Stockholms universitet

Introduction to stochastic epidemic models

Tom Britton

May 2023

References: e.g. textbooks
Andersson and Britton (2000) and
Diekmann et al (2013)

Stockholms universitet

Mathematical models

Aim of mathematical modelling: To describe some real world phenomenon mathematically in order to learn more about it

Main idea: Mathematical models describes some feature in a *simplified way*, keeping only the essential features

Trade-off between simple and complicated models: Simple models are easier to understand but don't mimick reality very well. Complicated models are harder to analyse and contain many parameters which may be hard to estimate

Stochastic models:

The discrepancy between model and reality may be contained in "random part" in model

Stochastic models enable uncertainty estimates (i.e. standard errors) when estimating parameters

Stockholms universitet

Background: Infectious disease models

We want to model the spread of a transmittable disease in a community of individuals

At a given time-point an individual may be *Susceptible*, infected but not yet infectious (*Latent* or *Exposed*), *Infectious*, or recovered and immune (*Removed*)

Different class of epidemic models: SIR, SEIR, SIS, SIRS, ...

Main focus: SIR (childhood diseases, STDs, influenza, covid-19...)

Short term outbreak vs endemic situation

Simplification for short term: fixed population, no waning immunity

Stockholms universitet

Notation

Some notation to be used

- $n = \#$ individuals ($n(t)$ if varying over time)
- $S(t) = \#$ "susceptibles" (susceptible individuals) at time t
- $I(t) = \#$ "infectives" (infectious individuals) at time t
- $R(t) = \#$ "removeds" (removed individuals) at time t
- $T =$ the time when the epidemic stops
- Z $(= R(T) - 1) = \#$ infected during the epidemic (excluding index case). Possible values: $0,1,\ldots,n-1$.

We start with the simplest situation: all individuals are "identical" (with respect to disease spreading) and all pairs of individuals have contact at equal rates.

Homogeneous community that mixes uniformly

The Reed-Frost stochastic epidemic model

Short term outbreak (fixed community), homogeneous community, uniform mixing, SIR, discrete time: "generations"

An epidemic model (Reed-Frost, 1928)

- Assume 1 index case (externally infected) the rest $n-1$ susceptible
- Anyone who gets infected infects other susceptibles independently with prob p and then recovers
- A recovered individual plays no further role in epidemic

The index case infects a random number ($\text{Bin}(n-1,p)$) of individuals, they in turn infect an additional random number, and so on. Once no new individuals are infected the epidemic stops

Think in "generations"

Stockholms universitet

Exercise 1

Suppose $n = 3$ (one index case and 2 susceptibles) and $p = 0.2$

Possible values for Z: 0,1,2.

$P(Z = 0)$? For this to happen the index can't infect anyone

$P(Z = 1)$? For this to happen the index must infect EXACTLY one AND this individual cannot infect anyone further

$P(Z = 2)$? Either the index infects exactly one AND this individual infects the last one, OR the index infects both

Stockholms universitet

Exercise 1

Suppose $n = 3$ (one index case and 2 susceptibles) and $p = 0.2$

Possible values for Z: 0,1,2.

$P(Z = 0)$? For this to happen the index can't infect anyone

$P(Z = 1)$? For this to happen the index must infect EXACTLY one AND this individual cannot infect anyone further

$P(Z = 2)$? Either the index infects exactly one AND this individual infects the last one, OR the index infects both

$P(Z = 0) = (1 - p)^2 = 0.64$
$P(Z = 1) = \binom{2}{1} p(1 - p) \times (1 - p) = 0.256$
$P(Z = 2) = \binom{2}{1} p(1 - p) \times p + p^2 = 0.104$
or ... $P(Z = 2) = 1 - P(Z = 0) - P(Z = 1)$

Stockholms universitet

What about larger communities?

General n, think in "generations"

Epidemic chains: $i \to 3 \to 2 \to 0$: the index infects 3, they infect 2 and these infect no further and the epidemic stops

$$P(Z = 0) = P(i \to 0) = (1-p)^{n-1}$$

$$P(Z = 1) = P(i \to 1 \to 0) = \binom{n-1}{1} p^1 (1-p)^{n-2} \times (1-p)^{n-2}$$

$$P(Z = 2) = P(i \to 2 \to 0) + P(i \to 1 \to 1 \to 0) = \ldots$$

$$P(Z = 3) = P(i \to 3 \to 0) + P(i \to 2 \to 1 \to 0) + P(i \to 1 \to 2 \to 0) + P(i \to 1 \to 1 \to 1 \to 0) = \ldots$$

$P_n(Z = z)$ gets very complicated when $n \geq 10$ and $z \geq 5$.

Underlying reason for the complication: individuals' outcomes are **dependent**! (As opposed to other diseases)

What to do then?

Stockholms universitet

Approximations when n large

When n large then often p (=per individual transmission probability) is small.

