Physics of Life Data Epidemiology

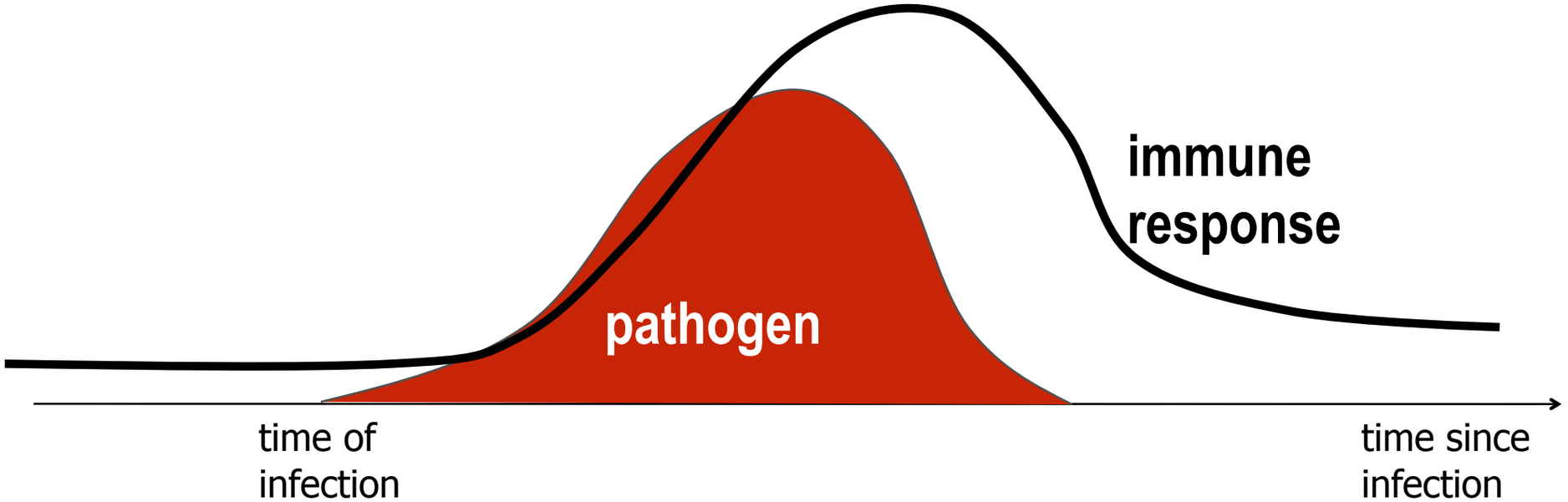
Lect 4: Infection time scales

Chiara Poletto
mail: chiara.poletto@unipd.it
web: chiara-poletto.github.io
bsky: @chpoletto.bsky.social

disease natural history

Take for computations of incubation periods, generation time and serial interval

- majority of infections have an incubation period
- the distribution of delays from being infected and starting causing new infections is Gamma/Negative Binomial/Weibull



[Keeling & Rohani, Modeling Infectious Diseases (2008)]

Exponential distribution of delays

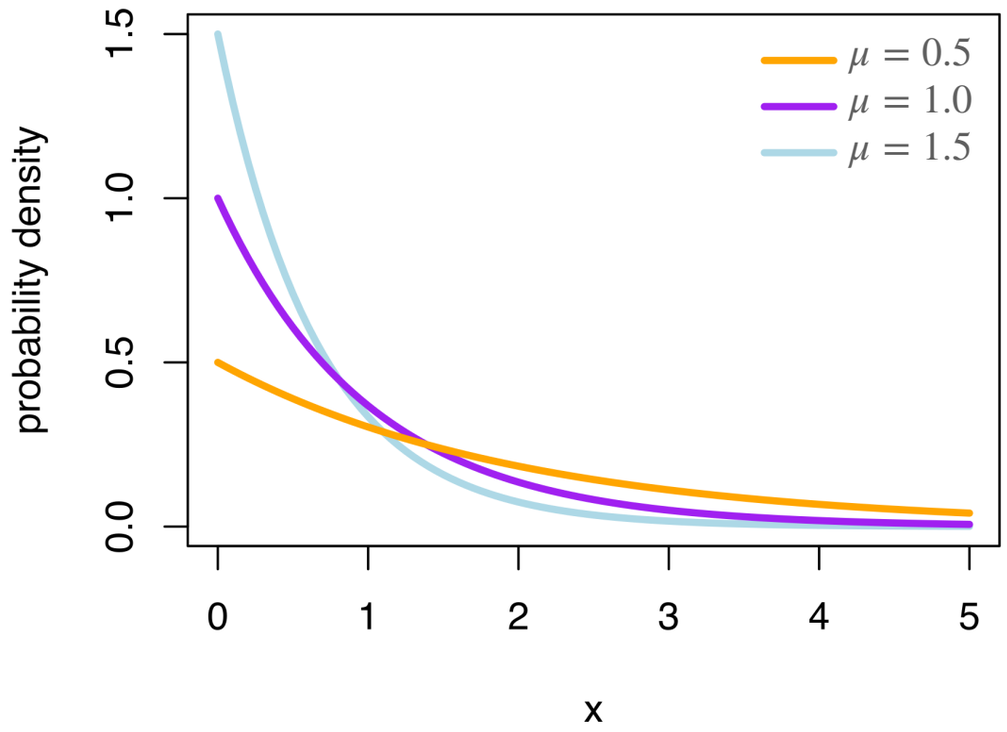
All the models till now assume that β and μ are constant rates → transition probabilities at time t do not depend on the time spent in that state. They depend only on the state of the system at time t and not on past states → **Memoryless - Markov process**

The transition is Poisson process → time from one transition to another exponential distribution

$$P(x) = \mu e^{-\mu x}$$

- Most probable duration of the disease = 0
- Probability decreases with time

Not realistic!



[By Newystats - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=80359746]

