Stockholms universitet

L4, Estimation of R_0; and Effective reproduction numbers

Tom Britton

June 2020

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:

Stockholms universitet

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)

Stockholms universitet

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

Stockholms universitet

Estimation of R_0 from one large outbreak

Assume a homogeneously mixing community and no preventive measures

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 4.2: Estimate R_0 if 20% were infected during an outbreak

Stockholms universitet

Estimation from one large outbreak, cont'd

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 4.3: 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 before outbreak has ended

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 ...

We now describe how to estimate R_0 from the initial phase of an epidemic

Stockholms universitet

The early stage of an epidemic outbreak

Consider a **large homogeneous** community and the **beginning** of an outbreak when few have been infected

Then it is very unlikely that two infectious people have infectious contact with the same person

$\Longrightarrow$ Infectious individuals infect new people $\approx$ **independently**

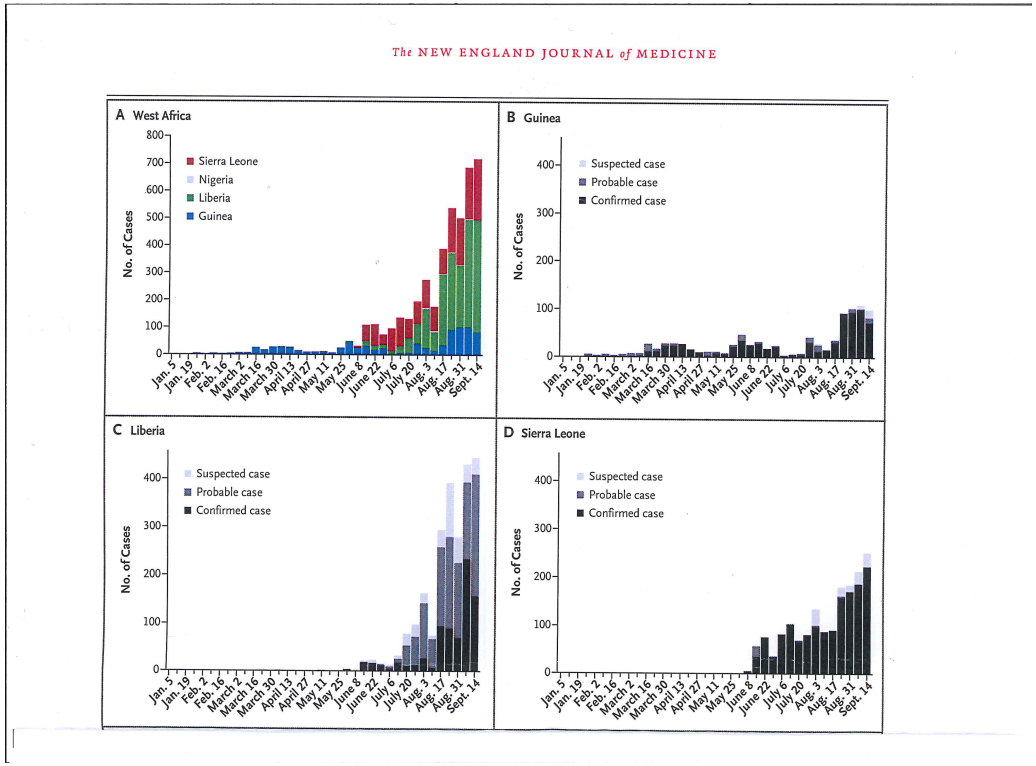
Branching process model: a model where individuals "give birth" (=infect in our case) independently of each other

Important result: Number of infectives $I(t)$ (as well as the number infected $n - S(t)$) grows exponentially: $I(t) \sim e^{rt}$

Exercise 4.4: Suppose the exponential growth rate r equals $r = 2.8$ (per week) and that there is one index case week 0. Compute the expected incidence ($\approx I(t)$) after 1, 2 and 3 weeks.

Stockholms universitet

Ebola (beginning): How many will get infected?