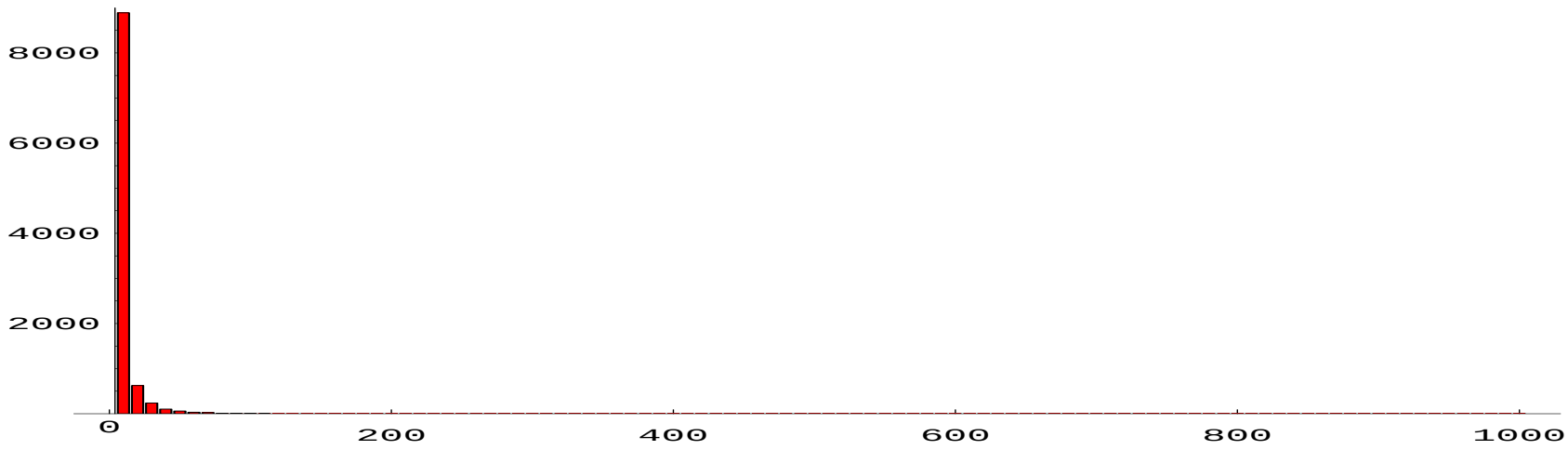
Expected number of infectious contacts: $(n-1)p \approx np =: R_0$

R_0 = basic reproduction number

Next page: Histogram of final outbreak sizes from 10 000 simulations in a community of $n = 1000$ individuals (both $R_0 < 1$ and $R_0 > 1$)

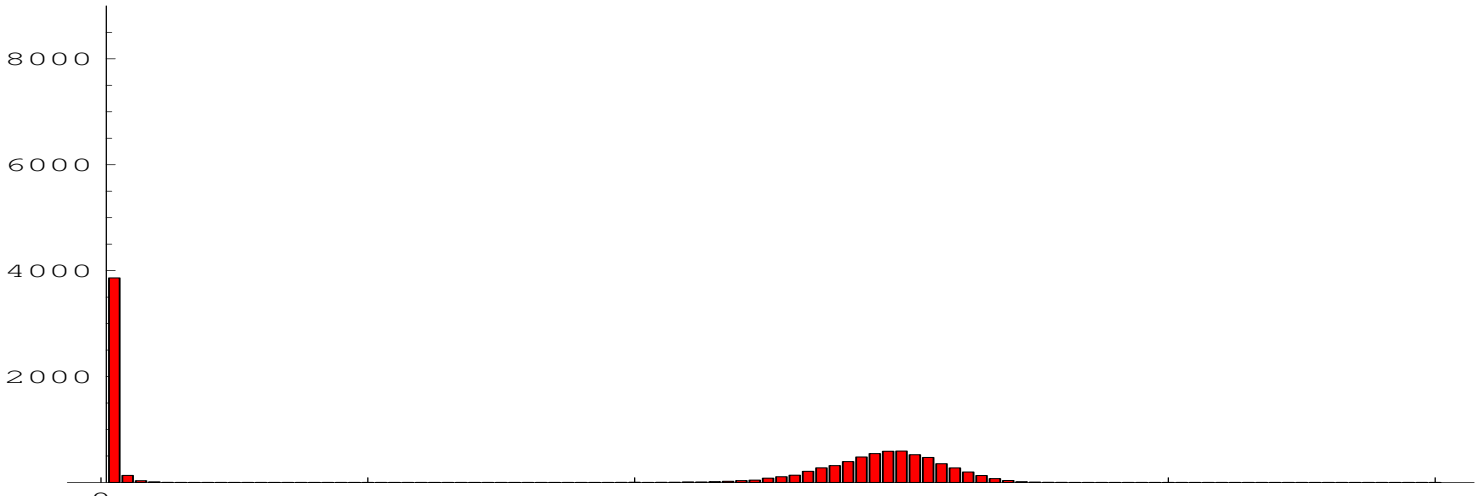
Stockholms universitet

Histogram of final size: $R_0 = 0.8$



Stockholms universitet

Histogram of final size: $R_0 = 1.5$



Stockholms universitet

An approximation for the final size

$R_0 = 1$ is "threshold value"

We now derive an equation for τ heuristically (recall $p = R_0/n$)

Assume n large and let $\tau = Z/n =$ final *fraction* infected

$$
\begin{aligned}
1 - \tau &= \text{proportion not infected} && (1)\\
&\approx \text{ probability not get infected} && (2)\\
&= \text{prob to escape inf from all infected} && (3)\\
&= (1-p)^Z && (4)\\
&= \left(1 - \frac{R_0}{n}\right)^{n\tau} && (5)\\
&\approx e^{-R_0\tau} \qquad \text{(using that } (1-x/n)^n \approx e^{-x}\text{)} && (6)
\end{aligned}
$$

