Demographic stochasticity

Disease spread in populations of individuals

Jonathan Dushoff

Princeton University Department of Ecology and Evolutionary Biology

McMaster University Department of Biology

US-Africa Advanced Study Institute on Mathematical Modeling of Infectious Diseases in Africa June 2007

Notes

- The printed notes are meant to assist you only
- They are incomplete
  - In fact, I have no idea what they say

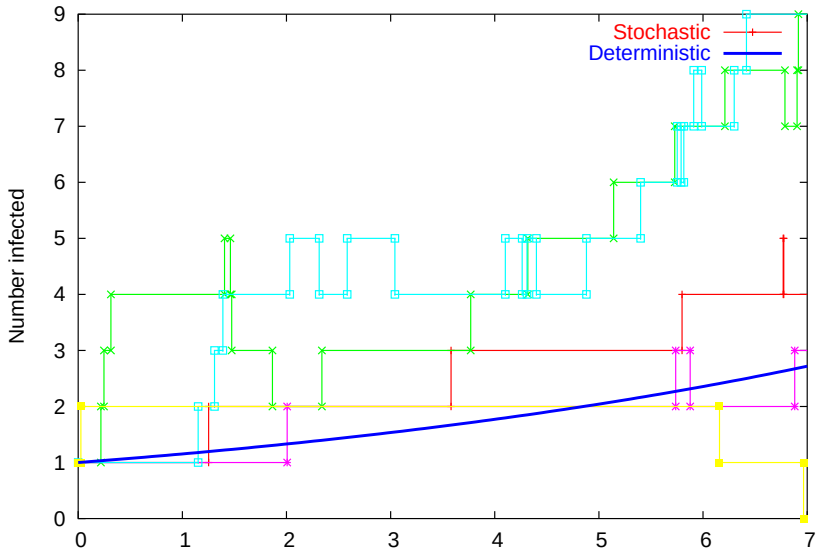
Modelling individual events

- Differential equations model continuous processes
- Disease spreads in the real world through discrete events
- Discrete events are fundamentally stochastic
  - Even in theory we don't know when the next event will occur, nor even what it will be

Types of stochasticity

- Demographic stochasticity is caused by the existence of individual people and discrete events
- Environmental stochasticity refers to events that affect more than one person at a time
  - Weather
  - Politics
  - Economics

Demographic spread

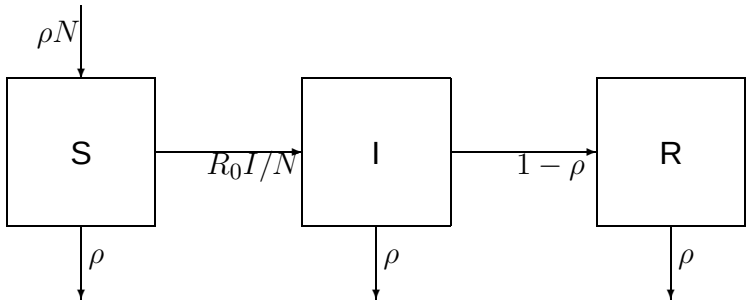


States and rates

- We describe our system in terms of the *probability rates* of events happening
  - If the rate of event E is $r_E(t)$, the probability of the event occurring in the time interval $(t, t+dt)$ is $r_E(t)dt$
- If the system is *Markovian*, $r_E(t)$ depends only on the state of the system at time t
  - The Markovian assumption is *extremely* convenient, but can have unwanted consequences
- If the system is *non-Markovian* $r_E(t)$ can depend (for example) on how long the system has been in a particular state

States and rates (Demographic)

SIR model with birth and death



Event	transition	rate
Infection	$S \to I$	$R_0 SI/N$
Recovery	$I \to R$	$(1-\rho)I$
Rebirth	$R \to S$	ρR
Rebirth	$I \to S$	ρI

Analogy

- The demographic model is an exact analogue of the deterministic one
  - Conceptually
  - In the limit as $N \to \infty$
- Choices may be required
  - Should we include stochasticity in population size?

