Stockholms universitet

Introduction to stochastic epidemic models

Tom Britton

September, 2023

Stockholms universitet

Advertisement

My (Math) Department announces open PostDoc positions in October/November every year

Contact me if you are interested to work on epidemics (and have some experience ...)

Stockholms universitet

A biased list of survey literature

Survey papers by me:

T Britton. (2010). Stochastic epidemic models: a survey. *Mathematical biosciences* 225 (1), 24-35

T Britton. (2020). Epidemic models on social networks—With inference. *Statistica Neerlandica* 74 (3), 222-241

Monographs on Stochastic epidemic models and inference by me and co-authors:

H Andersson and T Britton (2000). Stochastic epidemic models and their statistical analysis. *Springer*
T Britton and E Pardoux (2019). Stochastic epidemic models with inference. *Springer LNM*
O Diekmann et al (2013). Mathematical tools for understanding infectious disease dynamics. *Princeton UP*.

Also many contributions by others!

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,...,$n - 1$.

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

Stockholms universitet

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' outcome 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.

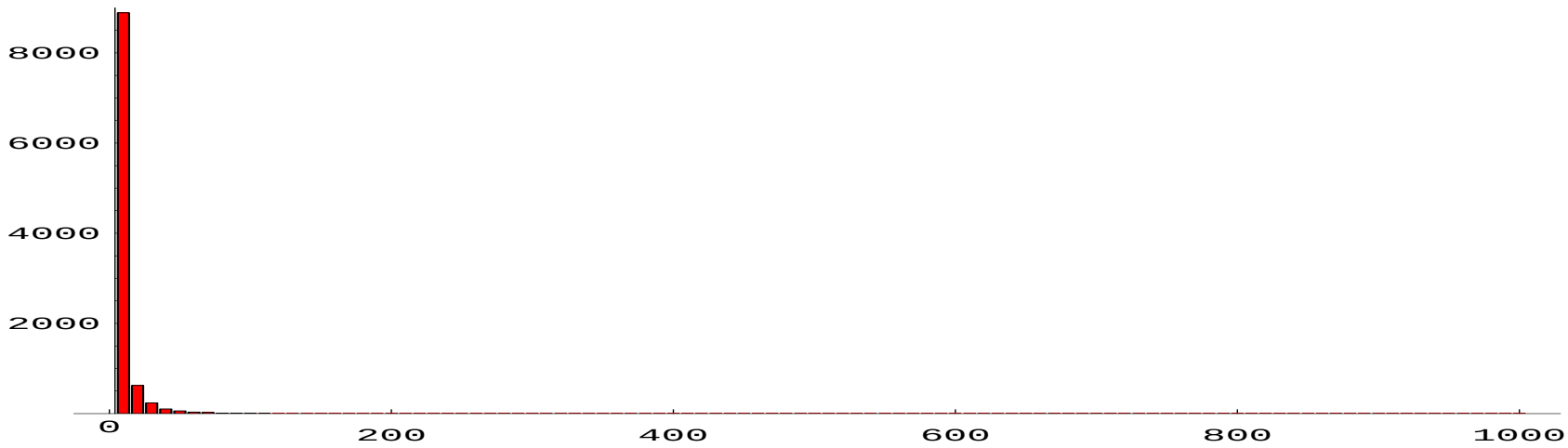
Expected number of infectious contacts: $(n-1)p \approx np =: R_0$

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 $R_0 > 1$)

