Stockholms universitet

L3, Inference on stochastic epidemic models

Tom Britton

July, 2017

Stockholms universitet

Statistical inference/estimation in general

Stochastic modelling can tell us (within a model and given some parameter values): what are the likely outcomes?

Example: Given R_0, about how many will get infected?

Statistical inference goes in the "opposite direction" (within a certain model): given an observed outcome, which parameter "fits" to the observation best?

Example: Suppose 20% were infected during an outbreak. What is R_0?

Estimation from outbreak sizes

Suppose an epidemic outbreak is observed and we want to estimate parameters, e.g. transmission probability p, or R_0

What is observed?

Final size: how many were infected and how many were not during outbreak

Important with additional knowledge of how many/what fraction were susceptible prior to outbreak!

If data comes from many small controlled experiments inference is quite easy:

Estimation from many small outbreaks

Example: suppose we have many (n) units of size 2 in which one was initially infected

If m out of the n households resulted in the second individual getting infected then we estimate the transmission probability p by the observed fraction of units in which infection took place:

$$\hat{p} = \frac{m}{n}$$

Note: Parameter estimates are equipped with "hat" (so $\hat{p}$ is an estimate of p)

Estimation from many small outbreaks

If units are isolated (independent) we have a binomial experiment and can easily give confidence bounds:

$$\hat{p} \pm \lambda_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$$

where $\lambda_{\alpha/2}$ is normal distribution quantile:

95% confidence interval ($\alpha = 0.05$) gives $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$

Exercise 13: Suppose 27 out of 100 units had the second individual infected. Give a 95% confidence interval for transmission probability p

More about small group outbreaks later

Estimation from one large outbreak

From before: in case of a large outbreak and assuming everyone was initially susceptible, the final fraction infected will be close to the positive solution of

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

Inference other way around: we observe that a fraction $\tilde{\tau}$ got infected. What is R_0?

Rewrite the equation: $R_0 = -\ln(1-\tau)/\tau$

Our estimate of R_0 is given by the corresponding observed value:

$$\hat{R}_0 = -\ln(1-\tilde{\tau})/\tilde{\tau}$$

Exercise 14: Estimate R_0 if 20% were infected during an outbreak

Estimation from one large outbreak

This estimate assumed everyone was initially susceptible!

If in fact a fraction r was initially immune we know from before that τ, the fraction *among the initially susceptible* who got infected approximately equals positive solution of

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

This leads to the estimate:

$$\hat{R}_0 = -\ln(1-\tilde{\tau})/(1-r)\tilde{\tau}$$

Note: The over all fraction infected equals $\tilde{\tau}(1-r)$

Exercise 15: Suppose as before that 20% were infected during an outbreak, but that only 50% were initially susceptible and the rest were immune. Compute first $\tilde{\tau}$ and then estimate R_0

Estimation of v_c from one large outbreak

It was shown earlier that: $v_c = 1 - 1/R_0$
By observing an outbreak we can hence also estimate v_c (for the same or similar community but not for any community!):

$$\hat{v}_c = 1 - \frac{1}{\hat{R}_0} = 1 - \frac{\tilde{\tau}}{-\ln(1-\tilde{\tau})}$$

If a fraction r was immune in the observed outbreak and $\tilde{\tau}$ of the initially susceptibles were infected this changes to

$$\hat{v}_c = 1 - \frac{1}{\hat{R}_0} = 1 - \frac{(1-r)\tilde{\tau}}{-\ln(1-\tilde{\tau})}$$

Estimation of v_c from one large outbreak

If vaccine not perfect but efficacy E known v_c estimated by

$$\hat{v}_c = \frac{1}{E}\left(1 - \frac{1}{\hat{R}_0}\right) = \frac{1}{E}\left(1 - \frac{(1-r)\tilde{\tau}}{-\ln(1-\tilde{\tau})}\right)$$

Exercise 16. Suppose as previous exercise that 20% of the community got infected but the initial fraction susceptible was 50% (so 40% of these susceptibles were infected). Estimate the critical vaccination coverage for a vaccine having 90% efficacy.

Stockholms universitet

Initial growth rate ρ

For new (so-called *emerging diseases*) and/or lethal diseases it is of course not desirable to wait until the outbreak is over in order to estimate R_0 and other parameters

From before we know $I(t) \approx e^{\rho t}$

So if we observe $I(t_1), \ldots, I(t_k)$ it follows that

$$\frac{I(t_k)}{I(t_1)} \approx e^{\rho(t_k - t_1)}$$

Initial growth rate ρ

This can be used to estimate ρ from data:

$$\ln(I(t_k)/I(t_1)) \approx \rho(t_k - t_1)$$

$$\Longrightarrow \hat{\rho} = \frac{\ln(I(t_k)/I(t_1))}{t_k - t_1}$$

(A more proper estimate would be based on logistic regression. Still, this estimator will be biased for various reasons, e.g. time discretization)

Exercise 17: Suppose the incidence ($\approx I(t)$) was observed the first three weeks and the numbers were: 7, 29 and 121 respectively. Estimate ρ.