Realizations and ensembles

- How do we think about the behavior of a stochastic process?
  - A single example of how the process could go (e.g., from a stochastic simulation) is called a *realization*
  - The universe of possible realizations is called the *ensemble*.
  - The probability distribution that describes what state we expect the population to be in at time t is called the *ensemble distribution*
    - Knowing how the ensemble distribution evolves is not the same as understanding the whole ensemble

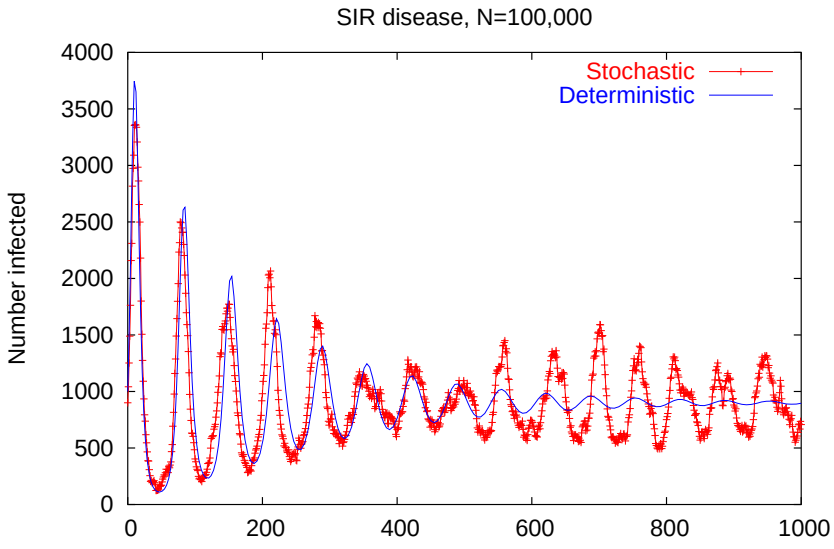
Some techniques

- Simulate one or many realizations
- Simulate the ensemble distribution
  - Requires one state variable for each possible state of the system
- Solve the ensemble distribution dynamics exactly!
  - Rarely possible
- Analytic approximations to the ensemble distribution

Simulating a realization

- Given a state of the system:
  - List possible events, and associated rates
  - Calculate the total rate r_T: this gives the rate at which the next event (whatever it is) will happen
    - i.e., an exponential waiting time with mean 1/r_T
  - The probability of event E is r_E/r_T
  - Randomly select the time and nature of the next event
  - Change the system state appropriately
  - Repeat forever
    - Or until system is extinct
    - Or until you are tired

Demographic model



Modelling the ensemble distribution

- We model the ensemble distribution by creating one conceptual ‘box’ for each possible state of the system, and asking what is the probability that the system is in each box.
  - This can be a lot of boxes
- The probabilities change as follows:
- $\dot{p}_S = \sum_{S'} p_{S'} r_{S' \to S} - p_S \sum_{S'} r_{S \to S'}$

Modelling the ensemble distribution

- If our system is small enough (particularly, if it has one state dimension) we might be able to simulate the ensemble distribution
- We might seek to solve these equations analytically
  - Only in special cases
- We might seek analytic approximations to increase our understanding
  - Moment approximations
  - Diffusion approximations

Questions

- What kind of questions do we want to ask with a stochastic model?
  - How does stochasticity affect disease dynamics?
    - Spatial distribution
    - Establishment
    - Persistence
  - How much variance do we expect stochasticity to cause?
  - Under what circumstances can we eliminate or eradicate a disease?