Gamma distribution of delays

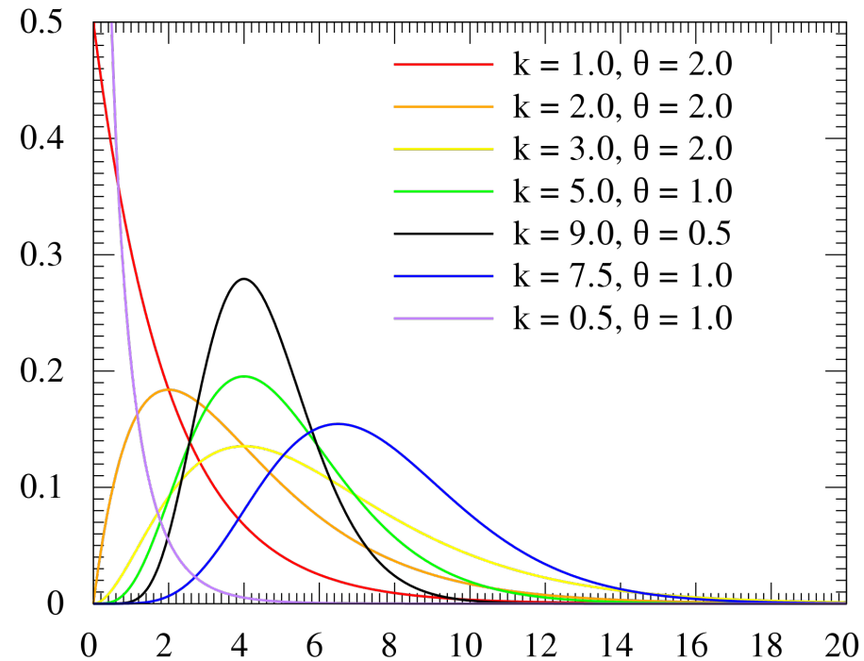
Gamma distribution

$$P(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

With $\Gamma(x)$ the gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$

k shape, θ scale

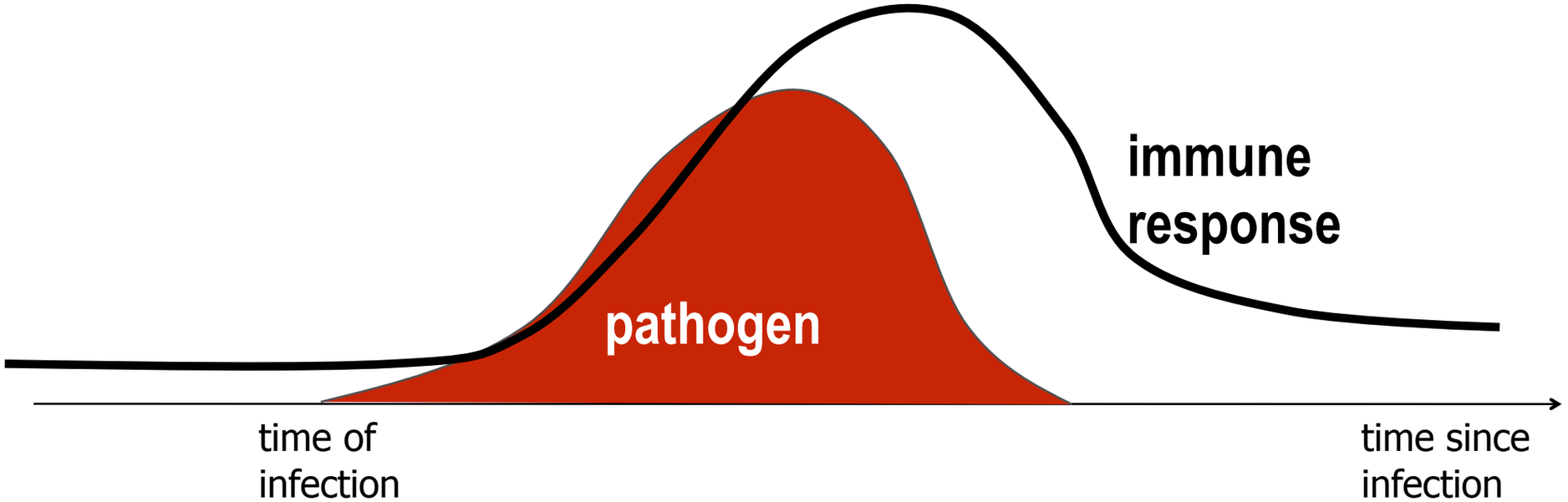
Mean $k\theta$, variance $k\theta^2$



disease natural history

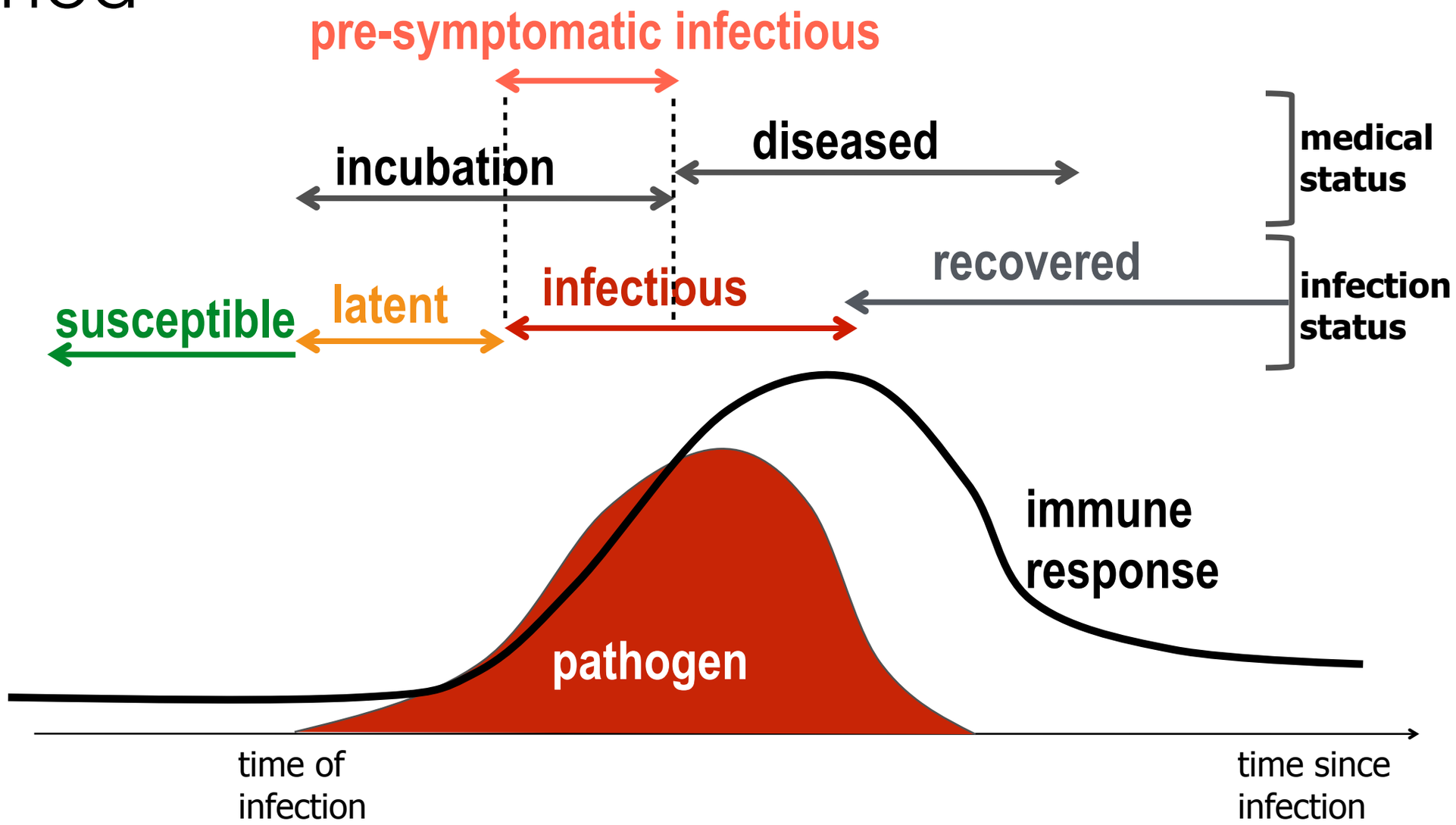
how to properly incorporate information on generation time, serial interval, incubation period into dynamical models?

- refined compartmental model
- different approach based on integral equations



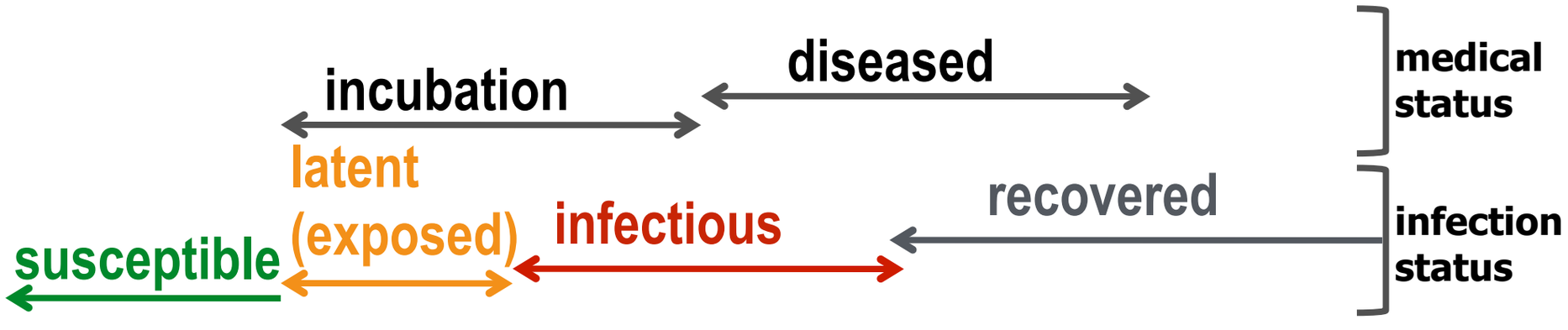
[Keeling & Rohani, Modeling Infectious Diseases (2008)]

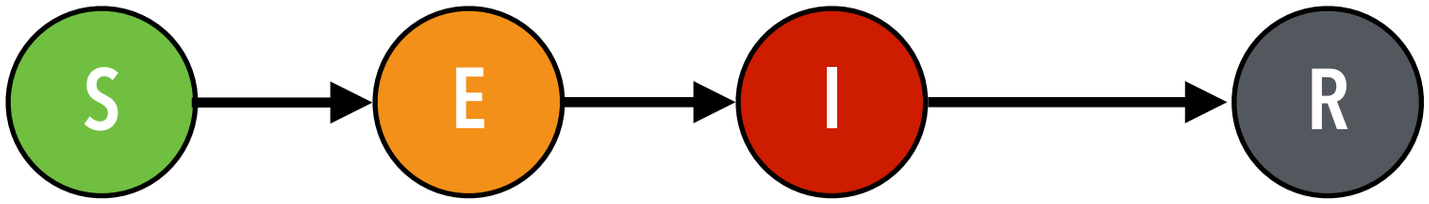
disease natural history: incubation period, latency period, infectious period



[Keeling & Rohani, Modeling Infectious Diseases (2008)]

SEIR





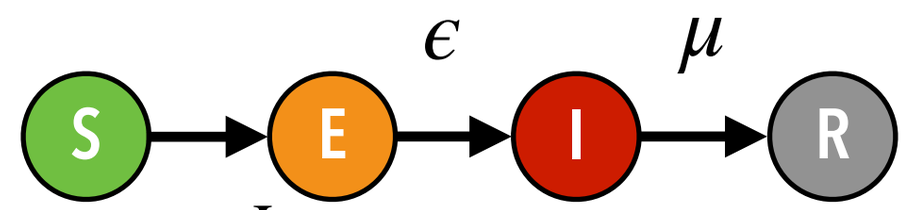
Latent stage: varies a lot according to the pathogen (few hours to years).

Sometimes can be neglected

[Keeling & Rohani, Modeling Infectious Diseases (2008)]

SEIR

$$\frac{ds}{dt} = -\beta\langle k\rangle si$$
$$\frac{de}{dt} = \beta\langle k\rangle si - \epsilon e$$
$$\frac{di}{dt} = \epsilon e - \mu i$$
$$\frac{dr}{dt} = \mu i$$



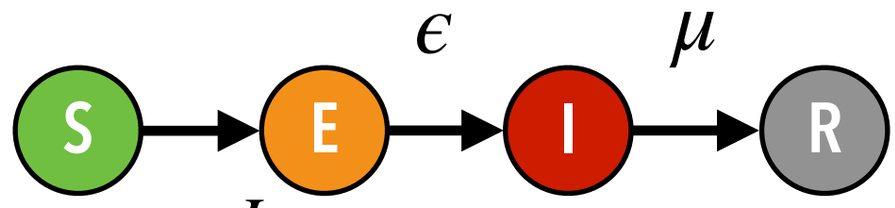
$$\beta\langle k\rangle\frac{I}{N} =$$

Force of Infection (FOI)

Definitions:

- $\tau_E = \epsilon^{-1}$ length of the exposed stage

SEIR



$$\beta\langle k\rangle\frac{I}{N} =$$

Force of Infection (FOI)

Definitions:

- $\tau_E = \epsilon^{-1}$ length of the exposed stage

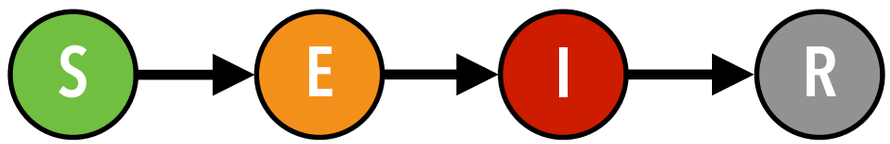
Same final state as the SIR

diving ds by dr we get

$$\frac{ds}{dr} = \frac{-\beta\langle k\rangle s}{\mu}$$

$$1 - r(\infty) = s_0 e^{-r(\infty)\frac{\beta\langle k\rangle}{\mu}}$$

SEIR



transient dynamics

Matrix equation around the disease-free equilibrium $(s^*, e^*, i^*, r^*) = (1,0,0,0)$, i.e. the initial state

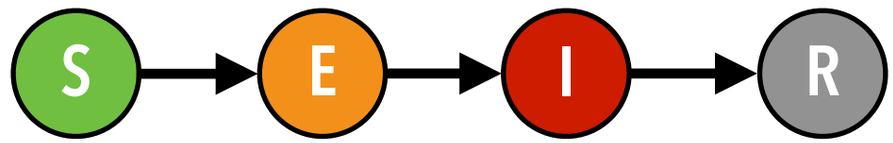
$$\begin{pmatrix} \frac{ds}{dt} \\ \frac{de}{dt} \\ \frac{di}{dt} \\ \frac{dr}{dt} \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\beta\langle k\rangle & 0 \\ 0 & -\epsilon & \beta\langle k\rangle & 0 \\ 0 & \epsilon & -\mu & 0 \\ 0 & 0 & \mu & 0 \end{pmatrix} \begin{pmatrix} s \\ e \\ i \\ r \end{pmatrix}$$

From the eigenvalues' analysis of the Jacobian we have

$$i(t) \simeq i_0 e^{\frac{1}{2}\left(\sqrt{4(R_0-1)\epsilon\mu + (\epsilon+\mu)^2} - (\epsilon+\mu)\right)t}$$

[Keeling & Rohani, Modeling Infectious Diseases (2008)]

SEIR



$$i(t) \simeq i_0 e^{\frac{1}{2}\left(\sqrt{4(R_0-1)\epsilon\mu+(\epsilon+\mu)^2}-(\epsilon+\mu)\right)t}$$