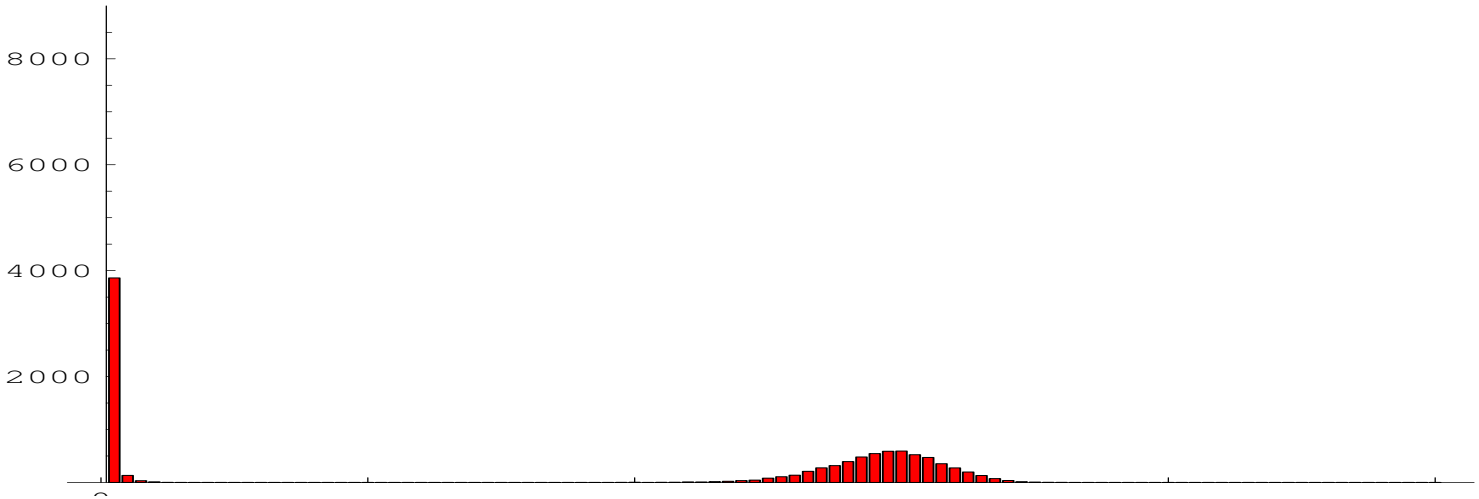
Stockholms universitet

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

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

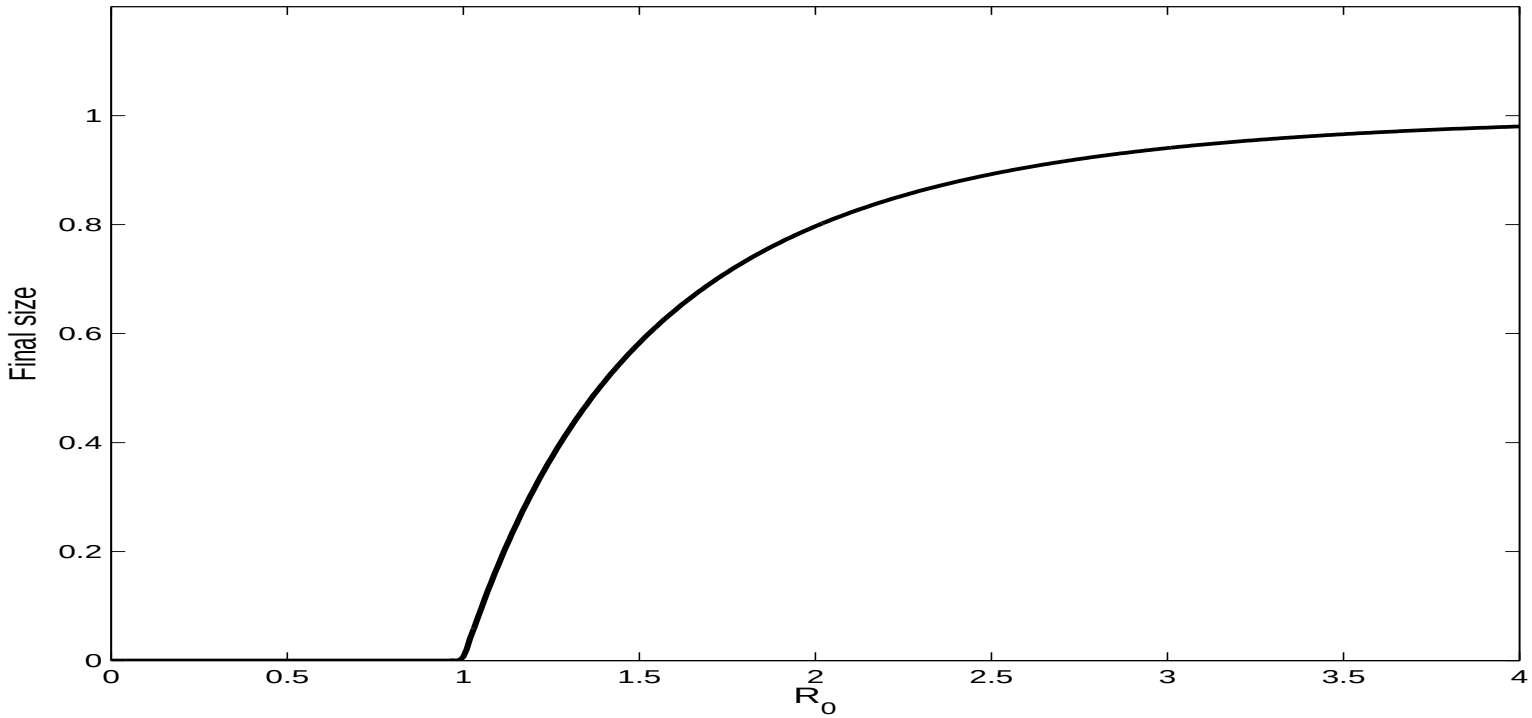
There are two solutions: $\tau = 0$ and (if $R_0 > 1$): $\tau = \tau^\star > 0$.