Equilibrium

- Define equilibrium as an ensemble distribution that does not change with time
- What are the equilibria of our stochastic SIR system?
  - Disease free equilibrium
  - Others?
- There is no equilibrium corresponding to the endemic equilibrium of the deterministic system!
  - As long as any populations not extinct, proportion extinct will increase.
  - This is deeply inconvenient

Quasi-equilibrium

- Consider the ensemble distribution confined to the subset of states where nothing is extinct
- This system can be described as a *fully connected* (you can get anywhere from anywhere else, *open* (you can leave the set) flow.
  - Intuition tells us that such a system has a unique, real, negative dominant eigenvalue, associated with a positive eigenvector
    - Linear algebra tells us this as well
  - *What does this mean?*
    - System will converge to a stable *relative* distribution of probabilities of being in each non-extinct box

Interpreting the quasi-equilibrium

- The quasi-equilibrium is the asymptotic distribution of system states given that nothing has gone extinct
- The eigenvalue λ_q associated with the quasi-equilibrium distribution is the rate at which the probability of non-extinction decays (exponentially)
  - The distribution of persistence times must be asymptotically exponential
  - The expected persistence time (looking forward) approaches $-1/\lambda_q$ as the system continues to persist

Modelling the quasi-distribution

- We can also model the probability of being in a particular state given that extinction has not occured
  - Computationally convenient
  - Can also keep track of cumulative extinction probability
- Define $q_S = p_S/(1 - p_N)$ where N is a 'null' state that we cannot escape from.
- $\dot{q}_S = \sum_{S'} q_{S'} r_{S' \to S} - q_S \sum_{S'} r_{S \to S'}$
  - $-q_S r_{S \to N}$
  - $+q_S \sum_{S'} q_{S'} r_{S' \to N}$

The fate of infectious disease

- Fizzle
  - Disease fails to "establish"
  - We will make this precise later
- Burn-out
  - Disease goes extinct after first epidemic
- Fade-out
  - Disease goes extinct after system approaches quasi-equilibrium
  - Can take a *long* time

Linearization

- Two of the most useful tools for understanding deterministic disease models are linearization:
  - **Disease-free equilibrium:** what factors control whether the disease can invade and persist?
  - **Endemic equilibrium:** tendency to cycle, damping or persistence of cycles
- Both of these methods have analogues in demographic models

Linear birth-death process

- We do an invasion analysis by asking how the number of infectives behaves in the limit where we assume that virtually the whole population is susceptible.
- This corresponds to a demographic model with the state determined by the number of infectious individuals I
- This system has only two events:
  - Infection at rate $R_0 I$
  - Recovery at rate I

Long-term behavior

- Unlike the finite systems discussed before, the probability of eventual extinction in this system is not one!
- Why not?
  - Probability of extinction given persistence goes to zero, as expected number of infectious individuals goes to ∞

Extinction probability

- Chains of infection are independent in this model
- We can use this fact to solve directly for the probability of extinction when starting from I infections, E_I
  - $E_I = R_0^{-I}$, when $R_0 > 1$
  - 1, otherwise
- We can define this as the 'fizzle' probability: the disease would have gone extinct even without depleting any susceptibles.

Moment calculations

- Ask: what is the expected behavior of the mean, variance, . . . of the ensemble?
- Define: $\mu = \sum_I I p_i$
- How does μ change through time?
  - $\dot{\mu} = \sum_I I \dot{p}_I$
  - $= \sum_I (b_I - m_I) p_I$, where $b(I) = R_0 I$ is the 'birth' rate, and $m(I) = I$ is the 'death' rate

Moment calculations

- How does μ change through time in the linear system?
  - $\dot{\mu} = \sum_I (b_I - m_I) p_I$
    - The birth rate is $b(I) = R_0 I$
    - The 'death' rate is $m(I) = I$
  - $\dot{\mu} = \sum_I (R_0 I - I) p_I$
  - $= (R_0 - 1) \sum_I I p_I$
  - $= (R_0 - 1)\mu$
- The expected value (average over realizations) grows exponentially!

Moment calculations

- How does μ change through time in a non-linear system?
  - $\dot{\mu} = \sum_I (b_I - m_I) p_I$
  - In general, we expect each moment to depend on higher moments:
    - e.g., mean depends on variance and skewness
- Moment-closure techniques, and linearization approximations, can help.

