

Modeling and estimating generation intervals

Tachikawa infectious boot camp, 2019

Jonathan Dushoff, McMaster University

https://github.com/dushoff/Generation_talks

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

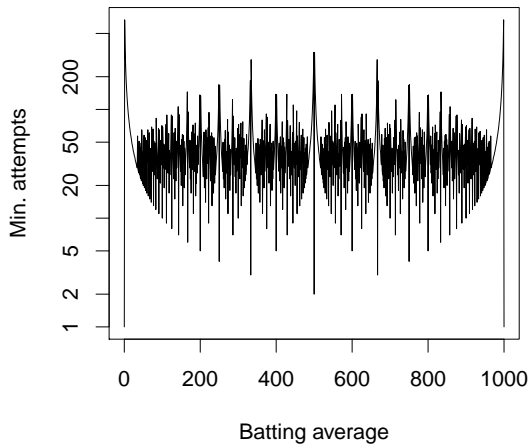
Conclusion

Renewal math (extra)

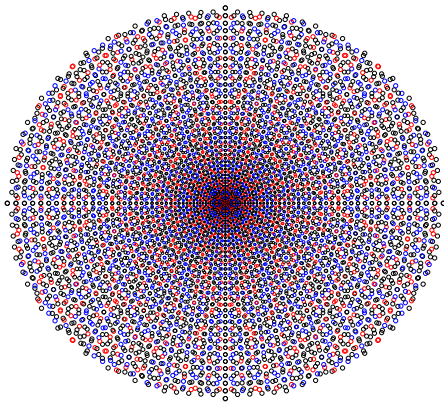
Introduction

- ▶ I talk too fast
- ▶ I have too much to tell you
- ▶ Interrupt me!

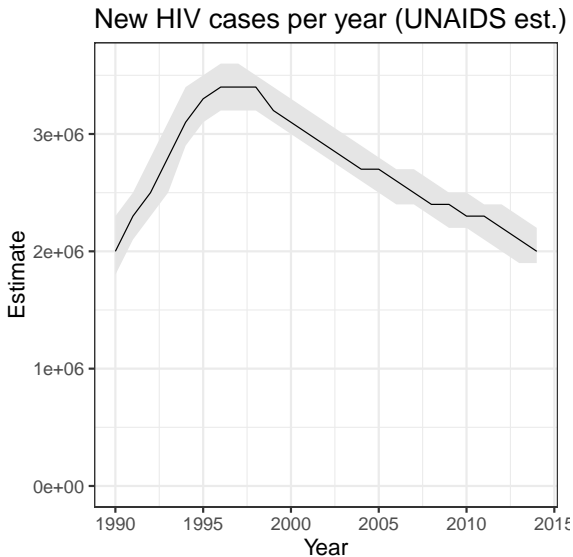
Who am I? Head (present)



Who am I? Heart (present)



Who am I? Stomach (present)

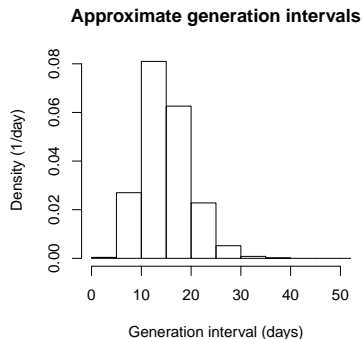


Who are you? (present)

- ▶ Math person
- ▶ Health person
- ▶ Biology person

How long is a disease generation?

- ▶ Introduced by Prof. Nakaoka
- ▶ If I am infected on day 0, when do I infect you?
- ▶ When do you infect Dr. Akhmetzhanov?

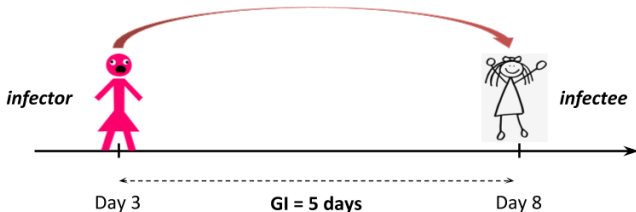


How long is a disease generation? (present)

Definition

Generation Interval:

Interval between the time that an individual is infected by an infector and the time this infector was infected



Goals

- ▶ Introduce a generation-based framework for modeling
- ▶ Discuss importance of generation intervals
- ▶ Discuss how generation intervals are defined and measured

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

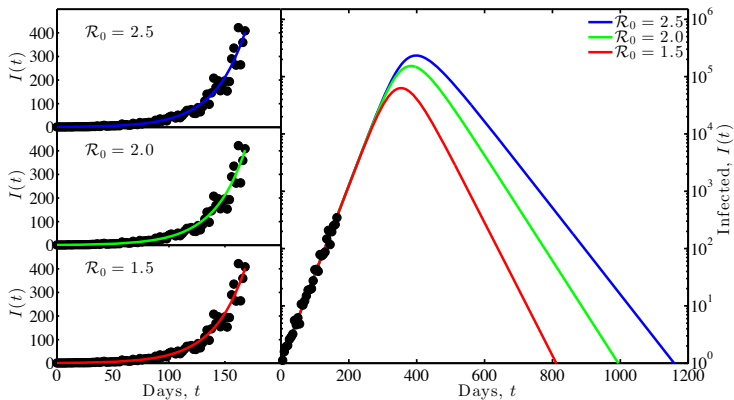
Conclusion

Renewal math (extra)

