exponentially distributed there can be a deterministic solution by solving a number of differential equations. N denotes the population size, $1/\beta$ the average time spent in the latent phase.

$$\begin{aligned} \frac{d\mathcal{S}}{dt} &= -\frac{\lambda}{N}\mathcal{S}\mathcal{I} \\ \frac{d\mathcal{E}}{dt} &= \frac{\lambda}{N}\mathcal{S} - \beta\mathcal{E} \\ \frac{d\mathcal{I}}{dt} &= \beta\mathcal{E} - \gamma\mathcal{I} \\ \frac{d\mathcal{R}}{dt} &= \gamma\mathcal{I} \end{aligned}$$

The downside is that with this setup, we are modeling what happens in average. It is crucial to decide if this is the quantity of interest before applying this approach.

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why the differential equations are scaled with $\frac{\mathcal{S}}{N}$.

The Basic reproductive distribution, $\boldsymbol{RD}$ with exponential distribution on the infectious time period will be a geometric distribution.

3.3 *Gamma distributed infectious time period*

The general epidemic model is modified so that the expected infectious time period of length $1/\gamma$ is divided into n independent blocks each having an exponentially distributed length of $1/n\gamma$.

$$
\begin{aligned}
\boldsymbol{I}_j &\sim \text{Exp}(\text{n}\gamma) \\
\boldsymbol{I} &= \sum_{j=1}^{a} \boldsymbol{I}_j \sim \Gamma(n, n\gamma) \\
\text{E}[\boldsymbol{I}_j] &= \frac{1}{n\gamma} \quad \text{Var}(\boldsymbol{I}_j) = \frac{1}{n^2\gamma^2} \\
\text{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \quad \text{Var}(\boldsymbol{I}) = \frac{1}{n\gamma^2} \\
\boldsymbol{\xi} &\sim \boldsymbol{RD} \\
\boldsymbol{R_0} &= \text{E}[\boldsymbol{\xi}] = \frac{\lambda}{\gamma}
\end{aligned}
$$

The main advantage is that the variation compared with the general epidemic model is reduced.

As with an exponential distribution, this can be modeled using differential equations, the difference being that the infectious time period $\mathcal{I}$ will be split into n different blocks.

$$
\begin{aligned}
\frac{d\mathcal{S}}{dt} &= -\frac{\lambda}{N}\mathcal{S}\sum_{j=1}^{n}\mathcal{I}_j \\
\frac{d\mathcal{E}}{dt} &= \frac{\lambda}{N}\mathcal{S}\sum_{j=1}^{n}\mathcal{I}_j - \beta\mathcal{E} \\
\frac{d\mathcal{I}_1}{dt} &= \beta\mathcal{E} - n\gamma\mathcal{I}_1 \\
\frac{d\mathcal{I}_2}{dt} &= n\gamma\mathcal{I}_1 - n\gamma\mathcal{I}_2 \\
&\vdots \\
\frac{d\mathcal{I}_n}{dt} &= n\gamma\mathcal{I}_{n-1} - n\gamma\mathcal{I}_n \\
\frac{d\mathcal{R}}{dt} &= n\gamma\mathcal{I}_n
\end{aligned}
$$

We would like to point out that in the latter calculations, not everyone in the population is assumed to be susceptible. This is why the differential equations are scaled with $\frac{\mathcal{S}}{N}$. The latent period is in the differential equation above assumed to be exponentially distributed.

The Basic reproductive distribution, $\boldsymbol{RD}$ with gamma distribution on the infectious time period will be a negative binomial distribution.

3.4 *Changes for the remaining of the article*

This far the latent period has been taken into account in the models. In the continuation of our report the latent period is abandoned from the model. The focus will be on the probability for a major outbreak and the time point at which that might be initiated.

4. Total number of infected during herd immunity

As stated in Chapter 2.2 the number of infected can become rather big. This even if the population is "protected" by herd immunity ($\boldsymbol{R_0} < 1$). The total number of infected is of interest. As shown in the Introduction, if this is a major outbreak or not, depends in which definition of outbreak to use.

Let $\boldsymbol{X}$ denote the total number of cases during the outbreak, including the index case(s).

Assume that the infectious contacts made by an infectious person is a Poisson process $\boldsymbol{K}$ with rate λ. The distribution of the infectious time period $\boldsymbol{I}$ is assumed to be known.

The progression of disease will be modeled by using techniques from the field of branching processes. The approximation is done by assuming that as infectious person possibly generates new cases, all of these will start new independent processes.

4.1 *Calculation of moments*

Let $\boldsymbol{\xi}_i$ denote the number of new cases made by an infectious individual, $\boldsymbol{\xi}_i \sim \boldsymbol{RD}$ i. The infectious contacts made by him is a Poisson process with rate λ. The probability generating function for $\boldsymbol{\xi} = g_{\boldsymbol{\xi}}(t)$.

$$\begin{aligned} g_{\boldsymbol{\xi}}(t) &= \psi_{\boldsymbol{I}}(\lambda(t-1)) \\ \psi_{\boldsymbol{I}}(t) &= \text{Moment generating function for } \boldsymbol{I} \\ S_x &= \sum_{j=1}^{x} \boldsymbol{\xi}_j \\ g_{S_x}(t) &= \psi_{\boldsymbol{I}}^x(\lambda(t-1)) \\ \boldsymbol{X} &= S_x \end{aligned}$$

As proved in (Sewastjanow (1975)):

$$\begin{aligned} P(\boldsymbol{X}=x) &= \frac{1}{x}P(S_x=x-1) \\ P(S_x=x-1) &= \frac{1}{(x-1)!}\frac{d^{x-1}}{dt^{x-1}}g_{S_x}(t)\Big|_{t=0} \\ P(\boldsymbol{X}=x) &= \frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}g_{S_x}(t)\Big|_{t=0} = \frac{1}{x!}\frac{d^{x-1}}{dt^{x-1}}\psi_{\boldsymbol{I}}^{x}(\lambda(t-1))\Big|_{t=0} \end{aligned}$$

$g_{\boldsymbol{X}}(t)$ =Probability generating function for $\boldsymbol{X}$.

$$\begin{aligned} g_{\boldsymbol{X}}(t) &= \mathrm{E}[t^{\boldsymbol{X}}] = \sum_{j=0}^{\infty} t^j \frac{1}{j!}\frac{d^{x-1}}{dt^{x-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \psi_{\boldsymbol{X}}(t) &= \mathrm{E}[e^{t\boldsymbol{X}}] = \sum_{j=0}^{\infty} e^{tj} \frac{1}{j!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \mathrm{E}[\boldsymbol{X}] &= \psi'_{\boldsymbol{X}}(0) = \sum_{j=0}^{\infty} \frac{1}{(j-1)!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \\ \mathrm{E}[\boldsymbol{X}^{\mathrm{v}}] &= \psi^{v}_{\boldsymbol{X}}(0) = \sum_{j=0}^{\infty} \frac{j^{v-1}}{(j-1)!}\frac{d^{j-1}}{dt^{j-1}}\psi_{\boldsymbol{I}}^{j}(\lambda(t-1))\Big|_{t=0} \end{aligned}$$

This enables to calculate moments of acquired value.

