L8, Estimation uncertainty + Herd immunity

Tom Britton

July, 2021

Advertisment

2-year Post Doc position in "Modelling and analyses of epidemics" at Stockholm University

Deadline: September 1, 2021

Write me an e-mail for more information. Or check www.math.su.se or the specific link https://www.su.se/department-of-mathematics/news/available-job-postdoc-1.564363

Repetition: Inference from large outbreaks

From lecture 3: basic reproduction number R_0 and critical vaccination coverage v_c were estimated by:

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

if outbreak takes place in a fully susceptible homogeneous community resulting in a fraction $\tilde{\tau}$ getting infected during the outbreak

How about uncertainty?

Uncertainty of previous estimate

Intuition: The larger community (and more getting infected) the less uncertainty

It was mentioned that final number infected $n\tilde{\tau} = Z$ in case of a major outbreak is normally distributed with mean $n\tau^*$ and standard deviation $\sqrt{n\sigma^2}$ where σ^2 depends on model parameters and shown two slides ahead

This result can be used to show that $\hat{R}_0$ and $\hat{v}_c$ are normally distributed with correct means (i.e. R_0 and v_c respectively) and standard errors to be derived using δ-method

The δ-method

Suppose random variable X has mean $\mu = E(X)$ and variance $V(X)$

Then the δ-method gives the following approximation for the mean and variance of $f(X)$, where $f(x)$ is a "nice function":

$$E(f(X)) \approx f(\mu) \qquad V(f(X)) \approx (f'(\mu))^2 V(X)$$

The approximation holds better the smaller variance X has (i.e. smaller $V(X)$)

The δ-method for $V(\hat{R}_0)$

Probabilists have proven that the asymptotic variance of $\tilde{\tau}$ equals:

$$V(\tilde{\tau}) \approx \frac{1}{n}\frac{\tau(1-\tau)}{(1-(1-\tau)R_0)^2}\left(1+c_v^2(1-\tau)R_0^2\right)$$

where τ and R_0 are the true parameter values related by $R_0 = -\ln(1-\tau)/\tau$, and c_v is the coefficient of variation of the infectious period.

We now apply the δ-method on $\hat{R}_0 = -\ln(1-\tilde{\tau})/\tilde{\tau}$, we hence have the function $f(x) = -\ln(1-x)/x$

After some algebra we get $V(\hat{R}_0) \approx \frac{1}{n\tau(1-\tau)}\left(1+c_v^2(1-\tau)R_0^2\right)$

For a standard error estimate we take square roots and replace unknown quantities with there estimates/observed values. The result, also for $\hat{v}_c$, is given by:

Uncertainty of previous estimate

$$s.e.(\hat{R}_0) = \sqrt{\frac{1 + c_v^2(1-\tilde{\tau})\hat{R}_0^2}{\tilde{\tau}(1-\tilde{\tau})}/n}$$

$$s.e.(\hat{v}_c) = \sqrt{\frac{1 + c_v^2(1-\tilde{\tau})\hat{R}_0^2}{\hat{R}_0^4\tilde{\tau}(1-\tilde{\tau})}/n}$$

$c_v^2 = V(I)/(E(I))^2$= squared coefficient of variation of infectious period of individuals (variance divided by the squared mean)

Larger *n* gives smaller standard deviation (as expected)!

Uncertainty of previous estimate

c_v^2 cannot be estimated from final outbreak size – possibly known from before

If not one has to insert a "conservative" bound. E.g. $c_v^2 = 1$: very rarely is standard deviation larger than mean

Exercise 25 Suppose that 239 out of 651 individuals in an isolated village were infected during an outbreak. Estimate R_0 and v_c and give 95% confidence interval for the estimates. Consider both the case when all individuals have the same length of infectious period (so no variation) and the case where its standard deviation is equal to the mean.

Exercise 26 Do the same thing assuming 2390 out of 6510 got infected.

More detailed data

Suppose that disease incidence is observed during outbreak – not only final number

Intuition: more detailed data should improve estimation

Answer: yes, in a couple of ways:

- estimate of R_0 and v_c becomes more complicated, but standard errors are (moderately) smaller
- enables estimation of more parameters: exponential growth rate ρ, latent and infectious period distributions, ...
- possible to detect deviations from model: changing behavior, non-homogeneity, ...

If also information about contacts are available: "transmission probability upon contact" can be estimated

Multitype epidemics

Suppose final size of a multitype epidemic observed: $\tilde{\tau}_1, \ldots, \tilde{\tau}_k$,
$\tilde{\tau}_i =$ observed proportion infected among i-types

Also assumed that community fractions $\pi_1, \ldots, \pi_k$ known.

We want to estimate R_0 which is largest eigenvalue of next generation matrix M

First estimate M. Impossible!! Data has dimension k and M has dimension k^2.

$\Longrightarrow$ M and R_0 cannot be estimated consistently!

Multitype epidemics, cont'd

Why? We can observe who was infected but not who "caused" the infections

Susceptibility easier to estimate than infectivity!