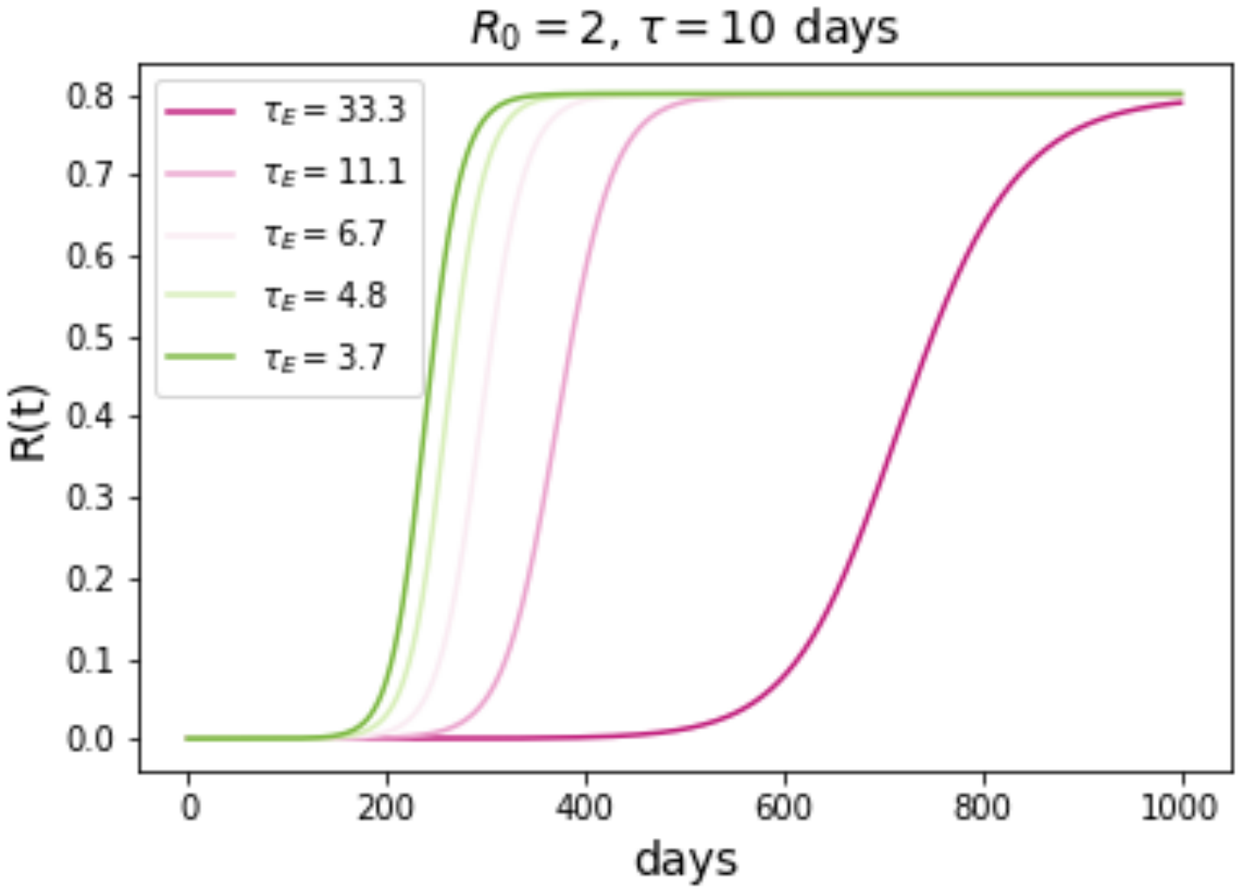
Exponential growth $G = \frac{1}{2}\left(\sqrt{4(R_0-1)\epsilon\mu+(\epsilon+\mu)^2}-(\epsilon+\mu)\right)$.

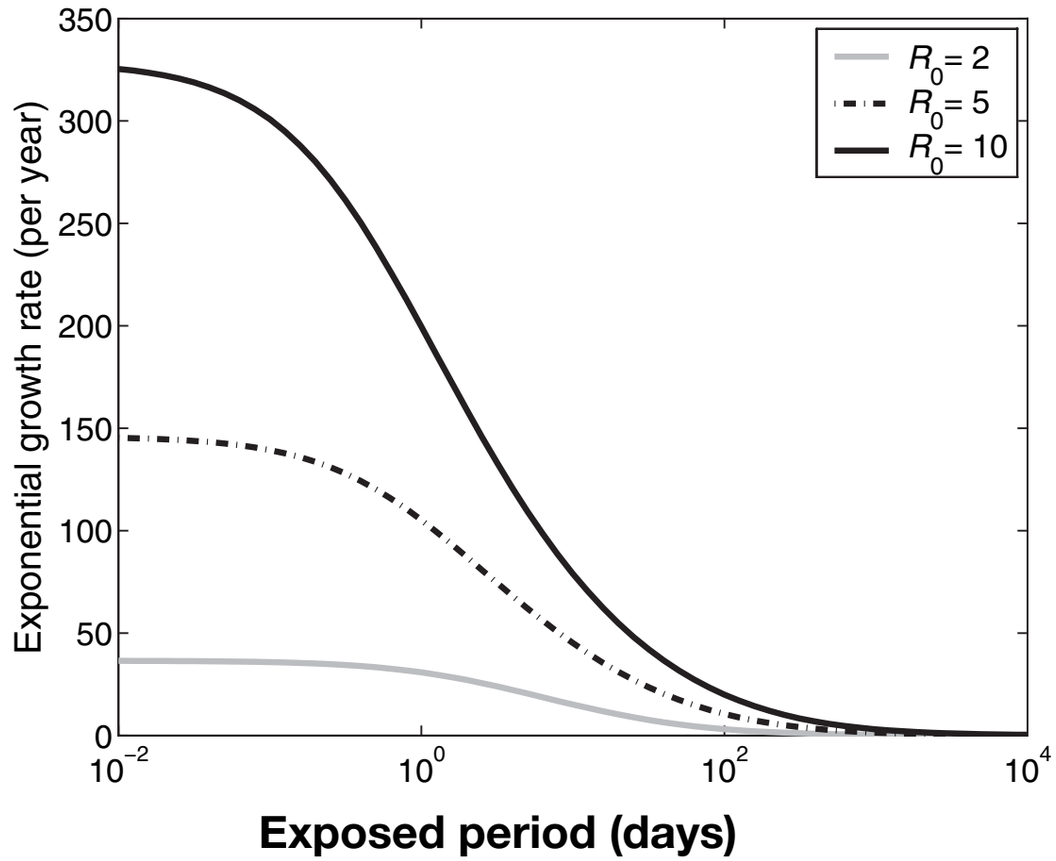
$G > 0$ if $R_0 > 1$, same threshold behaviour as the SIR

$G_{\mathbf{SEIR}} < G_{\mathbf{SIR}}$, the exposed stage adds a delay on the dynamics.

[Keeling & Rohani, Modeling Infectious Diseases (2008)]

SEIR





[Keeling & Rohani, Modeling Infectious Diseases (2008)]

Accounting for the incubation period –> slower growth

Fitting the data without considering the exposed period –> under-estimation of R_0

How to account for a more realistic distribution of delays

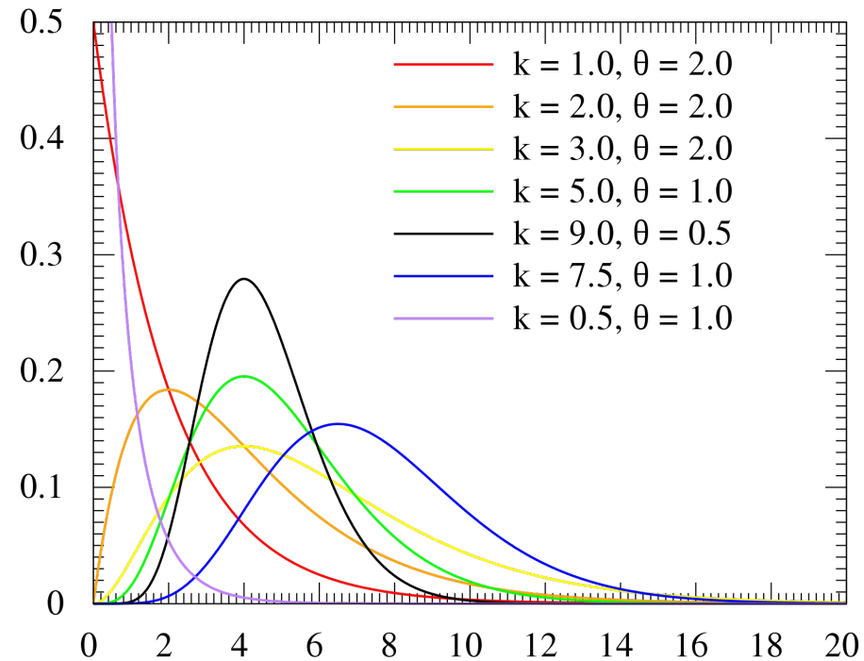
Gamma distribution

$$P(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

With $\Gamma(x)$ the gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$

k shape, θ scale

Mean $k\theta$, variance $k\theta^2$



[By Gamma_distribution_pdf.png: MarkSweep and Cburnettderivative work: Autopilot (talk) - Gamma_distribution_pdf.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=10734916]

Erlang distribution: $P(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$

Gamma when the shape k is an integer, with $\lambda = 1/\theta$ is the inverse of the scale

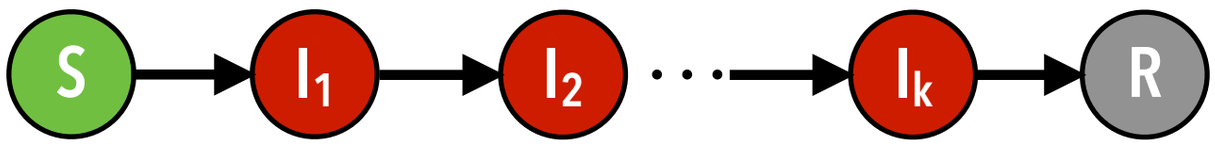
Erlang distribution of shape K is the convolution of K exponential distributions

the exponentials have mean $\theta = 1/\lambda$

Gamma trick

to obtain Erlang distributed incubation and infectious period within the rate equations framework

We neglect here for simplicity the exposed period



$$\frac{ds}{dt} = -\beta si$$

$$\frac{di_1}{dt} = \beta si - K\mu i_1$$

$$\frac{di_2}{dt} = K\mu i_1 - K\mu i_2$$

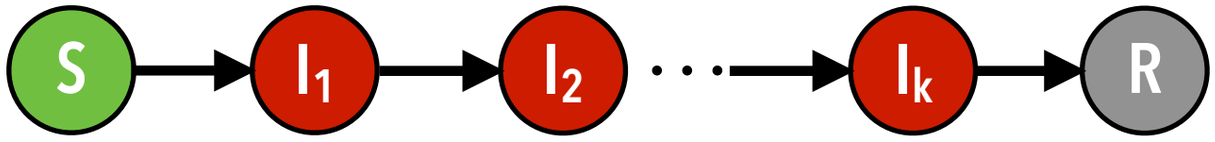
⋮ The rate of each *I* transition is $K\mu$

$$\frac{dr}{dt} = K\mu i_k$$

with $i = \sum_{k=1}^{K} i_k$

(*) In this chapter I will *adsorb* $\langle k \rangle$ in β, i.e. $\beta\langle k \rangle$ becomes β

