An introduction to models for disease dynamics

Murali Haran

Department of Statistics
Penn State University

Spatial Epidemiology
SAMSI, December 2009

Modeling infectious diseases: motivation

Useful to model and understand disease dynamics because:

- Important for basic science as well as public policy.
- Models for infectious diseases are helpful for prevention and control of emerging infectious diseases like SARS, HIV/AIDS, H5N1 (“bird flu”), H1N1 (“swine flu”).
- Useful for studying the development and spread of drug resistant strains e.g. malaria, tuberculosis, MRSA.

Modeling infectious diseases: outline

We consider here models for directly transmitted, microparasitic infectious diseases. Outline:

1. Discussion of some simple deterministic models.
2. Discussion of simple stochastic models.
3. Some simple spatial (dynamic) models.

Caveats:

- The discussion of statistical inference will be relatively short as there is relatively little statistical inference in this field; more time will be spent simply discussing the deterministic or stochastic models for 'forward simulation'.
- The focus is on providing a feel for relevant models, terminology and methodology. Details are in references.

Deterministic models for infectious diseases

Basics: classifying individuals

- **Susceptible** Initially, individual/host is susceptible to infection: no pathogen is present; just a low-level nonspecific immunity within the host.
- **Exposed** In early stages, the host may or may not exhibit obvious signs of infection and abundance of pathogen may be too low to allow further transmission.
- **Infectious** Host encounters infectious individual and becomes infected with a microparasite; abundance of the parasite grows with time.
- **Recovered** The host is either no longer infectious or ‘removed’ (dead).

Implications of SEIR classification

The SEIR classification is clearly a simplification.

- This classification solely depends on the host's ability to transmit the pathogen.
- The health status of the host is therefore irrelevant — it is not important whether the individual is showing symptoms; an individual who feels perfectly healthy can be excreting large amounts of pathogen.
- Note that in reality boundaries between exposed and infectious and infectious and recovered are fuzzy because ability to transmit is not binary (on-off).
- Also, complications due to variablity in responses between different individuals and the variability in pathogen levels over the infectious period.

Some simplifications and variants

- Often mathematically simpler and justifiable to ignore the exposed class, resulting in *SIR* dynamics.
- *SI* models useful for some plant infections since they remain infected until they die.
- *SIS* models are appropriate for some diseases, in particular sexually transmitted diseases like gonorrhoea because once recovered, the host is once again susceptible to infection.
- Some diseases require model specifications with more classes of individuals, so they will not fit into SEIR framework. e.g. meningitis, chlamydia, MRSA.

Compartment models

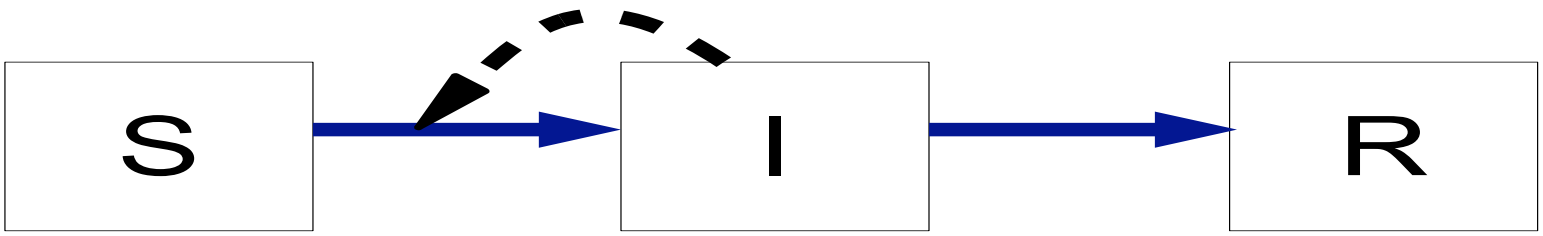
Flexible, interpretable framework:

- Vector-valued process: $X(t) = (X_1(t), \ldots, X_k(t))$ denoting integer (counts) or real-valued (e.g. proportions) in compartments $1, \ldots, k$.
- Define flow process: $N_{ij}(t) - N_{ij}(s)$ counts number of transitions from compartment i to compartment j between times s and t.
- N_{ij} is associated with a rate function $\nu_{ij}(t, X(t)) = \mu_{ij}(t, X(t))X_i(t)$.
- Interpretation using ordinary differential equations:

$$\frac{d}{dt} N_{ij}(t) = \nu_{ij}(t, X(t))$$

for $i \neq j$.