Speed

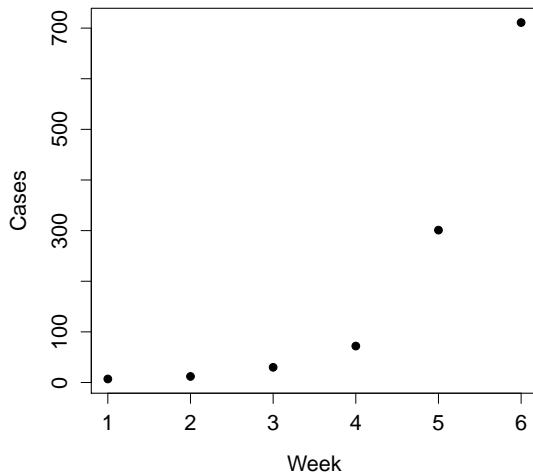
- ▶ We measure epidemic speed using little r :
 - ▶ The ratio of the *change* in disease impact to the *amount* of disease impact
 - ▶ *Units*: [1/time]
 - ▶ Disease increases like e^{rt}
- ▶ Time scale is $C = 1/r$

Ebola outbreak



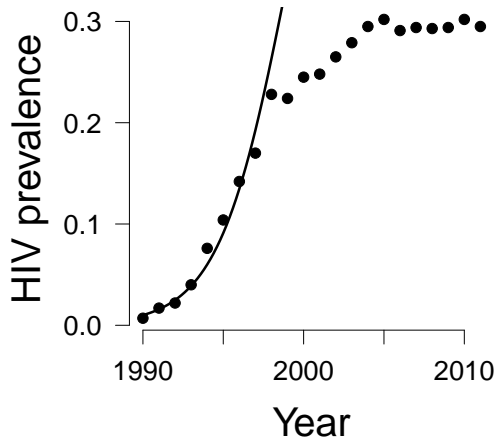
$C \approx 1$ month. Sort-of fast.

Mexican flu



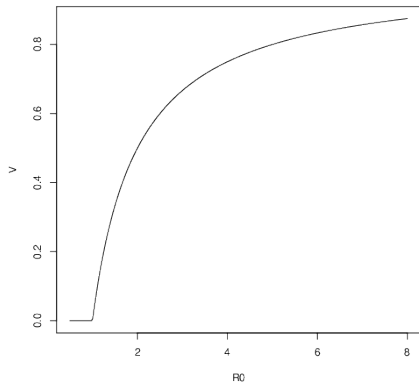
$C \approx 1$ week. Sort-of fast.

HIV in sub-Saharan Africa



$C \approx 18$ month. Horrifyingly fast.

Strength of epidemics



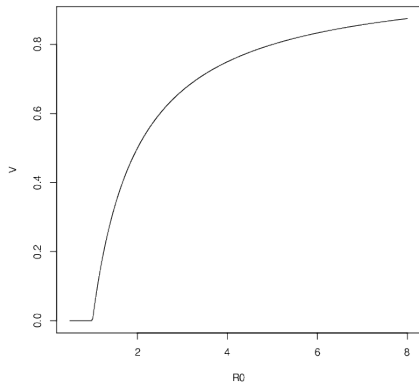
\mathcal{R} and control

- ▶ We describe epidemic strength with big \mathcal{R}
- ▶ Number of potential new cases per case
 - ▶ Not accounting for proportion susceptible
- ▶ To eliminate disease, we must:
 - ▶ Reduce effective reproduction by a factor of \mathcal{R}

\mathcal{R} and equilibrium

- ▶ If we have \mathcal{R} new cases per case when everyone is susceptible
- ▶ And 1 case per case (on average) at equilibrium:
 - ▶ Proportion susceptible at equilibrium is $S = 1/\mathcal{R}$
 - ▶ Proportion affected at equilibrium is $V = 1 - 1/\mathcal{R}$

DEFHEAD \mathcal{R} and control (present)



Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

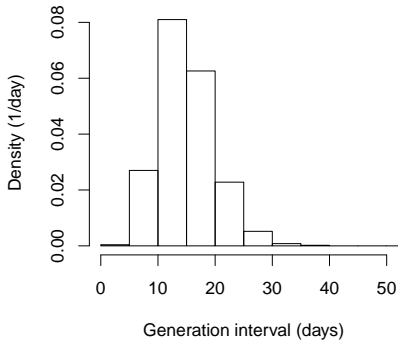
Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Approximate generation intervals

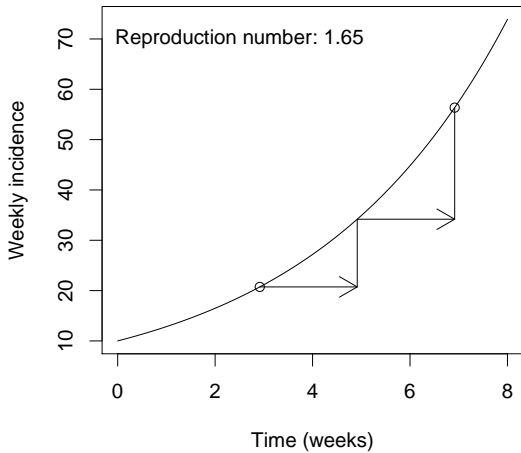


- ▶ The generation distribution measures generations of the disease
 - ▶ Interval between “index” infection and resulting infection
- ▶ Do fast disease generations mean more danger or less danger?

