Stochastic epidemic modelling and analysis: current perspective and future challenges

Frank Ball (University of Nottingham) and Tom Britton (Stockholm University)

Workshop on Disease Dynamics, Isaac Newton Institute for Mathematical Sciences

20 August 2013

Outline of presentation

- Orientation (FGB)
- Emerging epidemics – early stages of an outbreak (FGB)
- Established epidemics – beyond the initial phase (TB)
- Open problems (TB)

When are stochastic models needed?

When numbers are small

- Start of an outbreak
- End of an epidemic
  - time to extinction
  - fade out
- Small populations/subgroups*
- Smaller units within populations, e.g. households and workplaces
- Estimation, e.g. confidence intervals

* Not considered in talk.

Model types

- Focus on (simple) stochastic models for which some fully-rigorous analytic progress is possible.
- Complex agent-based simulation models
  - invaluable tool for informing public health policy, owing to their far greater realism
  - probably also miss some features
  - require numerous parameters to be estimated
  - more difficult to interpret and perform sensitivity analyses on.
- Deterministic models are appropriate when all numbers are large and are usually easier to analyse than stochastic models.
- Different modelling approaches complement each other.

Early stages of an outbreak

KEY QUESTIONS

- Is a major outbreak possible? (R_0)
- If so,
  - What is the probability of a major outbreak?
  - How fast will it spread? (exponential growth rate)
- Estimation in an emerging outbreak.
- Control measures, e.g. vaccination strategies.

All of these questions can be addressed using branching processes.

General branching process

- In a general Crump–Mode–Jagers branching process individuals have independent and identically distributed (iid) life histories $\mathcal{H} = (I, \xi)$, where I denotes a typical individual's age at death and ξ is a point process of ages at which she reproduces. [Note that $\xi((I, \infty)) = 0$.]
- Thus if an individual with life history $\mathcal{H} = (I, \xi)$ is born at time b and $0 < \tau_1 \leq \tau_2 \leq \cdots \leq I$ denote the points of ξ then she has one child at each of times $b + \tau_1, b + \tau_2, \cdots$.
- The life histories are pieced together in the obvious fashion to form the population process.
- Such a process approximates the process of infectives in the early stages of an epidemic in a large, closed homogeneously mixing population, where I denotes the sum of an infective's latent and infectious periods and ξ gives the infectious ages at which that infective creates new infectives.

General branching process

- In an epidemic setting, equate extinction/survival of the branching process with a minor/major outbreak.
- Let $R = \xi((0,\infty))$ be the number of offspring of a typical individual and $R_0 = \mathrm{E}[R]$. Suppose there are a initial individuals.
  - $\mathrm{P}(\text{extinction}) < 1 \iff R_0 > 1$.
  - $\mathrm{P}(\text{extinction}) = q^a$, where q is given by the smallest solution of $f(s) = s$ in $[0,1]$ and $f(s) = \mathrm{E}[s^R]$ is the PGF of R.
- Extends to multitype processes. E.g. if there are J types of individuals, labelled $1, 2, \cdots, J$, then R_0 is the maximal eigenvalue of the mean offspring matrix $M = [m_{ij}]$, where m_{ij} is the mean number of type-j individuals spawned by a type-i individual.

General branching process

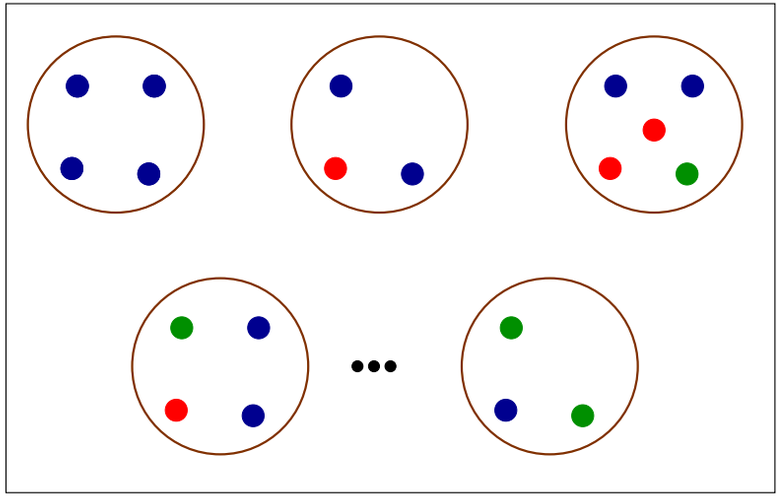
Let $Z(t)$ be the number of individuals alive at time t. Suppose that $R_0 > 1$ and there exists a Malthusian parameter $r > 0$ given by

$$\int_0^\infty \mathrm{e}^{-rt}\mu(\mathrm{d}t) = 1,$$

where $\mu(t) = \mathrm{E}[\xi((0,t])]$.

- $\mathrm{e}^{-rt}Z(t) \xrightarrow{\text{a.s.}} W$, where $W \geq 0$ is a random variable that equals 0 if and only if the process goes extinct.
- r is the real time growth rate for the corresponding epidemic model.