4.2 *Expected value of total number infected*

The expected value of total number infected can be calculated by letting $\boldsymbol{\xi}$ denote the number of infected by an infectious person, $\boldsymbol{\xi} \sim \boldsymbol{RD}$. Let $\boldsymbol{Y}$ denote the total number of infected caused by the index case. Use methods from branching processes. The index case will generate $\boldsymbol{\xi}$ new cases. Each of these $\boldsymbol{\xi}$ will start new independent processes of size $\boldsymbol{Y}$.

$$
\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} \boldsymbol{Y}+1 \\
\boldsymbol{Y}|\boldsymbol{\xi} &= \boldsymbol{\xi}+\sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{Y}_j \\
\mathrm{E}[\boldsymbol{Y}|\boldsymbol{\xi}] &= \mathrm{E}[\boldsymbol{\xi}]+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{Y}_{\mathrm{j}}] \\
\text{Since } \boldsymbol{Y}_j \stackrel{d}{=} \boldsymbol{Y} &\quad \text{and} \quad \mathrm{E}[\mathrm{E}[\boldsymbol{Y}|\boldsymbol{\xi}]] = \mathrm{E}[\boldsymbol{Y}] \\
\mathrm{E}[\boldsymbol{Y}] &= \mathrm{E}[\boldsymbol{\xi}]+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{Y}] \\
\mathrm{E}[\boldsymbol{Y}] &= \frac{\mathrm{E}[\boldsymbol{\xi}]}{1-\mathrm{E}[\boldsymbol{\xi}]} \\
\mathrm{E}[\boldsymbol{X}] &= 1+\mathrm{E}[\boldsymbol{Y}] = \frac{1}{1-\mathrm{E}[\boldsymbol{\xi}]}
\end{aligned}
$$

Alternative solution can be obtained by noting that $\boldsymbol{X}$ has the same distribution as the index case plus the total number infected by the offsprings of the index case.

$$
\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} 1+\sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{X}_j \\
\mathrm{E}[\boldsymbol{X}] &= 1+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{X}_{\mathrm{j}}] = 1+\mathrm{E}[\boldsymbol{\xi}]\mathrm{E}[\boldsymbol{X}] \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{1-\mathrm{E}[\boldsymbol{\xi}]} \qquad (1)
\end{aligned}
$$

The expected value of the total number of cases during the epidemic, will be independent of the distribution on the infectious time $\boldsymbol{I}$. As long as it has the same mean $1/\gamma$.

4.3 *Variance of total number infected*

The variance of total number infected can be calculated by letting $\boldsymbol{\xi}$ denote the number of infected by an infectious person, $\boldsymbol{\xi} \sim \boldsymbol{RD}$. Let $\boldsymbol{X}$ denote the total number of infected including the index case. This index case generates $\boldsymbol{\xi}$ new cases. Each of these will start a new independent branching process,

with the same distribution as $\boldsymbol{X}$.

$$\begin{aligned}
\boldsymbol{X} &\stackrel{d}{=} 1 + \sum_{j=1}^{\boldsymbol{\xi}} \boldsymbol{X}_j \\
\mathrm{Var}(\boldsymbol{X}) &= \mathrm{E}[\boldsymbol{\xi}]\mathrm{Var}(\boldsymbol{X}_\mathrm{j}) + (\mathrm{E}[\boldsymbol{X}_\mathrm{j}])^2\mathrm{Var}(\boldsymbol{\xi}) \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\mathrm{Var}(\boldsymbol{\xi})}{(1-\mathrm{E}[\boldsymbol{\xi}])^3} \qquad (2)
\end{aligned}$$

In contrast to the expected value, which is same for all distributions with the same mean value. The variance will depend on the variance of $\boldsymbol{\xi}$. In our case $\boldsymbol{\xi}$ will be a compound Poisson distribution.

4.4 *Constant value on the infectious time period*

The total number of cases $\boldsymbol{X}$ in a Reed-Frost process (constant infectious time period) can be shown to follow a Borel-Tanner distribution (Height and Breuer (1960), Tanner (1953), Borel (1942)). This distribution is derived from queuing theory, in the case for a constant service period and arrival of new individuals according to a Poisson process with rate λ.

$$\begin{aligned}
\boldsymbol{I} &\sim \delta(\frac{1}{\gamma}) \\
\psi_{\boldsymbol{I}}(t) &= e^{\frac{t}{\gamma}} \\
\psi_{\boldsymbol{I}}(\lambda(t-1)) &= e^{\frac{\lambda}{\gamma}(t-1)} = e^{\boldsymbol{R_0}(t-1)} \\
\psi^x_{\boldsymbol{I}}(\lambda(t-1)) &= e^{x\boldsymbol{R_0}(t-1)} \\
\frac{d^x}{dt^x} &= (x\boldsymbol{R_0})^x e^{x\boldsymbol{R_0}(t-1)} \\
P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{1}{x!}(x\boldsymbol{R_0})^{x-1}e^{x\boldsymbol{R_0}(t-1)}\Big|_{t=0} \\
P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{(x\boldsymbol{R_0})^{x-1}e^{-x\boldsymbol{R_0}}}{x!} \\
\boldsymbol{X} &\sim \text{Borel-Tanner}(\boldsymbol{R_0}) \quad ; \quad \boldsymbol{R_0} < 1 \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}}{(1-\boldsymbol{R_0})^3} \\
&\Longleftrightarrow
\end{aligned}$$

$$\begin{array}{rcl} & \Longleftrightarrow & \\ \lim_{\boldsymbol{R_0}\to 1-} \mathrm{E}[\boldsymbol{X}] & = & \infty \\ \lim_{\boldsymbol{R_0}\to 1-} \mathrm{Var}(\boldsymbol{X}) & = & \infty \end{array}$$

Alternative solution by using equation (2) and (1), side 19 and 18:

$$\begin{aligned} \boldsymbol{\xi} &\sim Po(\lambda/\gamma) = Po(\boldsymbol{R_0}) \\ \mathrm{E}[\boldsymbol{\xi}] &= \boldsymbol{R_0} \\ \mathrm{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\ \mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}}{(1-\boldsymbol{R_0})^3} \end{aligned}$$

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.5 *Exponential distribution on infectious times*

Since the infectious time is exponentially distributed, the offspring distribution in a branching process approach becomes geometric. This given that infectious contacts occur according to a Poisson process with rate λ. The total number of infected $\boldsymbol{X}$ including index cases has been investigated by (Farrington et al. (2003)), by using the Lagrangian generalized negative binomial distribution (or Lagrangian negative binomial distribution).

$$\begin{aligned} \boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\ \psi_{\boldsymbol{I}}^x(t) &= (1-\frac{t}{\gamma})^{-x} \\ \frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(t) &= \frac{(x+x-1)!}{(x-1)!}\frac{1}{\gamma^l}(1-\frac{t}{\gamma})^{-(x+l)} \\ \frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(\lambda(t-1)) &= \frac{(x+l-1)!}{(x-1)!}(\frac{\lambda}{\gamma})^l(1-\frac{\lambda(t-1)}{\gamma})^{-(x+l)} \\ P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(x+(x-1)-1)!}{(x-1)!}(\frac{\lambda}{\gamma})^{x-1}(1-\frac{\lambda(t-1)}{\gamma})^{-(x+(x-1))}\Big|_{t=0} \\ P(\boldsymbol{X}=x|\boldsymbol{R_0}) &= \frac{(2x-2)!}{x!(x-1)!}\frac{{\boldsymbol{R_0}}^{x-1}}{(1+\boldsymbol{R_0})^{2x-1}} \end{aligned}$$

Rewriting the expression as $\boldsymbol{Z} = \boldsymbol{X} - 1 \sim$ Lagrangian binomial distribution (Kotz and Johnson (1983)), also known as Lagrangian negative binomial distribution (Jain and Consul (1971)).