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

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

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 ν (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 $= \beta\nu$

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

$$s'(t) = -\beta s(t) i(t) \tag{7}$$

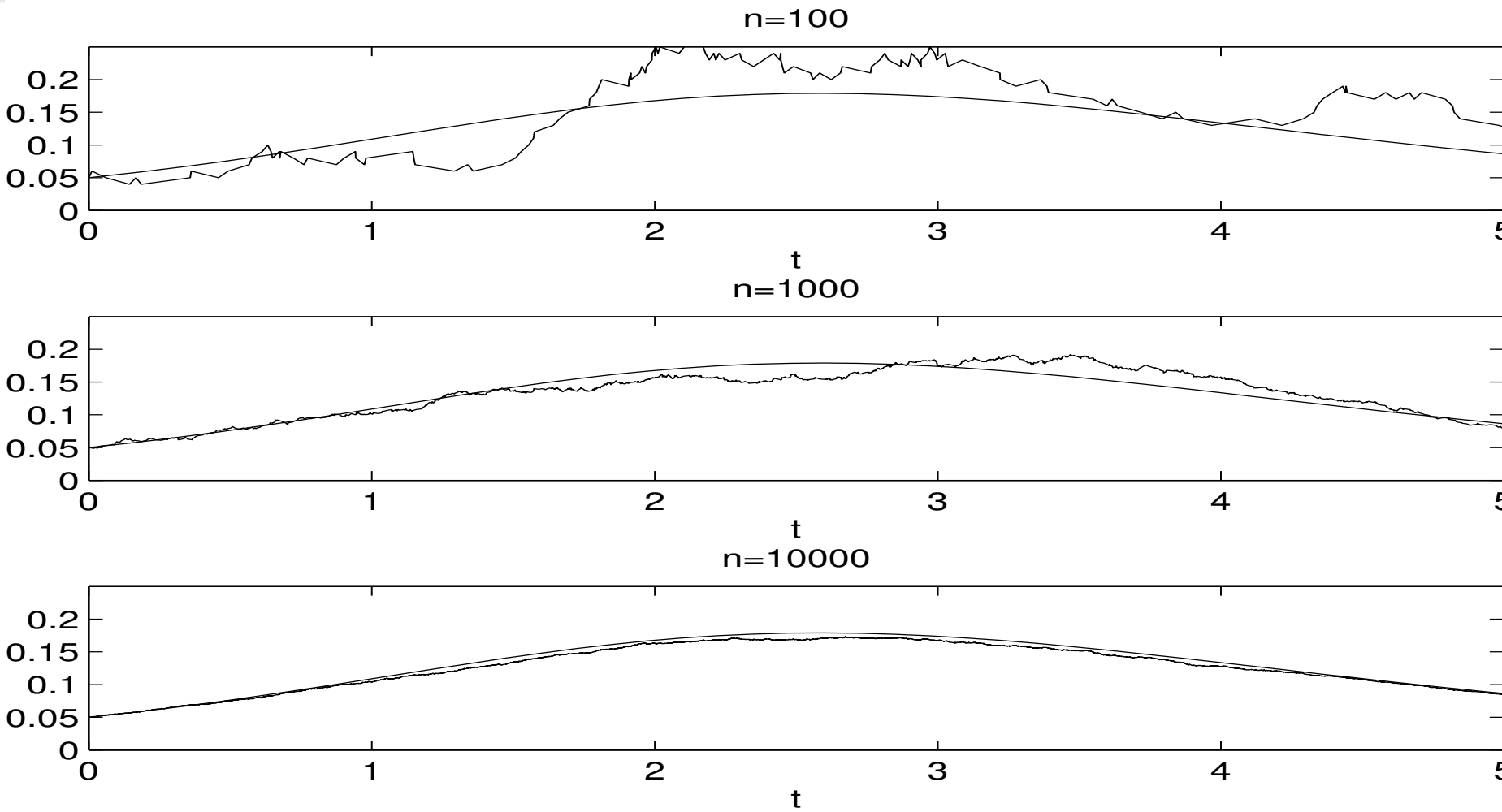
$$i'(t) = \beta s(t) i(t) - \frac{1}{\nu} i(t) \tag{8}$$

$$r'(t) = \frac{1}{\nu} i(t) \tag{9}$$

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: $I(t) \sim e^{rt}$, r solves Euler-Lotka eq.