SIR model flow diagram



Blue/solid arrows: movement between classes.

Black/dotted arrow: level of infectious disease influences the rate at which a susceptible individual moves into the infected class. Assumption: disease confers lifelong immunity. e.g. measles.

SIR models

Basic SIR models make the following assumptions:

- Individuals are born into the susceptible class.
- Susceptible individuals have never come into contact with the disease and are able to catch the disease, after which they move into the infected class.
- Infected individuals spread the disease to susceptibles, and remain in the infected class (the infected period) before moving into the recovered class.
- Individuals in the recovered class are assumed to be immune for life.

SIR model formulation (without demography)

Classical SIR model assumes that individuals that leave one class must enter another. For a closed population (no births, deaths, migration):

- S, I, R are proportions susceptible, infectious and recovered. $S + I + R = 1$, so if S, I are known, R is known.

$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I.$$
$$\frac{dR}{dt} = \gamma I.$$

- Initial conditions: $S(0), I(0) > 0$, $R(0) = 0$.
- γ = removal or recovery rate. $1/\gamma$ is average infectious period. Generally well determined from epidemiological data.

Basic SIR model formulation [cont'd]

- β is a transmission parameter.
- Analytical expressions for dynamics of S, I through time are not available in closed form. Need to solve numerically.
- Derivation of transmission term (first equation) is possible from first principles.

Basic SIR model: force of infection

- *Force of infection*, λ is the per capita rate at which susceptible individuals contract the infection. Rate at which new infecteds are produced = λX where X=number of susceptibles.
- Intuitively, $\lambda \propto$ number of infectious individuals.
- Kinds of transmission:
  - Frequency-dependent or 'mass action transmission': $\lambda = \beta Y/N$, where Y=# of infectious,$N =$ total population size, β= product of contact rates and transmission probability. Assumption: # of contacts is independent of N.
  - Density-dependent transmission: $\lambda = \beta Y$. Assumption: as the density of individuals increases, so does contact rate.
- If we assume host population size stays constant, let $S = X/N$, $I = Y/N$.

The threshold phenomenon and R_o

Consider initial stages after $I(0)$ infectives introduced into a population of $S(0)$ susceptibles.

- Will an epidemic occur or will the invasion fade?
- Rewrite equation: $\frac{dI}{dt} = I(\beta S - \gamma)$.
  - If $S(0) < \gamma/\beta$ then $\frac{dI}{dt} < 0$ and infection dies out.
- Known as 'threshold phenomenon' (Kermack and McKendrick, 1927) since susceptibles must exceed a critical threshold for an infection to invade.

The threshold phenomenon and R_o

- Another important interpretation: need γ/β, the relative removal rate, to be small enough.
- **Basic reproductive ratio**, $R_o = \beta/\gamma$ = average number of secondary cases arising from an average primary case in a completely susceptible population.
- If everyone in the population is initially susceptible ($S(0) = 1$), a pathogen can invade only if $R_0 > 1$.
- R_o = rate at which new cases are produced by an infectious individual (when population is completely susceptible), multiplied by the average infectious period.
- Examples of R_o: 1.1-1.5 for FIV (Smith, 2001), 3-4 for influenza (Murray, 1989), 10-12 for chickenpox and 16-18 for measles (Anderson and May, 1991, 1982).

The importance of R_o

For an infectious disease with average infectious period $1/\gamma$ and transmission rate β, $R_o = \beta/\gamma$:

- For a closed population, an infectious disease can only invade if there is a threshold fraction of susceptibles greater than $1/R_o$.
- Vaccination policy: if proportion of susceptibles is reduced to below $1/R_o$, can eradicate the disease.

Epidemic burnout

What happens in the long run (asymptotically?)

- Dividing equation (1) by (3), $\frac{dS}{dR} = -\beta S/\gamma = -R_0 S$.
- Integrating w.r.t. R, $S(t) = S(0)e^{-R(t)R(0)}$, assuming $R(0) = 0$.
- Note that $S(t) \geq e^{-R_0}$ since $R(t) < 1$. Hence, always have a fraction of susceptibles in the population.
- Hence, chain of transmission eventually breaks down due to lack of number of infecteds, not lack of number of susceptibles (counter-intuitive!)

Epidemic burnout [cont'd]

- Since $R + S + I = 1$ and epidemic dies if $I(t) = 0$, long term behavior of (4): $S(\infty) = 1 - R(\infty) = \exp(-R(\infty)R_0)$ or $1 - R(\infty) - \exp(-R(\infty)R_0) = 0$.
- Numerical solution to this suggests that for large R_0, say $R_o > 5$, virtually everyone in a well mixed population will eventually be infected.
- Can also obtain previous result with a simple probabilistic argument.