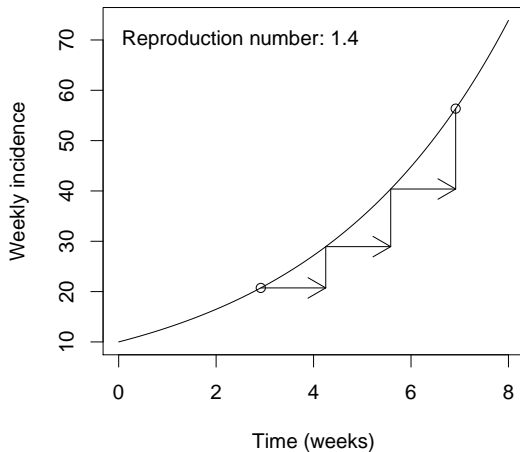
Conditional effect of generation time

- ▶ *Given* the reproductive number \mathcal{R}
 - ▶ faster generation time G means faster growth rate r
 - ▶ More danger
- ▶ *Given* the growth rate r
 - ▶ faster generation time G means *smaller* \mathcal{R}
 - ▶ Less danger

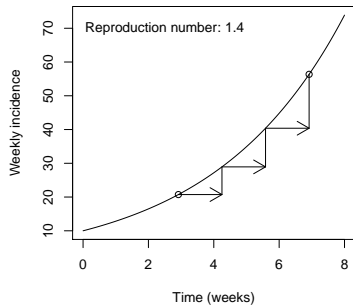
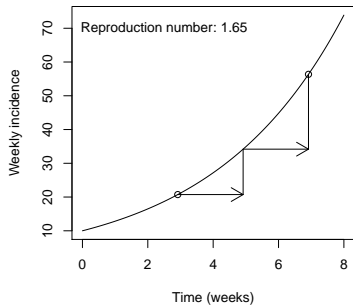
Generations and $\mathcal{R}(\text{present})$



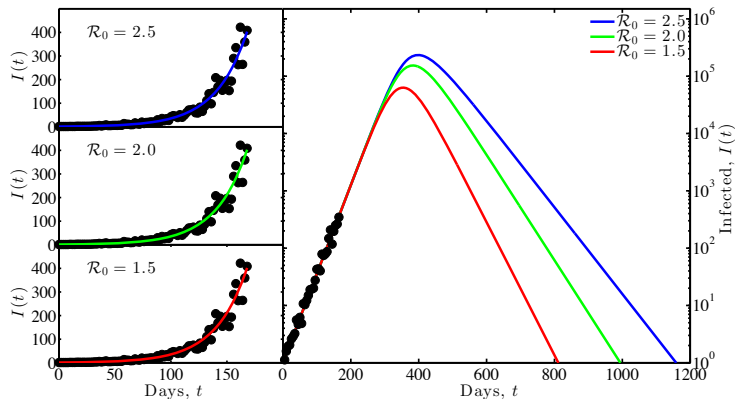
Generations and \mathcal{R}



Generations and \mathcal{R} (present)

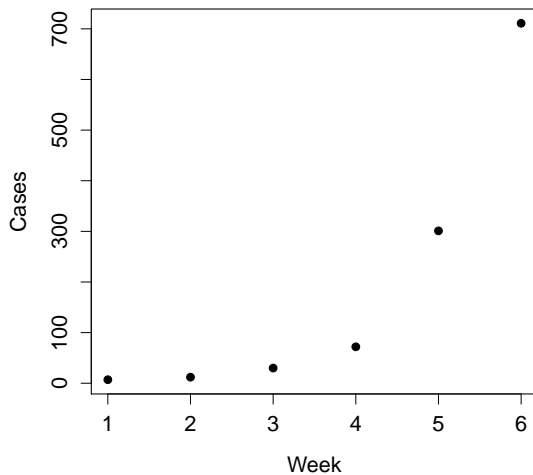


Ebola outbreak



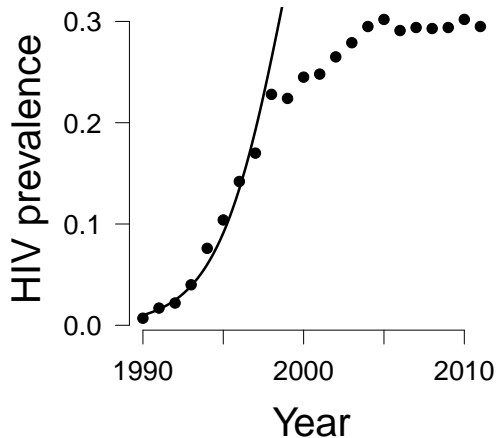
$C \approx 1$ month, $G \approx 2$ week

Mexican flu



$C \approx 1$ week, $G \approx 3$ day

HIV in sub-Saharan Africa



$C \approx 18$ month, $G \approx 4$ years

Linking framework

- ▶ Epidemic speed (r) is a *product*:
 - ▶ generation speed \times
 - ▶ epidemic strength
- ▶ WRONG

- ▶ Epidemic speed (r) is a *product*:
 - ▶ (something to do with) generation speed \times
 - ▶ (something to do with) epidemic strength

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

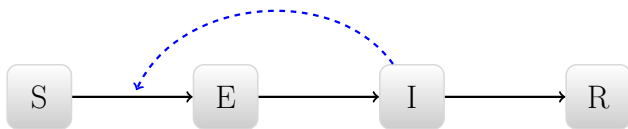
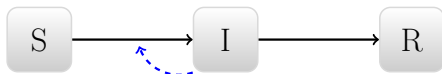
Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Box models

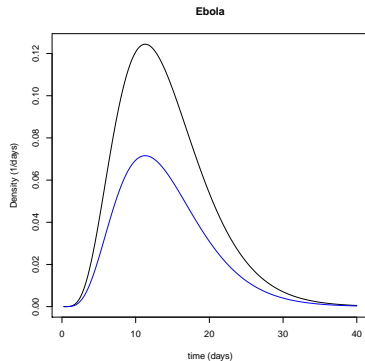


Renewal equation

- ▶ A broad framework that covers a wide range of underlying models
- ▶ $i(t) = S(t) \int k(\tau) i(t - \tau) d\tau$
 - ▶ $i(t)$ is the *rate* of new infections (per-capita incidence)
 - ▶ $S(t)$ is the *proportion* of the population susceptible
 - ▶ $k(\tau)$ measures how infectious a person is (on average) at time τ after becoming infected
- ▶ For invasion, treat S as constant