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

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

Estimation: If we know $g(s)$ and observe "early" growth rate r Euler-Lotka can be used to estimate R_0! (More in last lecture)

Stockholms universitet

The basic reproduction number

Recall: R_0 = expected number individuals a typical infected person infects when everyone is susceptible

R_0 depends both on disease (infectious agent) and on community!!

$R_0 < 1$ or $R_0 > 1$ makes a very big difference!

Next page: R_0 for some diseases (and communities and time periods), Anderson and May, 1991

Stockholms universitet

R_0 for some diseases, communities and time periods (Anderson & May, 1991)

Table 4.1 Estimated values of the basic reproductive rate, R_0, for various infections (data from Anderson (1982*b*), Anderson and May (1982*d*, 1985*c*, 1988), Anderson *et al.* (1988), Nokes and Anderson (1988)).

Infection	Geographical location	Time period	R_0
Measles	Cirencester, England	1947–50	13–14
	England and Wales	1950–68	16–18
	Kansas, USA	1918–21	5–6
	Ontario, Canada	1912–13	11–12
	Willesden, England	1912–13	11–12
	Ghana	1960–8	14–15
	Eastern Nigeria	1960–8	16–17
Pertussis	England and Wales	1944–78	16–18
	Maryland, USA	1943	16–17
	Ontario, Canada	1912–13	10–11
Chicken pox	Maryland, USA	1913–17	7–8
	New Jersey, USA	1912–21	7–8
	Baltimore, USA	1943	10–11
	England and Wales	1944–68	10–12
Diphtheria	New York, USA	1918–19	4–5
	Maryland, USA	1908–17	4–5
Scarlet fever	Maryland, USA	1908–17	7–8
	New York, USA	1918–19	5–6
	Pennsylvania, USA	1910–16	6–7
Mumps	Baltimore, USA	1943	7–8
	England and Wales	1960–80	11–14
	Netherlands	1970–80	11–14
Rubella	England and Wales	1960–70	6–7
	West Germany	1970–7	6–7
	Czechoslovakia	1970–7	8–9
	Poland	1970–7	11–12
	Gambia	1976	15–16
Poliomyelitis	USA	1955	5–6
	Netherlands	1960	6–7
Human Immunodeficiency Virus (Type I)	England and Wales (male homosexuals)	1981–5	2–5
	Nairobi, Kenya (female prostitutes)	1981–5	11–12

Exercise 6: Why is $R_0 > 1$ for all diseases above?

Stockholms universitet

Extensions (within homogeneous mixing)

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

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?

Stockholms universitet

Modelling vaccination (prior to epidemic!)

Why is modelling of disease spread important?

Stockholms universitet

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 (discussed in previous lecture)

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 $v > 1 - 1/R_0$

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!

Stockholms universitet

v_c for some diseases (Anderson & May, 1991)

Fig. 5.1. The dependence of the critical level of vaccination coverage required to halt transmission, p_c, on the basic reproductive rate R_0, or, equivalently, on the average age at infection, A (see eqns (5.2) and (5.3)).

Table 5.1 Approximate estimates of the vaccination coverage (the degree of herd immunity) required to eradicate a variety of viral, bacterial, and protozoan infections in developed and developing countries (eqn (5.2) in the main text)

Infectious disease	Critical proportions (p_c) of the population to be immunized for eradication
Malaria (*P. falciparum* in a hyperendemic region)	99%
Measles	90–95%
Whooping cough (pertussis)	90–95%
Fifths disease (human parvovirus infection)	90–95%
Chicken pox	85–90%
Mumps	85–90%
Rubella	82–87%
Poliomyelitis	82–87%
Diphtheria	82–87%
Scarlet fever	82–87%
Smallpox	70–80%

Stockholms universitet

Endemic diseases (deterministic only)

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")?