SIR model with demography

More general formulation of SIR model:

- Assume rate at which individuals in any class suffer natural mortality = μ. Historically, μ is also crude birth rate (to keep total population constant: $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$.

$$\begin{aligned}\frac{dS}{dt} &= \mu - \beta SI - \mu S \\ \frac{dI}{dt} &= \beta SI - \gamma I - \mu I \\ \frac{dR}{dt} &= \gamma I - \mu R.\end{aligned}$$

- If we assume entire population is susceptible,

$$R_0 = \frac{\beta}{\gamma + \mu}$$

(R_0= transmission rate* infectious period)

SIR model with demography: equilibrium

- Inclusion of host demographic dynamics may permit a disease to persist in a population in the long term.
- *Disease-free equilibrium*: pathogen has suffered extinction and everyone in the population is susceptible.
  (S^*, I^*, R^*)=(1,0,0). How likely is this? Consider $\frac{dI}{dt}$:

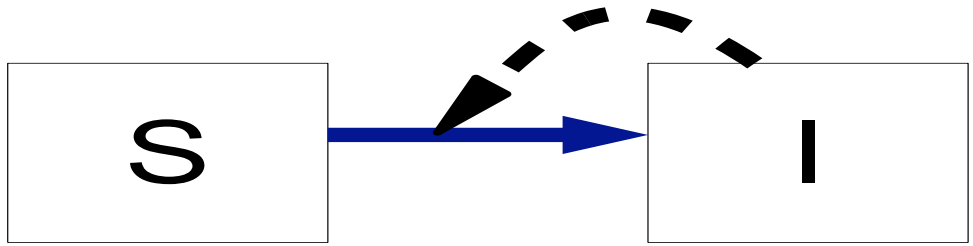
$$\beta SI - (\gamma + \mu)I = 0.$$

  After factoring for I, $I(\beta S - (\gamma + \mu)) = 0$.
- Above is satisfied when $I^* = 0$ or $S^* = \frac{\gamma+\mu}{\beta}$.
- $I^* = 0$ is disease-free equilibrium.

SIR model equilibrium [cont'd]

- $S^* = \frac{\gamma+\mu}{\beta} = \frac{1}{R_0}$. Hence, *endemic equilibrium* (where the disease is always present without any re-introduction) is characterized by the fraction of susceptibles in the population being the inverse of R_o. This equilibrium is only feasible when $R_0 > 1$ (otherwise positivity of variables not always satisfied.)
- Since $S^* = 1/R_0$, can solve for $\left(S^*, I^*, R^*\right) = (\frac{1}{R_0}, \frac{\mu}{\beta}(R_0 - 1), 1 - \frac{1}{R_0} - \frac{\mu}{\beta}(R_0 - 1)\Big)$.
- Endemic equilibrium is stable if $R_0 > 1$ otherwise disease-free equilibrium is stable.

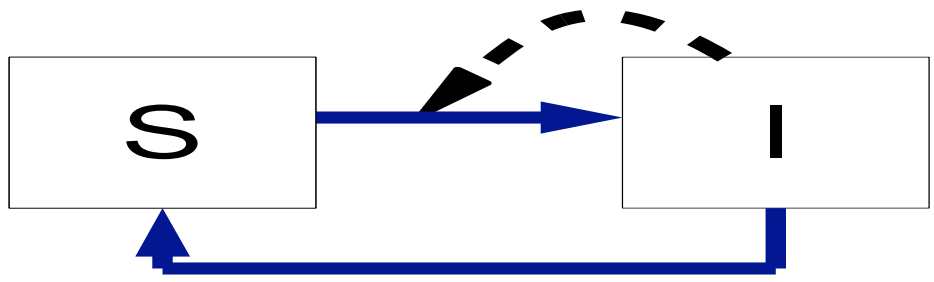
SI model flow diagram



Blue/solid arrows: movement between classes.

Black/dotted arrow: level of infectious disease influences the rate at which a susceptible individual moves into the infected class. Assumption: infections are fatal. e.g. highly pathogenic avian influenza (H5N1), feline infectious peritonitis (FIP).