$\Longrightarrow$ only possible to obtain bounds on R_0: lower bound assuming all infections caused by least infected type – upper bound assuming all infections caused by most infected type

Inference in networks

Inference can be performed without an outbreak: estimation of network properties: $E(D)$, $V(D)$, clustering c, ...

R_0, potential outbreak size τ and v_c can then be estimated as a function of transmission probability p

Typical conclusion: Outbreaks are only possible for a disease having higher transmission probability than $p = 0.13$

Or: An STD with $p = 0.08$ can only become endemic in core-groups with average number of partners higher than $E(D) = 4.2$ per year

Inference in more complicated models

More complicated model $\Longrightarrow$ harder inference and more detailed data need

Inference of spread of infections extra hard:

- There are strong dependencies because infections are not independent events (likelihood complicated)
- Many things unobserved: infectious contacts, latent period, infectious period, ...

Inference with more detailed data gives higher precision

Illustration

Suppose an infected infects each susceptible independently with prob p

Data = epidemic chain: $1 \to 2 \to 2 \to 0$

Initially 1 index and 9 susceptible

Likelihood: $L(p) =$
$$\binom{9}{2} p^2 (1-p)^7 \cdot \binom{7}{2} \left(1-(1-p)^2\right)^2 \left((1-p)^2\right)^5 \cdot \binom{5}{0} \left((1-p)^2\right)^5$$
Maximum-likelihood (ML) estimate $\hat{p}$ maximizes $L(\cdot)$:
$\Longrightarrow$ quite easy for a computer

If we instead only know that 5 out of 10 were infected likelihood is much more complicated (a sum over all possible chains)

Alternative approach for complicated models

Basic idea: If likelihood complicated for available data we can "pretend" as if we had more detailed data, estimate parameters under this assumption, recompute some likely more detailed data, re-estimate parameters, ...

This is underlying idea in both EM-algorithm and recently very popular *MCMC*

MCMC: here parameters are treated as outcomes of random variables (Bayesian framework) and even very complicated likelihoods (posterior probabilities) can be evaluated numerically with arbitrary high precision

MCMC: Very computer intensive. Treated specifically in other Modules

Herd immunity (Britton, Ball, Trapman, 2020+2021)

Classical result: Critical vaccination coverage (= herd immunity level) when immunity/vaccination is uniformly distributed equals

$$V_c = 1 - \frac{1}{R}$$

But last year (before vaccine arrival) first wave was stopped by mitigation/suppression (and summer effects)

Infected people (later immune) are not uniformly distributed – more immunity among socially active and highly susceptible!

This should lead to a **smaller overall immunity level** required for herd immunity!!

Scientific task: Investigate and quantify this effect

A model for COVID-19 allowing for heterogeneities

SIR epidemic with four types of heterogeneities:

- **Age cohorts**: with mixing and community fractions taken from empirical study (Wallinga et al, 2006)
- **Variable social activity**: assumed independent of other heterogeneities
- **Variable susceptibility**: assumed independent of other heterogeneities
- **Variable infectivity**: assumed independent of other heterogeneities

Simple model for social activity, susceptibility and infectivity:

50% have medium level, 25% have low (=half this level) and 25% have high (=double this level)

A model for COVID-19 allowing for heterogeneities, cont'd

Model of heterogeneity quite arbitrary but:

no left or right tails, and coefficient of variation $= 0.48$

Age-distribution gives a next generation matrix including mixing features, age-differences and population fractions (6 age-groups)

"On top" of this individuals are categorized according to social activity, susceptibility and infectivity independently

First result: **Variable infectivity has no effect** (on deteterministic model)

Model: Deterministic Multitype epidemic: $6 * 3 * 3 = 54$ types

$R_0 =$ largest eigenvalue to 54*54 next generation matrix

Final size equations exist

Including prevention and vaccine-induced immunity

Preventive measure assumption: **all** contact rates are reduced with the same factor p (restrictive assumption!)

Suppose a fraction $\hat{i}$ are immunized from (uniform) vaccination

Effective reproduction number

$$R_E = R_0(1-\hat{i})(1-p)$$

Same expression as homogeneous case!

$\Longrightarrow$ Same herd immunity level $\hat{i}_{Vac} = 1 - 1/R_0$

and same $p_{Min}^{(Vac)} = 1 - 1/(R_0(1-\hat{i}))$ as in homogeneous case

where p_{Min} = minimal amount of preventive measures to avoid an outbreak

Including prevention and disease-induced immunity

Suppose instead that a fraction $\hat{i}$ are immune from a suppressed or mitigated outbreak

Then immunity is not uniformly distributed: socially active and highly susceptible individuals are over-represented

$\Longrightarrow$ This immunity is more "effectively distributed"

$$\Longrightarrow R_t < R_0(1-p)(1-\hat{i})$$

so $\Longrightarrow \hat{i}_{Dis} < 1 - 1/R_0$

and $p_{Min}^{(Dis)} < p_{Min}^{(Vac)} = 1 - 1/(R_0(1-\hat{i}))$

$\Longrightarrow$ The minimal effect of preventive measures is lower
a) if immunity comes from disease spreading vs vaccination
b) if acknowledging heterogeneities vs homogeneous model

Herd immunity levels (B+B+T *Science* 2020)

$p_{Min} = 0 \Longrightarrow$ Herd-immunity.