Moment-generating functions

- We can use a very slick mathematical trick: define $\phi = \sum_I p_I x^I$.
  - If we know the function $\phi(x)$ at a given time, we know all of the probabilities p_I.
- We can use our state/rate equations to write an expression for $\dot{\phi}$, which can be converted into a partial differential equation and analyzed, approximated, and (rarely) solved.
- For example, we can calculate the exact ensemble distribution through time of the linear birth-death process
  - Negative binomial with mean $I_0 \exp((R_0 - 1)t$ and parameter I_0

Invasion analysis

Summary

- We understand exact behavior of linearized invasion model
- This allows us to understand the properties of the SIR model, in the absence of effects due to depletion of susceptibles:
  - Fizzle probability
  - Size distribution of fizzled epidemics
  - Mean and variance in epidemic growth rate

Diffusion approximations

- We can approximate the discrete-valued demographic system with a real-valued system that reflects the mean *and variance* of the demographic system
  - Thus we can incorporate demographic stochasticity in a continuous system
  - An excellent approximation except when some values are very small
- In a linear (or linearized) system, we can solve the equilibrium distribution of the continuous equations, and thus approximate the quasi-equilibrium distribution
  - Disease persistence
  - Size of demographic fluctuations

Diffusion approximations

- We linearize about the endemic equilibrium, in exact analogy to Jacobian methods for stability in deterministic models
- Diffusion (and thus demographic stochasticity) is relatively unimportant when the square of the number infected is large compared to the demographic variance
  - Number infected at equilibrium: $\frac{(R_0-1)\rho N}{R}$
    - $\approx \rho N$
  - Demographic variance: $\approx N$
- Diffusion index $\approx \rho^2 N$. If ρ is small, demographic stochasticity can be important even for very large populations.

Diffusion approximations

Summary

- Linearization tells us *local* behavior of system
- Diffusion approximation is approximate
- Good estimate of the size of small oscillations
- Not a good estimate of extinction threshold
  - Extinction is a global problem
  - Small diffusion index implies no extinction
  - Large diffusion index implies non-local behavior
- Reference: Nasell 1999 (on ASI site)

Simulating distributions

- Practical for medium-sized “one-dimensional” systems or small “two-dimensional” systems
- Write differential equations and simulate
- The “quasi-” equations may work better, numerically
- There are probably tricks for finding approximate answers without modelling the whole state space

Lagrangian and Eulerian views

- In fluid mechanics:
  - Eulerian means a streambed-based viewpoint
    - Analogous to ‘box’ models of disease: how many people are here?
  - Lagrangian refers to fluid-based viewpoint
    - Analogous models of disease which follow each individual
- Does individual-based mean:
  - Demographic?
  - Lagrangian?

Lagrangian and Eulerian views

- Box view (Eulerian)
  - Usually more convenient
    - Fast simulation
    - Analytic techniques
- Individual view (Lagrangian)
  - Allows for realistic detail
    - Time distributions
    - Infection history
- Combined view
  - We can also use simple Eulerian models to *approximate* more realistic Lagrangian simulations

Exponential distributions

- The Markovian assumption is equivalent to assuming exponential time distributions of events
  - Can be spectacularly unrealistic
    - Does this mean that classical models are stupid?
    - The assumption needs to be evaluated in terms of the question the model is asking: how much does the assumption effect the dynamics?

Practical simulation modelling

- Eulerian (Markovian) approach
  - Erlang and pseudo-Erlang distributions
- Lagrangian extensions
  - Queueing
  - Dynamic updating

Markovian

- Given a state of the system:
  - List possible events, and associated rates
  - Calculate the total rate r_T: this gives the rate at which the next event (whatever it is) will happen
    - i.e., an exponential waiting time with mean 1/r_T
  - The probability of event E is r_E/r_T
  - Randomly select the time and nature of the next event
  - Change the system state appropriately
  - Repeat forever