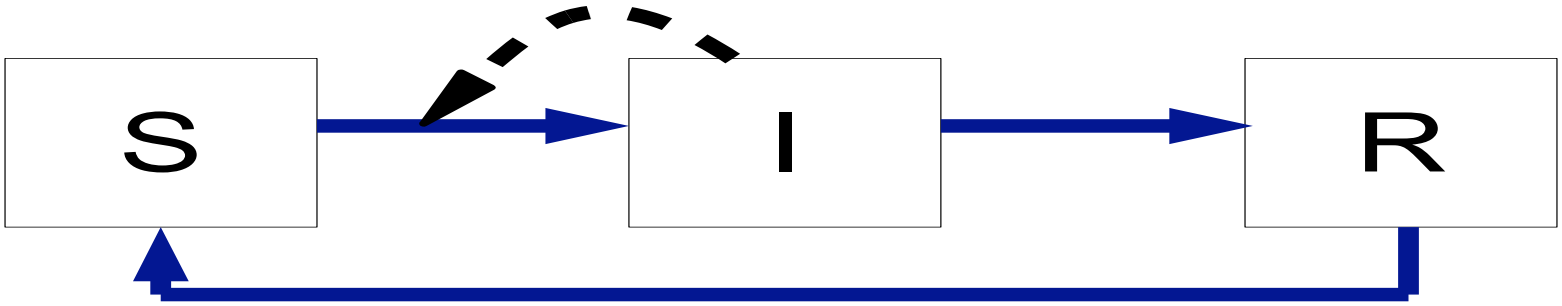
SIS model flow diagram



Blue/solid arrows: movement between classes.

Black/dotted arrow: level of infectious disease influences the rate at which a susceptible individual moves into the infected class. Assumption: disease does not confer long-lasting immunity. e.g. rotaviruses, STDs.

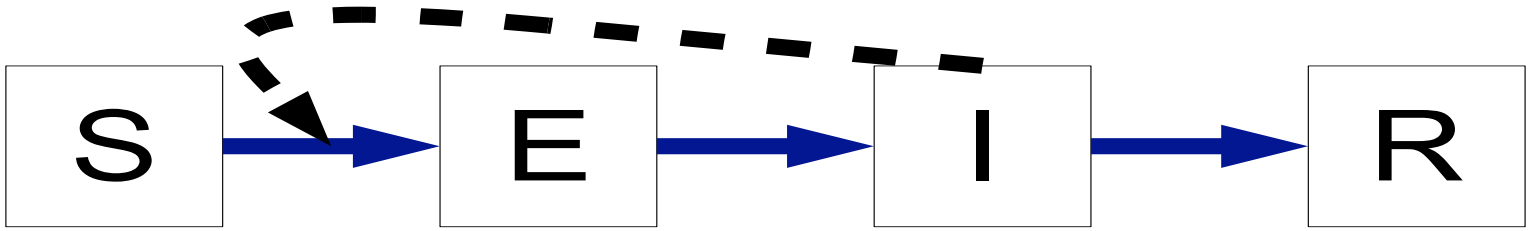
SIR waning immunity model flow diagram



Blue/solid arrows: movement between classes.

Black/dotted arrow: level of infectious disease influences the rate at which a susceptible individual moves into the infected class. Assumption: immunity lasts for a limited period before waning such that the individual is once again susceptible. e.g. syphillis.

SEIR model flow diagram



Blue/solid arrows: movement between classes.

Black/dotted arrow: level of infectious disease influences the rate at which a susceptible individual moves into the infected class. Assumption: account for latent period when pathogen is present in individual (infected) but not transmittable to other susceptibles (not infectious).

Important features/parameters in SEIR models

- R_0: is the basic reproductive ratio.
- Timescale of infection: measured by the infectious period for SIS and SIR infections or by a mixture of exposed and infectious periods in diseases with SEIR dynamics.
- Infectious diseases are highly variables in terms of infectious profiles, parameter values, and timescales.

Many other variants of these models, specialized to each of several infectious diseases.

Discrete-time models: example

- An alternative to continuous-time models.
  - Statistical inference may be easier.
  - Data are observed at discrete intervals.
  - Issue: may lead to severely biased conclusions (Glass et al., 2003).
- For e.g. assume that disease has latent and infectious periods of exactly one week, and:

$$\begin{aligned} S_{t+1} &= \mu - S_t \exp(-\beta I_t) \\ E_{t+1} &= S_t(1 - \exp(-\beta I_t)) \\ I_{t+1} &= E_t. \end{aligned}$$

with μ = weekly per capita births. First equation = per capita probability of not contracting the infection given I_t infectives with transmission β.