Gamma trick

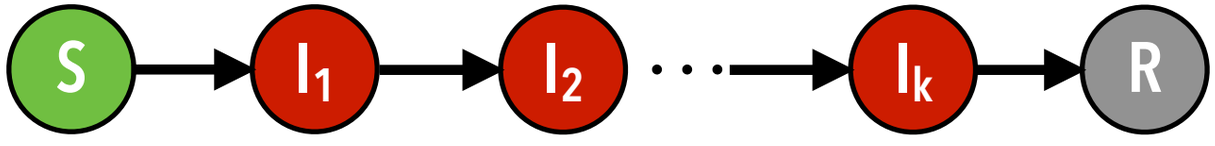


- Infectious period distribution

$$P(\tau) = \frac{(\mu K)^K}{\Gamma(K)} \tau^{K-1} e^{-\mu K \tau}$$

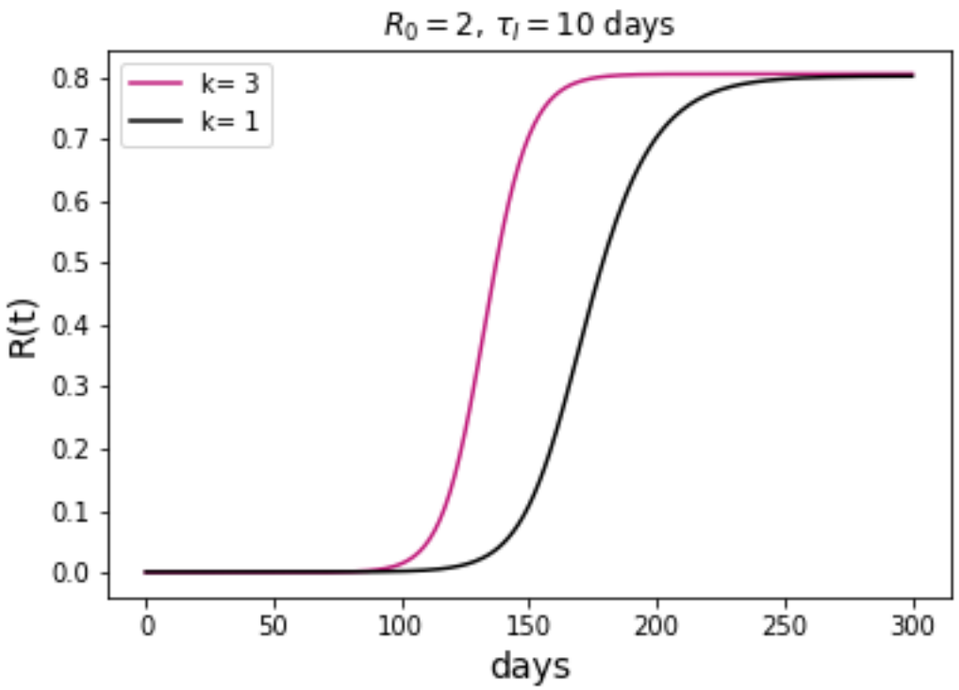
- Mean: still $\frac{1}{\mu}$ but shape totally different
- Special cases:
  - $K = 1$ Exponential
  - $K \to \infty$ Fixed (delta function)

Gamma trick



- We still have $R_0 = \beta/\mu$
- Final epidemic always satisfies $r_\infty = 1 - e^{-R_0 r_\infty}$
- Early growth:
  - Growing faster and shorter epidemic duration
  - $i(t) \simeq i_0 e^{Gt}$, relationship between G and R_0 changes. From the solution of the characteristic equation we have

$$R_0 = \frac{G}{\mu\left(1 - \left(\frac{G}{K\mu} + 1\right)^{-K}\right)}$$



Gamma trick

SARS Epidemic of 2003

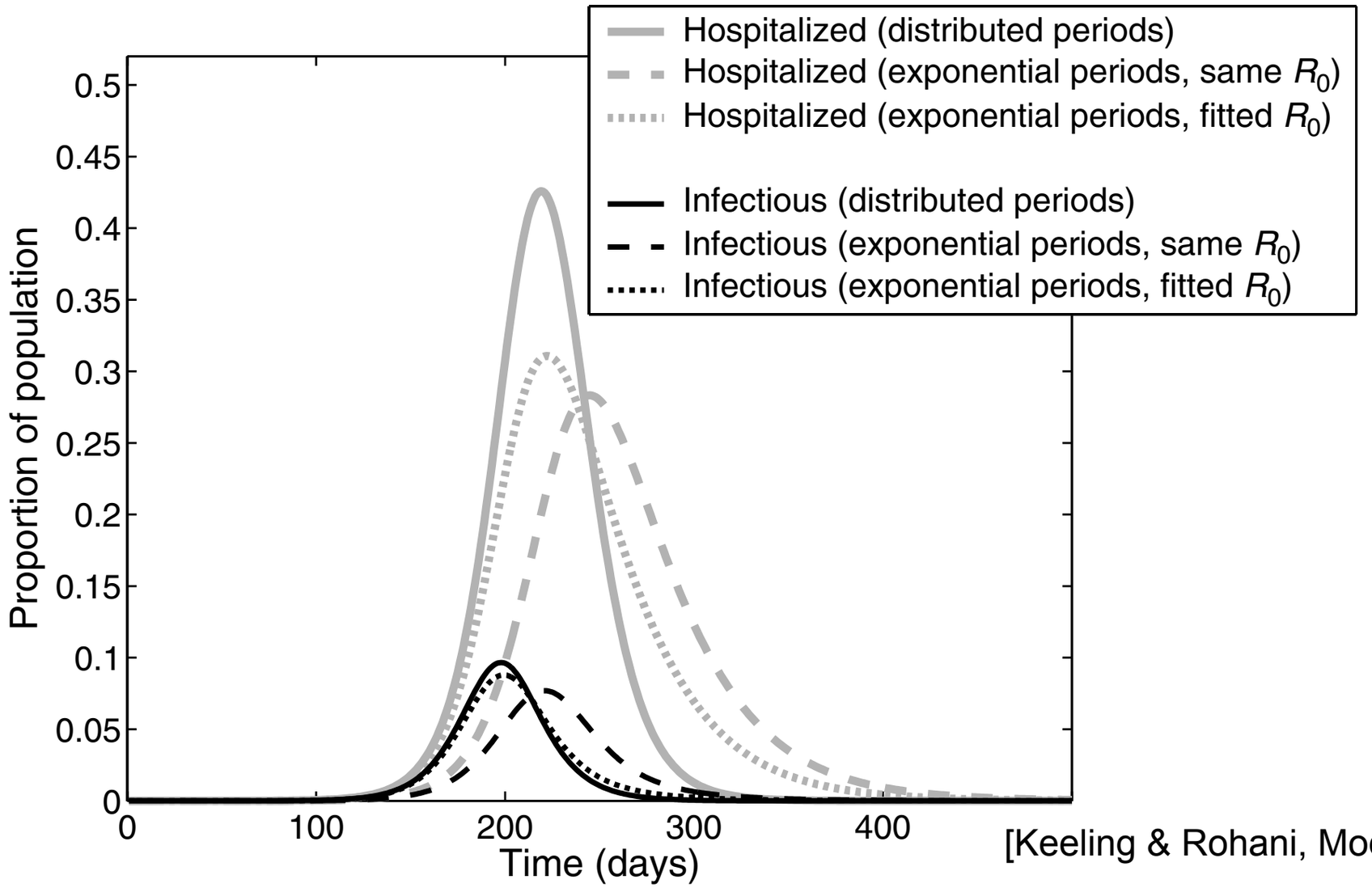
Incubation period SARS= latency period= 6 days

Duration of infectious at home= 4 days

Duration of hospitalisation= 23 days if they recover, 36 days if they die

Number of hospitalised? Risk of hospital saturation?

From observations you get that the delays are Gamma distributed



[Keeling & Rohani, Modeling Infectious Diseases (2008)]

Renewal equation

change of notation!

generation time, stochastic variable: t_g -> τ

average generation time: τ -> T_c

incidence at time t (number of individuals): $i(t)$

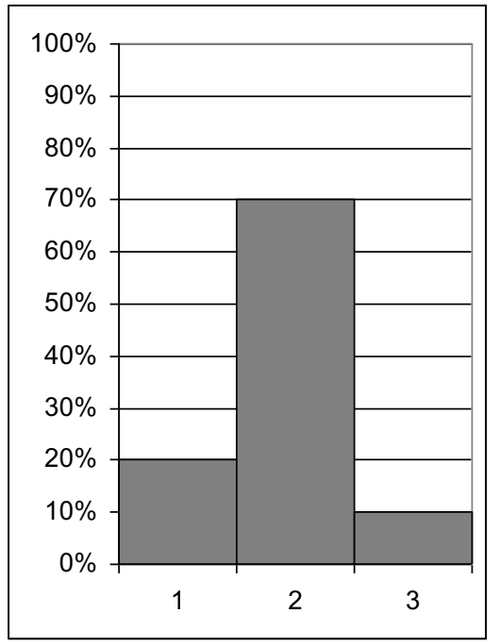
Wallinga, Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers Proc. R. Soc B (2007) 274, 599

Grassly, N., Fraser, C. Mathematical models of infectious disease transmission. Nat Rev Microbiol 6, 477–487 (2008)

- Less convenient than compartmental models to simulate an outbreak

- provides mathematical understandings on the relationship between exponential growth and basic reproductive ratio