Infection kernel

- ▶ $k(\tau)$ is the expected rate at which you infect at time τ after being infected
- ▶ $\int_{\tau} k(\tau) d\tau$ is the expected number of people infected:
 - ▶ \mathcal{R} the effective reproductive number
- ▶ $k(\tau)/\mathcal{R}$ is a distribution:
 - ▶ $g(\tau)$, the *intrinsic* generation distribution



Renewal equations

- ▶ More flexible than ODEs
 - ▶ Non-exponential distributions, variation in infectiousness through time
 - ▶ The ODEs we've seen can be rewritten as renewal equations!
- ▶ Can be parameterized by observing generation intervals
 - ▶ Contact tracing (realized intervals)
 - ▶ infectiousness of studied individuals (intrinsic distribution)

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Euler-Lotka equation

- ▶ Model



$$i(t) = S \int k(\tau) i(t - \tau) d\tau$$

- ▶ If we neglect changes in S , we expect exponential growth

- ▶ Exponential phase

- ▶ Disease grows with characteristic time $C = 1/r$



$$i(t) = i(0) \exp(rt)$$

Euler-Lotka equation



$$i(t) = S \int k(\tau) i(t - \tau) d\tau$$

- ▶ Substitute:

$$i(t) = i(0) \exp(rt)$$

- ▶ $1 = \int k(\tau) \exp(-r\tau) d\tau$

- ▶ i.e., the total of *discounted* contributions is 1

- ▶ $1/\mathcal{R} = \int g(\tau) \exp(-r\tau) d\tau$

Interpretation: generating functions

- ▶ $1/\mathcal{R} = \int g(\tau) \exp(-r\tau) d\tau$
- ▶ *J Wallinga, M Lipsitch; DOI:
10.1098/rspb.2006.3754*



Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Interpretation: “effective” generation times

- ▶ Define the effective generation time so that



$$\mathcal{R} = \exp(r\hat{G})$$

- ▶ Then:



$$1/\mathcal{R} = \int g(\tau) \exp(-r\tau) d\tau$$



$$\exp(-r\hat{G}) = \langle \exp(-r\tau) \rangle_g.$$

- ▶ A filtered mean:

- ▶ The discounted value of \hat{G} is the expectation of the discounted values across the distribution

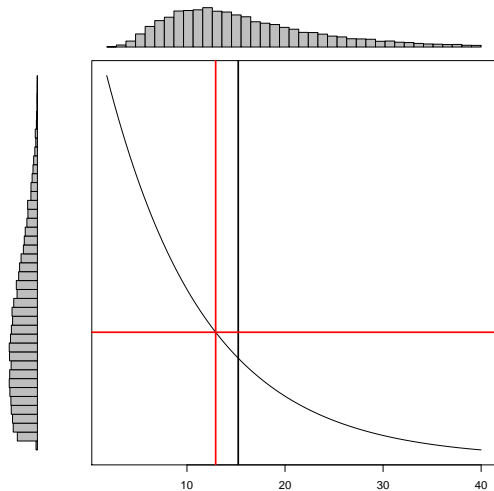
Filtered means

- ▶ Many things we know about are examples of filtered means
 - ▶ Geometric mean (log function)
 - ▶ Harmonic mean (reciprocal function)
 - ▶ Root mean square (square)

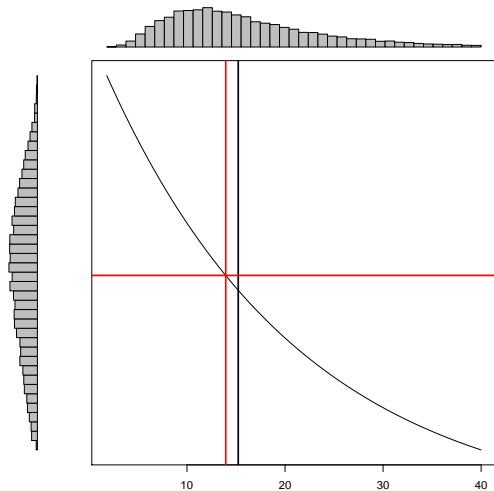
Linking framework

- ▶ Epidemic speed (r) is a *product*:
 - ▶ (something to do with) generation speed \times
 - ▶ (something to do with) epidemic strength
- ▶ In particular:
 - ▶ $r = (1/\hat{G}) \times \log(\mathcal{R})$
 - ▶ \hat{G} is the effective mean generation time

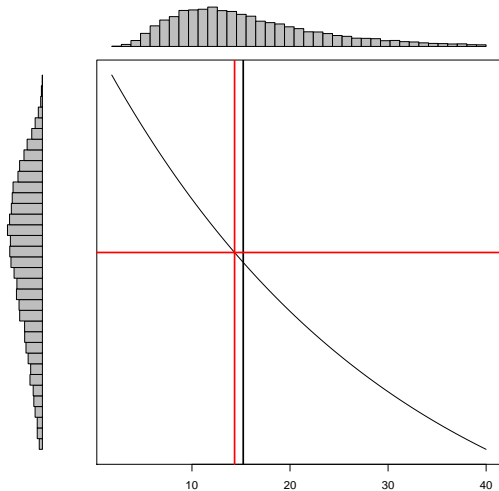
$$C = 1/r = 10d$$



$$C = 1/r = 20d$$



$$C = 1/r = 30d \text{ (present)}$$



Filtered means have intuitive properties

- ▶ Shifts in distribution shift the mean about how you would expect
 - ▶ More late transmission means longer \hat{G}
 - ▶ Longer \hat{G} means higher \mathcal{R} for a given r
- ▶ As distribution gets narrower, \hat{G} increases toward the mean \bar{G}
- ▶ As distribution gets wider, \hat{G} decreases
 - ▶ Scientific interpretation?

