

L1, Mathematics and statistics for infectious disease outbreaks

Tom Britton

June, 2020

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 parameter values 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

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

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: $0, 1, \dots, n - 1$.

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"

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

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

$$P(Z = 0) = (1 - p)^2 = 0.64$$

$$P(Z = 1) = \binom{2}{1} p(1 - p) \times (1 - p) = 0.256$$

$$P(Z = 2) = \binom{2}{1} p(1 - p) \times p + p^2 = 0.104$$

$$\text{or ... } P(Z = 2) = 1 - P(Z = 0) - P(Z = 1)$$

What about larger communities?

General n , think in "generations"

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

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

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

$$P(Z = 2) = P(i \rightarrow 2 \rightarrow 0) + P(i \rightarrow 1 \rightarrow 1 \rightarrow 0) = \dots$$

$$P(Z = 3) = P(i \rightarrow 3 \rightarrow 0) + P(i \rightarrow 2 \rightarrow 1 \rightarrow 0) + P(i \rightarrow 1 \rightarrow 2 \rightarrow 0) + P(i \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 0) = \dots$$

$P_n(Z = z)$ 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?

Approximations when n large

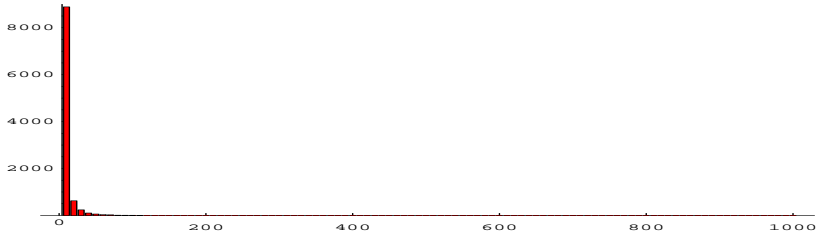
When n large then often p (=per individual transmission probability) is small.

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

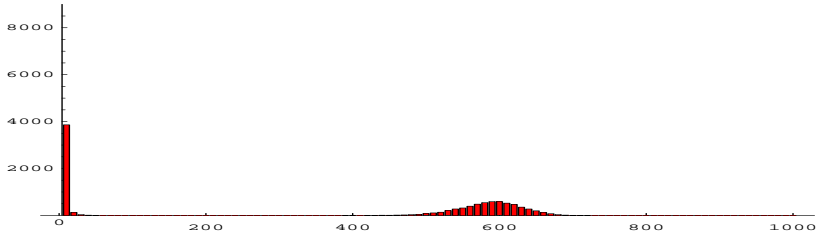
R_0 = basic reproduction number

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

Histogram of final size: $R_0 = 0.8$



Histogram of final size: $R_0 = 1.5$



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

$$1 - \tau = \text{proportion not infected} \quad (1)$$

$$\approx \text{probability not get infected} \quad (2)$$

$$= \text{prob to escape inf from all infected} \quad (3)$$

$$= (1 - p)^Z \quad (4)$$

$$= \left(1 - \frac{R_0}{n}\right)^{n\tau} \quad (5)$$

$$\approx e^{-R_0\tau} \quad (\text{using that } (1 - x/n)^n \approx e^{-x}) \quad (6)$$

Approximation for final size

τ should hence (approximately) solve

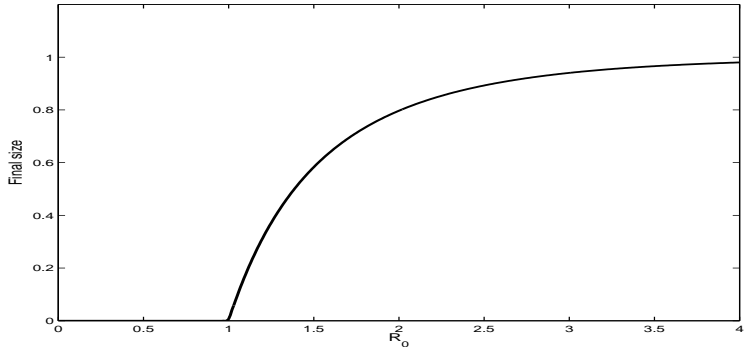
$$1 - \tau = e^{-R_0 \tau}$$

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

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

Plot of final outbreak size as function of R_0



Approximation, cont'd

Seen from simulations: strong dichotomy: minor outbreak – major outbreak

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

For Reed-Frost model: $P(\text{major outbreak}) = \tau^*$!!!

Normal distribution for major outbreak:

$$\sqrt{n} \left(\frac{Z}{n} - \tau^* \right) \approx \text{Normal}(0, \sigma^2)$$

σ^2 depends on model parameters

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)

$$R_0 = \text{expected number of infectious contacts} = \beta\nu$$

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

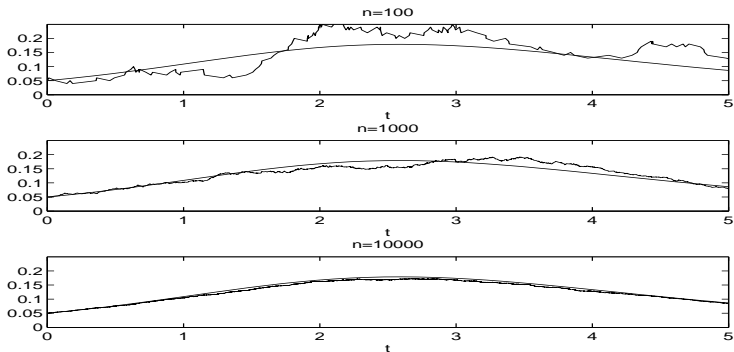
$$s'(t) = -\beta s(t)i(t) \quad (7)$$

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

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

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

Plots of simulated stochastic epidemic and deterministic curve



Summary

Exact distribution of Z (final size) is complicated

When n large two things may happen:

- either very few get infected ($Z/n \approx 0$), or
- a close to deterministic fraction $Z/n \approx \tau^*$ get infected

$R_0 = np = \beta\nu$ = expected number of infections by one individual during his/her infectious period

Second scenario only possible if $R_0 > 1$

$P(\text{major outbreak}) = \tau^*$ for Reed-Frost model

Extensions

Random infectious force (e.g. length of infectious period): affects $P(\text{outbreak})$ 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 $P(\text{outbreak})$ but negligible effect on final size τ in large outbreak

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

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 initially 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?