Households SIR epidemic model



m_n households of size n $(n = 1, 2, \dots)$

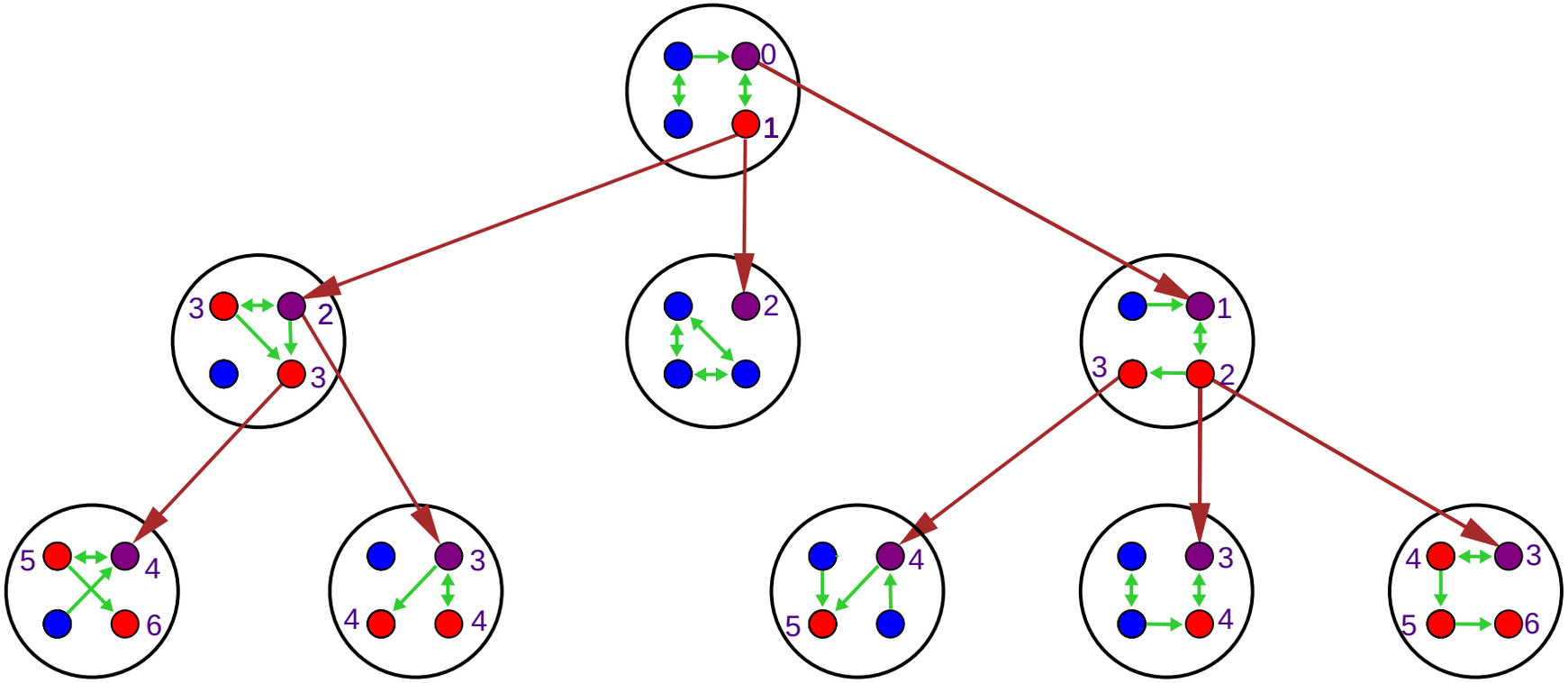
total no. of households $m = \sum_{n=1}^{\infty} m_n$

total no. of individuals $N = \sum_{n=1}^{\infty} n m_n < \infty$

- Let μ_G be the expected number of global contacts made by a typical infective.
- Let μ_H be the (household-size-biased) mean final size (including the initial infective) of a typical single-household epidemic.

(Bartoszyński (1972), Becker and Dietz (1995), Ball, Mollison and Scalia-Tomba (1997))