Stockholms universitet

Approximation for final size

τ should hence (approximately) solve

$$1 - \tau = e^{-R_0 \tau}$$

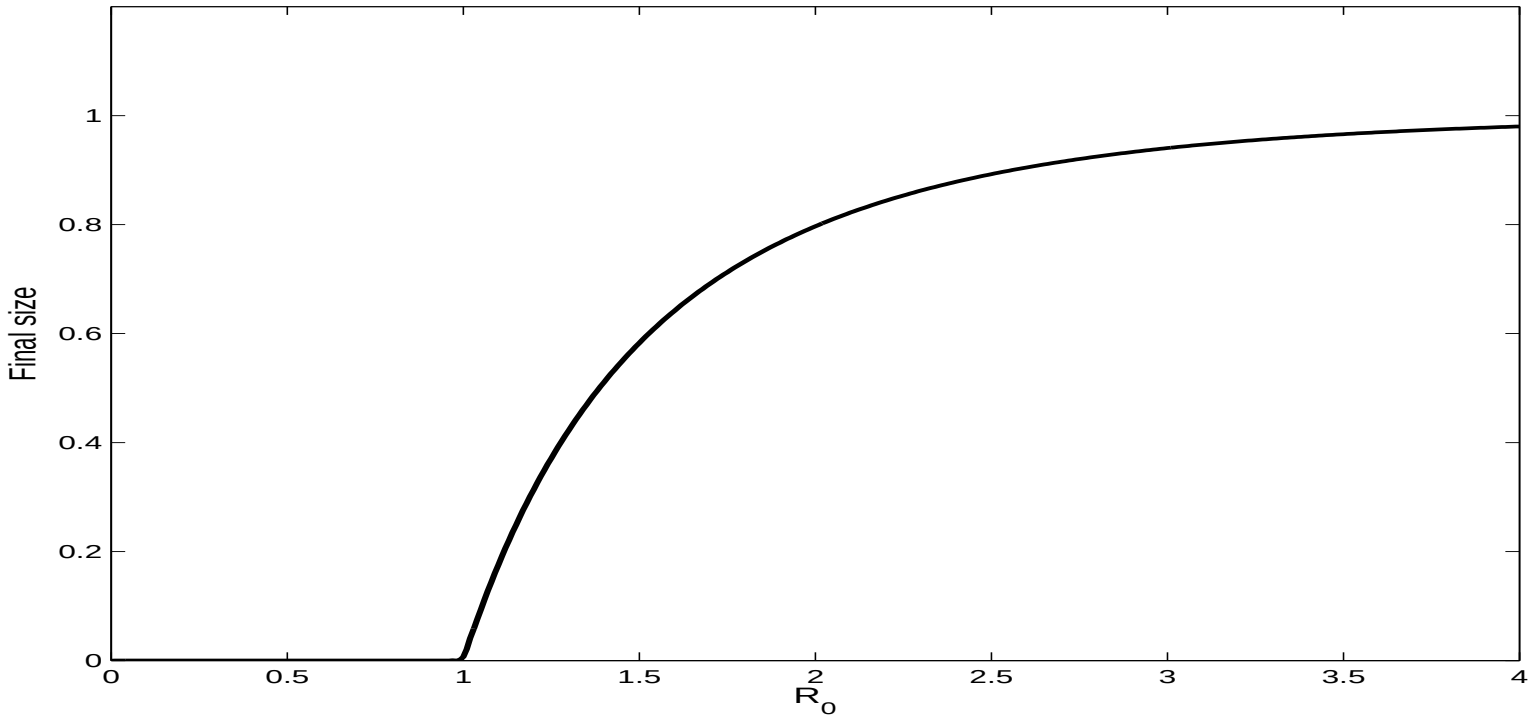
There are two solutions: $\tau = 0$ and (if $R_0 > 1$): $\tau = \tau^\star > 0$.

Exercise 2 Compute τ^* numerically when $R_0 = 1.5$, 3 and 6.

On next page is a plot of final size as function of R_0

Stockholms universitet

Plot of final outbreak size as function of R_0



Stockholms universitet

Approximation, cont'd

Strong dichotomy: minor outbreak – major outbreak

$P(\text{major outbreak}) = 1 - P(\text{minor outbreak})$ can be determined using *branching process* theory (random graph theory):

Final size = size of connected component of a randomly selected node in an Erdös-Renyi random graph

$\Longrightarrow P(\text{major outbreak}) = \tau^* = \text{relative size of giant !!!}$

CLT for major outbreak: $\sqrt{n}\left(\frac{Z}{n} - \tau^*\right) \approx N(0, \sigma^2)$

σ^2 depends on model parameters

Estimation: $1 - z = e^{-R_0 z} \Longleftrightarrow R_0 = -\log(1-z)/z$

So if outbreak size $\tilde{z}$ observed $\hat{R}_0 = -\log(1-\tilde{z})/\tilde{z}$

+ explicit st.err. from CLT

Stockholms universitet

What about epidemic over time?

A related stochastic epidemic model (the "General stochastic epidemic") can be defined in continuous time:

- During the infectious period an individual has "infectious contacts" randomly in time at the average rate β, each time individual is chosen randomly
- A susceptible who receives an infectious contact becomes infectious and remains so for a exponentially distributed time with mean $1/\gamma$ (other contacts have no effect)

Fundamental difference to Reed-Frost: Infectious period random implies that infection events from an individual become dependent!
$\Longrightarrow$ undirected E-R random network no longer applicable

$R_0 =$ expected number of infectious contacts $= \beta/\gamma$

Stockholms universitet

What about epidemic over time?

