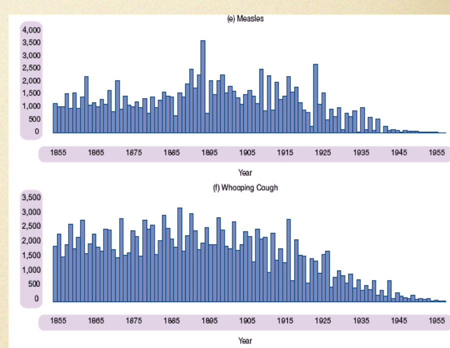
Stochastic Models

Epidemiological data

- 3 Main kinds of stochasticity
  1. Observation noise
     - Likelihood of detecting and reporting a case
  2. Environmental noise
     - “good” versus “bad” years
  3. Demographic noise
     - Individual-level chance events



Motivation

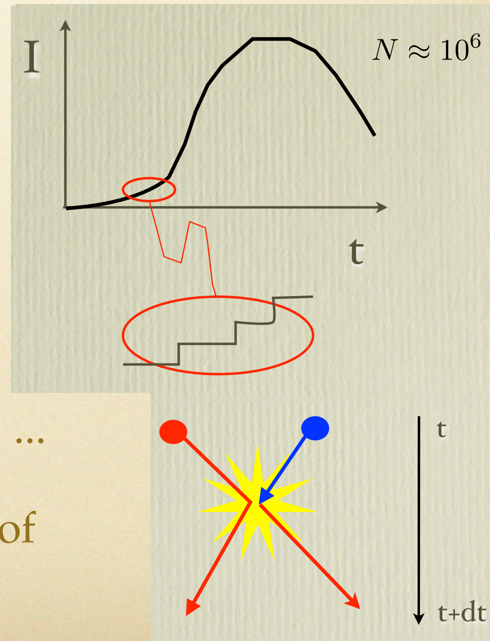
- SIR model

$$\frac{dI}{dt} = \beta X \frac{Y}{N} - \gamma I$$

- Which means

$$dY = \beta X \frac{Y}{N} dt - \gamma Y dt$$

- But, in reality, dY = 0, 1, 2, ...
- Need to consider physics of individuals bumping into each other



Demographic Stochasticity

- Defined as fluctuations in population processes arising from random nature of events at level of individual
- Baseline probability associated with each event is fixed, but because of chance events, individuals experience differing fates
- Need to modify ODEs in two ways to incorporate demographic stochasticity:
  - Make state variables integer-valued (X, Y, Z)
  - Introduce transition probabilities
- Some analytical methods *possible*

Demographic Stochasticity

- Analytical approaches involving Master Equation potentially powerful, but often too difficult to implement
- Alternatively, results from 'branching process' theory can be used
- And then, there's always brute-force simulation (which is what we'll start with)

Demographic Stochasticity

- Good news: very straightforward methods exist for exact simulation of these stochastic processes
- To simulate, we need to answer two fundamental questions, starting with a specified system state at time t:
  - *When* is next event?
  - *What* is next event?

Deriving time to next event

- Can derive "inter-event" times from fundamental population principles
- Let's assume we have a population of size N at time t, then we define $G_N(s)$ as probability that no event occurs in subsequent time interval of length s. So,

$G_N(s+\delta s)$=Pr{no event in time interval (t,t+s+δs)}
=Pr{no event in (t,t+s)} x Pr{no event in (t+s,t+s+δs)}

- So, substituting C(N)=sum of frequencies of all events, we get

$$G_N(s+\delta s) = G_N(s) \times \{1 - C(N) \times \delta s\}$$

Deriving time to next event

$$G_N(s+\delta s) = G_N(s) \times \{1 - C(N) \times \delta s\}$$

After tidying up and re-arranging:

$$\frac{\left[G_N(s+\delta s) - G_N(s)\right]}{\delta s} = -C(N) \times G_N(s)$$

By now, you shouldn't be surprised by what comes next. We let δs -> 0, which gives

$$\frac{dG_N}{ds} = -C(N) \times G_N(s)$$

Deriving time to next event

Solution is exponential equation:

$$G_N(s) = e^{-C(N)s}$$

Naturally, probability that next event occurs in (t,t+s) is therefore:

$$F_N(s) = 1 - G_N(s) = 1 - e^{-C(N)s}$$

Note: independent of starting time *t*

So, F_N is exponentially distributed

Deriving time to next event

To simulate a random inter-event time, draw a random number Y_1 from a uniform distribution ($1 \geq U_1 \geq 0$) and equate with $F_N(s)$

$$U_1 = F_N(s) = 1 - e^{-C(N)s}$$

Now, solve for *s*:

$$s = -\frac{\log(U_1)}{C(N)}$$

Can now move on to our exact stochastic simulation algorithm

Demographic Stochasticity

- Many ways to implement such an approach (K&R pp200-205)
- A popular (and mathematically rigourous) method is called Gillespie's Direct Algorithm (Gillespie 1977)

Gillespie's Direct Method

1. Label all possible events $E_1, \ldots, E_n$
2. Calculate rate at which each event occurs $R_1, \ldots, R_n$
3. Rate at which *any* event occurs is $R_{sum}=\Sigma_i R_i$
4. Calculate time until next event

$$\delta t = \frac{-1}{R_{sum}} \log(U_1)$$

5. Generate uniform deviate, U_2, and set $P = U_2 \times R_{sum}$
6. Event *p* occurs if $$\sum_{m=1}^{p-1} R_m < P \leq \sum_{m=1}^{p} R_m$$
7. Update state, time $t=t+\delta t$ and return to step 2

Gillespie's Direct Method

- Let's implement this for SIR model, with demography
- Event rates are **R**=(μN, βXY/N, μX, γY, μY, μZ)
- Sum of event frequencies is

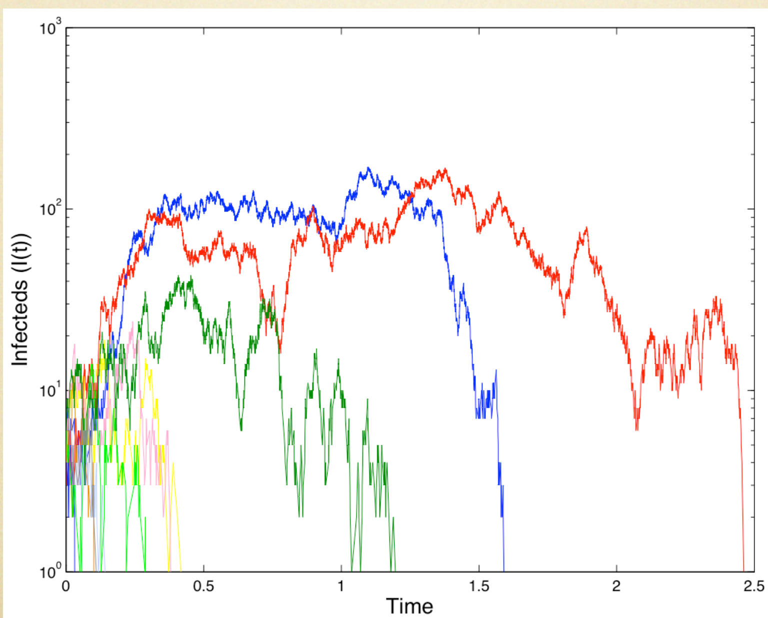
$$R_{sum} = 2 \times \mu N + \beta XY + \gamma Y$$

- Time until next event:

$$\delta t = \frac{-1}{R_{sum}} \log(U_1)$$

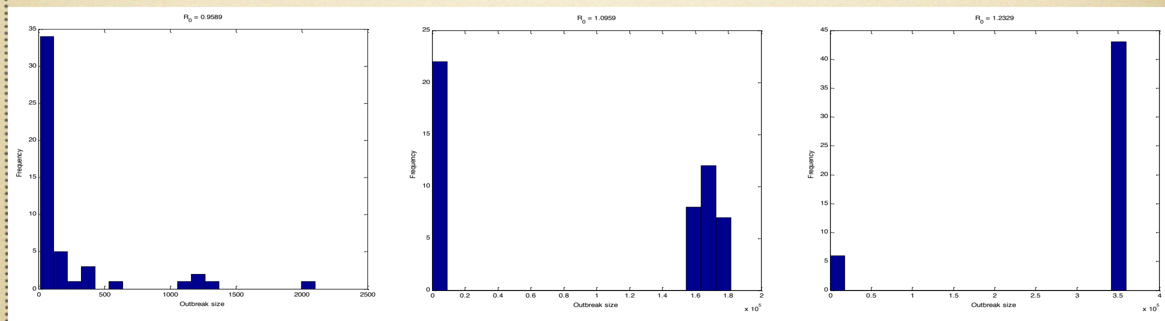
- Set P = U_2 x R_{sum} and find *p* $\sum_{m=1}^{p-1} R_m < P \leq \sum_{m=1}^{p} R_m$
- Update variables, time (t = t + δt) and repeat

Stochastic Epidemics



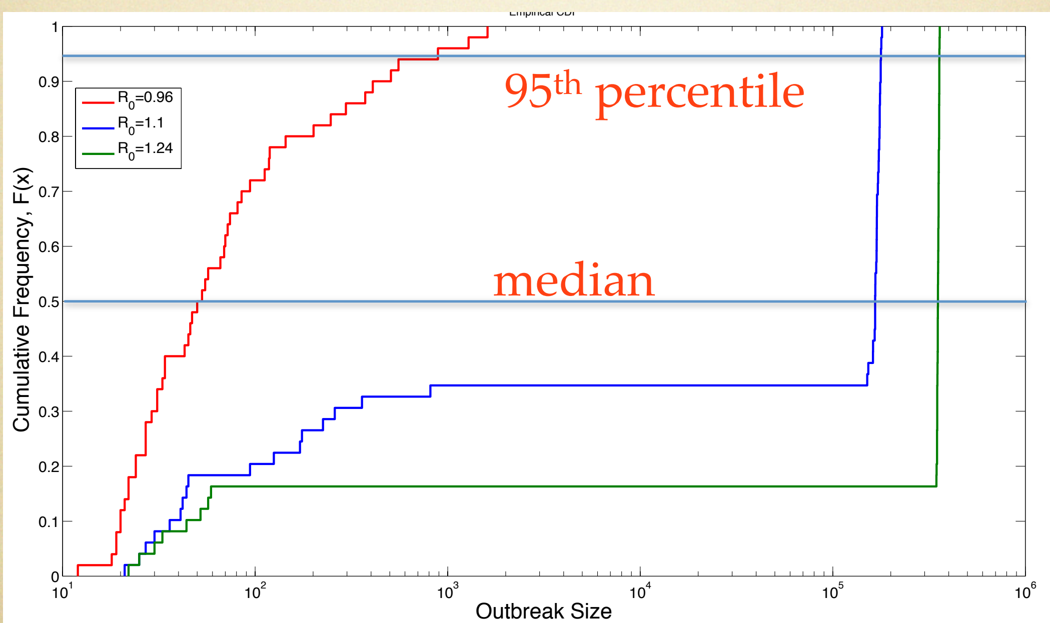
Characterising variability

- Can repeat invasion of pathogen into virgin population a number of times and examine