Equilibrium for discrete-time models

- Note: β is now analogous to maximum reproductive potential (R_0) of the infection.
- Can still perform equilibrium analysis. For this discrete-time model, obtain solutions:

$$
\begin{aligned}
S^* &= \frac{\mu}{1 - \exp(-\beta\mu)}, \\
E^* &= \mu, \\
I^* &= \mu.
\end{aligned}
$$

Stochastic models

Deterministic to stochastic models

There are many strong arguments for moving away from deterministic to stochastic models:

- Real processes are stochastic. At the very least, adding stochasticity to a model gives it a great deal of added flexibility so the model better fits real data.
- A stochastic framework can allow for important parameters of a model to be identified even when they are unidentifiable in a deterministic setting.
- Can understand *distributions* associated with characteristics of a process. Often of great interest.
- Simulation of such stochastic models may be easier (though expensive).
- Rigorous procedures may be available for inference.

Important statistical challenges

Major issues raised by Bjornstad and Grenfell (2001) in the context of ecological modeling and inference (not just disease modeling):

- Combining measurement error and process error.
- Including covariates in mechanistically plausible ways.
- Continuous-time models.
- Modeling and estimating interactions in coupled systems.
- Dealing with unobserved variables.
- Modeling spatial-temporal dynamics.

Sound statistical inference for even fairly simple models is still generally challenging.

Statistical inference

- Re-examining our notion of a model: Scientists working in the physical and natural sciences are often interested in learning about the mechanisms or ‘laws’ and processes underlying physical phenomena. (Not enough to simply fit statistical models to observations to get a good fit.)
- Time series data sets, often with spatially explicit information, are now increasingly available. Hence, it is possible to learn about disease dynamics from data though incomplete data/information continues to pose challenges.
- However, statistical inference for such models can be very challenging since likelihoods associated with such models may be computationally very expensive or not even available in tractable form.

Chain binomial models

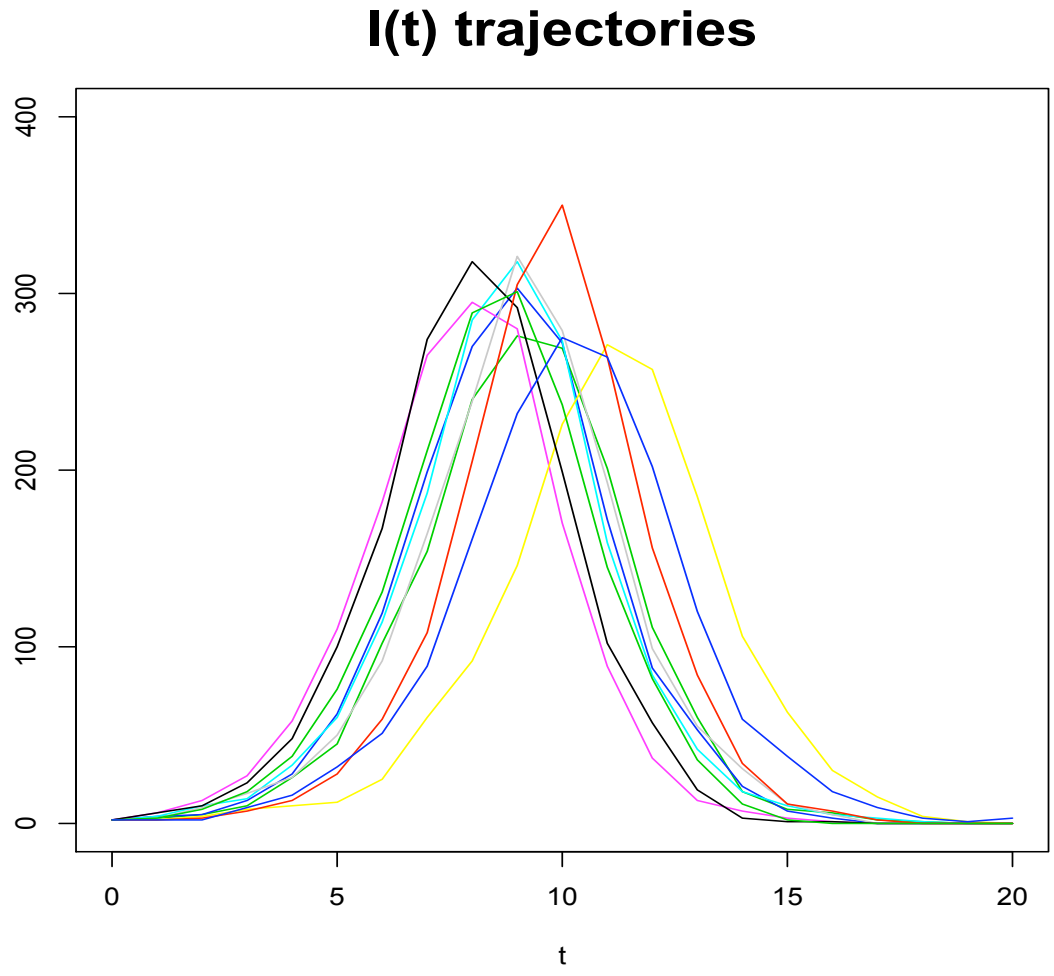
- A popular discrete-time stochastic model is the simple chain binomial model.