$$
\begin{aligned}
\boldsymbol{Z} &= \boldsymbol{X} - 1 \sim \text{ Lagrangian generalized negative binomial distribution} \\
\mathrm{P}(\boldsymbol{Z} = z) &= \frac{n}{n+\alpha t}\binom{n+\alpha z}{z} p^z q^{n+\alpha z - z} \quad , n > 0 \quad , p < \alpha p < 1 \quad , z = 0, 1, 2, \ldots \\
\mathrm{E}[\boldsymbol{Z}] &= \frac{np}{1-\alpha p} \\
\mathrm{Var}(\boldsymbol{Z}) &= \frac{npq}{(1-\alpha p)^3} \\
&\vdots \\
\mathrm{E}[\boldsymbol{Z}] &= \frac{\frac{\boldsymbol{R_0}}{1+\boldsymbol{R_0}}}{1-\frac{2\boldsymbol{R_0}}{1+\boldsymbol{R_0}}} = \frac{\boldsymbol{R_0}}{1+\boldsymbol{R_0}-2\boldsymbol{R_0}} = \frac{\boldsymbol{R_0}}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{Z}) &= \frac{\frac{\boldsymbol{R_0}}{(1+\boldsymbol{R_0})^2}}{(1-\frac{2\boldsymbol{R_0}}{1+\boldsymbol{R_0}})^3} = \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3} \\
\mathrm{E}[\boldsymbol{X}] &= \mathrm{E}[\boldsymbol{Z}] + 1 = \frac{\boldsymbol{R_0}}{1-\boldsymbol{R_0}} + 1 = \frac{1}{1-\boldsymbol{R_0}} \\
\mathrm{Var}(\boldsymbol{X}) &= \mathrm{Var}(\boldsymbol{Z}) = \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

Since the offspring is geometrically distributed, the variance of $\boldsymbol{X}$ can also be calculated by equation (2) and (1), side 19 and 18:

$$
\begin{aligned}
\boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\
\boldsymbol{\xi} \sim \boldsymbol{RD} &\sim \mathrm{Po}(\lambda \boldsymbol{I}) \\
\boldsymbol{\xi} &\sim \mathrm{Ge}(p) \\
\mathrm{Var}(\boldsymbol{\xi}) &= \frac{q}{p^2} = \boldsymbol{R_0}(1+\boldsymbol{R_0}) \\
\mathrm{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\
\mathrm{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}(1+\boldsymbol{R_0})}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.6 *Gamma distributed infectious times*

Dividing the expected time $1/\gamma$ into n separate parts. Each of these are exponentially distributed with rate $n\gamma$.

$$\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, \frac{1}{n\gamma}) \\
\psi_{\boldsymbol{I}}^x(t) &= (1 - \frac{t}{n\gamma})^{-nx} \\
\frac{d^l}{dt^l}\psi_{\boldsymbol{I}}^x(t) &= \frac{(nx+l-1)!}{(nx-1)!}(\frac{1}{n\gamma})^l(1 - \frac{t}{n\gamma})^{-(nx+l)} \\
\frac{d^l}{ds^l}\psi_{\boldsymbol{I}}^x(\lambda(t-1)) &= \frac{(nx+l-1)!}{(nx-1)!}(\frac{\lambda}{n\gamma})^l(1 - \frac{\lambda(t-1)}{n\gamma})^{-(nx+l)} \\
P(\boldsymbol{X} = x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(nx+(x-1)-1)!}{(nx-1)!}(\frac{\lambda}{n\gamma})^{x-1}(1 - \frac{\lambda(t-1)}{n\gamma})^{-(nx+(x-1))}\Big|_{t=0} \\
P(\boldsymbol{X} = x|\boldsymbol{R_0}) &= \frac{1}{x!}\frac{(x(n+1)-2)!}{(nx-1)!}(\frac{\boldsymbol{R_0}}{n})^{x-1}\frac{1}{(1+\frac{\boldsymbol{R_0}}{n})^{x(n+1)-1}}
\end{aligned}$$

When having a compound Poisson distribution with a gamma distribution, the resulting distribution will be a negative binomial.

$$\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, \frac{1}{n\gamma}) \\
\boldsymbol{Y} &\sim \boldsymbol{K}|\boldsymbol{I} = i \sim \mathrm{Po}(\lambda i) \\
\mathrm{P}(\boldsymbol{\xi} = \rho) &= \int_0^\infty \mathrm{P}(\boldsymbol{\xi} = \rho|\boldsymbol{I} = i)\mathrm{P}(\boldsymbol{I} = i)di \\
&= \int_0^\infty e^{-\lambda i}\frac{(\lambda i)^\rho}{\rho!}\frac{1}{\Gamma(n)}i^{n-1}(n\gamma)^n e^{-in\gamma}di = \ldots = \\
&= \frac{(n\gamma)^n\gamma^\rho}{\rho!\Gamma(n)}(\rho+n-1)!(\frac{1}{\lambda+n\gamma})^n(\frac{1}{\lambda+n\gamma})^\rho = \\
&= \frac{(\rho+n-1)!}{\rho!(n-1)!}(\frac{\lambda}{\lambda+n\gamma})^n(\frac{n\gamma}{\lambda+n\gamma})^\rho
\end{aligned}$$

Set $p = \frac{n\gamma}{\lambda+n\gamma}$. Using equation (2) and (1), side 19 and 18:

$$
\begin{aligned}
\boldsymbol{\xi} &\sim \text{NegBin(n, p)} \\
\text{E}[\boldsymbol{\xi}] &= n\frac{q}{p} = n\frac{\lambda}{n\gamma} \\
\text{Var}(\boldsymbol{\xi}) &= n\frac{q}{p^2} = n\frac{\frac{\lambda}{(\lambda+n\gamma)^2}}{\frac{n\gamma}{\lambda+n\gamma}} = \frac{\lambda(\lambda+n\gamma)}{n\gamma^2} = \boldsymbol{R_0}(\frac{\boldsymbol{R_0}}{n}+1) \\
\text{E}[\boldsymbol{X}] &= \frac{1}{(1-\boldsymbol{R_0})} \\
\text{Var}(\boldsymbol{X}) &= \frac{\boldsymbol{R_0}(\frac{\boldsymbol{R_0}}{n}+1)}{(1-\boldsymbol{R_0})^3}
\end{aligned}
$$

If $n = 1$ we will get the variance for the exponential distribution on the infectious time period. If $n = \infty$ we will get the variance for the constant infectious time period.

See Figure 3 for the progress and the increase in the number of infected as $\boldsymbol{R_0} \to 1-$.

4.7 *Comparison between different epidemic models*

All models show that the uncertainty in the number of infected rapidly increases when approaching the herd immunity threshold limit ($\boldsymbol{R_0} = \text{E}[\boldsymbol{\xi}] = 1$). In all cases the mean number will not differ. The variance is shifting between the models, see Figure 3. Assuming exponential distribution, shows the largest distribution in our case. This can be explained due to that some individuals have very long infectious periods.

[!ht]