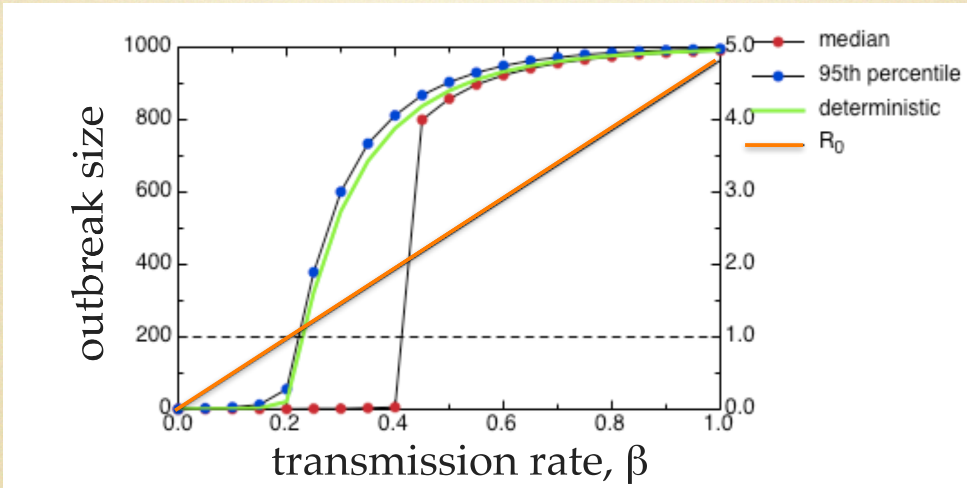


This is referred to as the J-to-U transition

Quantifying distribution

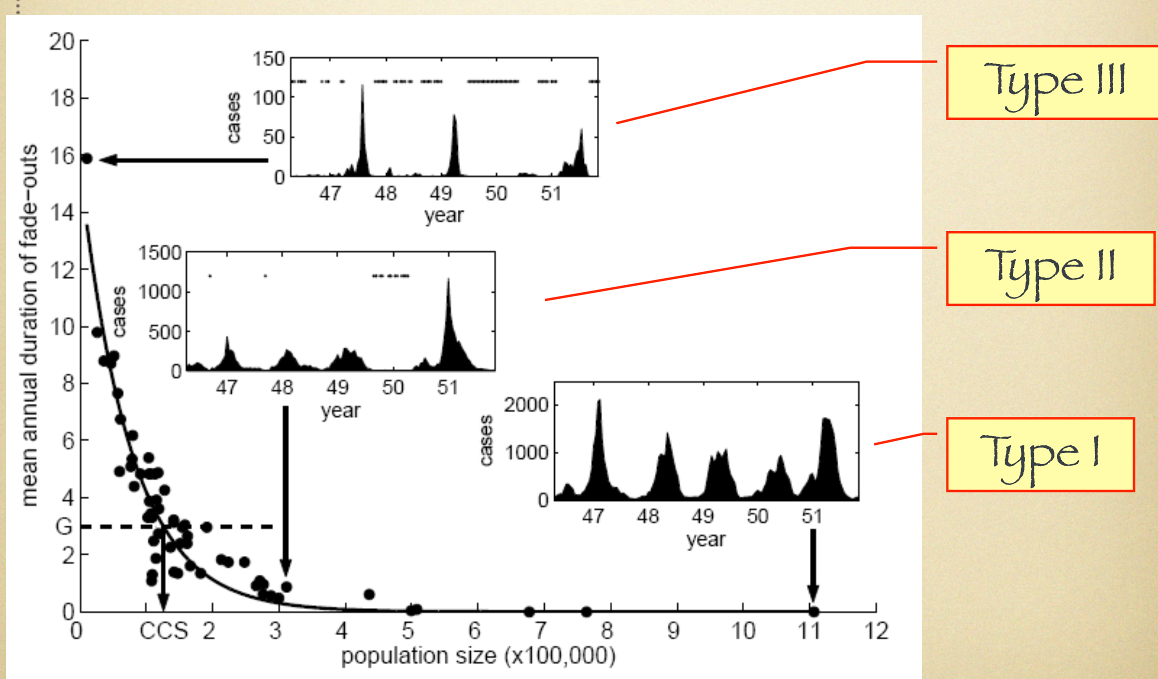


Stochastic Invasion Threshold

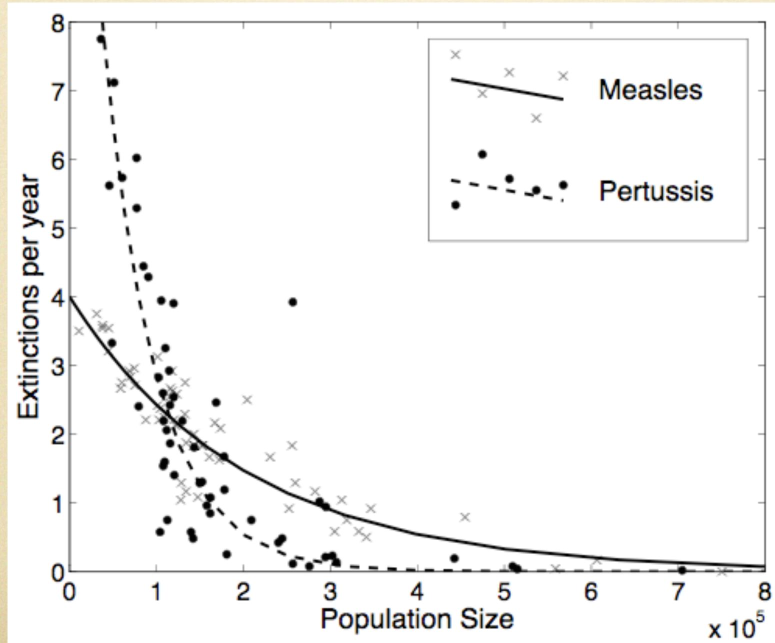


Important predictions for emerging pathogens: when R_0 only slightly above 1, "likely" outcome (median) very different from "worst-case-scenario" (95th prctile)

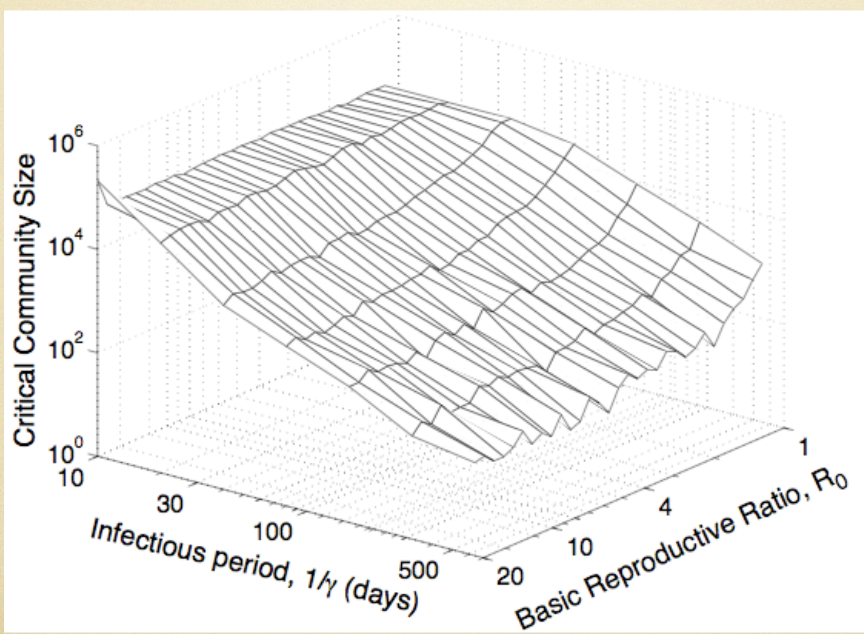
Different kinds of epidemics



Critical Community Size



Stochastic SIR simulations



Pros & Cons

- Gillespie's Direct Method has many virtues:
  - It's exact
  - Straightforward to implement
  - Widely known and used

- But it has one main drawback:
  - It's computationally costly when
  1. Population size is large
  2. There are lots of events (eg models with loss of immunity, many species etc)

Approximations

- Most commonly used approximation to GDM is so-called τ-leap method (also due to Gillespie)
- Here, system is updated at fixed time steps of length τ
- Assuming *rate* at which an event occurs is r_i, then number of times this event occurs in a small time interval τ is given by Poisson($r_i \tau$)
- [Can show GDM recovered as τ -> 0]

Example

- For SIR system without demography
  - Set parameter values and select τ
  - Set ICs and set time = 0
  - while time < Tmax
    - Infections = poissrnd(τ*beta*X(i)*Y(i)/N)
    - Recovery = poissrnd(gamma*Y(i)*τ)
    - X (i+1) = X(i) – Infections
    - Y(i+1) = Y(i) + Infections – Recovery
    - time = time + τ
  - end
- As $\tau \rightarrow 0$, this approximation approaches GDM
- Choice of τ is critical
  - too small and it's not much faster than GDM
  - yet too big can lead to imprecision and negative variables

Lecture Summary ...

- In reality, inferences about transmission process from data masked by 3 different kinds of noise: observation, environmental & demographic
- Can use Gillespie's Direct Method
- Demographic noise important, especially when making predictions in settings with R_0 near 1