The filtering function

- ▶ $\exp(-r\hat{G}) = \langle \exp(-r\tau) \rangle_g,$
- ▶ \hat{G} is the mean of the generation distribution $g(\tau)$...
- ▶ Filtered by the discount function associated with the rate of exponential growth of the epidemic
 - ▶ i.e., the relative importance of a contribution at that time

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

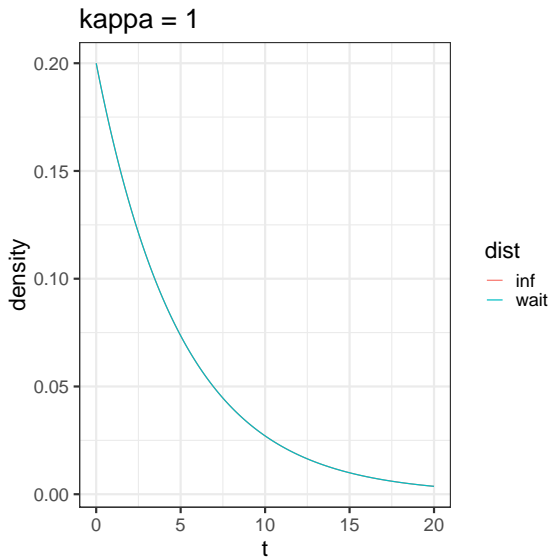
Problems

- ▶ The filtered mean has drawbacks
- ▶ \hat{G} depends on r as well as G
- ▶ How is
 - ▶ $\mathcal{R} = \exp(r\hat{G})$
- ▶ Consistent with the result from ODEs
 - ▶ $\mathcal{R} = 1 + r\bar{G}$?

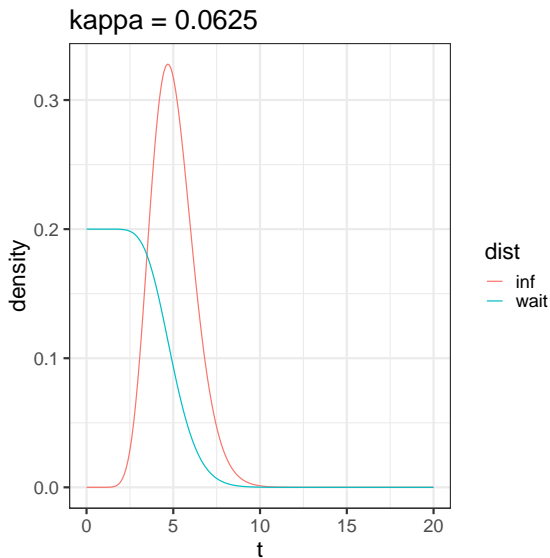
Infectious times and waiting times

- ▶ If the average infectious time is 5 days, what is the average generation time?
 - ▶ The average latent period plus the infectious-waiting period
- ▶ What is the average infectious-waiting period?
 - ▶ $5d(1 + \kappa)/2$
 - ▶ κ measures the relative variation of the infectious period
 - ▶ The waiting period is not the infectious period
 - ▶ The exponential distribution is trying to trick you!

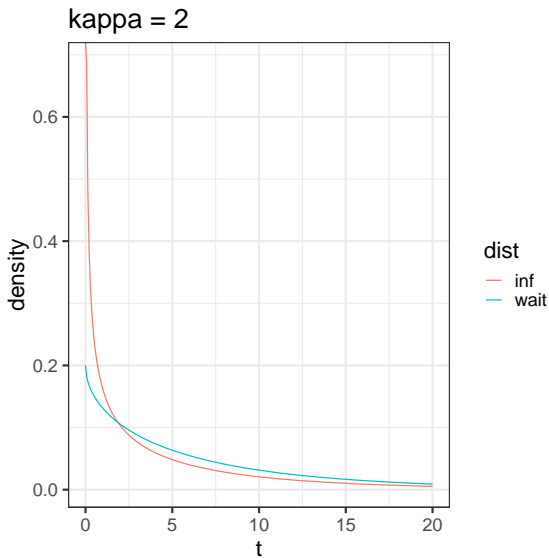
Infectious and waiting periods (present)



Infectious and waiting periods



Infectious and waiting periods



An approximation

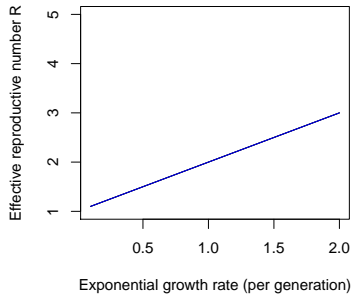
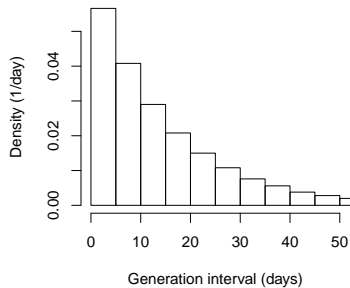
- ▶ We make the $r\mathcal{R}$ link with a moment approximation
- ▶ Define $\kappa = \sigma_G^2 / \mu_G^2$ – the squared coefficient of variation of the generation distribution
- ▶ $\mathcal{R} \approx (1 + r\kappa\bar{G})^{1/\kappa}$
 - ▶ Equal when $g(\tau)$ has a gamma distribution
 - ▶ Simple and straightforward
 - ▶ When is it a useful approximation?