NEJM Sept 24, 2014. WHO Ebola response team: (There had been 4 500 cases by Sept 14, 2014): "If no preventive measures: >20 000 will have been infected by Nov 2, 2014"

Stockholms universitet

Estimating the initial growth rate r

Branching process theory. $I(t) \sim e^{rt}$ and $n - S(t) \sim e^{rt}$.
Cumulative number of infected $n - S(t)$ is more robust

$$\ln((n - S(t_k))/(n - S(t_0))) \approx r(t_k - t_0)$$

$$\Longrightarrow \hat{r} = \frac{\ln((n - S(t_k))/(n - S(t_0)))}{t_k - t_0}$$

(A more proper estimate would be based on logistic regression)

Exercise 4.5: Suppose the incidence ($\approx I(t)$) was observed the first three weeks and the numbers were: 7, 29 and 121 respectively. Estimate r using estimator above (as well as with logistic regression).

Stockholms universitet

Estimation of R_0 from initial phase

Suppose we have estimated the growth rate r from an emerging outbreak

How about estimating R_0?

Unfortunately the connection between r and R_0 is week (see next slide): r is a rate (depends on time) and R_0 is dimensionless (indep of time)

Stockholms universitet

Illustration that R_0 and r not very related

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

Consider now another disease having $\beta = 1$ and $1/\gamma = 2$ (less infectious but longer infectious period). Clearly this new disease also has the same $R_0 = \beta/\gamma = 2$. How about r?

The latter is twice as slow $\Longrightarrow$ new r is half of the former: $r_{\text{new}} = r_{\text{old}}/2$. So same R_0 but different r

However, branching process theory connect r and R_0 be means of the generation time distribution!

Stockholms universitet

Initial growth rate

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

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

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

(Problems with estimating $g(s)$ treated in L3!, see consequences a few slides ahead)

Exercise 4.6: Show that if $g(s) \sim \Gamma(\alpha, \beta)$ then Euler-Lotka gives that

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

Stockholms universitet

Covid-19: R_0 estimates from different countries

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 r from reported cumulative case fatalities: starting first day with > 50 cumulative case fatalities (c_0) and two weeks later c_{14} case fatalities: $\hat{r} = \ln(C_{14}/C_0))/14$ (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 (except Norway and Denmark: start instead when > 10 have died)

Stockholms universitet

Covid-19: R_0 estimates from different countries

Country	C_0	C_{14}	$\hat{r}$	$\hat{R}_0$	$\hat{h}_C$
"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, more later)

$\Longrightarrow$ 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

Effects of bias in estimates of $g(s)$

$I(t)$ = incidence day t = # infected day t (now discrete time)

How many that get infected day t depends on: R_0 =, basic reproduction number and $\{g(s)\}$ = Generation time

– how many that got infected s days ago? Answer: $= I(t-s)$

Model definition (common model)

$$I(t) \sim \text{Pois}\left(R_0 \sum_{s=1}^{t} g(s) I(t-s)\right), t = 1, 2 \ldots, \qquad (*)$$

"Pois( )" means Poisson distribution, and the mean equals the parameter, $R_0 \sum_{s=1}^{t} g(s) I(t-s)$

Exercise 4.7: Show that this is more or less identical to the Euler-Lotka equation (Hint: replace the Poisson random variable by its mean)

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

$$I(t) \sim \text{Pois}\left(R_0 \sum_{s=1}^{t} g(s) I(t-s)\right), t = 1, 2 \ldots, \qquad (*)$$

If $\{g(s)\}$ known (or estimated), Eq. $(*)$ can be used for:

1: Estimating R_0 (from observed incidence $I(1), \ldots, I(t)$), or
2: Predicting outbreak incidence $I(1), \ldots, I(t)$ (if R_0 known before-hand)

Both 1 and 2 require knowledge about $\{g(s)\}$

Main question: How to estimate generation time distribution $\{g(s)\}$ and what happens to estimates of R_0 (or predictions $I(1), I(2), \ldots$) if $\{g(s)\}$ is estimated incorrectly?