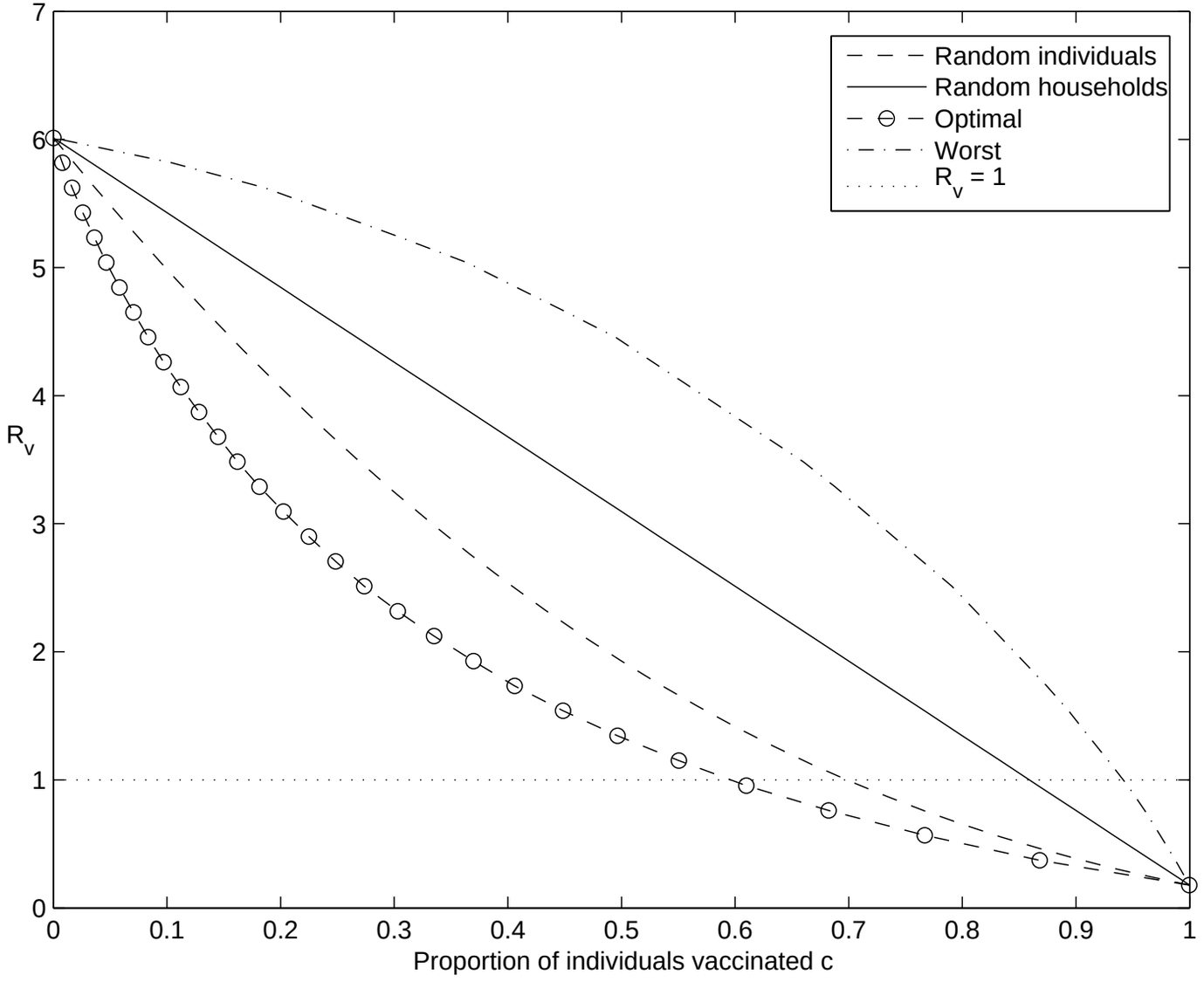
Households model



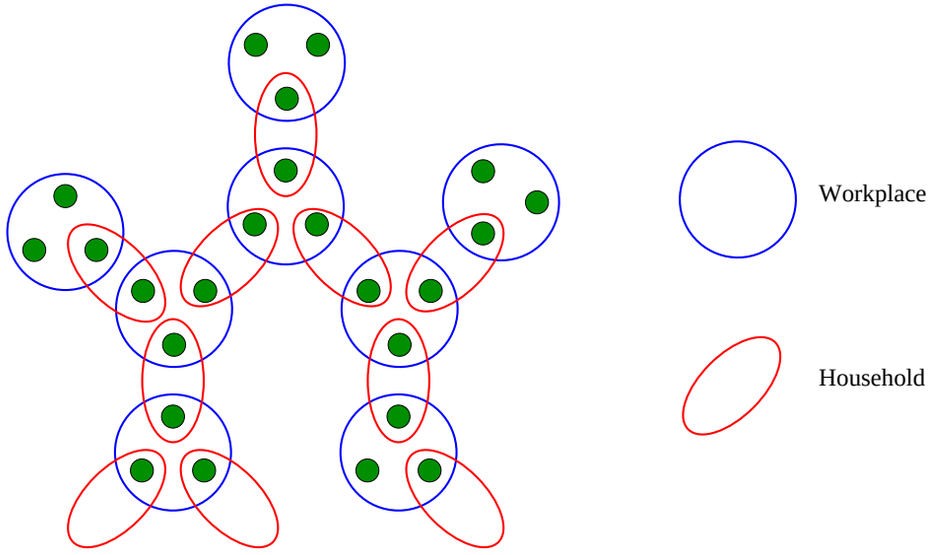
Branching process of infected households

Threshold parameter $R_* = \mu_G \mu_H$

Comparison of vaccination strategies



Households-workplaces model



- Three types of infectious contacts: global, within-household and within-workplace.
- Let μ_H (μ_W) be the household (workplace)-size-biased mean final size (including the initial infective) of typical single-household (single-workplace) epidemic.

(Ball and Neal (2002), Pellis et al. (2009), (2011), (2012), cf. Andersson (1997))

Households-workplaces model

- Approximate early stages of an epidemic by a three-type branching process, with types being
  - double-primary – infected by a global contact
  - household-primary – infected through a workplace
  - workplace-primary – infected through a household

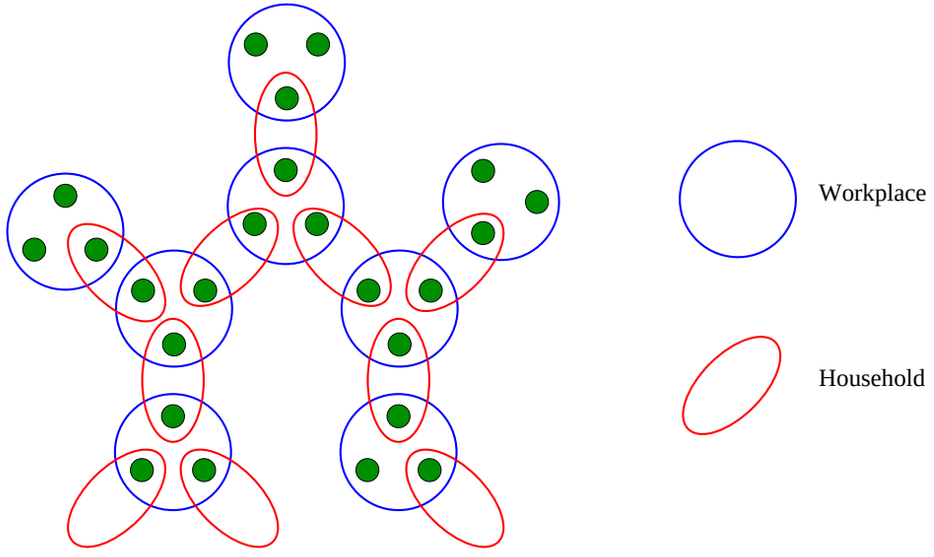
  in which all secondary cases in a single-household (or -workplace) epidemic are attributed to the initial case in that household (workplace).
- The mean offspring matrix of this branching process is