Compound-interest interpretation

- ▶ Define $\mathcal{R} \approx (1 + r\kappa\bar{G})^{1/\kappa} \equiv X(r\bar{G}; 1/\kappa)$
- ▶ X is the compound-interest approximation to the exponential
 - ▶ Linear when $\kappa = 1$ (i.e., when g is exponential)
 - ▶ Approaches exponential as $\kappa \rightarrow 0$

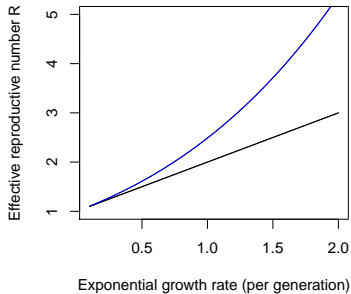
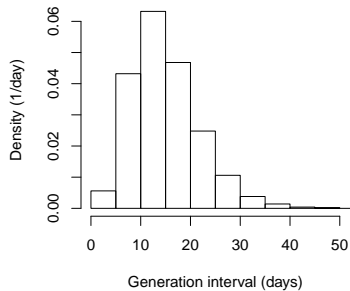
Moment approximation (present)

Approximate generation intervals



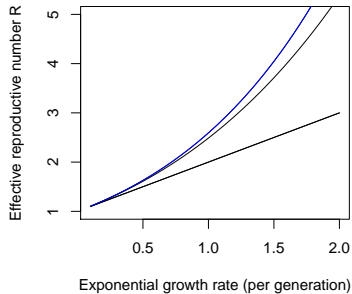
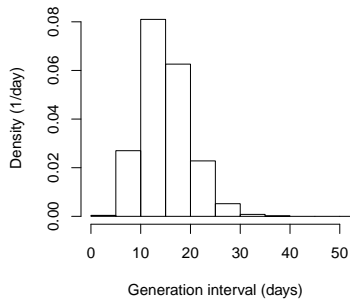
Moment approximation (present)

Approximate generation intervals



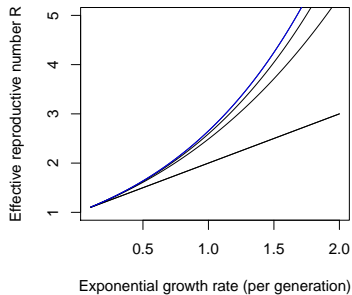
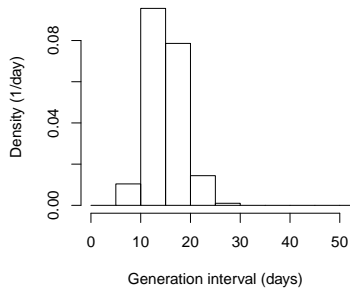
Moment approximation (present)

Approximate generation intervals



Moment approximation

Approximate generation intervals



Qualitative response

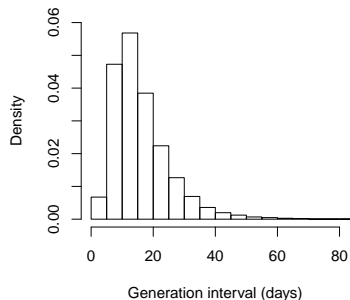
- ▶ For a given value of \bar{G} , smaller values of κ mean:
 - ▶ less variation in generation interval
 - ▶ less compounding of growth
 - ▶ greater \mathcal{R} required for a given r

Fitting to Ebola

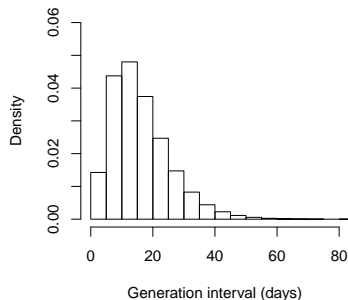
- ▶ Simulate generation intervals based on data and approach from WHO report
- ▶ Use both lognormals and gammas
 - ▶ WHO used gammas
 - ▶ Lognormals should be more challenging

Approximating the distribution

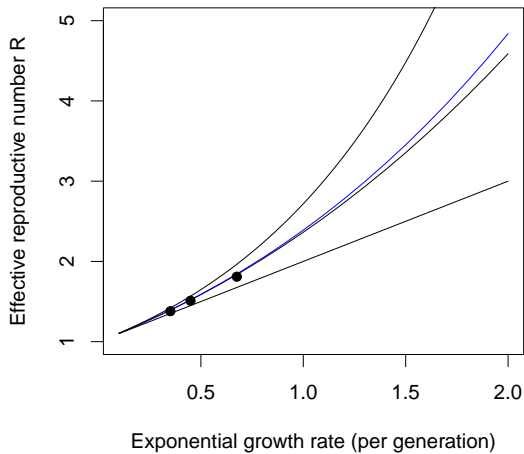
Lognormal SEIR



Single-gamma approximation



Approximating the curve



Linking framework

- ▶ Epidemic speed (r) is a *product*:
 - ▶ (something to do with) generation speed \times
 - ▶ (something to do with) epidemic strength
- ▶ In particular:
 - ▶ $r \approx (1/\bar{G}) \times \ell(\mathcal{R}; \kappa_g)$
 - ▶ ℓ is the inverse of X

Other diseases

- ▶ This approximation works suspiciously well for measles parameters
- ▶ Noticeably less well for rabies parameters
 - ▶ Can be improved using gamma-based estimates of the moments

Summary

- ▶ For many practical applications:
 - ▶ Estimating the mean generation interval is not enough
 - ▶ But estimating the mean and CV may be enough
 - ▶ This can also allow us to address our uncertainty
- ▶ Filtered mean is useful for qualitative explanations
 - ▶ e.g., Ebola burial