Renewal equation

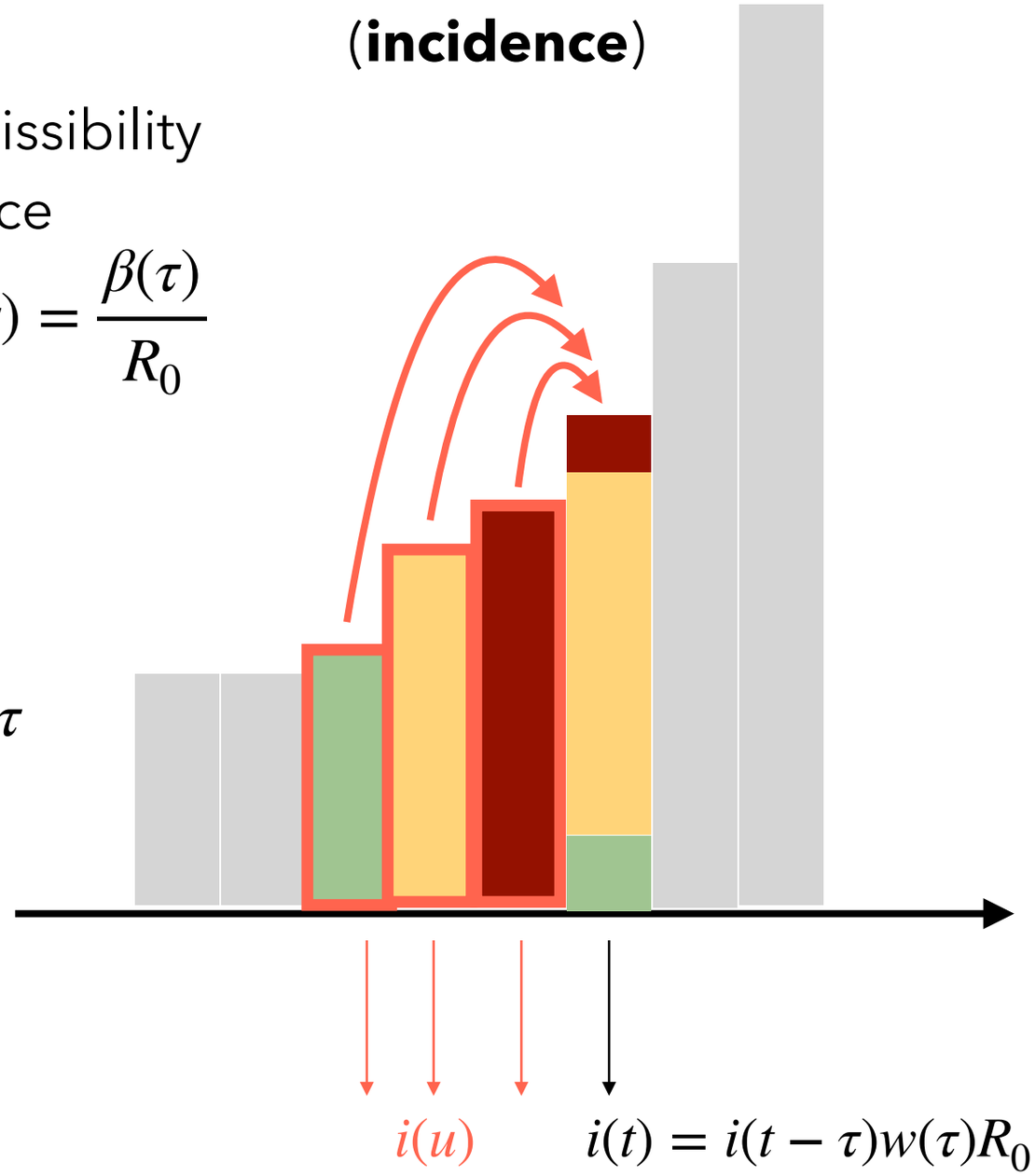


generic generation time distribution $w(\tau)$, i.e. transmissibility

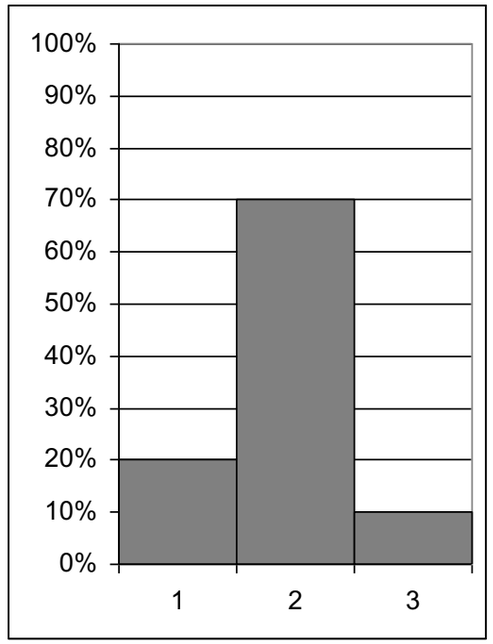
generic function of time since infection, $\beta(\tau)$, such that $w(\tau) = \dfrac{\beta(\tau)}{R_0}$

number of individuals they will infect at $t = u + \tau$ in a time interval $\delta\tau$: Poisson distributed with mean $i(u)\beta(\tau)\delta\tau = i(t-\tau)\beta(\tau)\delta\tau$

$i(u)$, individuals newly infected at time u (**incidence**)



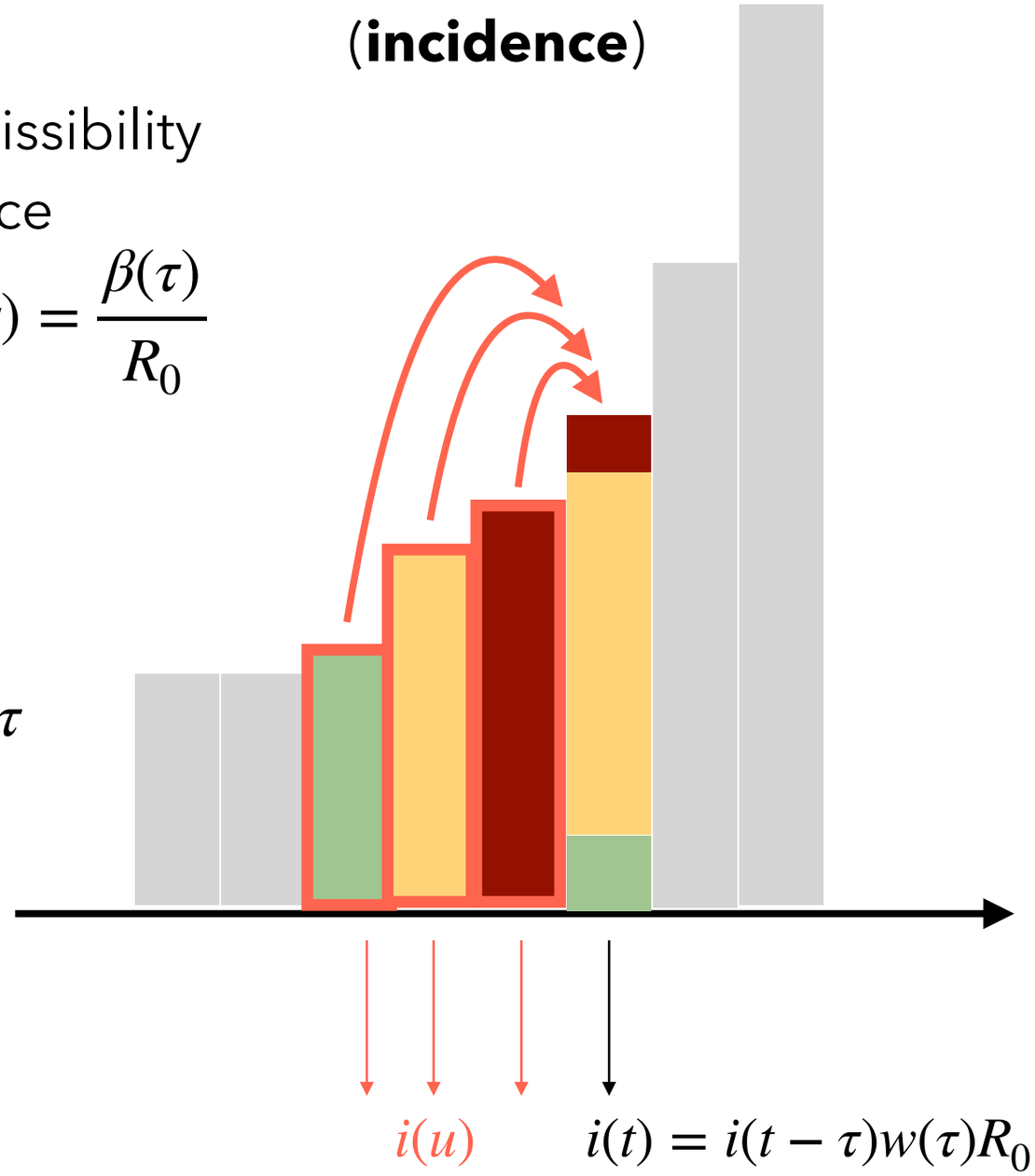
Renewal equation



generic generation time distribution $w(\tau)$, i.e. transmissibility

generic function of time since infection, $\beta(\tau)$, such that $w(\tau) = \dfrac{\beta(\tau)}{R_0}$

$i(u)$, individuals newly infected at time u (**incidence**)



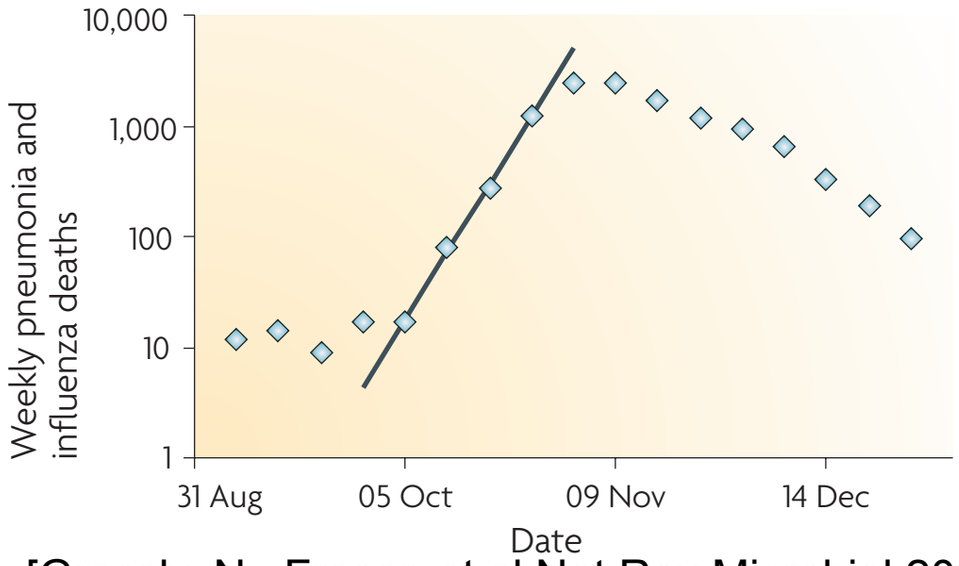
number of individuals they will infect at $t = u + \tau$ in a time interval $\delta\tau$: Poisson distributed with mean $i(u)\beta(\tau)\delta\tau = i(t-\tau)\beta(\tau)\delta\tau$

Renewal equation:

$$i(t) = \int_0^\infty i(t-\tau)\beta(\tau)d\tau$$

Lotka Euler equation (from population dynamics)

Renewal equation



[Grassly, N., Fraser, et al Nat Rev Microbiol 2008]

How do the generation time distribution affect the exponential growth?