$$M = \begin{bmatrix} \mu_G & \mu_W - 1 & \mu_H - 1 \\ \mu_G & 0 & \mu_H - 1 \\ \mu_G & \mu_W - 1 & 0 \end{bmatrix}.$$

- Threshold parameter R_I given by maximal eigenvalue of M.

Households-workplaces model

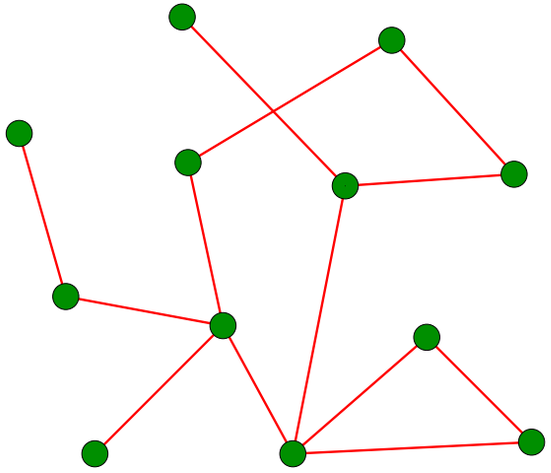
- Impact of second mixing group is significant (e.g. model is supercritical without global infection if $(\mu_H - 1)(\mu_W - 1) \geq 1$) BUT analysis makes unrealistic assumptions about how households and workplaces interact, in that they locally form a tree.



Households-workplaces model

- Impact of second mixing group is significant (e.g. model is supercritical without global infection if $(\mu_H - 1)(\mu_W - 1) \geq 1$) BUT analysis makes unrealistic assumptions about how households and workplaces interact, in that they locally form a tree.
- Similar unrealistic assumptions are made in mathematically tractable models of random networks with clustering, such as random intersection graphs.
- Challenge is to develop more realistic models that are still susceptible to analysis.

Network SIR epidemic model



D= degree of typical individual

$p_k = \mathrm{P}(D = k) \quad (k = 0, 1, \dots)$

$\tilde{D}$= degree of typical neighbour

$\mathrm{P}(\tilde{D} = k) = k p_k / \mathrm{E}[D] \quad (k = 0, 1, \dots)$

- $R_0 = p\mathrm{E}[\tilde{D} - 1] = p\left(\mathrm{E}[D] + \frac{\mathrm{Var}(D)}{\mathrm{E}[D]} - 1\right)$,

  where p is the probability that an infective infects a given neighbour.

- $R_0 = \infty$ (so critical vaccination coverage is 1) if $\mathrm{Var}(D) = \infty$, even if $\mathrm{E}[D] < \infty$.

- What does this mean for finite populations?

(Andersson (1999), Newman (2002))