Queueing

- One practical way to extend a model to allow non-Markovian (and non-Eulerian) complications is to keep of 'queue' of events that are slated to happen
  - Queue is sorted chronologically
  - At each step perform the next event, and add events triggered by it to the queue
    - When somebody is born, add their death to the queue
    - When somebody is infected, add their recovery, potentially infectious contacts to the queue

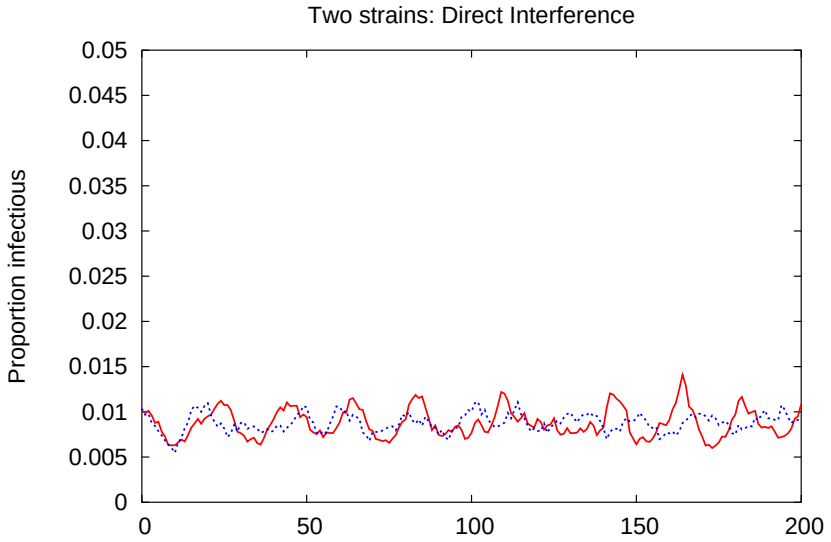
Dynamic updating

- If we assume that potentially infectious contacts are Markovian, we can model efficiently using dynamic updating:
  - Choose a potential infector at random and update their state
  - If the infector is infectious, choose a potential infectee at random and update their state

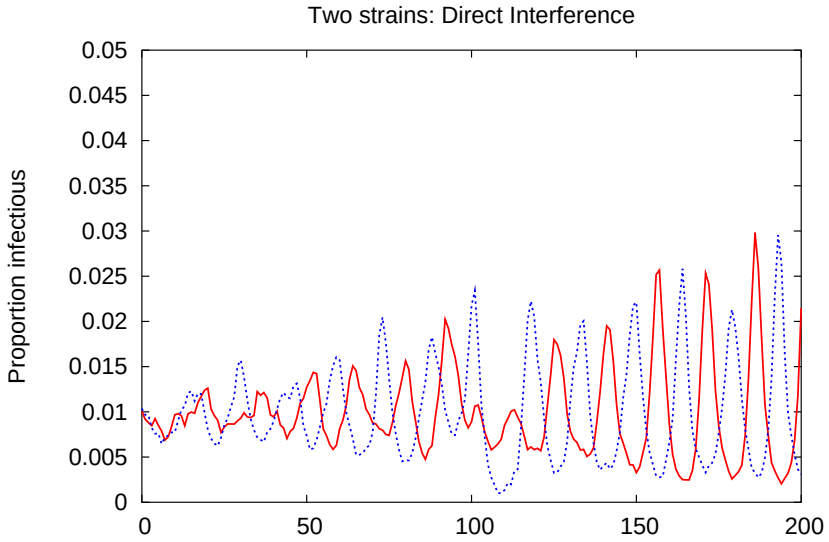
Multiple strains

- Many pathogens have different competing strains, and often new ones evolve on relatively should time scales
- This is another area where including demographic stochasticity has important effects

Markov (exponential) waiting times



Realistic (Gaussian) waiting times



Conclusions

- Treating individuals as individuals can have dramatic effects on models of disease transmission
  - Acquired immunity is an important part of this phenomenon
- Stochastic models are hard, and we usually combine techniques to understand them:
  - Analytic approximation
  - Simulating ensemble distributions
  - Simulating realizations
- Remember: demographic stochasticity is real!