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

Recall, $I(t) \sim \text{Pois}\left(R_0 \sum_{s=1}^{t} g(s) I(t-s)\right)$

where $I(0), \ldots, I(t)$ grows, typically exponentially

How are estimates of R_0 (or predictions $I(1), \ldots, I(t)$) affected by the generation time distribution $\{g(s)\}$?

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

Recall, $I(t) \sim \text{Pois}\left(R_0 \sum_{s=1}^{t} g(s) I(t-s)\right)$

where $I(0), \ldots, I(t)$ grows, typically exponentially

How are estimates of R_0 (or predictions $I(1), \ldots, I(t)$) affected by the generation time distribution $\{g(s)\}$?

It is easy to show that the mean parameter

$R_0 \sum_{s=0}^{t} g(s) I(t-s)$ **increases** if:

– $g(s)$ is replaced by $\hat{g}(s)$ which has smaller mean

– $g(s)$ is replaced by $\hat{g}(s)$ which has same mean and larger variance

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

Recall, $I(t) \sim \text{Pois}\left(R_0 \sum_{s=1}^{t} g(s) I(t-s)\right)$

where $I(0), \ldots, I(t)$ grows, typically exponentially

How are estimates of R_0 (or predictions $I(1), \ldots, I(t)$) affected by the generation time distribution $\{g(s)\}$?

It is easy to show that the mean parameter

$R_0 \sum_{s=0}^{t} g(s) I(t-s)$ **increases** if:

– $g(s)$ is replaced by $\hat{g}(s)$ which has smaller mean

– $g(s)$ is replaced by $\hat{g}(s)$ which has same mean and larger variance

So, if our estimate of $\{g(s)\}$ has mean biased from below we will **under-estimate** R_0

And if we estimate $\{g(s)\}$ by something with the correct mean but larger variance we will **under-estimate** R_0

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

In Lecture 3 we showed three problems when estimating $g(s)$ from **contact tracing**:

1) Looking backwards rather than forward in time: $g(s)$ was biased from below ($E(G)$ under-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

2) What if multiple infector candidates: $g(s)$ was biased from below ($E(G)$ under-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

3) Observing Serial intervals instead of Generation times $g(s)$ has too large standard deviation ($V(G)$ over-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

Stockholms universitet

Effects of bias in estimates of $g(s)$ (cont'd)

In Lecture 3 we showed three problems when estimating $g(s)$ from **contact tracing**:

1) Looking backwards rather than forward in time: $g(s)$ was biased from below ($E(G)$ under-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

2) What if multiple infector candidates: $g(s)$ was biased from below ($E(G)$ under-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

3) Observing Serial intervals instead of Generation times $g(s)$ has too large standard deviation ($V(G)$ over-estimated)
$\Longrightarrow R_0$ will be **under-estimated**

Conclusion: Unless taken account for, all three problems make R_0 *under-estimated*. See Britton & Scalia-Tomba (Interface, 2019)

Stockholms universitet

Preventive measures: homogeneous case – initial phase

Suppose **preventive measures** (put in place very early) reduce contact rate β by a factor c

$\Longrightarrow$ new *Effective* reproduction number equals $R_E = (1-c)R_0$

No outbreak possible if $R_E \leq 1$ which is equivalent to $c \geq 1 - 1/R_0$

More focus on preventive measures in e.g. Lecture 7, now focus is on Effective reproduction nmbers

Stockholms universitet

Preventive measures and immunity

If $R_E \geq 1$ the epidemic still grows initially and immunity builds up: only infectious contacts with not yet infected individuals result in infection:

$R_E(t) = R_0(1-c)s(t)$, where $s(t)$ is *fraction* susceptible

If initially $R_E = R_E(0) > 1$ then $R_E(t)$ decays and for t large enough $R_E(t) < 1$ (because $s(t)$ becomes small) and then the epidemic starts declining

Currently $R_E(t) < 1$ in all (?) countries of Europe for Covid-19 (but probably $R_E = R_E(0) > 1$ in Sweden)

Terminology: some use "effective reproductiion number" for R_E and others for $R_E(t)$ (i.e. also including immunity). $R_E(t)$ also denoted "current", "instantaneous" or "daily" reproduction number