Finite populations

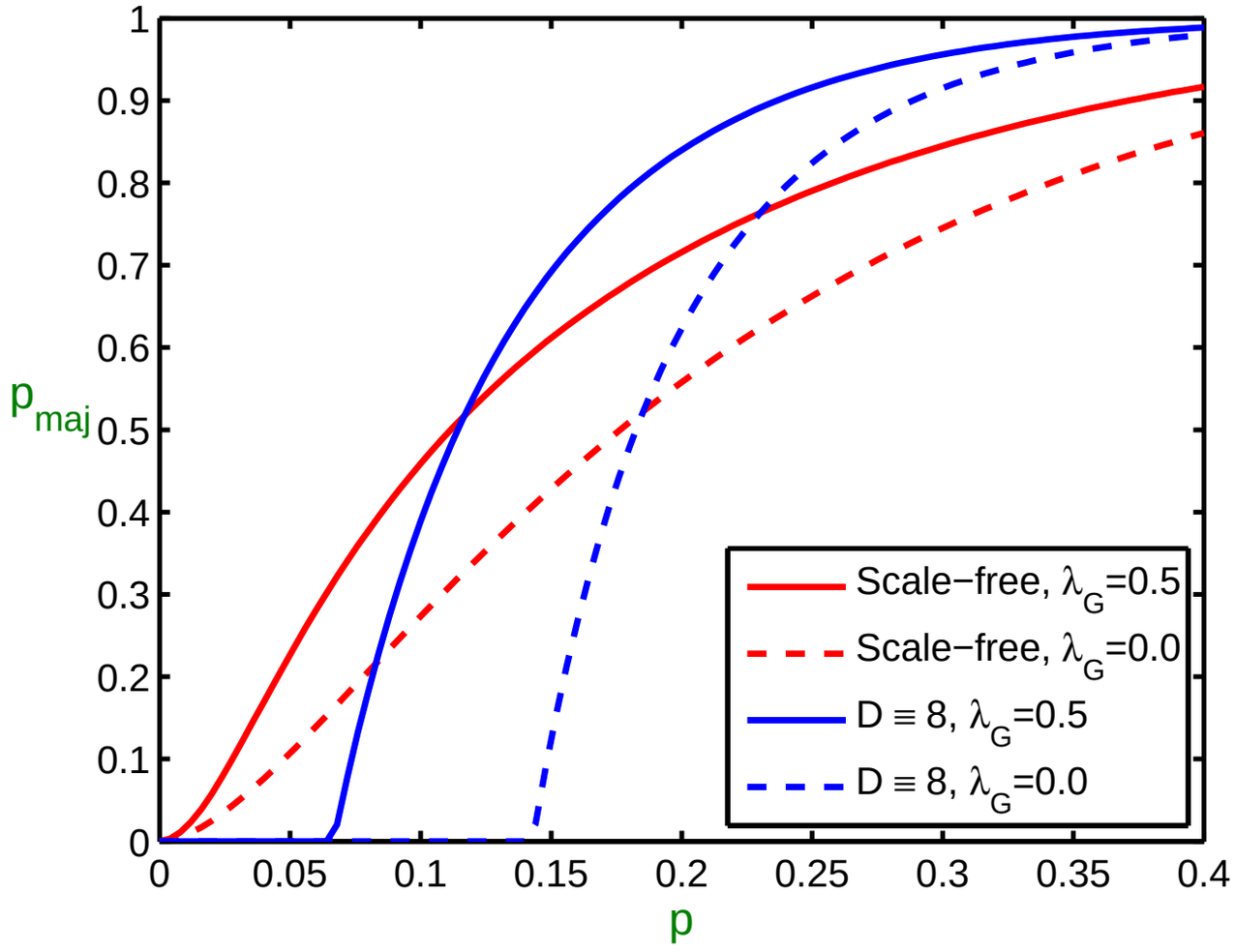
- For homogeneously mixing finite populations, Ross (2011) defines R_0 to be the expected number of infections (rather than infectious contacts) made by a typical infective in an otherwise fully susceptible population.
- This yields a sharper upper bounding branching process but the interpretation of minor and major outbreaks still essentially requires the population size $n \to \infty$.
- Aim of control is often to reduce R_0 to 1 but how appropriate is this for finite populations? As $n \to \infty$,
  - When $R_0 = 1$, mean size of epidemic $\to \infty$.
  - For $R_0 = 1 + \varepsilon$, mean fraction infected by major outbreak $\approx \frac{2\varepsilon}{(1+\varepsilon)^2}$ is small.
- Nasell (1995) defined the threshold value of R_0 to be when the final size distribution changes from being unimodal to bimodal and conjectured that this threshold $\approx 1 + \rho/n^{\frac{1}{3}}$ for large n, though a formal proof is still open.
- No sharp change at threshold, so control $\mathrm{P}(\text{size} > \text{specified value})$ instead?

Network epidemic models

Areas in need of further work include

- Dynamic networks, including effect of epidemic on network.
- Clustered networks
  - Mathematically tractable models have tree-like local structure at some level
  - There exist distinct network models having the same degree distribution, clustering coefficient and degree correlation but yielding differing epidemic properties, so caution is required when extrapolating results to real-life networks.
- Non-SIR models, e.g. SIS, SIRS.
- Casual contacts.

Effect of casual contacts



Asymptotic probability of a major outbreak, p_{maj}, for constant-degree and scale-free ($\mathrm{P}(D=k) \propto k^{-2.466956}\ (k=3,4,\ldots)$) networks with $\mathrm{E}[D]=8$ when $I \equiv 1$ and infectives make casual contacts at rate λ_G. (Ball and Neal (2008))

Estimation in an emerging epidemic

- Suppose that infectives have iid infectiousness profiles $\{\Lambda(a) : a \geq 0\}$; an infective with infectious age a creates new infectives at rate $\Lambda(a)$.
- Thus $\Lambda(a)$ is the (possibly random) intensity of the point process ξ governing births in the general branching process that approximates the homogeneously mixing epidemic.
- Real time growth rate r satisfies $\int_0^\infty \mathrm{e}^{-ra}\mathrm{E}[\Lambda(a)]\mathrm{d}a = 1$.
- $R_0 = \int_0^\infty \mathrm{E}[\Lambda(a)]\mathrm{d}a$.
- "Infectious contact interval" W has pdf $w(t) = \mathrm{E}[\Lambda(t)]/R_0$ $(t > 0)$.
- Let $M_W(\theta) = \mathrm{E}[\mathrm{e}^{-\theta W}] = \frac{1}{R_0}\int_0^\infty \mathrm{e}^{-\theta t}\mathrm{E}[\Lambda(t)]\mathrm{d}t$. Then $R_0 M_W(r) = 1$.
- Thus if the distribution of W is known and the real time growth rate r is observed, an estimate of R_0 is given by $1/M_W(r)$.