$$I_{t+1} \sim \text{Bin}(S_t, 1 - \exp(-\beta I_t)).$$
$$S_{t+1} = S_t - I_{t+1}.$$

- Assumes that generations are perfectly synchronized. Major assumption, though may be reasonable for some diseases.

Chain binomial models: simulations

Easy to simulate from the chain binomial model. Multiple realizations below from same parameter setting.



Spatial and temporal models

Spatial Modeling for infectious diseases

- Space can play a significant role in disease dynamics.
- The spatial models that describe dynamics are different from ‘typical’ spatial models that are primarily focused on modeling dependence.
- We will discuss spatial models that focus on:
  - Determining the rate of spatial spread of a pathogen.
  - Calculating the influence of large populations on smaller ones.
- These models work by partitioning the population according to the spatial position of the hosts: nearby hosts are grouped together and interact more strongly.

Spatial modeling concepts

- Heterogeneity: either describe underlying differences between two populations at two different locations or emerging dynamic differences in the population levels (e.g. proportion of population that are infectious).
- Correlations between subpopulations: positive/negative correlation indicates that epidemics are spatially synchronous/asynchronous.
- Spatial coupling allows infection to spread and acts to synchronize epidemic dynamics at the two locations. Often use a *transmission kernel* K, a function of the distance between the populations, to modify coupling term. For e.g. exponential $K \propto \exp(-Ad)$ where $d =$distance between populations.

Spatial modeling concepts [cont'd]

- Localized extinctions are common in smaller subpopulations, while large-scale eradication is prevented by coupling between subpopulations leading to reintroduction of infection into disease-free areas.
- Scale is critical: both how the population can be subdivided and the scale at which subpopulations interact. A common issue is striking a balance between known scales for interaction for a disease and computational considerations.

Metapopulations

- Metapopulations: provide a powerful framework for modeling disease dynamics for hosts that can be naturally partitioned into spatial subunits.
- Spread of human diseases in metapopulations is best captured by rapid commuter movements of individuals from their home subpopulation to another subpopulation and back again.

A simple two-patch spatial model

Consider two patches where hosts move between the two patches at some rate independent of disease status. Simple closed SIR two-patch model:

$$\begin{aligned}
\frac{dS_1}{dt} &= -\beta S_1 I_1 + m(S_2 - S_1) \\
\frac{dI_1}{dt} &= \beta S_1 I_1 - \gamma I_1 + m(I_2 - I_1) \\
\frac{dS_2}{dt} &= -\beta S_2 I_2 + m(S_1 - S_1) \\
\frac{dI_2}{dt} &= \beta S_2 I_2 - \gamma I_2 + m(I_1 - I_2).
\end{aligned}$$

Subscripts 1 and 2 indicate the two patches. m is a migration parameter.

Categories of flexible spatial models

- Coupled lattice-based models: subpopulations are arranged on a grid and coupling is generally to the nearest neighbors only.
- Individual-based models: account for spatial interaction between individual hosts distributed on a spatial landscape.
- Spatial networks: consider nature of individual disease transmission — two individuals are linked if infections can pass between them.
- Continuous-space, continuous-time models: use partial differential equations (PDEs) or integro-differential equations (IDEs). Flexible but mathematically complex.

Time series SIR model

A very simple model where we assume the discrete time step equals the generation time of the pathogen (for e.g. 2 weeks for measles), ignoring the latent period:

- $S_{t+1} = S_t + B_t - I_{t+1}$, and $\lambda_{t+1} = \beta S_t I_t^{\alpha}$, where
  - S_t, I_t are the numbers of susceptibles and infecteds in (pathogen) generation time t.
  - B_t is the number of births (into susceptible class).
  - β is the transmission rate.
  - λ_{t+1} is the expectation for the new number of infecteds (I_{t+1}) in the next generation.
  - α is a parameter to try to 'adjust' partially for discretizing the underlying continuous process.
- Number of infecteds may also be stochastic, for e.g. $I_{t+1} \sim \text{Poi}(\lambda_{t+1})$ or $I_{t+1} \sim \text{NegBin}(I_t, \lambda_t)$.