$$i(t) = \int_0^\infty i(t-\tau)\beta(\tau)d\tau$$

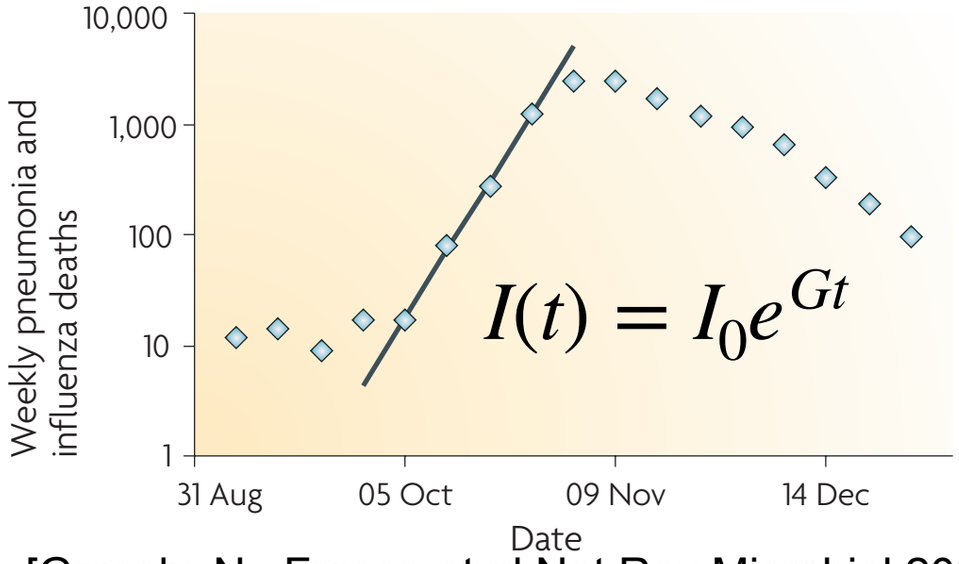
$$I(t) = I_0 e^{Gt}$$

$$i(t) = i(t-\tau)e^{G\tau} \quad \Rightarrow \quad i(t) = \int_0^\infty i(t)e^{-G\tau}\beta(\tau)\,d\tau$$

$$1 = \int_0^\infty e^{-G\tau}\beta(\tau)\,d\tau \quad \Rightarrow \quad \frac{1}{R_0} = \int_0^\infty e^{-G\tau}w(\tau)\,d\tau$$

$$w(\tau) = \frac{\beta(\tau)}{R_0}$$

R_0 and exponential growth



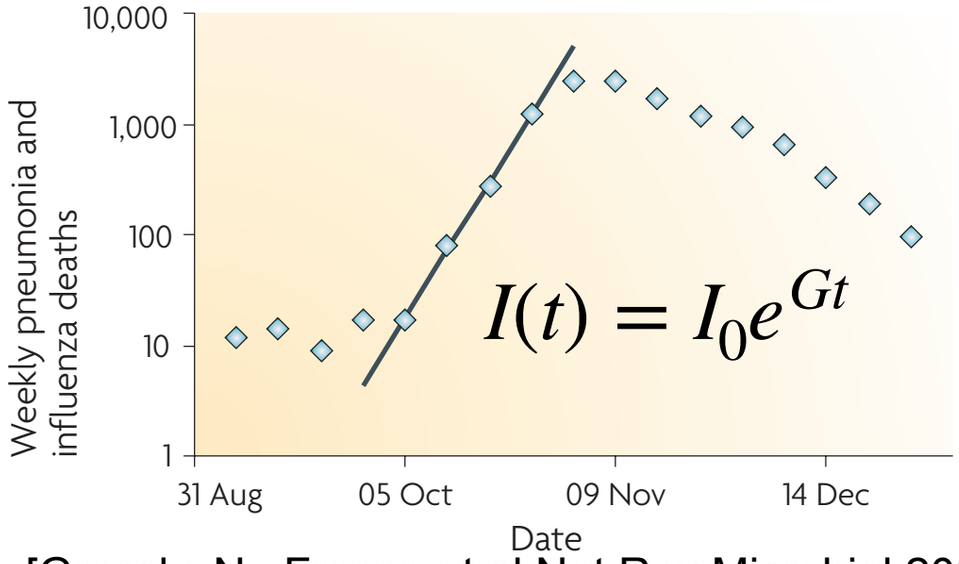
[Grassly, N., Fraser, et al Nat Rev Microbiol 2008]

$$\frac{1}{R_0} = \int_0^\infty e^{-G\tau} w(\tau)\, d\tau$$

Laplace transform of $w(\tau)$, i.e. moment generating function of $w(-\tau)$.

Moment generating function of $w(\tau)$: $\mathscr{M}_{w(\tau)}(z) = \int_0^\infty e^{z\tau} w(\tau)\, d\tau$

R_0 and exponential growth



[Grassly, N., Fraser, et al Nat Rev Microbiol 2008]

$$\frac{1}{R_0} = \int_0^\infty e^{-G\tau} w(\tau)\, d\tau$$

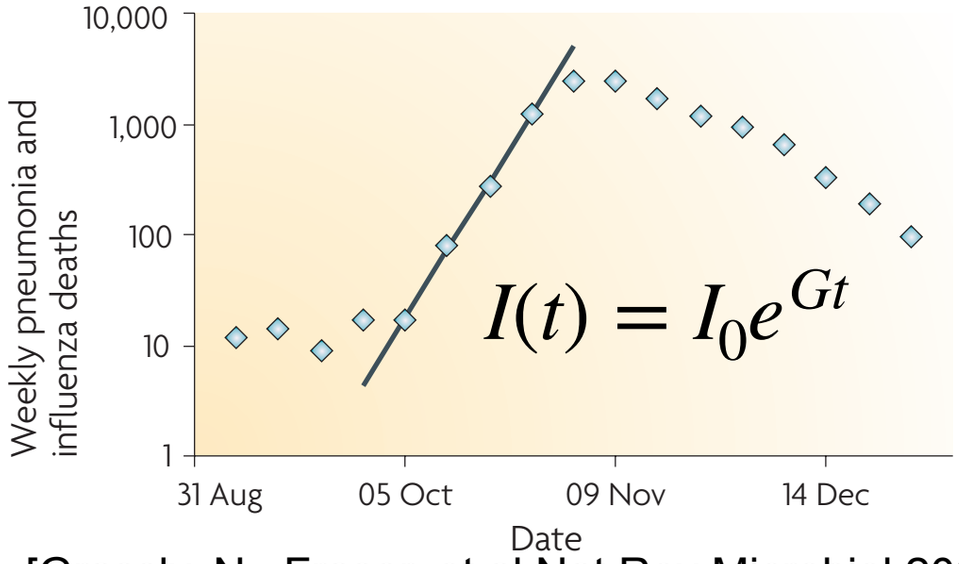
Laplace transform of $w(\tau)$, i.e. moment generating function of $w(-\tau)$.

Moment generating function of $w(\tau)$: $\mathscr{M}_{w(\tau)}(z) = \int_0^\infty e^{z\tau} w(\tau)\, d\tau$

Moment generating function:
alternative specification of a random variable probability distribution

$$\left.\frac{d^k \mathscr{M}_{w(\tau)}(z)}{dz^k}\right|_{z=0} = E(\tau^k)$$

R_0 and exponential growth



[Grassly, N., Fraser, et al Nat Rev Microbiol 2008]

$$\frac{1}{R_0} = \int_0^\infty e^{-G\tau} w(\tau)\, d\tau$$

Laplace transform of $w(\tau)$, i.e. moment generating function of $w(-\tau)$.

Moment generating function of $w(\tau)$: $\mathscr{M}_{w(\tau)}(z) = \int_0^\infty e^{z\tau} w(\tau)\, d\tau$

$$\Rightarrow R_0 = \frac{1}{\mathscr{M}_{w(\tau)}(-G)}$$

R_0 and exponential growth

$$R_0 = \frac{1}{\mathscr{M}_{w(\tau)}(-G)}$$

$w(\tau) = \mathcal{N}(T_c, \sigma^2) \Rightarrow R_0 = e^{G\,T_c - (1/2)\,G^2\,\sigma^2}$

$w(\tau) = \delta(T_c) \Rightarrow R_0 = e^{G\,T_c}$ **Upper bound of** R_0

If I add some variability on the generation time, a given G will correspond to a lower R_0

(If I add some variability on the generation time, a given R_0 will correspond to a higher G)

R_0 and exponential growth

$$R_0 = \frac{1}{\mathscr{M}_{w(\tau)}(-G)}$$

$w(\tau) = \mathcal{N}(T_c, \sigma^2) \Rightarrow R_0 = e^{G\,T_c - (1/2)\,G^2\,\sigma^2}$

$w(\tau) = \delta(T_c) \Rightarrow R_0 = e^{G\,T_c}$ **Upper bound of** R_0

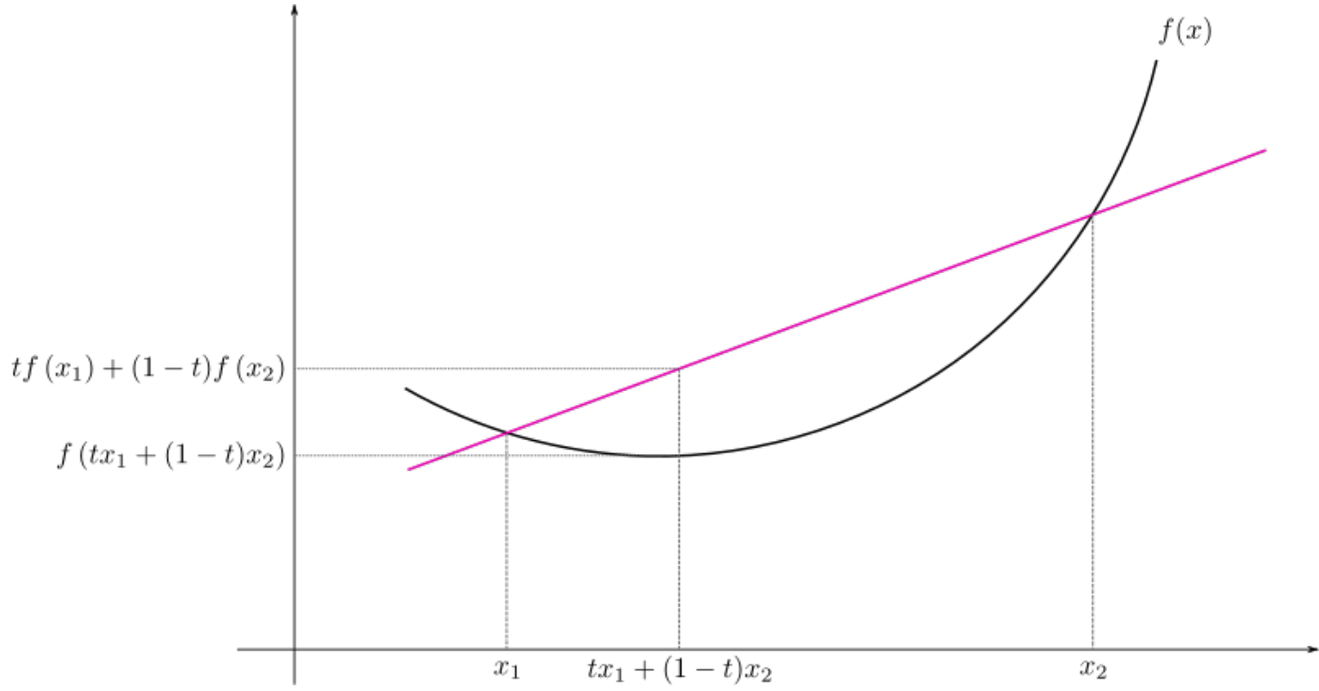
If I add some variability on the generation time, a given G will correspond to a lower R_0