(Wallinga and Lipsitch (2007))

Households model

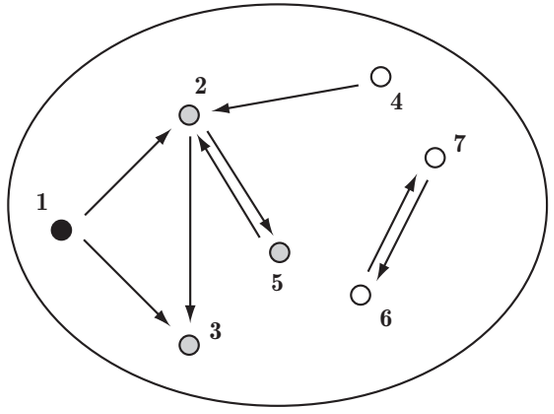
- Infectiousness profile $\{\Lambda_0(a) : a \geq 0\}$ satisfying $\mathrm{E}[\int_0^\infty \Lambda_0(a)\mathrm{d}a] = 1$.
- Infective with infectious age a makes global contacts at rate $\mu_G\Lambda_0(a)$ and contacts any given susceptible in its household independently at rate $\lambda_L\Lambda_0(a)$.
- Global contacts emanate from a typical single-household epidemic t time units after that household was infected at expected rate $\beta_H(t) = \mu_G w_H(t)$, where $w_H(t)$ depends only on household structure and within-household disease dynamics.
- For $\theta \geq 0$, let $M_H(\theta) = \int_0^\infty \mathrm{e}^{-\theta t} w_H(t)\mathrm{d}t$. Then

$$R_* = \int_0^\infty \beta_H(t)\mathrm{d}t = \mu_G M_H(0) \qquad \text{and} \qquad \mu_G \int_0^\infty \mathrm{e}^{-rt} w_H(t)\mathrm{d}t = 1.$$

- Thus if $M_H(\theta)$ $(\theta \geq 0)$ is known and the real time growth rate r is observed, $\hat{\mu}_G = 1/M_H(r)$ and $\hat{R}_* = M_H(0)/M_H(r)$. (Fraser (2007))

Two challenges

- Estimating $(\lambda_L, \Lambda_0(\cdot))$ in an emerging epidemic.
- Determining $M_H(\theta)$ from λ_L and distribution of $\{\Lambda_0(a) : a \geq 0\}$.
  - matrix methods exist for Markov models (e.g. with exponential latent and infectious periods) – Goldstein et al. (2009) and Pellis et al. (2011).
  - simulate single-household epidemics – can be computationally expensive
  - Fraser (2007) developed a closed-form approximate method which assumes that a typical household-generation-k infective is infected at time $W_1 + W_2 + \cdots + W_k$ after the household was infected, where $W_1, W_2, \cdots, W_k \overset{\text{iid}}{\sim} W$ and W has pdf $w(t) = \mathrm{E}[\Lambda_0(t)]$ $(t > 0)$. Works well if household size or var(W) is sufficiently small.