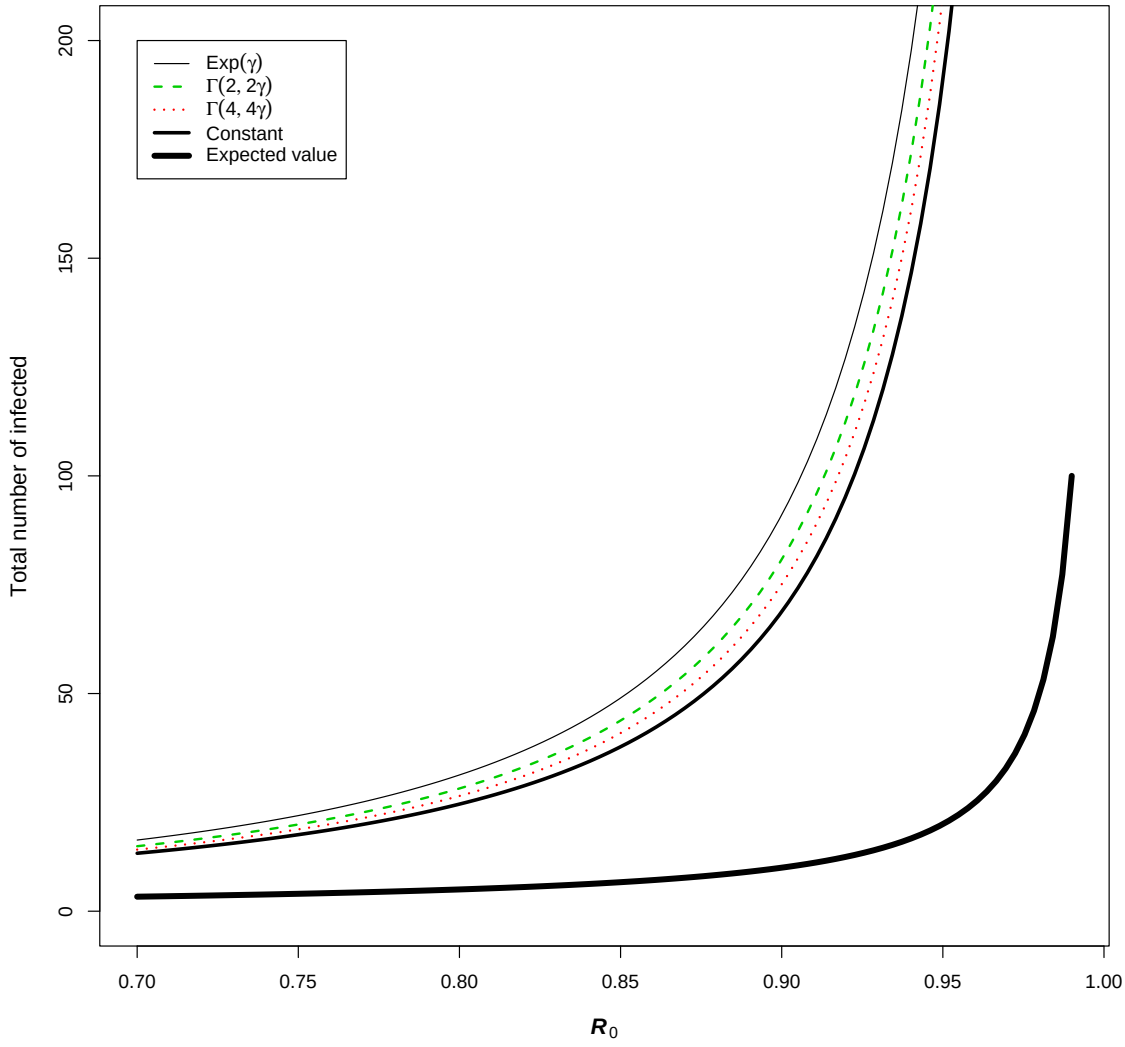

Figure 3. Number of infected as a function of $\boldsymbol{R_0}$, the mean value of the infectious time period is $1/\gamma$. Figure showing the expected value and 95% upper prediction interval.

5. Probability for a major outbreak when no herd immunity

5.1 *Background*

A common problem with different infectious diseases is that the population might not be protected by herd immunity. This can be caused by decreasing vaccination coverage. Or in the case of whooping cough, the vaccine is not efficient enough. Even if everyone would be vaccinated it would not be possible to obtain herd immunity.

When the herd immunity threshold has been reached, there will be a positive probability for a major outbreak. This can happen every time a new index case enters the population. The probability for a major outbreak can be calculated from branching process theory (Sewastjanow (1975)).

$$\begin{aligned} \boldsymbol{\xi} &= \text{The number of new infectious contacts} \\ \boldsymbol{\xi} &\sim \text{Po}(\lambda) \\ g_{\boldsymbol{\xi}}(t) &= \text{probability generating function for } \boldsymbol{\xi} \\ t &= g_{\boldsymbol{\xi}}(t) \end{aligned}$$

The smallest solution t is the probability for not having a major outbreak (epidemic).

Knowing the distribution of the infectious time period $\boldsymbol{I}$ gives a way to calculate the probability for not having a major outbreak.

We will calculate this in different epidemic models. The assumption is that the probability is for one index case. The probability generating function for $\boldsymbol{\xi}$ can be written as:

$$g_{\boldsymbol{\xi}}(t) = \mathrm{E}[t^{\boldsymbol{\xi}}] = \mathrm{E}[\mathrm{E}[t^{\boldsymbol{\xi}}|\boldsymbol{I}]] = \mathrm{E}[e^{-\lambda \boldsymbol{I}(1-t)}] = \mathrm{E}[e^{\lambda \boldsymbol{I}(t-1)}] = \Psi_{\boldsymbol{I}}(\lambda(t-1)) \quad (3)$$

Where $\Psi_{\boldsymbol{I}}(t)$ is the moment generating function for $\boldsymbol{I}$. In equation (3) $\boldsymbol{\xi} \sim \boldsymbol{RD}$, which is the basic reproductive distribution. See Figure 4 for different random distributions on $\boldsymbol{I}$.

[!ht]

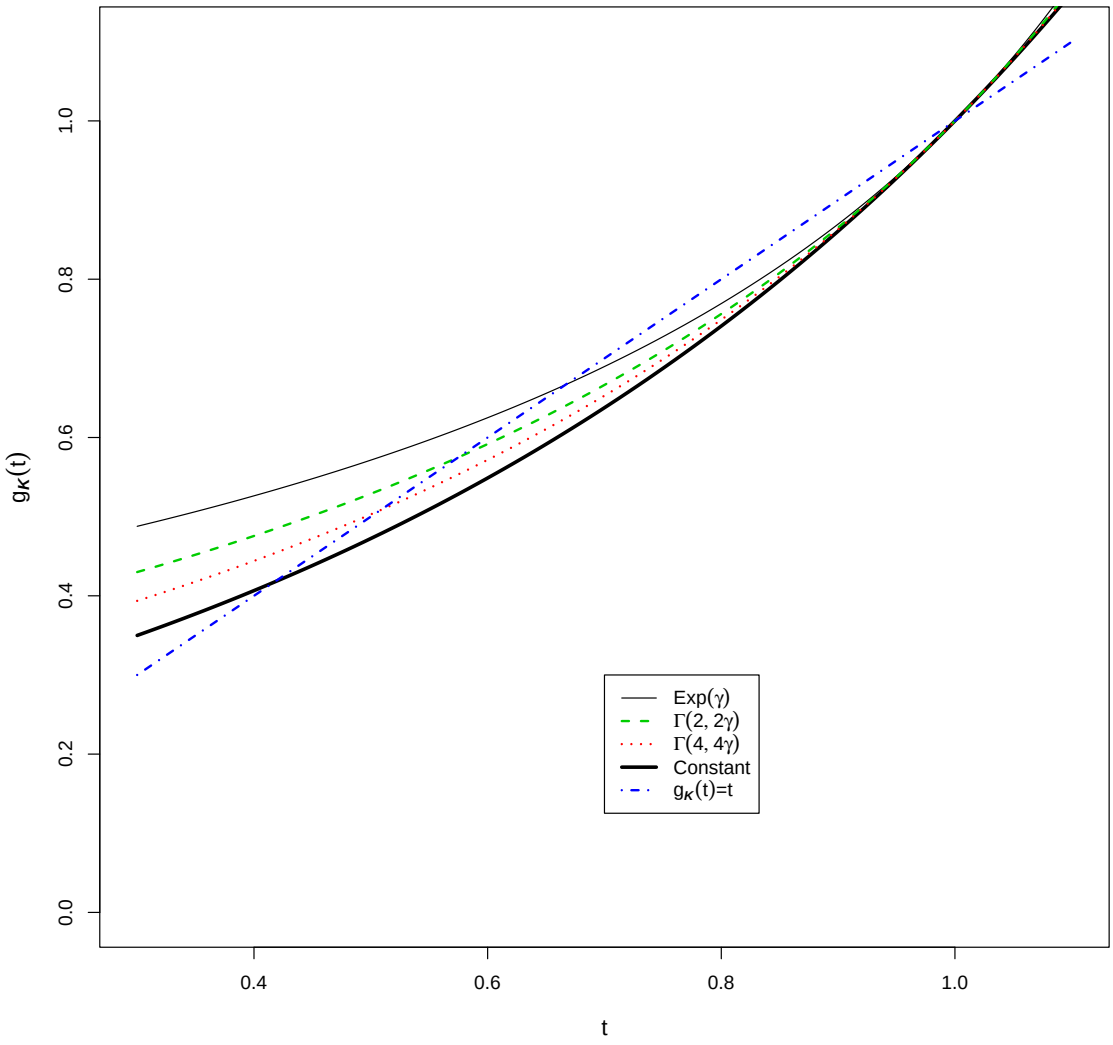

Figure 4. Probability for an outbreak depending on the random distribution of infectious times, $\boldsymbol{R_0} = 1.5$