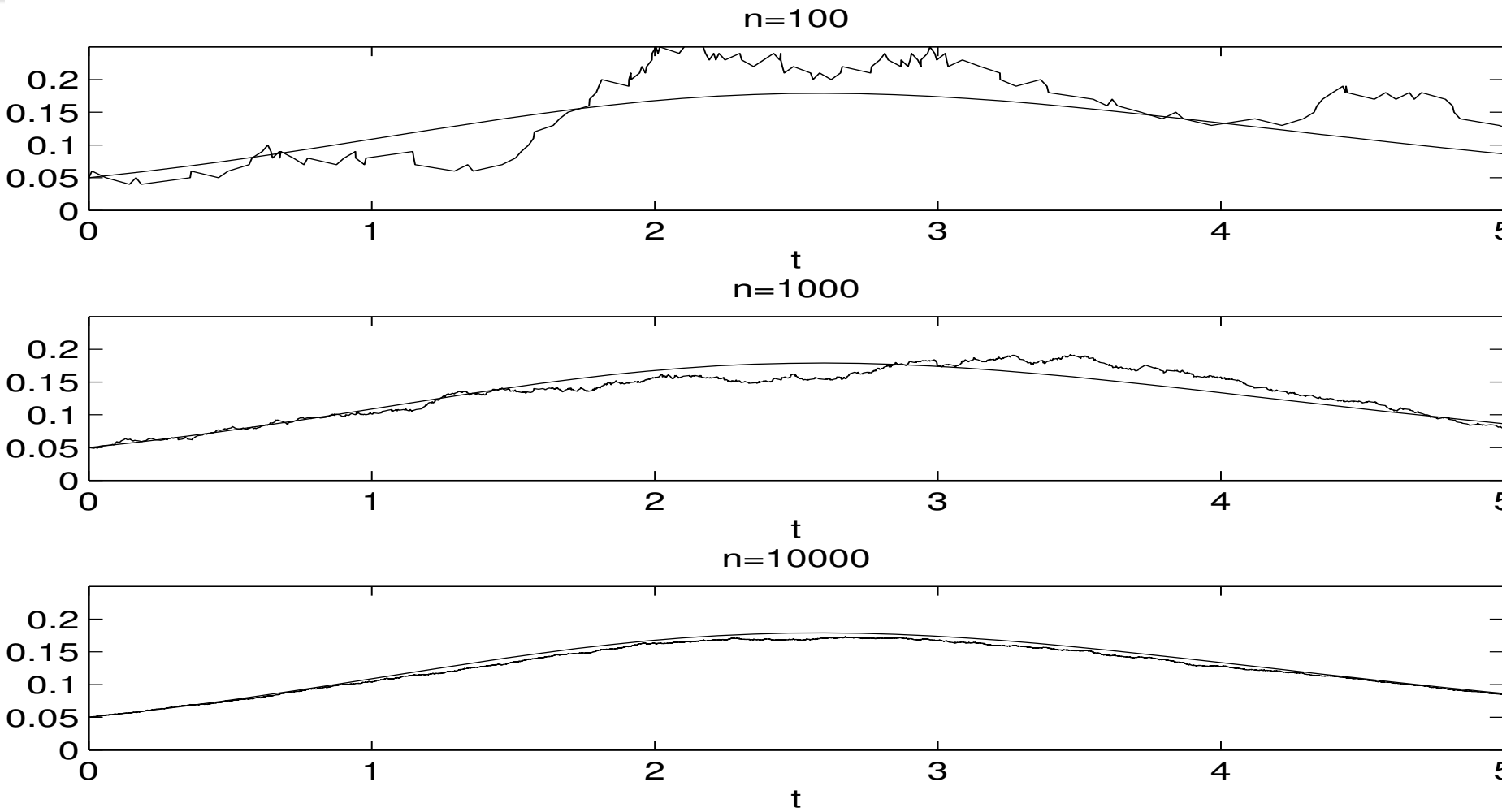
When n is large the process $(S(t)/n, I(t)/n)$ is close to deterministic limit $(s(t), i(t))$ which solves differential system

$$s'(t) = -\beta s(t) i(t) \tag{7}$$
$$i'(t) = \beta s(t) i(t) - \gamma i(t) \tag{8}$$
$$r'(t) = \gamma i(t) \tag{9}$$

Next page: plot of $I(t)/n$ for one (typical) simulated epidemic and deterministic limit $i(t)$, for a few different n

Stockholms universitet

Plots of simulated stochastic epidemic and deterministic curve



Stockholms universitet

Beginning of outbreak

Infectious individuals infect new individual at rate $\beta * (S(t)/n)$ and recover at rate γ

In beginning of outbreak in large community $S(t)/n \approx 1$, so more or less constant and equal rate for all infectives!

Infecting $\rightarrow$ "give birth", recover $\rightarrow$ "die" $\Longrightarrow$ branching process paradigm

$\Longrightarrow$ exponential growth rate r which solves Euler-Lotka equation

$$\int_0^\infty e^{-rs} g(s) ds = \frac{1}{R_0}$$

where $g(s) =$ is the *generation time ditribution* ($g(s) = \gamma e^{-\gamma s}$ for this model)

Estimation: If we know $g(s)$ (or estimate from other data source) and observe "early" growth rate r Euler-Lotka can be used to estimate R_0!

Stockholms universitet

Initial growth rate

The growth rate parameter $\rho(= r)$ is called the **Malthusian parameter** and depends both on R_0 and the generation time distribution $g(s)$. Branching process theory: ρ is the solution to the Euler-Lotka equation

$$R_0 \int_0^\infty e^{-\rho s} g(s) ds = 1$$

So if we know the generation time distribution $g(\cdot)$ we can estimate R_0 from observing the exponential growth ρ!