T-SIR model with environmental drivers

- Can allow for environmental factors to influence a number of variables in the T-SIR model. For example: through variable numbers of births of susceptibles, i.e., B_t. Or through the transmission rate to vary.
- A possible simplification is to have the number of infecteds, I_t, be non-stochastic, i.e., $I_t = \lambda_t$.
- This is a simple model to simulate in principle (though simulation can become computationally expensive.) Thus, it is possible, in principle, to study the behavior of the model for various values of the parameters.

T-SIR model with environmental drivers: simple inference

- Consider well known data on measles:
  - The UK Registrar General's data for 952 cities in England and Wales for years 1944-1966 of biweekly incidences of measles. Very rich spatio-temporal data.
  - Data for number of susceptibles from standard susceptible reconstruction algorithms (Fine and Clarkson 1982a, Schenzle 1984, Ellner et al. 1998, Bobashev et al. 2000, Finkenstadt and Grenfell 2000).

T-SIR model simple inference [cont'd]

- Susceptible reconstruction done using approach in Bobashev et al. and Finkenstadt and Grenfell (1998).
- Once susceptibles are known, a rough approximation for the parameters β, α may be obtained by setting I_{t+1} to its expectation:

$$\log(I_{t+1}) = \log(\beta) + \log(S_t) + \alpha \log(I_t),$$

then estimating α, β by regressing $\log(I_{t+1})$ on $\log(I_t)$ with an offset of $\log(S_t)$. Intercept and slope estimates are estimates of $\log(\beta)$ and α respectively (Bjornstad and Grenfell, 2002).

State-space models

State-space models allow a very flexible framework for incorporating dynamics/mechanisms based on scientific understanding.

- Unobserved/latent state process: x_t.
- Observations: y_t.
- Assumption: observation process y_t is conditionally independent of the past given x_t.
- Model for x_t can be scientifically based. Can incorporate spatial dynamics, for example.
- Can incorporate measurement error in model for observations.

Forward and inverse problems

- Forward problem: given (i) dynamical system that describes how some process $\{y_t\}$ evolves over time (t), (ii) parameters θ that specify this system, and (iii) initial value of the system, y_{t0}, what is y_t for $t > t_0$?
- Inverse problem: given (i) dynamical system that describes how some process $\{y_t\}$ evolves over time (t), and (ii) values of the system, $y_{t0}, \ldots, y_t$, what is θ?
- Huge literature on this: applied mathematics, physics, engineering etc. Serious identifiability issues, so often apply regularization approaches.
- Lots of approaches based on least squares fits of model to observations. Often, assumption: model has no biases, no stochasticity, model (function) is easy to evaluate.

Estimating parameters in deterministic models

- Parameter estimation in deterministic models is often carried out using numerical techniques.
- Since learning about the influence of the parameters on the behavior of the model is of much interest, it is critical to carry out a sensitivity analysis.
- “Sensitivity analysis is often used to study how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation, and of how the given model depends on the information fed into it.” (see Saltelli et al., 2000, 2008)
- Some likelihood-based approaches are available, but can be very expensive since a large number of missing parameters need to be integrated out.

Plug-and-play inference for disease dynamics

- Since statistical inference for mechanistic models for disease dynamics is very complicated, models fit tend to be rather simplistic.
- Examples of desired flexibility (beyond allowing for stochasticity): non-stationarity, measurement error, and unobserved (latent) variables.
- 'Plug-and-play' methods (e.g. He et al., 2009) allow for more flexible models since they require only simulations from the model. Sometimes called 'equation-free' (Kevrekidis et al., 2004).

Approximate Bayesian computation (ABC)

- ABC (Tavare et al., 1997; Beaumont, Zhang, Balding, 2002) is an approach to perform approximate Bayesian inference without evaluating the likelihood.
- General approach: use simulations from complicated probability model in place of evaluating the likelihood.
- The basic idea and variants are useful in classical Monte Carlo, Markov chain Monte Carlo (Marjoram, Molitor, Plagnol, Tavare, 2003), and sequential Monte Carlo (Sisson et al., 2007).
- For approach to be feasible, need simulation from probability model to be very cheap.