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Generation intervals through time

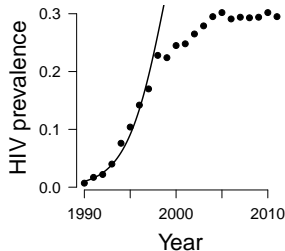
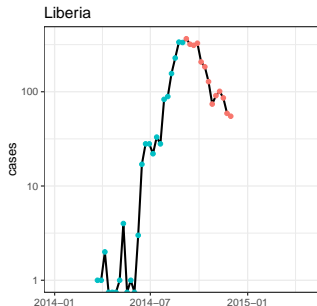
- ▶ Generation intervals can be estimated by:
 - ▶ Observing patients:
 - ▶ How long does it take to become infectious?
 - ▶ How long does it take to recover?
 - ▶ What is the time profile of infectiousness/activity?
 - ▶ Contact tracing
 - ▶ Who (probably) infected whom?
 - ▶ When did each become ill (serial interval)?

Types of interval

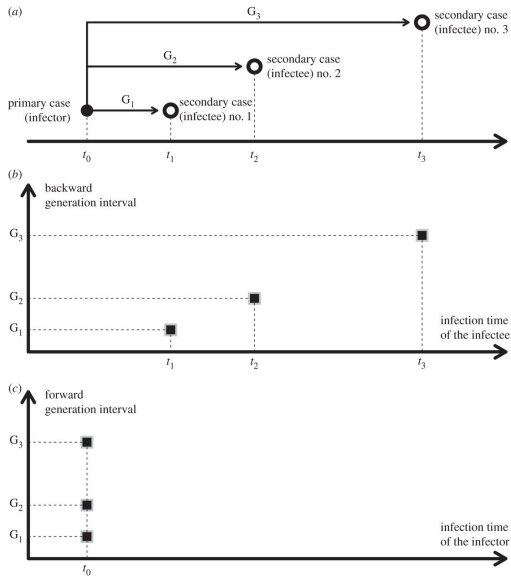
- ▶ Contact-tracing intervals look systematically different, depending on when you observe them.
- ▶ Define:
 - ▶ *Intrinsic interval*: How infectious is a patient at time τ after infection?
 - ▶ *Forward interval*: When do people infected at a particular time infect others?
 - ▶ *Backward interval*: When were the people who infect at a particular time infected?

Growing epidemics

- ▶ Generation intervals look *shorter* at the beginning of an epidemic
 - ▶ A disproportionate number of people are infectious right now
 - ▶ They haven't finished all of their transmitting
 - ▶ We are biased towards observing faster events



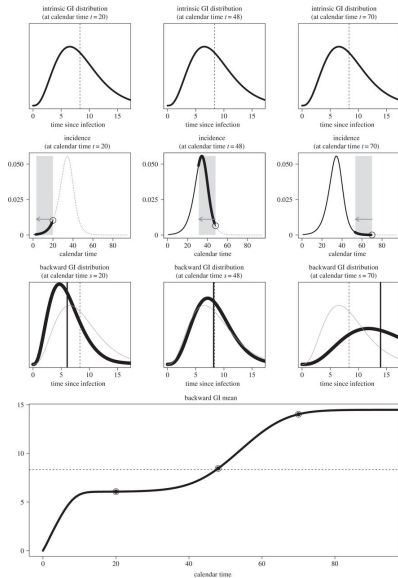
Forward and backward intervals



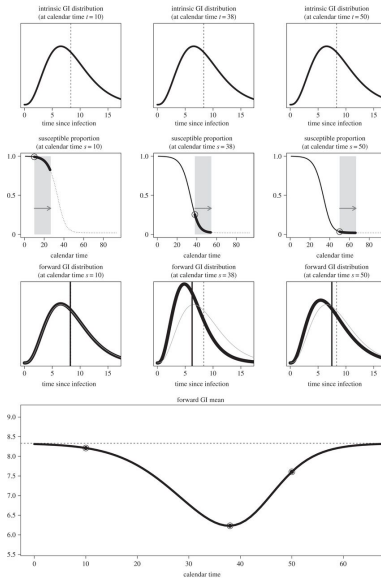
Correcting

- ▶ Infection events: someone infected at time s is infecting someone at time t
 - ▶ $i_s(t) = S(t)k(t-s)i(s)$
- ▶ Backward intervals
 - ▶ Who infected the people infected at time t ?
 - ▶ $\propto k(t-s)i(s)$
 - ▶ Depends on k , but also on changes in $i(s)$
- ▶ Forward intervals
 - ▶ Who did the people infected at time s infect?
 - ▶ $\propto S(t)k(t-s)$
 - ▶ Depends on k , but also on changes in $S(t)$

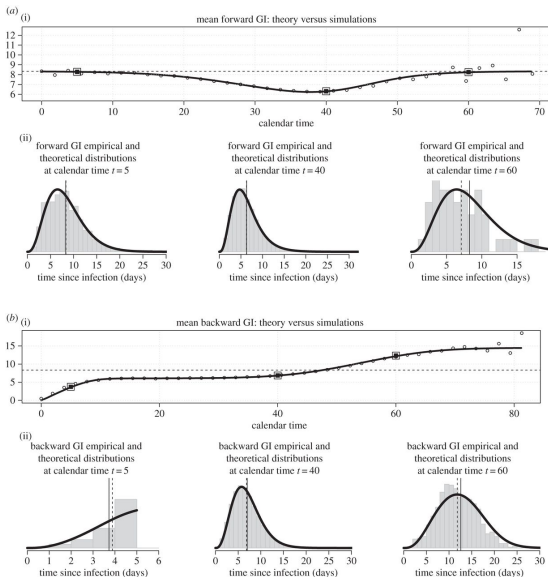
What changes backward intervals? (present)