It is easy to show that if $g(s) \sim \Gamma(\alpha, \beta)$ then Euler-Lotka gives that

$$R_0 = \left(\frac{\rho}{\beta} + 1 \right)^\alpha$$

Stockholms universitet

Covid-19: R_0 estimates, **first wave** (original strain)

Covid-19: A common estimate is that $g(s) \sim \Gamma$ with mean 6.5 days and s.d. 4 days. We assume this to apply to all countries!

We estimate "country" specific ρ from reported cumulative case fatalities: starting first day with > 50 cumulative case fatalities (C_1) and two weeks later C_{15} case fatalities: $\hat{\rho} = \ln(C_{15}/C_1))/14$ (Data: Worldometer)

Common dates: first half of March to end of March (before effects of lockdown)

When 50 have died, between 5 000 and 20 000 had been infected so not VERY early in epidemic which is usually atypical and faster (Norway and Denmark: start instead when > 10 have died)

Stockholms universitet

Covid-19: R_0 estimates, cont'd

Country	C_1	C_{11}	$\hat{\rho}$	$\hat{R}_0$	$\hat{h}_C$
"Norway"	12	89	0.14	2.2	54%
"Denmark"	13	161	0.18	2.6	62%
"Sweden"	62	687	0.17	2.5	60%
"Germany"	68	1275	0.21	3.0	67%
"Belgium"	67	1283	0.21	3.0	67%
"UK"	65	2043	0.25	3.5	71%
"Spain"	55	3647	0.30	4.3	77%

(h_C = critical vaccination coverage for herd immunity)

$\Longrightarrow$ There is not one correct R_0 for covid-19!!

Big differences also within countries!
(Sweden starting when > 10 had died gave $\hat{R}_0 = 3.1$)

Stockholms universitet

Extensions (within homogeneous mixing and individuals)

Random infectious force (e.g. length of infectious period): affects P(outbreak) but hardly final size τ

Latent period: big effect on timing of epidemic peak and duration of epidemic but no effect on final size (unless control measures are initiated)

More than one index case: big effect on P(outbreak) but negligible effect on final size τ in large outbreak

Exercise 3. If infectious period deterministic (=R-F) then P(major outbreak)$= \tau^*$. If infectious period is exponentially distributed then P(major outbreak)$= 1 - 1/R_0$. Compute the latter probability for $R_0 = 1.5$, 3 and 6 and compare with Reed-Frost model.

Stockholms universitet

Extensions (homogeneous mixing)

Initial fraction of immunes. If there is a fraction r of initially immunes the same methodology can be used. The difference is that R_0 is replaced by $R_0(1-r)$ since initially only the fraction $(1-r)$ is susceptible. The final fraction infected *among the initally susceptible* then solves

$$1 - \tau = e^{-R_0(1-r)\tau}$$

Major outbreak possible only if $R_0(1-r) > 1$

Exercise 4. Compute τ^* if initially only 50% were susceptible (and 50% were immune), for $R_0 = 1.5$, 3 and 6.

Exercise 5. What are the *overall* fractions infected during outbreak in later case?

Modelling vaccination (prior to epidemic!)

Why is modelling of disease spread important?

Increase understanding and *prevention* (e.g. **vaccination**)

Modelling vaccination (prior to epidemic!)

Why is modelling of disease spread important?

Increase understanding and *prevention* (e.g. **vaccination**)

Suppose that a fraction v are vaccinated prior to outbreak

Assume first a perfect vaccine (100% immunity)

$\Longrightarrow$ a fraction v are initially immune (treated two pages earlier!)

R_v is the reproduction number after a fraction v has been vaccinated

$\Longrightarrow R_v = R_0(1-v)$

$R_v < 1$ equivalent to $R_0(1-v) < 1$ equivalent to $v > 1 - 1/R_0$

Stockholms universitet

Modelling vaccination cont'd

So, if $v > 1 - 1/R_0$ there will be no major outbreak: "Herd immunity"

$v_c = 1 - 1/R_0$ is called the *critical vaccination coverage*

Exercise 8: Compute v_c for a disease having $R_0 = 1.5$, 3 and 6

Stockholms universitet

Modelling vaccination cont'd

If vaccine is not perfect but relative risk of getting infected from an infectious contact for vaccinees is $1 - E$, $0 < E \leq 1$ (E for "efficacy", later to be called VE_S), then

$$v_c = \frac{1}{E}\left(1 - \frac{1}{R_0}\right)$$

For a highly infectious disease (R_0 large) and a not so effective vaccine (E not too close to 1) v_c might exceed 1. This means vaccination alone cannot prevent an outbreak!

Endemic diseases

When interest is on long-term situation (as opposed to short term outbreaks) the assumption of a fixed population must be relaxed

Consider an SIR disease in a population where individuals die and new are born. Assume:

- SIR disease (life long immunity)
- population at "equilibrium" (in terms of size and incidence)
- disease endemic (constantly present, no big fluctuations)
- $\tilde{s}$, $\tilde{i}$ and $\tilde{r}$ denote the average fractions susceptible, infectious and removed
- R_0 = average number of infections caused by one individual – if everyone was susceptible!

Think of childhood diseases (e.g. chicken-pox)

Stockholms universitet

Endemic diseases, expression for $\tilde{s}$

When disease is in endemic equilibrium each infected individual on average infects exactly 1 new person!