<u>formally:</u>

Jensen's inequality, i.e. the average of transformed stochastic variables is at least equal to the transformed average of those variables when the transformation is convex

$$\int_0^\infty e^{z\tau} w(\tau) d\tau \geq e^{z\int_0^\infty \tau w(\tau) d\tau}$$

$$\Rightarrow \quad \mathscr{M}_{w(\tau)}(z) \geq e^{zT_c} \quad \Rightarrow \quad R_0 \leq e^{GT_c}$$



[wikipedia: Jensen's inequality]

R_0 and exponential growth

$$R_0 = \frac{1}{\mathscr{M}_{w(\tau)}(-G)}$$

$w(\tau) = \mathcal{N}(T_c, \sigma^2) \Rightarrow R_0 = e^{G\,T_c - (1/2)\,G^2\,\sigma^2}$

$w(\tau) = \delta(T_c) \Rightarrow R_0 = e^{G\,T_c}$ **Upper bound of** R_0

If I add some variability on the generation time, a given G will correspond to a lower R_0

$\Rightarrow$ If I over-estimate the dispersion in the generation time I will under-estimate R_0
problem if I use the serial interval as a proxy for the generation time

R_0 and exponential growth

$$R_0 = \frac{1}{\mathscr{M}_{w(\tau)}(-G)}$$

τ distributed according to $w(\tau)$

$\tau = \tau_E + \tau_I$

τ_E distributed according to $g(\tau_E)$

τ_I distributed according to $h(\tau_I)$

$\Rightarrow w(\tau) = g(\tau_E) * h(\tau_I)$ (w is the convolution between g and h)

$\Rightarrow \mathscr{M}_{w(\tau)} = \mathscr{M}_{g(\tau_E)} \times \mathscr{M}_{h(\tau_I)}$

R_0 and exponential growth

$$R_0 = \frac{1}{\mathscr{M}(-G)}$$

τ distributed according to $w(\tau)$

$\tau = \tau_E + \tau_I$

τ_E distributed according to $g(\tau_E)$

τ_I distributed according to $h(\tau_I)$

$\Rightarrow w(\tau) = g(\tau_E) * h(\tau_I)$ (w is the convolution between g and h)

$\Rightarrow \mathscr{M}_{w(\tau)} = \mathscr{M}_{g(\tau_E)} \times \mathscr{M}_{h(\tau_I)}$

useful if we consider the generation time as the sum of

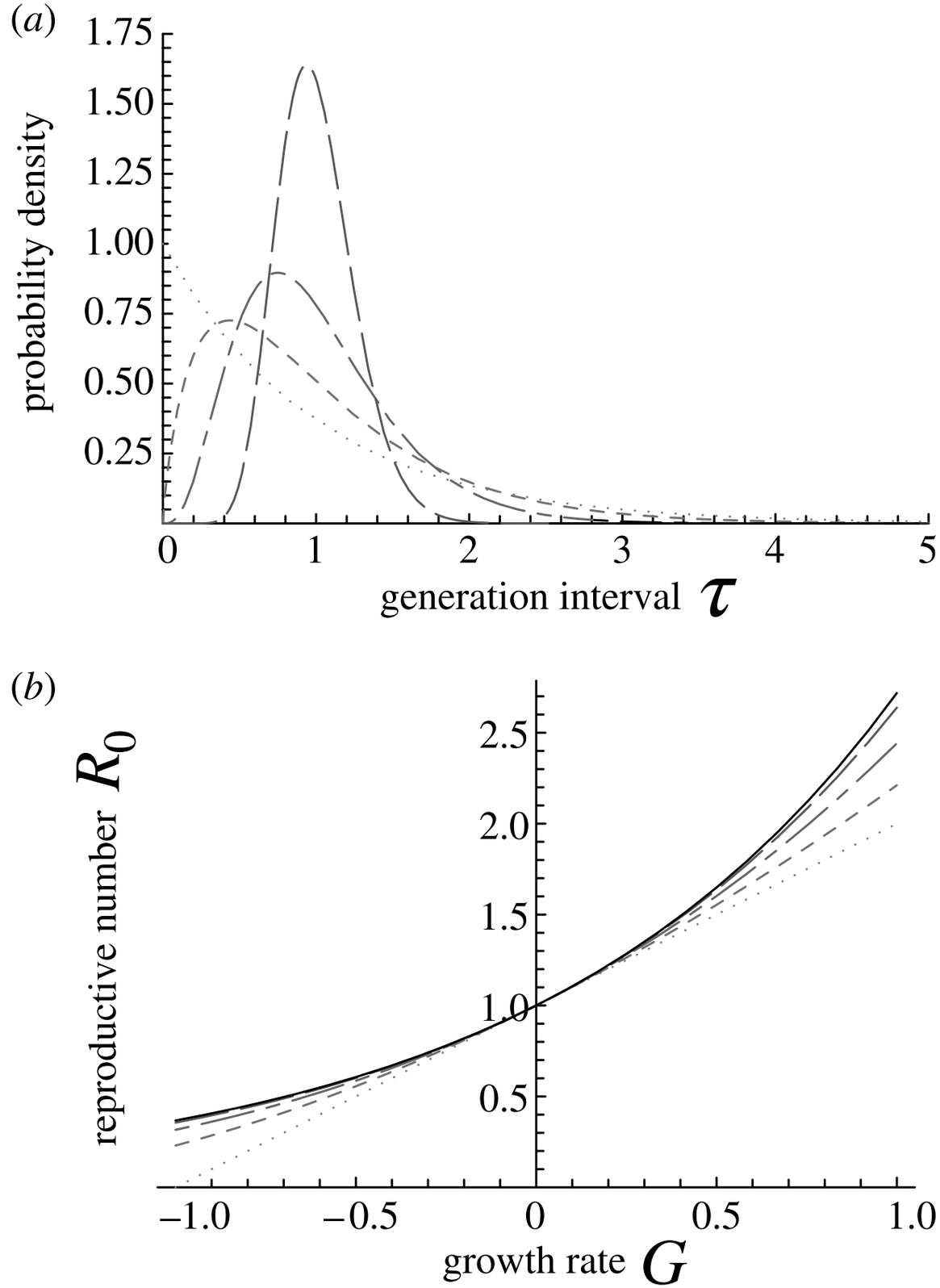
- a latent period τ_E (where no infection can be generated) distributed as $g(\tau_E)$
- an infection period τ_I (where infections can be generated) distributed as $h(\tau_I)$

e.g.: $\tau_E \sim Exp(\epsilon);\ \tau_I \sim Exp(\mu)$

$$R_0 = \frac{1}{\mathscr{M}(-G)} = \left(1 + \frac{G}{\epsilon}\right)\left(1 + \frac{G}{\mu}\right) \simeq 1 + G\left(\epsilon^{-1} + \mu^{-1}\right)$$

→ **this is the SEIR**

basic reproductive ratio



basic reproductive ratio estimation

In practice:

we have an incidence time series $\{y_t\}$

1.Identify a time window where the growth is exponential

2.Fit incidence points, e.g. Poisson regression:

$$\log(E(y_t \mid t)) = \alpha_0 + Gt$$

$$\mathscr{L}(\alpha_0, G \mid \{y_t\}) = \prod_{t=0}^{t_M} \frac{e^{y_t(\alpha_0+Gt)} e^{-e^{\alpha_0+Gt}}}{y_t!}$$

I use available information on the generation time distribution to compute R_0. In case I have only T_c and σ, I can use the Gaussian approximation $R_0 \simeq e^{G\,T_c - (1/2)\,G^2\,\sigma^2}$

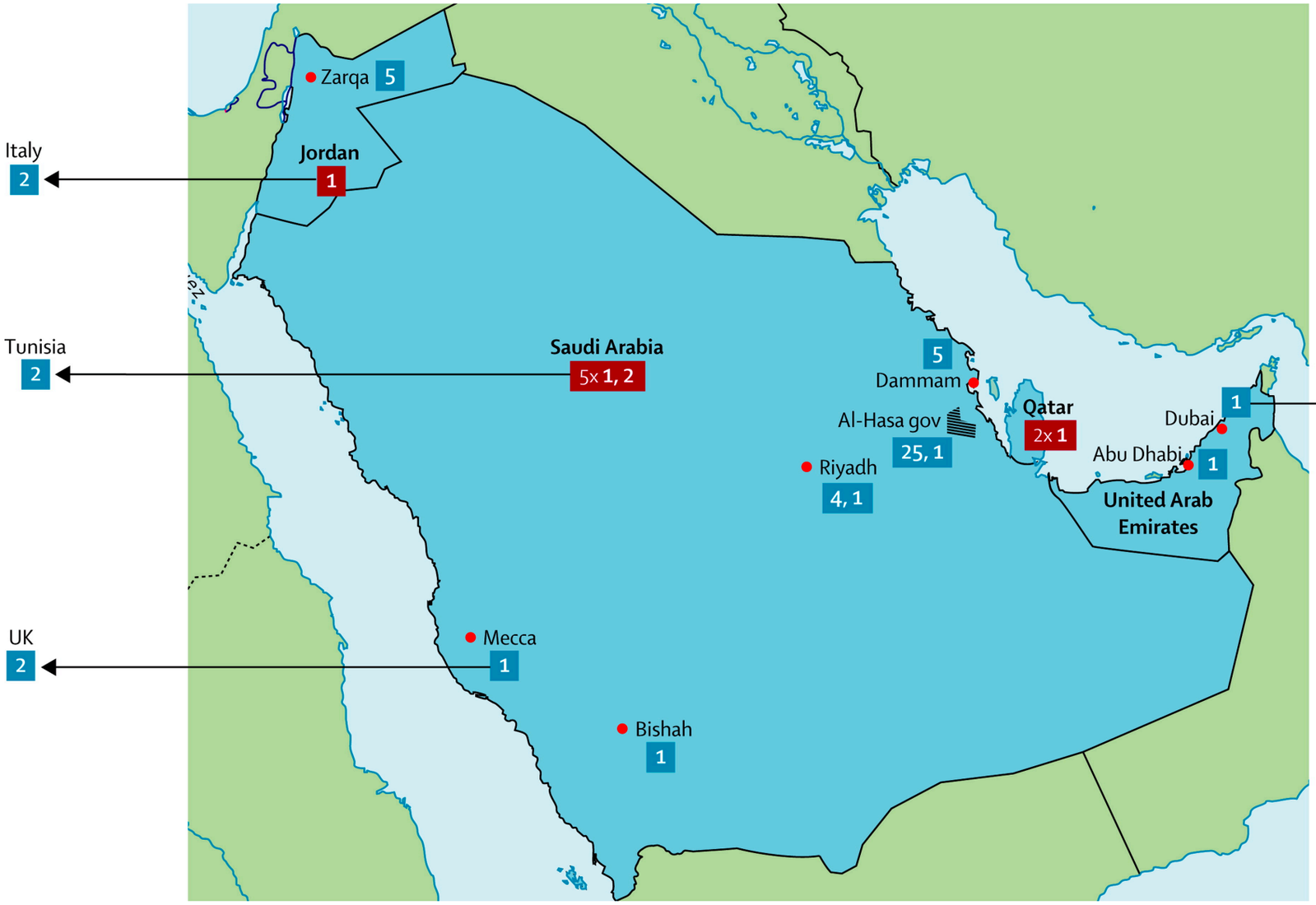
basic reproductive ratio estimation

what if I do not have an exponential growth?