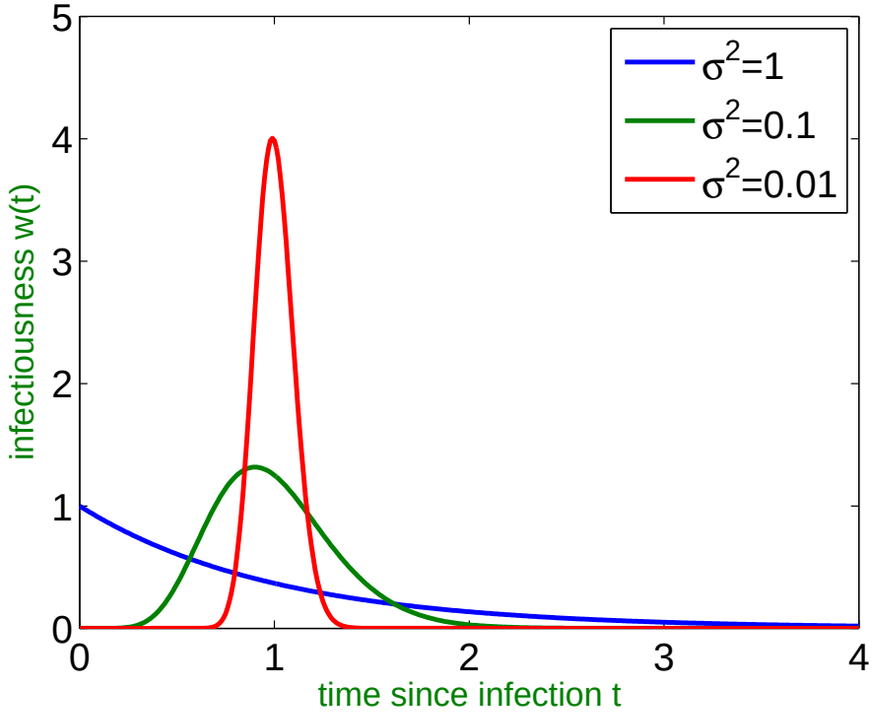


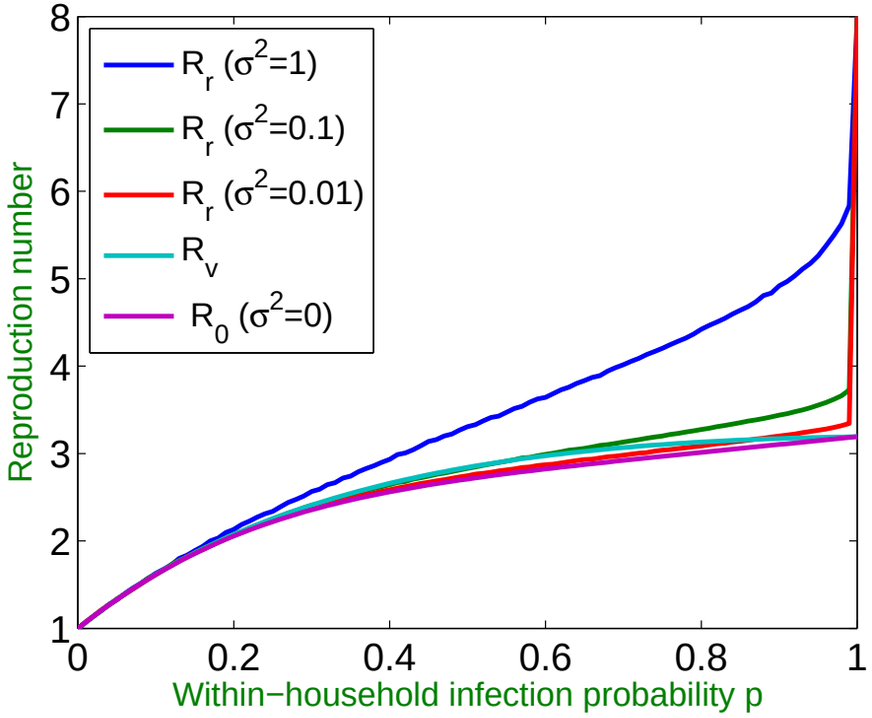
Households model

Goldstein et al. (2009) introduced the following reproduction numbers:

- Exponential-growth-associated reproduction number $R_r = 1/M_W(r)$
  - yields R_0 if model is homogeneously mixing,
  - yields households model R_0 (Pellis et al. (2012)) if Fraser (2007) approximation is used.
- Perfect-vaccine-associated reproduction number $R_V = (1 - c_V)^{-1}$, where c_V is the critical vaccination coverage assuming individuals are chosen uniformly at random for vaccination. (Note that $c_V = 1 - R_V^{-1}$).

Households model reproduction numbers





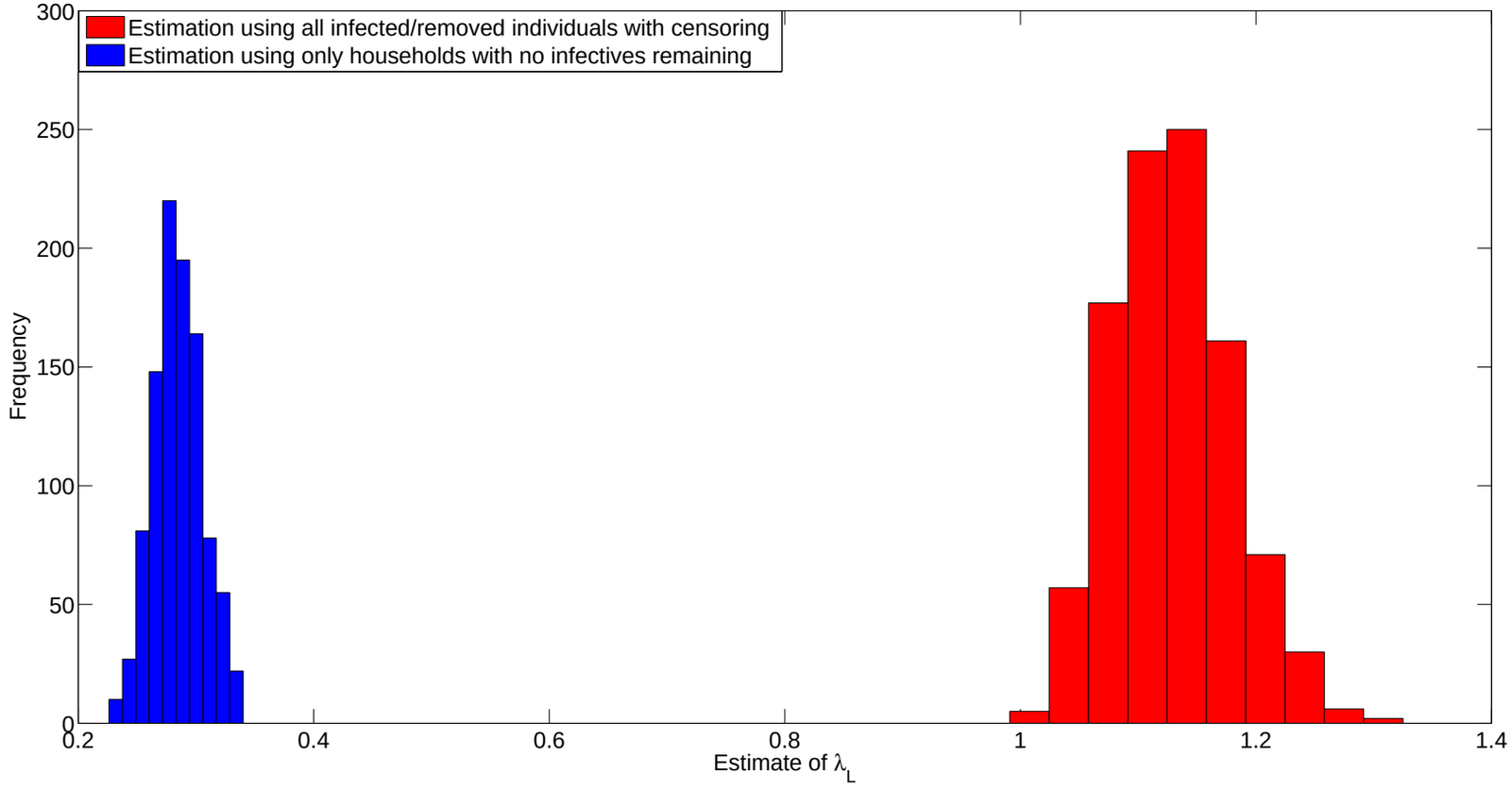
Reproduction numbers (right panel) for households model with $\mu_G = 1$, household size $= 8$ and non-random infectiousness profile $\Lambda_0(t) \equiv w(t)$ (left panel) that follows a gamma distribution with mean 1.