Given R_0 and $\tilde{s}$ an infected individual infects on average $R_0\tilde{s}$ new individuals

Stockholms universitet

Endemic diseases, expression for $\tilde{s}$

When disease is in endemic equilibrium each infected individual on average infects exactly 1 new person!

Given R_0 and $\tilde{s}$ an infected individual infects on average $R_0\tilde{s}$ new individuals

$\Longrightarrow R_0\tilde{s} = 1$!!

$\tilde{s} = \frac{1}{R_0}$

$\tilde{s}$ = average fraction susceptible $= \frac{\text{average age at infection}}{\text{average life-length}}$

Exercise 9 Suppose $R_0 = 1.5$, 3 and 6 respectively, compute $\tilde{s}$.

Estimation:

$$\hat{R}_0 = \frac{1}{\tilde{s}} = \frac{\text{average life-length}}{\text{average age at infection}}$$

Stockholms universitet

Endemic diseases, expression for $\tilde{i}$

If ι is the average length of infectious period and ℓ average life-length, then ι/ℓ is the average time of the life an individual is infectious

Since population/disease in equilibrium this is also the population fraction of infectives

$$\tilde{i} = \frac{\iota}{\ell}$$

Average *number* of infectives: $n\tilde{i}$

Stockholms universitet

Exercises

Exercise 10 Consider an endemic disease with one week infectious period and a population with 75 years expected life-length. Compute the average fraction infective $\tilde{i}$.

Exercise 11 Consider the disease in the previous exercise and consider the Icelandic population ($n = 250\ 000$). What is the average *number* of infectives? How about England ($n = 60\ 000\ 000$)?

Exercise 12 What do you think will happen with the disease in the two countries (remember that if the number of infectives drops to 0 the disease goes extinct - until it is "re-imported")?

Stockholms universitet

More general extensions

Many solved as well as open problems for various extentions

- Considering different types of individual (Multitype epidemic)
- Including other preventive measures
- Including social structures: network epidemics, household epidemics, ...
- SEIR, SIRS, ,,,
- Dynamic population and dynamic behaviour
- Spatial aspects and mobility
- Virus evolution and immunity waning
- Estimation!!!
- ...

Stockholms universitet

How to lock down optimally in time and magnitude

The standard SIR epidemic

$$s'(t) = -\beta s(t) i(t)$$
$$i'(t) = \beta s(t) i(t) - \gamma i(t)$$
$$r'(t) = \gamma i(t)$$

Stockholms universitet

The SIR epidemic with time-varying prevention

Introduce a (non-pharmaceutical) intervention strategy $P = \{p(t); 0 \leq t < \infty\}$:

Contacts reduced by fraction $p(t)$ at t: $\beta \to \beta(1 - p(t))$

$$
\begin{aligned}
s'_P(t) &= -\beta(1 - p(t))s_P(t)i_P(t) \\
i'_P(t) &= \beta(1 - p(t))s_P(t)i_P(t) - \gamma i_P(t) \\
r'_P(t) &= \gamma i_P(t)
\end{aligned}
$$

Final size: $z_P = r_P(\infty) = 1 - s_P(\infty)$

Total cost of prevention strategy: $C(P) := \int_0^\infty p(t)dt$

Optimization problem: Which preventive strategy P, with cost satisfying $\int_0^\infty p(t)dt \leq C_{\max}$, *minimizes* final size z_P?

Stockholms universitet

The SIR epidemic with time-varying prevention

Introduce a (non-pharmaceutical) intervention strategy $P = \{p(t); 0 \leq t < \infty\}$:

Contacts reduced by fraction $p(t)$ at t: $\beta \to \beta(1 - p(t))$

$$
\begin{aligned}
s'_P(t) &= -\beta(1 - p(t))s_P(t)i_P(t) \\
i'_P(t) &= \beta(1 - p(t))s_P(t)i_P(t) - \gamma i_P(t) \\
r'_P(t) &= \gamma i_P(t)
\end{aligned}
$$

Final size: $z_P = r_P(\infty) = 1 - s_P(\infty)$

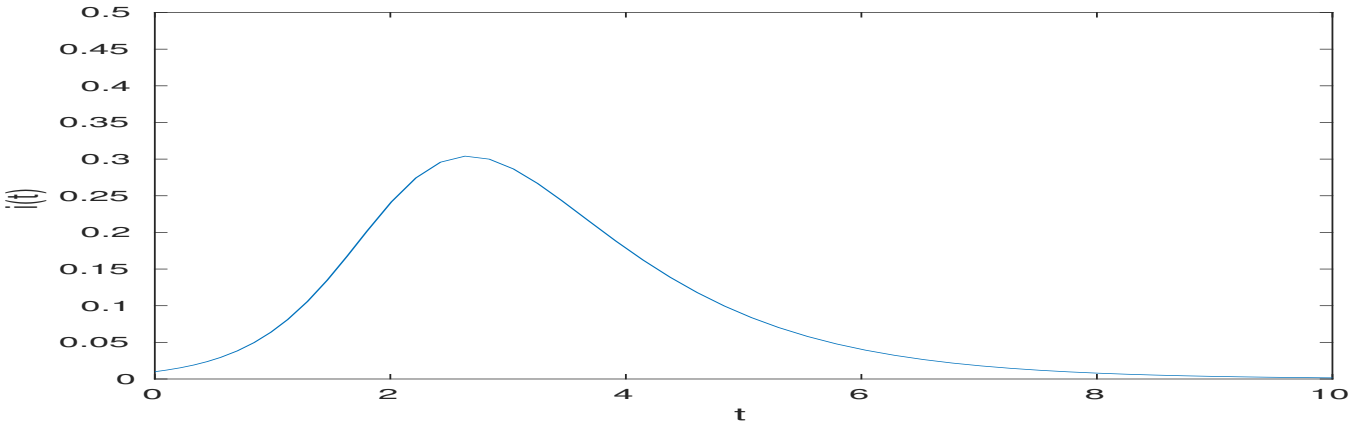
Total cost of prevention strategy: $C(P) := \int_0^\infty p(t)dt$