Analysis of cluster size distribution

R_0 in the subcritical case

infections with $R_0 < 1$: e.g. zoonotic infections, or infections close to the eradication threshold (measles). How far are we from the epidemic threshold ? Important because there is a risk of mutation (for MERS) or a decrease in vaccine uptake (measles)

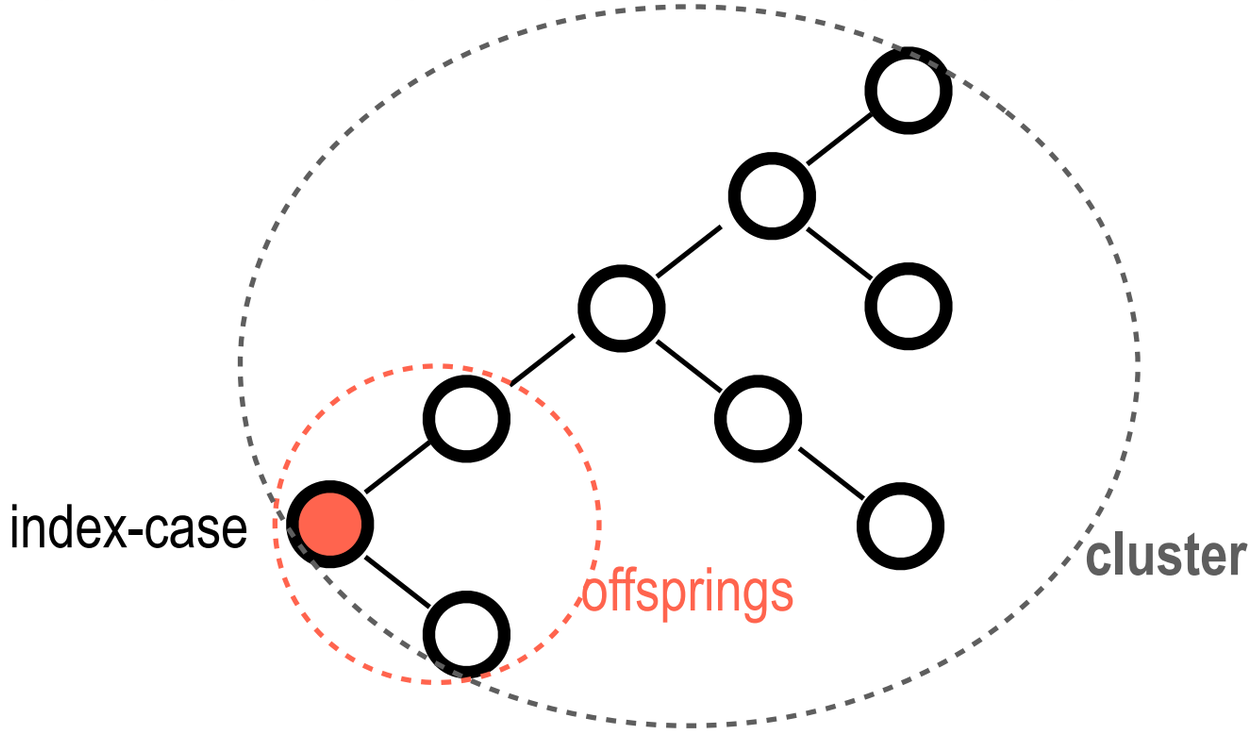


Middle East Respiratory Syndrome (MERS) coronavirus, 2013
[Breban, et al. The Lancet 2013]

Size of cluster (number of cases)	Baseline (all laboratory-confirmed cases in the ME region[b] as of 31 August 2013)
1[c]	42
2	8
3	2
5	2
10	0
22	1
Total number of cases	96

[Poletto, et al. Eurosurveillance 2014]

R_0 in the subcritical case



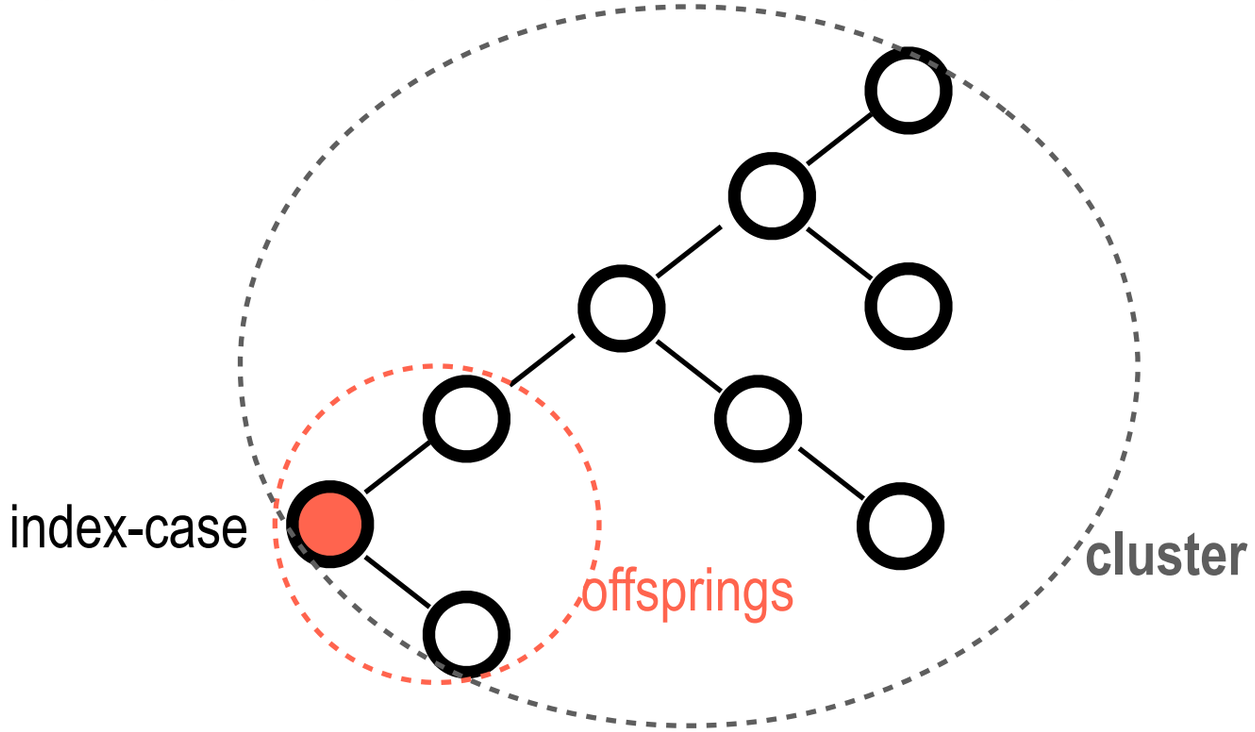
index case: infection caused by an external source

offsprings: cases infected by the index case

cluster: all cases originated by the index case

If we have different clusters, the distribution of their size depends on R_0: $P(s \mid R_0)$

R_0 in the subcritical case



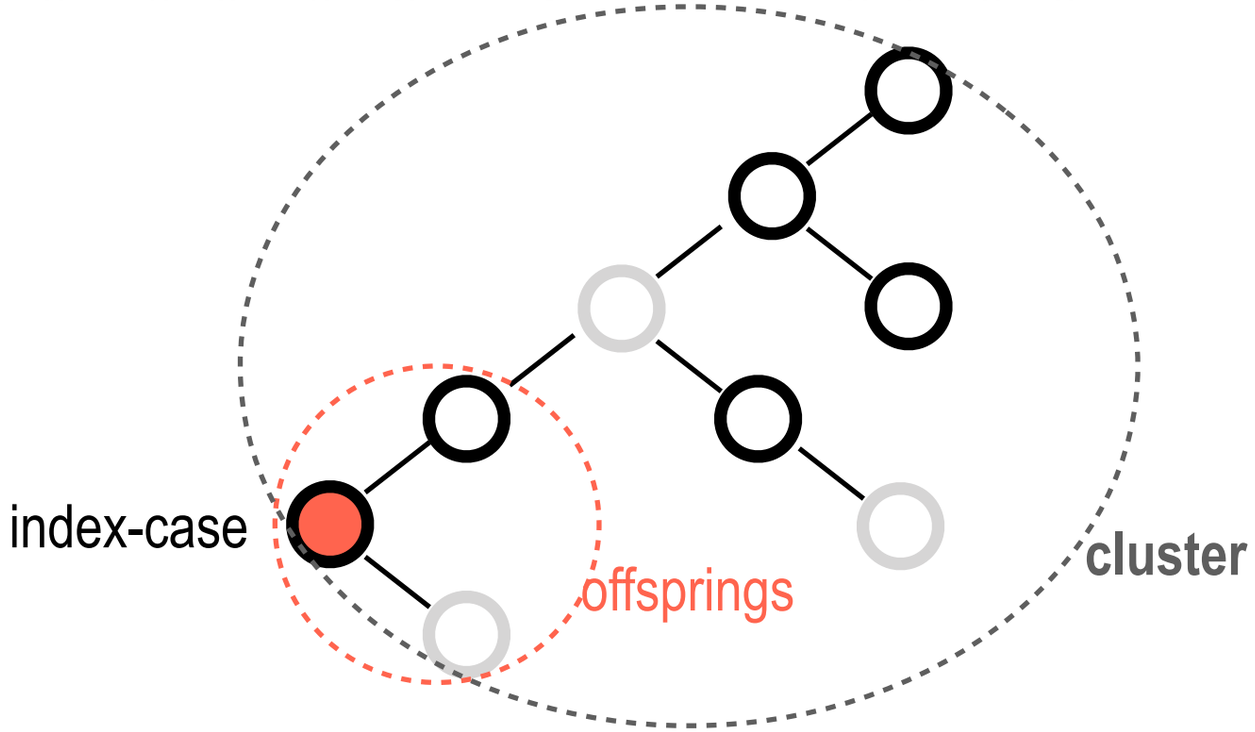
Continuous-time SIR dynamics: being R_0 below 1 stochastic effect are important

let's assume each case generate a number of secondary infections i, that is Poisson distributed: $p(i \mid R_0) = \frac{R_0^i e^{-R_0}}{i!}$

then $P(s \mid R_0) = \frac{s^{s-2} R_0^{s-1} e^{-sR_0}}{(s-1)!}$

[Farrington, et al Biostatistics 2003]

R_0 in the subcritical case



We need to account for under-reporting

each case may go unobserved with probability p_{miss}. If a cluster has size s, I observe a cluster of size $o \leq s$

$$P(o \mid R_0, p_{\text{miss}}, o \geq 1) = \frac{\sum_{s \geq o} P(s \mid R_0) \binom{s}{o} p_{\text{miss}}^{s-o} (1 - p_{\text{miss}})^o}{1 - P(o = 0 \mid R_0, p_{\text{miss}})}$$

$$\mathscr{L}(R_0, p_{\text{miss}} \mid o_i) \propto \prod_i P(o \mid R_0, p_{\text{miss}}, o \geq 1)$$