Estimation of R_0 from initial phase

Suppose we could estimate the growth rate ρ from an emerging outbreak

How about estimating R_0?

Unfortunately the connection between ρ and R_0 is week (see next slide)

Information about latency period L and infectious period I also needed to estimate R_0

Estimation of L and I hard for two reasons:
1) These periods are rarely observed
2) Even if they were: during the early stages of outbreak short periods are over-represented

Stockholms universitet

Illustration that R_0 and ρ not very related

Illustration. Consider a disease with contact intensity $\beta = 2$ contacts per week and mean infectious $\nu = 1$ week. Then $R_0 = \beta\nu = 2$ and some exponential growth rate ρ.

Consider now another disease having $\beta = 1$ and $\nu = 2$ (less infectious but longer infectious period). Clearly this new disease also has the same $R_0 = \beta\nu = 2$. How about ρ?
The latter is twice as slow $\Longrightarrow$ new ρ is half of the former:
$\rho_{\text{new}} = \rho_{\text{old}}/2$

Pitfalls when estimating ρ in emerging outbreaks

From an epidemic model it is possible to derive

$\lambda(s) =$ the average rate of infecting new individuals s time units after infection (during early stage)

e.g. for Gen epid $\lambda(s) = \beta P(\text{still infectious at } s) = \beta e^{-s/\nu}$

The following is known (has been proven mathematically):
1) $R_0 = \int \lambda(s) ds$
2) $f_G(s) := \lambda(s)/R_0 =$ density of generation times
3) ρ is the unique solution to $\int e^{-\rho s} R_0 f_G(s) ds = 1 \quad (*)$

(*) can be used to obtain estimates of ρ and predict future growth

Ebola analysis: "In 6 weeks 20K individuals will be infected if no prevetive measures"

Stockholms universitet

Pitfalls when estimating ρ in emerging outbreaks, cont'd

By analysing (*) it can be shown that

- ρ decreases with $E(G)$
- ρ increases with $V(G)$
- ρ less affected by R_0 (but increases with R_0)

How to estimate $f_G(s)$?

Pitfalls when estimating ρ in emerging outbreaks, cont'd

By analysing (*) it can be shown that

- ρ decreases with $E(G)$
- ρ increases with $V(G)$
- ρ less affected by R_0 (but increases with R_0)

How to estimate $f_G(s)$? Contact tracing: look up infectors of infected people and compare onset of symptoms

Three problems with this:

1) Serial times instead of infection times
G = time between infection times
S = time between onset of symptoms
$\Longrightarrow S = G + (I_1 - I_2)$ (I_1 and I_2 =incubation periods of infector and infectee)

Pitfalls when estimating ρ in emerging outbreaks, cont'd

So, if incubation times are independent and independent of G, then $E(S) = E(G)$, and $V(S) \geq V(G)$
So, if we estimate S instead of G and plug this into (*) our estimate ρ will be *over-estimated*

2) Looking backwards rather than forward in time

G was defined as time between infection of an individual and time of infection of a random person he/she infects

Contact tracing looks backward in time

As a consequence: long generation times will not have occurred and short generation times will be over-represented

$\Longrightarrow E(G)$ will be under-estimated $\Longrightarrow \rho$ will be *over-estimated*

Stockholms universitet

Pitfalls when estimating ρ in emerging outbreaks, cont'd

3) Multiple infector candidates

Sometimes more than one infector candidate exists – what to do?

If multiple candidates, earlier ones are more likely (in simple models)

Easiest to discard them – what is effect?

Remaining generation times tend to be shorter

$\Longrightarrow E(G)$ will be under-estimated $\Longrightarrow \rho$ will be *over-estimated*

Stockholms universitet

Pitfalls when estimating ρ in emerging outbreaks, cont'd

Three different "problems"

- Serial times instead of generation times
- Looking backwards instead of forwards in time
- Discarding infections with several infector candidates

All of them lead to *over-estimation* of growth rate ρ

Magnitude of combined effect: Approximately: "20K infected within 6 weeks" $\rightarrow$ "10K infected within 6 weeks"

Stockholms universitet

Endemic diseases

Consider an *endemic disease* and that $\tilde{s}$ observed

$\tilde{s}$ = average fraction of susceptibles = average relative time spent in susceptible state = average age at infection/average life-length

From before we know $\tilde{s} \approx 1/R_0$

$\Longrightarrow \hat{R}_0 = \frac{1}{\tilde{s}}$

By only knowing the typical infection-age and life-length gives estimate of R_0!

Stockholms universitet

Endemic diseases: estimation of v_c

Same data: $\tilde{s}$ = average age of infection divided by average life-length (= average fraction susceptible in community)

We know that $v_c = 1 - 1/R_0$ (or $v_c = E^{-1}(1 - 1/R_0)$ if vaccine has known efficacy E)

$$\Longrightarrow \hat{v}_c = \frac{1}{E}(1 - \tilde{s})$$

Exercise 18 Suppose (as with measles) average age of infection is 5 years and average life-length is 75 years. Estimate R_0 and v_c assuming a vaccine having efficacy $E = 0.95$. (How about if $E = 0.90$?)