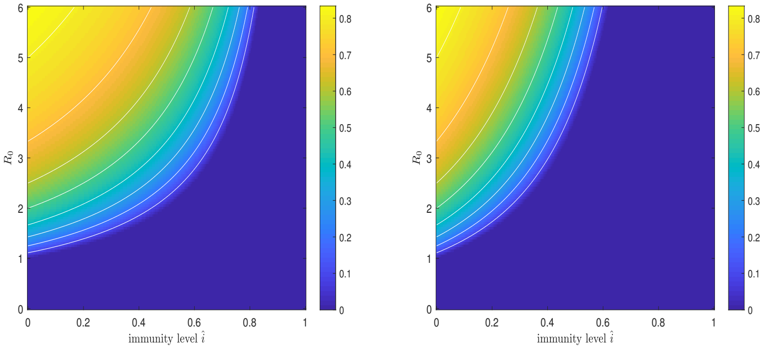
Tabell: Disease-induced herd immunity level $\hat{i}_{Dis}$ and vaccine-induced herd immunity level $\hat{i}_{Vac} = 1 - 1/R_0$, for $R_0 = 2.0$, 2.5 and 3.0. Levels correspond to percentages.

Population structure	$R_0 = 2.0$		$R_0 = 2.5$		$R_0 = 3.0$	
	$\hat{i}_{Dis}$	$\hat{i}_{Vac}$	$\hat{i}_{Dis}$	$\hat{i}_{Vac}$	$\hat{i}_{Dis}$	$\hat{i}_{Vac}$
Homogeneous	50.0	50.0	60.0	60.0	66.7	66.7
Age structure	46.0	50.0	55.8	60.0	62.5	66.7
Activity structure	37.7	50.0	46.3	60.0	52.5	66.7
Age & Activity structure	34.6	50.0	**43.0**	**60.0**	49.1	66.7

Herd immunity level is lower than earlier believed! (Unclear exactly how much lower!!)

Heatmap of minimal preventive measure p_{Min} (BTB, 2021)

Left: Vaccine-induced immunity and/or homogeneous model
Right: Disease-induced immunity + heterogeneous model



Example: $R_0 = 2.5$, $\hat{i} = 25\%$: $p_{Min}^{(Vac)} = 47\%$ and $p_{Min}^{(Dis)} = 29\%$

Illustration: Country estimates of R_0 taken from Flaxman et al (2020) and tweeked within country from country specific analyses

Region	R_0	Deaths/100k	$\hat{i}$ (%)	$p_{Min}^{(start)}$(%)	$p_{Min}^{(Dis)}$	$p_{Min}^{(Vac)}$
Madrid	4.7			78.7		
Cataluna	4.5			77.8		
Lombardy	3.4			70.6		
Lazio	3.4			70.6		
New York	4.9			79.6		
Wash D.C.	2.5			60.0		
Stockholm	3.9			**74.4**		
Copenhagen	3.5			**71.4**		
Oslo	3.0			**66.7**		

Illustration: Immunity estimates taken from case fatality numbers September 2020 and assuming the **same** $ifr = 0.5\%$ in all regions.

Region	R_0	Deaths/100k	$\hat{i}$ (%)	$p_{Min}^{(start)}$ (%)	$p_{Min}^{(Dis)}$	$p_{Min}^{(Vac)}$
Madrid	4.7	145	29.0	78.7	58.3	70.0
Cataluna	4.5	77.4	15.5	77.8	68.9	73.7
Lombardy	3.4	168	33.6	70.6	34.7	55.7
Lazio	3.4	16.2	3.2	70.6	68.6	69.6
New York	4.9	169	33.8	79.6	54.4	69.2
Wash D.C.	2.5	89.4	17.9	60.0	40.8	51.3
Stockholm	3.9	102	20.4	**74.4**	**59.7**	67.8
Copenhagen	3.5	20.0	4.0	**71.4**	**69.0**	70.2
Oslo	3.0	11.4	2.3	**66.7**	**65.1**	65.9

Conclusions

Vaccine-induced immunity is less efficiently distributed compared with disease-induced immunity

$\Longrightarrow$ more individuals need to be immunized with vaccination to obtain herd immunity

$\Longrightarrow$ more preventive measures needed (for a fixed overall immunity level) if immunity comes from vaccination compared to disease-induced immunity

(The exact size differences need to be investigated further – we use a toy model)

Important result, but **NOT** an argument for aiming for disease-induced herd immunity **OR** to skip vaccination!

Over-all summary

General advice: Complement more advanced statistical analysis with simple model analysis. If similar conclusions: reassuring. If very different: mistake or understanding needed

Some important messages

- Prior (partial) immunty makes big difference for estimates
- Inference for emerging epidemics is hard
- Heterogeneities usually makes R_0 larger but not necessarily bigger outbreak!

Important but not treated:
– Changing behaviour over time
– Selection bias
– Asymptomatics and other under-reporting