5.2 *Constant value on the infectious time period*

Also known as the Reed-Frost model.

$$\begin{aligned} \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \Psi_{\boldsymbol{I}}(t) &= e^{\frac{\lambda(t-1)}{\gamma}} \\ t &= e^{\frac{\lambda(t-1)}{\gamma}} \\ s &= 1-t \\ 1-s &= e^{\frac{-\lambda s}{\gamma}} = e^{-\boldsymbol{R_0}s} \end{aligned}$$

Here s will denote the probability of having a major outbreak. It can be solved numerically for different values of $\boldsymbol{R_0}$.

5.3 *Exponentially distributed infectious time period*

Also known as the general epidemic model. The probability for a major outbreak will be less than for the case with a constant infectious time period.

$$\begin{aligned} \boldsymbol{I} &\sim \mathrm{Exp}(\gamma) \\ \mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\ \Psi_{\boldsymbol{I}}(t) &= \frac{1}{1-\frac{t}{\gamma}} = \frac{\gamma}{\gamma - t} \\ t &= \Psi_{\boldsymbol{I}}(\lambda(t-1)) \\ t &= \frac{\gamma}{\gamma - \lambda t + t} \\ t &= \frac{\gamma}{\lambda} \\ s &= 1-t \\ s &= \frac{\lambda-\gamma}{\lambda} = \frac{\boldsymbol{R_0}-1}{\boldsymbol{R_0}} = 1 - \frac{1}{\boldsymbol{R_0}} \end{aligned}$$

Here s will denote the probability of having a major outbreak. The general epidemic model gives an easy way to calculate the probability for a major outbreak.

5.4 *Gamma distributed infectious time period*

$$
\begin{aligned}
\boldsymbol{I} &\sim \Gamma(n, n\gamma) \\
\mathrm{E}[\boldsymbol{I}] &= \frac{1}{\gamma} \\
\mathrm{Var}[\boldsymbol{I}] &= \frac{1}{n\gamma^2} \\
\Psi_{\boldsymbol{I}}(t) &= \frac{1}{(1-\frac{t}{n\gamma})^n} \\
t &= \Psi_{\boldsymbol{I}}(\lambda(t-1)) \\
&\vdots \\
1 &= t^{1/n}\Big(1 - \frac{\lambda(t-1)}{n\gamma}\Big) \\
1 &= t^{1/n}\Big(1 - \frac{\boldsymbol{R_0}(t-1)}{n}\Big) \\
s &= 1 - t \\
1 &= (1-s)^{1/n}\Big(1 + \frac{\boldsymbol{R_0}s}{n}\Big)
\end{aligned}
$$

Here s will denote the probability of having a major outbreak. It can be solved numerically for different values of $\boldsymbol{R_0}$.

5.5 *Comparison between the different epidemic models*

The probability for a major outbreak given by the Reed-Frost, gamma distributed infectious time period and the general epidemic model are shown in Figure 5. For low values in $\boldsymbol{R_0}$ the difference can be remarkably large. For example with $\boldsymbol{R_0} = \lambda \mathrm{E}[\boldsymbol{I}] = 2$, the general epidemic has a probability of 0.50 while the Reed-Frost has 0.80. The ordering seems to be that constant distribution > gamma distribution > exponential distribution, when it comes to the probability for having a major outbreak.

5.6 *Order of probabilities for major outbreaks with different random distributions on the infectious time periods*

Assume that $\boldsymbol{R_0}$ has the same value in all models.

[!ht]

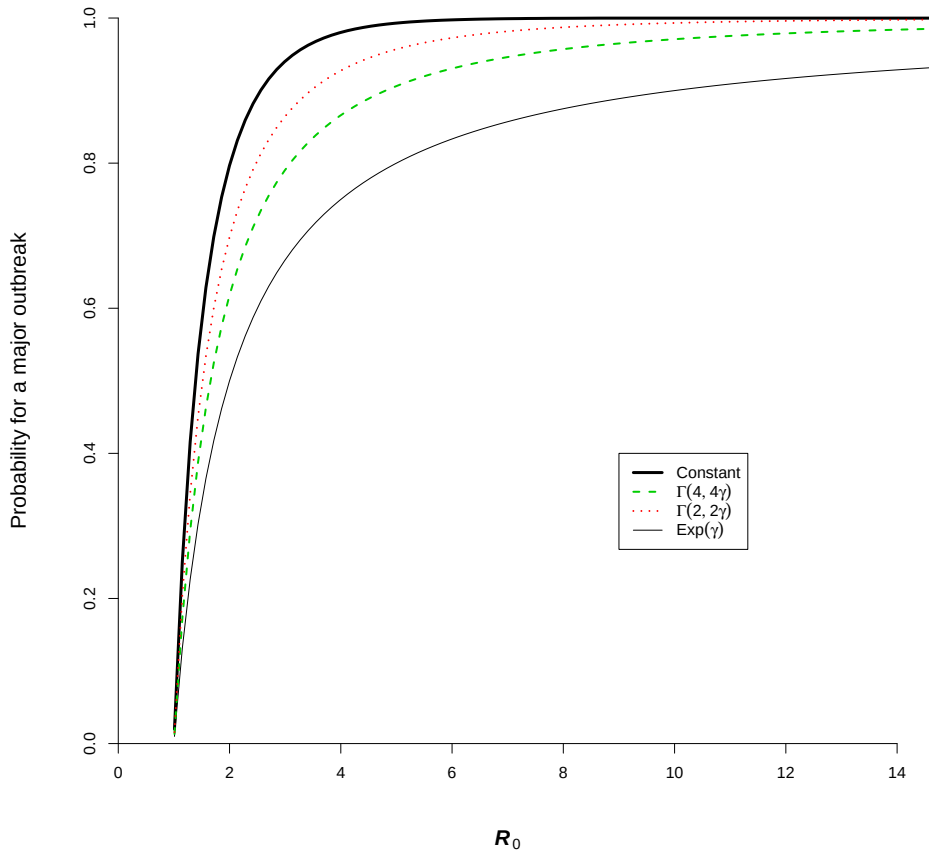

Figure 5. Probability for a major outbreak as a function of $\boldsymbol{R_0}$

Let f be a convex function and $\boldsymbol{U}$ be a stochastic variable.

$$\begin{aligned}
\text{Jensen's inequality} &\Rightarrow f(\mathrm{E}[\boldsymbol{U}]) \leq \mathrm{E}[f(\boldsymbol{U})] \\
\text{Let } f(t) &= e^{-\lambda t} \\
e^{-\lambda \mathrm{E}[\boldsymbol{I}]} &\leq \mathrm{E}[e^{-\lambda \boldsymbol{I}}] \\
\mathrm{P}(\text{Infected during time period } \boldsymbol{I}) &= 1 - e^{-\lambda \boldsymbol{I}} \\
1 - \mathrm{E}[e^{-\lambda \boldsymbol{I}}] &\leq 1 - e^{-\lambda \mathrm{E}[\boldsymbol{I}]}
\end{aligned}$$

In the constant case there is an equality while for the exponential and gamma distributions on the infectious time period there can be an inequality. This gives that the constant distribution has the highest probability for a major outbreak.