Estimation in an emerging epidemic

- Consider an SIR epidemic among $5,000,000$ households, with sizes distributed uniformly on 1 to 5, in which a typical infective makes global contacts at overall rate λ_G and contacts a given susceptible in its own household at rate λ_L, throughout and infectious period that has an exponential distribution with mean 1.
- After $1,000$ individuals have become infected
  - estimate real time growth rate empirically
  - estimate λ_L by fitting final size distribution of single-household epidemics, with one initial infective, to observed household outbreaks
  - estimate λ_G and hence R_* using (exact) Fraser's method.

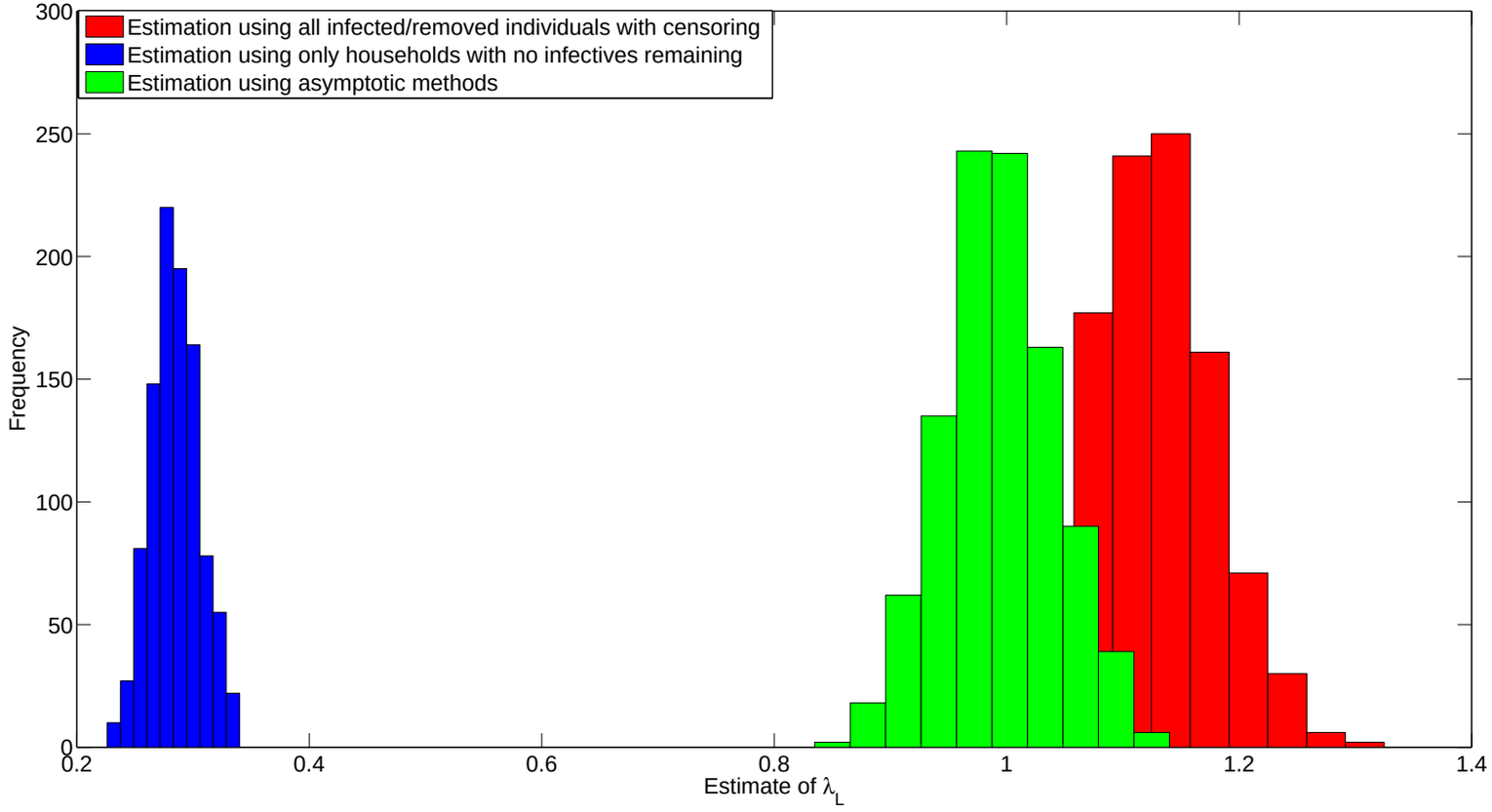
(Ball and Shaw (2013))

Estimation in an emerging epidemic



Histograms of estimates of within-household infection rate λ_L based on $1,000$ simulated epidemics with $\lambda_L = \lambda_G = 1$ that took off.

Estimation in an emerging epidemic



Histograms of estimates of within-household infection rate λ_L based on $1,000$ simulated epidemics with $\lambda_L = \lambda_G = 1$ that took off.

Estimation in an emerging epidemic

- Observed household final sizes do not follow the usual distribution owing to emerging nature of epidemic among population of households.
- “Correct" distribution can be obtained using asymptotic theory of multitype continuous time branching processes (Ball and Shaw (2013)).
- Similar issues arise more generally (i.e. outside of simple households models) in statistical analysis of emerging epidemics; data collected need to be modelled very carefully taking due account of the emerging nature of epidemic.

Closing remark

Branching processes have a very rich and well-developed theory, owing in part to their linear nature, and provide a fertile ground for investigating the early stages of epidemics in large populations.