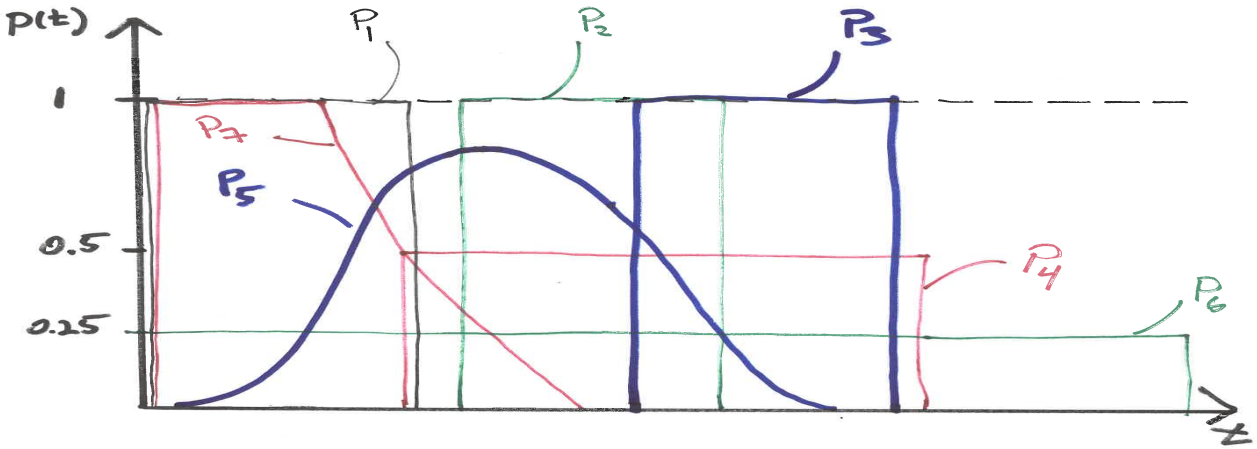
Optimization problem: Which preventive strategy P, with cost satisfying $\int_0^\infty p(t)dt \leq C_{\max}$, *minimizes* final size z_P?

Of course **many simplifications**. Most crucial for conclusions:
– No vaccine available (or expected to arrive in near future)

Stockholms universitet

Uncontrolled incidence (top), some preventions (bottom)