Next we show that the probability for a major outbreak with the general epidemic model, is less or equal to the case with a gamma distribution with an integer as the form parameter. This can be shown by methods from (Daley (1990)). $\phi_{\boldsymbol{I}}(t)$ is the Laplace transform for $\boldsymbol{I}$, $\phi_{\boldsymbol{I}}(t) = \mathrm{E}[e^{-t\boldsymbol{I}}]$. Let p_{mo} = probability for a major outbreak.

$$\begin{aligned}
\boldsymbol{I}_1 \sim \mathrm{Exp}(\gamma) \qquad & \qquad \boldsymbol{I}_2 \sim \Gamma(n, n\gamma) \\
\phi_{\boldsymbol{I}_1}(t) = \frac{1}{1+\frac{t}{\gamma}} \qquad & \qquad \phi_{\boldsymbol{I}_2}(t) = \frac{1}{(1+\frac{t}{n\gamma})^n} \\
\phi_{\boldsymbol{I}_1}(t) \quad &\geq \quad \phi_{\boldsymbol{I}_2}(t) \text{ , since } \frac{\phi(\boldsymbol{I}_1)}{\phi(\boldsymbol{I}_2)} = \frac{(1+\frac{t}{n\gamma})^n}{1+\frac{t}{\gamma}} \\
&\Longrightarrow \\
\boldsymbol{I}_1 \quad &\leq_L \quad \boldsymbol{I}_2 \\
&\Longrightarrow \\
p_{mo}(\boldsymbol{I}_1) \quad &\leq \quad p_{mo}(\boldsymbol{I}_2)
\end{aligned}$$

For $n = 1$ there is equality, since it gives the exponential distribution. While for $n > 1$ there is an inequality. Showing that gamma distribution has greater or equal probability for an outbreak than the exponential distribution.

Letting $n \to \infty$ gives $\frac{\phi(\boldsymbol{I}_1)}{\phi(\boldsymbol{I}_2)} = \frac{e^{\frac{t}{\gamma}}}{1+\frac{t}{\gamma}}$.

6. Time until a major outbreak

In Chapter 5 we calculated the probability for having a major outbreak. In real life the probability is not always the most wanted concept. People are more interested in what time span outbreaks might happen. This giving aid

in the planning procedures, such as producing the necessary amounts of anti viral or vaccine doses.

Assume that there is a flow of imported new infectious individuals (index cases) $\boldsymbol{K_i}$ who arrive according to a Poisson process with rate λ_i. If the population is protected by herd immunity, ($\boldsymbol{R_0} < 1$), this contributes with a probability 0 a.s. for a major outbreak. If $\boldsymbol{R_0} > 1$, there is a probability $k(\boldsymbol{R_0}(t))$ for a major outbreak. This at a certain time point t, where index cases enters the population.

The probability function $k(\boldsymbol{R_0}(t))$ has been calculated in Chapter 5. Let $\boldsymbol{T}$ denote the time point for a major outbreak.

$$
\begin{aligned}
\boldsymbol{R_0}(t) < 1 &\Leftrightarrow k(\boldsymbol{R_0}(t)) = 0. \\
\mathrm{P}(\mathrm{t} \leq \boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) &= \lambda_i h \mathrm{P}(\boldsymbol{T} > \mathrm{t}) \mathrm{k}(\boldsymbol{R_0}(\mathrm{t})) \\
\frac{\mathrm{P}(\boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) - \mathrm{P}(\boldsymbol{T} \leq \mathrm{t})}{h} &= \lambda_i \mathrm{P}(\boldsymbol{T} > \mathrm{t}) \mathrm{k}(\boldsymbol{R_0}(\mathrm{t})) \\
\lim_{h \to 0} \frac{\mathrm{P}(\boldsymbol{T} \leq \mathrm{t}+\mathrm{h}) - \mathrm{P}(\boldsymbol{T} \leq \mathrm{t})}{h} &\to \mathrm{P}'(\mathrm{t}) \\
\mathrm{P}(\boldsymbol{T} > \mathrm{t}) &= 1 - \mathrm{P}(\mathrm{t}) \\
\frac{\mathrm{P}'(\mathrm{t})}{1 - \mathrm{P}(\mathrm{t})} &= \lambda_i k(\boldsymbol{R_0}(t)) \\
\ln(1 - \mathrm{P}(\mathrm{t}))' &= -\lambda_i k(\boldsymbol{R_0}(t)) \\
\ln(1 - \mathrm{P}(\mathrm{t})) &= -\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) \mathrm{d}s \\
\mathrm{P}(\mathrm{t}) &= 1 - e^{-\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) ds} \\
\mathrm{P}(\mathrm{t}) &= \mathrm{F}_{\boldsymbol{T}}(t) \\
f_{\boldsymbol{T}}(t) &= \lambda_i k(\boldsymbol{R_0}(t)) e^{-\lambda_i \int_0^t k(\boldsymbol{R_0}(s)) ds} \qquad (4)
\end{aligned}
$$

Looking at the expression for P(t), shows that the probability for a major outbreak increases when $k(\boldsymbol{R_0})$ increases. This is also what to expect. Letting $k(\boldsymbol{R_0})$ increase is similar to that $\boldsymbol{R_0}$ is increasing. Knowing the function $k(\boldsymbol{R_0}(t))$, gives now a way to calculate the probability for a major outbreak at any given time point.

6.1 *Time for an outbreak when $\boldsymbol{R_0}$ is constant*

Assume that $\boldsymbol{R_0}$ will be constant for all time points and $\boldsymbol{R_0} > 1$. It is possible to give an analytic expression the expected time $\boldsymbol{T}$ for a major outbreak and its variance. Equation (4) shows this is the density function for an

exponential distribution with rate $\lambda_i k(\boldsymbol{R_0}(t))$, where $k(\boldsymbol{R_0}(t))$ is constant.

$$\begin{aligned} \boldsymbol{T} &\sim \text{Exp}(\lambda_i k(\boldsymbol{R_0}(t))) \\ \text{E}[\boldsymbol{T}] &= \frac{1}{\lambda_i k(\boldsymbol{R_0}(t))} \\ \text{Var}[\boldsymbol{T}] &= \frac{1}{(\lambda_i k(\boldsymbol{R_0}(t)))^2} \end{aligned}$$

Depending on which assumptions are on the infectious period of time $\boldsymbol{I}$, we can make comparisons between different models.

Figure 6 shows how the distribution of time depends on the value of $\boldsymbol{R_0}$. Figure 7 shows the expected time and the 95% prediction interval for a major outbreak with $\boldsymbol{R_0} = 1.1, 2, 4, 10$. For small values on $\boldsymbol{R_0}$ there is a certain difference.

6.2 *Time for an outbreak when $\boldsymbol{R_0}$ is not constant*

In this situation there rarely exists an exact solution, instead the calculations will be done numerically.

Two typical situations arise with this setting. First situation is diseases like measles which in many countries is controlled by herd immunity. But decreased vaccination coverage can make $\boldsymbol{R_0} > 1$. The second situation is when a new unknown disease enters the population. For example SARS or pandemic influenza. With diseases like these $\boldsymbol{R_0}$ might be far from 1 already at the beginning.

To answer these questions of when after crossing the herd immunity threshold level a major outbreak can occur, two different scenarios are calculated with different assumptions on the development of $\boldsymbol{R_0}$. Assume that one index case enters the population each year. In the first scenario $\boldsymbol{R_0}$ increases with 0.01 units each year, see Figure 8. The second scenario is when $\boldsymbol{R_0}$ first increases, followed by a decrease, see Figure 9.