What changes forward intervals? (present)



Theory and simulation



Champredon and Dushoff, 2015. DOI:10.1098/rspb.2015.2026

Conclusion

- ▶ Backward intervals change if the number of infectious individuals is changing as you look back
- ▶ Forward intervals change if the number of *susceptible* individuals is changing as you look forward
- ▶ Lack of care in defining generation intervals can lead to bias
 - ▶ In particular, generation intervals look short during an epidemic outbreak
 - ▶ Makes diseases look less dangerous!
 - ▶ These biases can be corrected

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

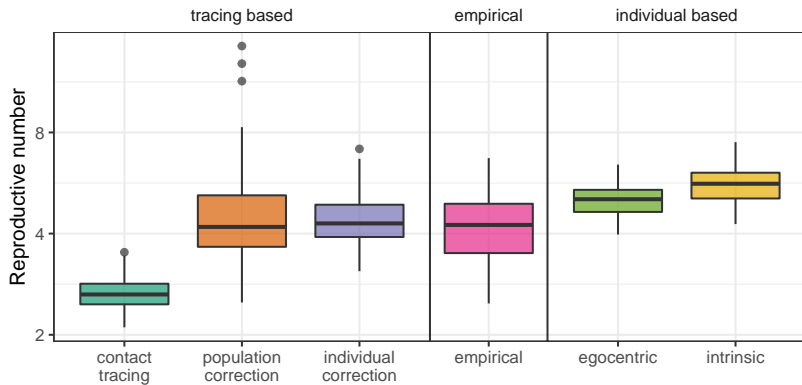
Moment approximations

Generation intervals through time

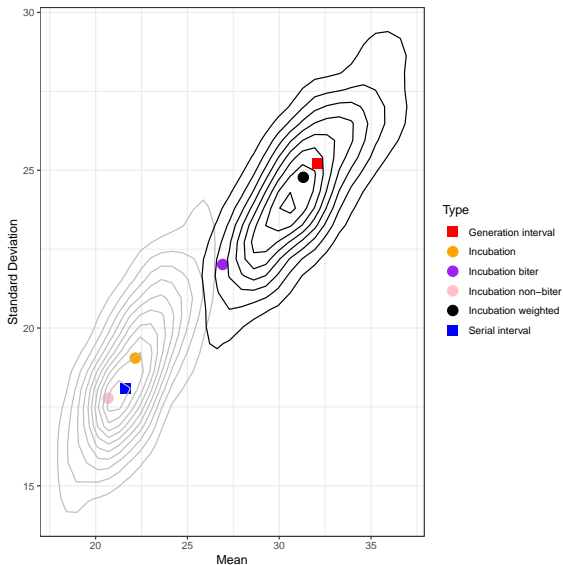
Conclusion

Renewal math (extra)

Spatial struture



Individual-level heterogeneity



Summary

- ▶ Generation intervals are often taken for granted
- ▶ We need better methods for defining different measures of disease generations
 - ▶ We need to consider our *uncertainty* about generations when making conclusions
- ▶ Generation intervals are fun conceptually, mathematically and practically

Outline

Introduction

Linking strength and speed

The link

Renewal-equation models

Estimating \mathcal{R}

Effective generation times

Moment approximations

Generation intervals through time

Conclusion

Renewal math (extra)

Compartments vs. cohorts

- ▶ We have a some simplified biological assumptions about how a disease spreads
- ▶ We can implement these assumptions into a model:
 - ▶ Track **compartments**: S, I, R, ...
 - ▶ *or* track **cohorts**: a group of people infected at the same time

Cohort approach

- ▶ Model



$$\frac{dS}{dt} = \mu S - \beta SI/N$$



$$\frac{dI}{dt} = \beta SI/N - \gamma I$$

- ▶ What happens to a cohort infected at time 0?



$$\frac{dI}{d\tau} = -\gamma I$$



$$I(\tau) = I(0) \exp(-\gamma\tau)$$

- ▶ We can write cohort equations for more complicated models as well

Another view of the model

- ▶ Model **incidence** i :



$$\frac{dS}{dt} = \mu S - i(t)$$



$$\frac{dI}{dt} = i(t) - \gamma I$$



$$i(t) = \beta SI/N$$

Cohort approach

- ▶ We can use standard methods for the differential equation:



$$\frac{dI}{dt} = i(t) - \gamma I$$

- ▶ or we can just write down the answer using a cohort approach:



$$I(t) = \int I(t - \tau, \tau) d\tau$$



$$= \int i(t - \tau) \exp(-\gamma\tau) d\tau$$

- ▶ This answer makes *biological* sense

Cohort-based equation

- ▶ We can eliminate I and write:



$$\frac{dS}{dt} = \mu S - i(t)$$



$$i(t) = \frac{S}{N} \int \beta i(t - \tau) \exp(-\gamma \tau) d\tau$$

- ▶ This is the *same model*

- ▶ Same assumptions, same dynamics

- ▶ We can generalize our compartmental assumptions:



$$i(t) = \frac{S}{N} \int i(t - \tau) k(\tau) d\tau$$

Renewal equation



$$i(t) = \frac{S}{N} \int i(t - \tau)k(\tau)d\tau$$

- ▶ $k(\tau)$ is the infection “kernel” – it describes how an incident (new, occurring) case tends to cause other incident cases over time
 - ▶ As a function of time since infection
- ▶ What are the advantages or disadvantages of this cohort-based approach, compared to a general compartmental model?
- ▶ How would you estimate an infection kernel?