The statistics of generation times in epidemic spread

Gianpaolo Scalia Tomba
Dept of Mathematics
Univ. of Rome Tor Vergata

In collaboration with
Åke Svensson, Jacco Wallinga

The simplest model for an epidemic is SIR

- Differential equation form

$S(0)=N-1$, $I(0)=1$, $R(0)=0$

$S' = -\beta SI/N$

$I' = \beta SI/N - \mu I$

$R' = \mu I$

=> .... (scaled form: s=S/N, i=I/N, r=R/N)

$s(0)=1-\varepsilon$, $i(0)=\varepsilon$, $r(0)=0$

$s' = -\beta si$

$i' = \beta si - \mu i$

$r' = \mu i$

- Markov process form

$S(0)=N-1$, $I(0)=1$, $R(0)=0$

$(S,I,R) \rightarrow (S-1,I+1,R)$ with intensity $\beta IS/N$

$(S,I,R) \rightarrow (S,I-1,R+1)$ with intensity μI

Basic quantities

(infectious) contact rate β (number/time unit)

duration infectious period $1/\mu$

=>

R_0 = (initial) reproduction number = β/μ

r = (initial) growth rate of #infected = $\beta-\mu$

=> threshold: $\beta \leq \mu$ = large outbreaks not possible,

usually given as $R_0 \leq 1$

if $R_0 > 1$, ODE model always grows, Markov allows for early extinction with probability $1/R_0$

=> herd immunity: immunize proportion $1-1/R_0$, all "protected"

Difficult to observe infections and infectious period

=>

generation time (serial time) = time between successive cases in a transmission chain, usually observed as "clinical onset interval".

Let us start by considering these as times between infections.

The possible addition of time from infection to symptoms (latent time) to both individuals should preserve the average length of GT.

How should we treat observed generation times...?

Example: two individuals,

- IP=1 and secondaries in 1/4, 2/4, 3/4 -> av. GT = 1/2

- IP=2 and secondaries in 1,2,3,4,5,6,7/4 -> av. GT = 1

Should we calculate the general average GT as 3/4 (infector based) or 34/40 (infectee based)?

The situation reminds us of

- "straight" sampling of infected individuals (forwards)

- size biased sampling (length of IP or # of secondaries) ->

$tf(t)/\mu$

Theoretical relations between β, μ and GT...

based on the Markov SIR model..., i.e. given T = infectious period with Exp(μ) distribution, cases are created according to a Poisson process on [0,T], thus S= number of secondary cases will be Po(βT)...

a) Given S=k, T is $\Gamma(k+1,\mu+\beta)$

b) $P(S=k) = \mu\beta^k/(\mu+\beta)^{k+1}$

d) $E(T|S>0) = 1/(\mu+\beta) + 1/\mu$

e) Given T and S=k>0, the k timepoints of events are iid U[0,T]

f) Let G be a randomly chosen point among these timepoints.

Then E(G|S>0) = E(T|S>0)/2

Return to observing generation times...

- "straight" sampling of infected individuals (forwards) should lead to an estimate of $1/\mu$, since the average "within individual" estimates his T/2...but we should include individuals without secondary cases...
- the "backwards" sampling should lead to an estimate of $(1/(\mu+\beta) + 1/\mu)/2$

The household sampling scheme

- The (infectious) contact intensity β may be different within households, compared the the "general" contact intensity, but there is also a sampling effect due to the depletion of susceptibles.

Consider a household with two susceptibles. If only one is infected, this happens at expected time $1/(2\beta+\mu)$ and the total expected length of IP is $1/(2\beta+\mu) + 1/(\beta+\mu)$.

If the second is also infected, this happens at time $1/(2\beta+\mu)$ + $1/(\beta+\mu)$, but the total interval now has expected length $1/(2\beta+\mu)$ + $1/(\beta+\mu)$ + $1/\mu$.

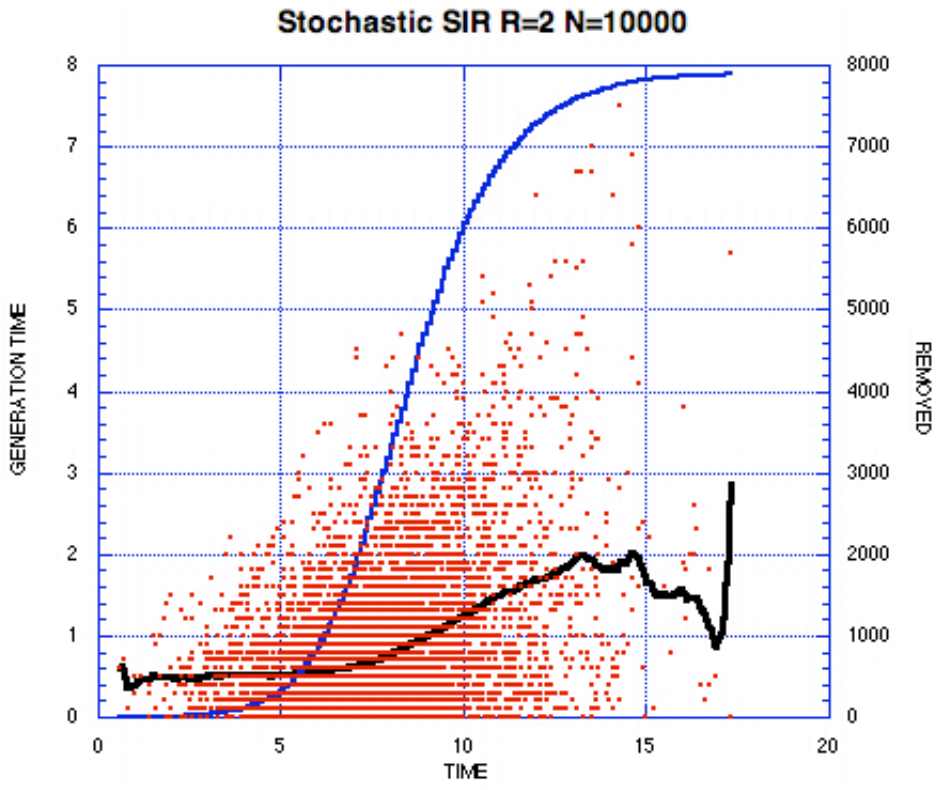
For the case β=2 and μ=1, we get probabilities of 0,1,2 secondaries = 3,4 and 8/15 and expected GTs = 3/15 within 8/15 in the second case and 3 and 8/15 within 23/15 in the third case. Thus GT tends to be shorter than half the infectious period.

Are there more problems?

Event based (linked pointer list) implementation of stochastic SIR (called SIRmulator...) that allows vision of all individual quantities...
because the usual simulation only updates total (S,I,R) numbers...

and...

Result of simulating backwards defined GTs



Natural history av. GT=2/3

Sizebiased av. GT= 1

“Malthus” biased av. GT $\in$ [1/2,2]

This effect can be understood by thinking about demographic models, exponential growth and the famous Malthusian parameter (our previous $r = \beta - \mu$). There are always more young individuals (short IP) "than there should be", in an exponentially growing population...

However, next step: systematic review of literature on generation times in order to better understand the different sampling schemes used and how these have affected the statistical analysis...