7. Results and Discussion

When using different epidemic models it is important to consider how the infectious time period is distributed. Both for the case of emerging new infections and for future outbreaks of diseases known which are known today.

The number of infected when the population is protected by herd immunity can also be considerably high. This shows the danger of using deterministic models which will predict the expected value, and this will be very low.

[!ht]

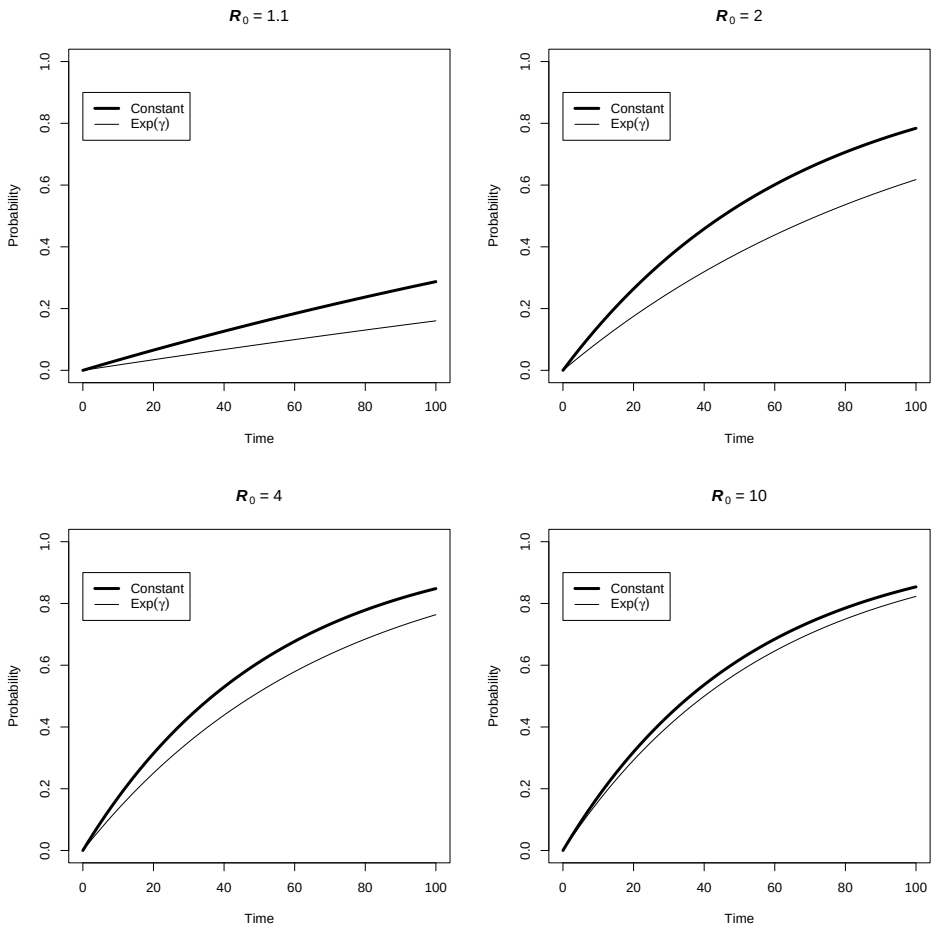

Figure 6. Probability for a major outbreak as a function of $\boldsymbol{R_0}$. The index cases are arriving with a rate of 1/52. This is equal to one new case arriving each year if the time scale is in weeks.

[!ht]

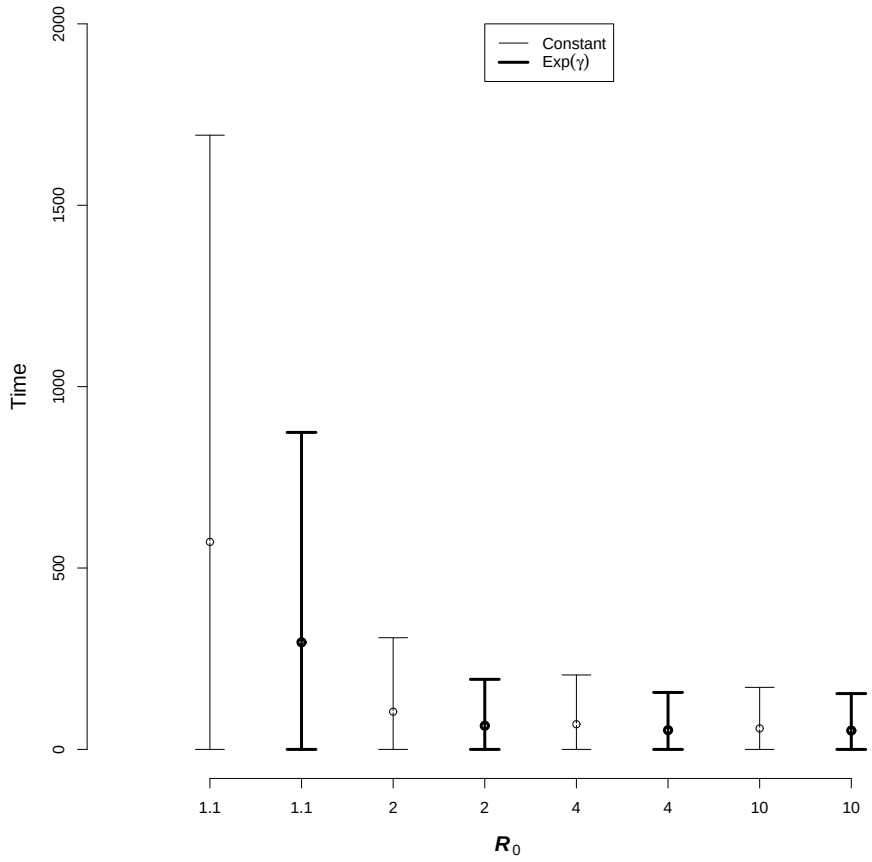

Figure 7. Time until a major outbreak as a function of $\boldsymbol{R_0}$, when $\boldsymbol{R_0}$ is constant. Index cases are arriving at a rate of $1/52$ each week.

[!ht]