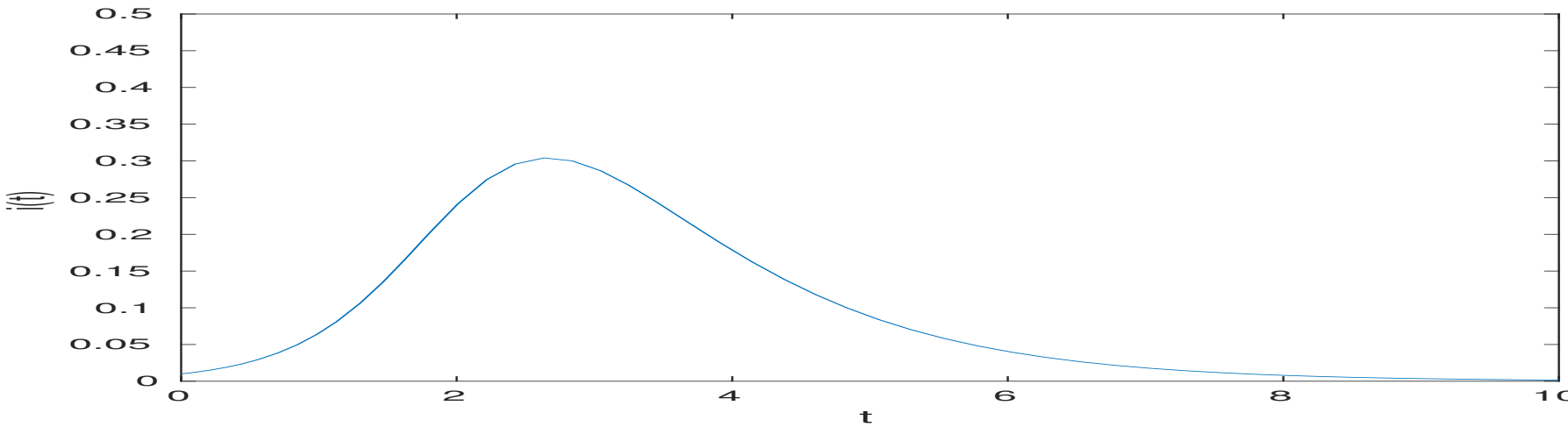


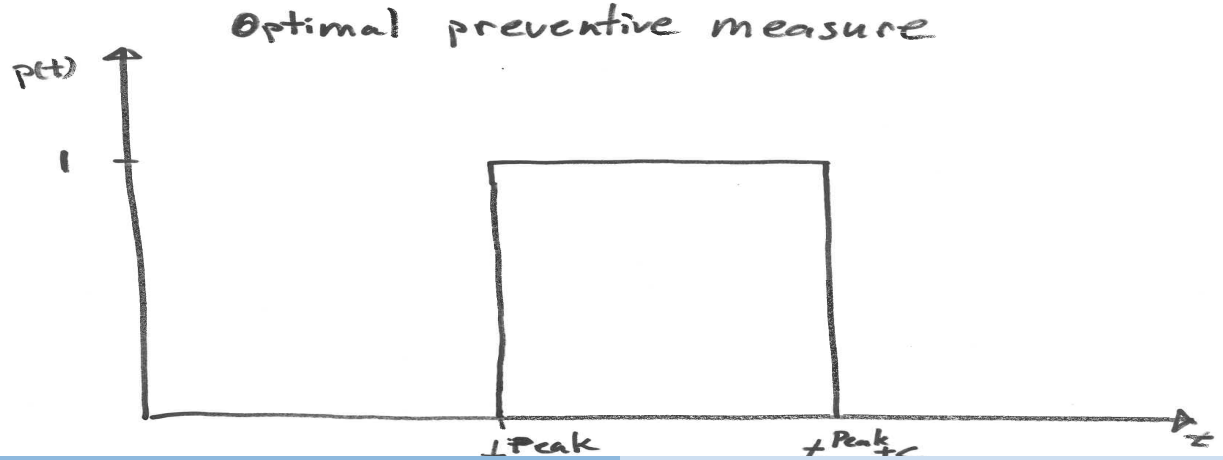


Which prevention reduces cumulative number of infected the most?

Stockholms universitet

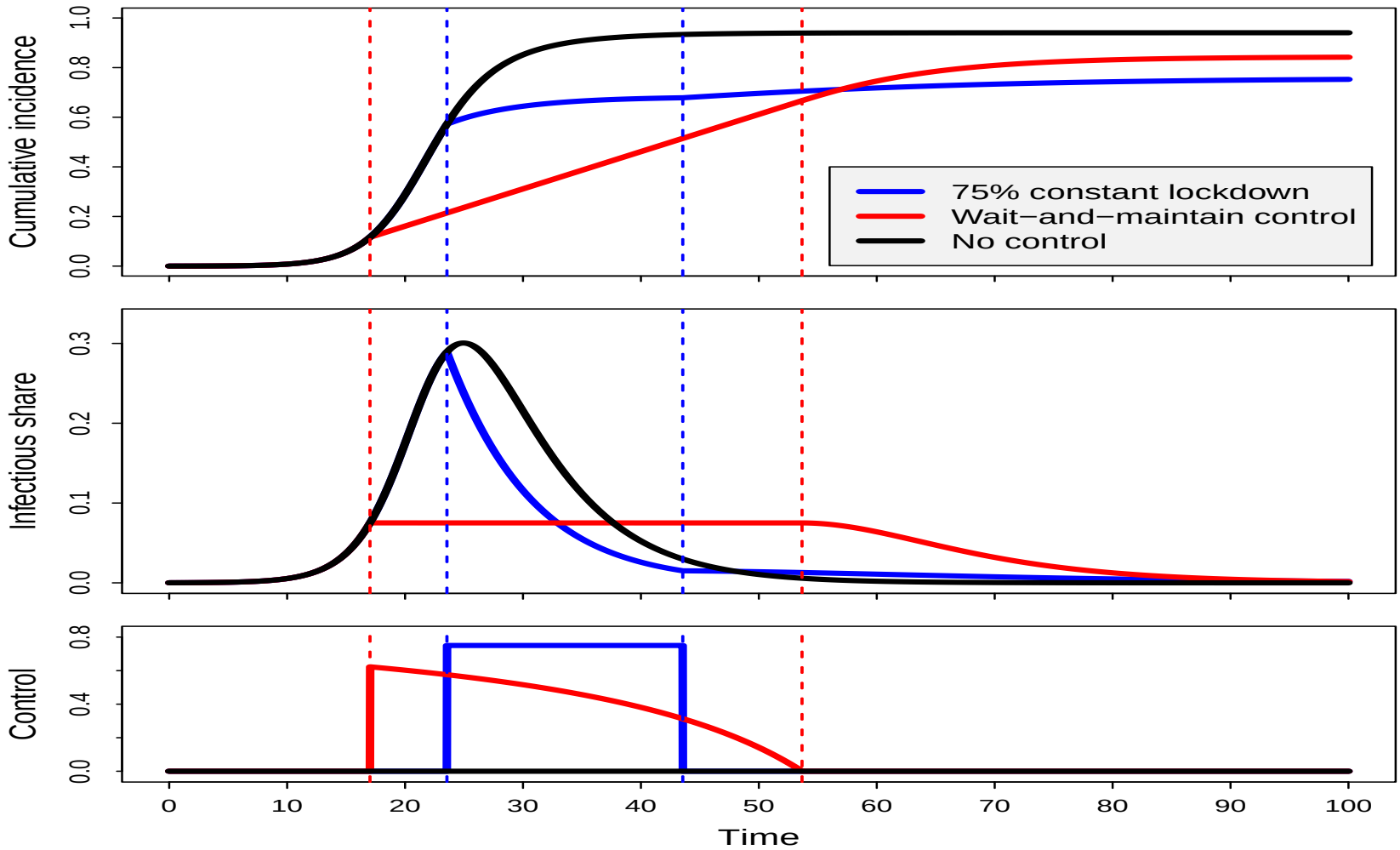
Best strategy: complete lockdown starting at peak





Stockholms universitet

Minimizing final size vs minimizing maximum peak



Stockholms universitet

Adding prevention before optimal may **increase** final size!