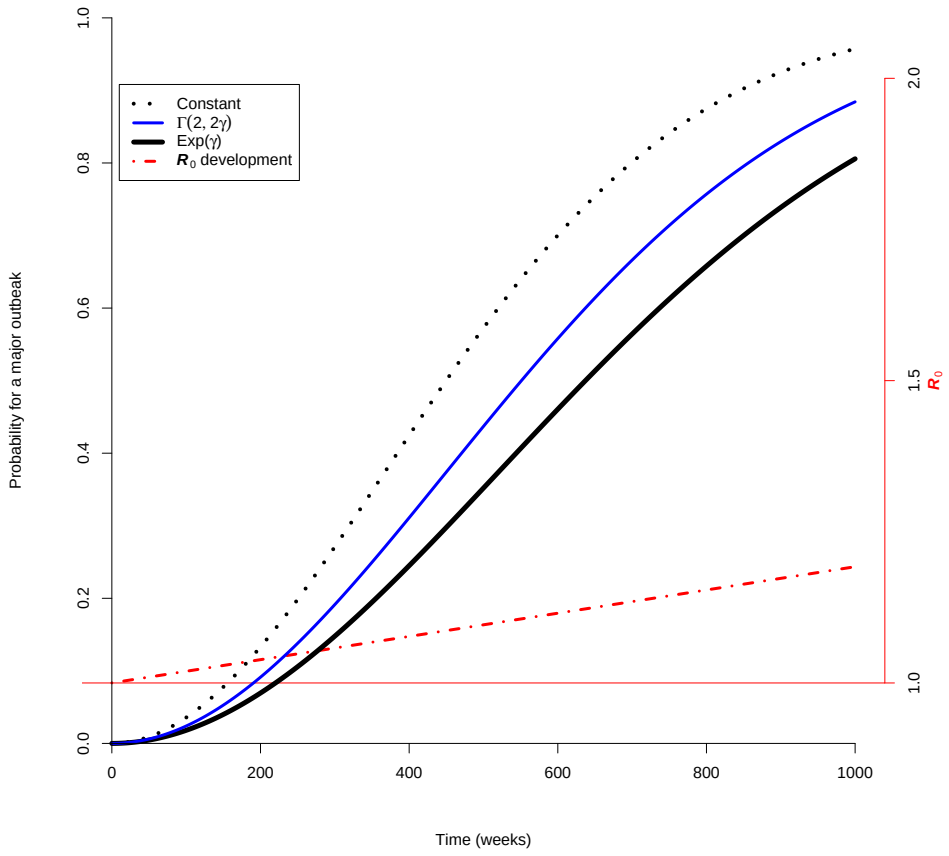

Figure 8. Time until a major outbreak as $\boldsymbol{R_0}$ increases with 0.01 units each year.Index cases are arriving at a rate of $1/52$ each week.

[!ht]

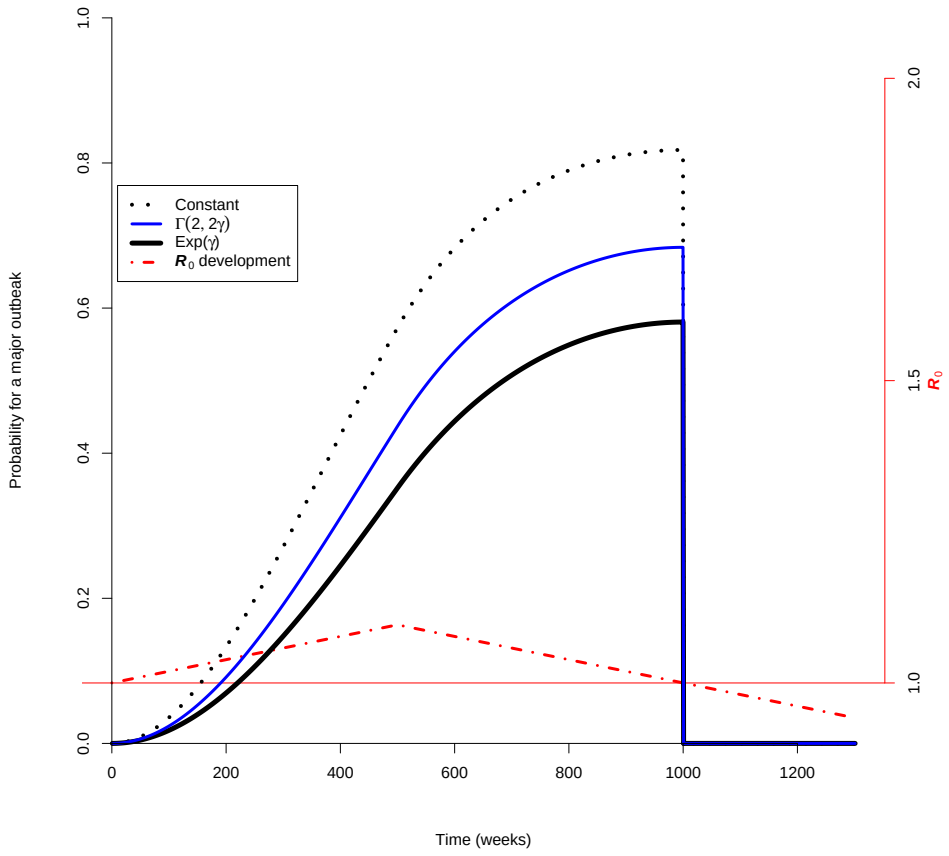

Figure 9. Time until a major outbreak when $\boldsymbol{R_0}$ increases followed by a decrease. Index cases are arriving at a rate of $1/52$ each week.

The time for major outbreaks when crossing the herd immunity level is very sensitive for the assumption on the distribution of the infectious time period.

There are limitations in this report. One is that the population is assumed to be homogeneously mixing. The aim is to concentrate on the random distribution on the infectious time period. Constant, exponentially distributed and gamma distributed time periods have been used. Although the methods can be used for other distributions.

In the future other distributions will be considered. Also the case with different heterogeneity in the population will be considered.

8. Acknowledgments

This work has been financially supported by DG Sanco project MODELREL, Swedish Institute for Infectious Disease Control, Karolinska Institute and Martin Rinds foundation. Thanks also to the epidemic modeling PhD students Elinore Bengtsson and Patricia Geli at Stockholm University for the support to get this report ready. Thanks also to all the staff at Department of epidemiology at the Swedish Institute for Infectious Disease Control for support and help. And making me aware that the problems addressed in this report are based on real problems. Thanks also to my two supervisors, Åke Svensson at Stockholm University and the Swedish Institute for Infectious Disease Control and Johan Giesecke from the European Centre for Disease Prevention and Control.

References

Abbey, H. (1952). An examination of the Reed-Frost theory of epidemics. *Human Biology* **24**, 201–33.

Asikainen, T. (2005). Lessons learned from a smallpox outbreak in stockholm 1963. Technical Report 2005:10, Department of mathematics.

Asikainen, T., Giesecke, J. and Svensson, Å. (2003). Reduced vaccination will result in epidemics, about 25 000 cases are expected to occur annualy within a 25 year period. *Läkartidningen* **40**, 3126–30.

Borel, E. (1942). Sur l'emploi du theoreme de bernoulli pour faciliter le calcul d'un infinite de coefficients. application au probleme de l'attente a un guichet. *C. R. Acad. Sci., Paris* **214**, 452–56.

Daley, D. (1990). The size of epidemics with variable infectious periods. Technical Report SMS-012-90, Statistics Research Section, School of Mathematical Sciences, Australian National University.

Farrington, C., Kanaan, M. and Gay, N. (2003). Brancing process mod-

els for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* **4**, 279–95.

Gani, R. and Leach, S. (2001). Transmission potential of smallpox in contemporary populations. *Nature* **414**, 748–751.

Giesecke, J. (2002). *Modern Infectious Disease Epidemiology*. Arnold, 338 Euston Road, London NW1 3BH, second edition.

Height, F. and Breuer, M. (1960). The Borel-Tanner distribution. *Biometrika* **47**, 143–50.

Heymann, D. L., editor (2004). *Control of communicable diseases manual*. APHA, 18 edition.

Jain, G. C. and Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics* **21**, 501–513.

Kermack, W. and McKendrick, A. (1927). Contributions to the mathematical theory of epidemics, part i. *Proc. Roy. Soc. Edin.* **115**, 700–21.

Kotz, S. and Johnson, N. L. (1983). *Encyclopedia of statistical sciences*, volume 4. Wiley.

Pounder, R., Wakefield, A., Thompson, N. and et al, S. M. (1995). Is measles vaccination a risk factor for inflammatory bowel disease. *Lancet* **34**, 1071–4.

Sewastjanow, B. (1975). *Verzweigungsprozessse*. Akademie-Verlag, 108 Berlin, Leipziger Str. 3-4, second edition.

Ström, J., editor (1964). *Akuta infektionssjukdomar*. Bröderna Ekstrands Tryckeri AB.

Tanner, J. (1953). A problem of inference between two queues. *Biometrika* **40**, 58–69.

Wakefield, A., Munch, S., Anthony, A., Linnell, J., Casson, D. and et al, M. B. (1998). Heallumphoid-nodular, non-specific colitis and pervasive developmental disorder in children. *Lancet* **251**, 637–41.