Note: there is still a likelihood (it is just not evaluated) so perhaps not ideal to call it 'likelihood-free'.

ABC

Simple set-up:

- Let $\mathbf{Y} \mid \theta \sim f(\mathbf{Y} \mid \theta)$ be the probability model for data $\mathbf{Y}$. θ is the parameter (often a vector of parameters) of interest.
- Let $p(\theta)$ be a prior distribution for θ.
- Inference for θ based on posterior $\pi(\theta \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \theta)p(\theta)$.
- Challenge: $f(\mathbf{Y} \mid \theta)$ is either analytically intractable or too expensive to evaluate repeatedly.

ABC rejection sampling

- Simplest rejection sampler:
  - Generate $\theta^* \sim p(\cdot)$.
  - Accept θ^* with probability $h(\theta^*; \mathbf{Y}) = f(\mathbf{Y} \mid \theta^*)$.
  - Repeat above: accepted θ^*s have distribution $\pi(\theta^* \mid \mathbf{Y})$.

  Since $h(\theta^*; \mathbf{Y})$ is intractable or too expensive, this approach is not practical.

ABC rejection sampling [cont'd]

- Alternative rejection sampler:
  - Generate $\theta \sim p(\cdot)$.
  - Simulate $\mathbf{Y}^* \sim f(\cdot \mid \theta^*)$.
  - Accept θ^* if $\mathbf{Y}^* = \mathbf{Y}$.
  - Repeat above: accepted θ have distribution $\pi(\theta \mid \mathbf{Y})$.

  This is usually not practical since the probability that $\mathbf{Y}^* = \mathbf{Y}$ is generally very small (and for a continuous state space, it has probability 0).

ABC rejection sampling [cont'd]

- Approximate rejection sampler:
  - Generate $\theta \sim p(\cdot)$.
  - Simulate $\mathbf{Y}^* \sim f(\cdot \mid \theta^*)$.
  - Accept θ^* if $\rho(\mathbf{Y}^*, \mathbf{Y}) < \epsilon$, where $\rho(\mathbf{Y}^*, \mathbf{Y}), \epsilon > 0$ are a distance and threshold defined by the user.
- As $\epsilon \to \infty$, this algorithm generates observations from the prior. As $\epsilon \to 0$, this algorithm generates from $\pi(\theta \mid \mathbf{Y})$.
- Often, the distance is defined on same summary statistics on $\mathbf{Y}$, say $S(\mathbf{Y})$, rather than on $\mathbf{Y}$ itself. That is, $\rho(\mathbf{Y}^*, \mathbf{Y}) = \rho(S(\mathbf{Y}^*), S(\mathbf{Y}))$. This is particularly useful when $\mathbf{Y}$ is high dimensional.

Likelihood-free MCMC

The previous algorithm is not very general. $p(\theta)$ will generally work poorly as a proposal for $\pi(\theta \mid \mathbf{Y})$ and it may also be very difficult to find another reasonable proposal, especially if θ has more than a few dimensions.

- Recall that the Metropolis-Hastings algorithm constructs a Markov chain with stationary distribution $\pi(\theta \mid \mathbf{Y})$ by generating the next state of the Markov chain as follows:
  - If current state is θ, propose a move to θ^* according to a transition kernel $q(\cdot \mid \theta)$.
  - Calculate acceptance probability, $\alpha(\theta, \theta^*) = \min\left(1, \frac{h(\theta^*;\mathbf{Y})}{h(\theta;\mathbf{Y})} \frac{p(\theta^*)}{p(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\right)$.
  - Accept θ^* as the next state with probability $\alpha(\theta, \theta^*)$.

Likelihood-free MCMC [cont'd]

- Again, $h(\theta^*; \mathbf{Y})$ is either intractable or too expensive.
- Likelihood-free MCMC:
  - If current state is θ, propose a move to θ^* according to a transition kernel $q(\cdot \mid \theta)$.
  - Generate $\mathbf{Y}^* \sim f(\cdot \mid \theta^*$.
  - If ($\mathbf{Y}^* \neq \mathbf{Y}$), reject θ^* (stay at θ). If ($\mathbf{Y}^* = \mathbf{Y}$) calculate acceptance probability, $\alpha(\theta, \theta^*) = \min\left(1, \frac{p(\theta^*)}{p(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\right)$. Accept θ^* as the next state with probability $\alpha(\theta, \theta^*)$.
- Avoided evaluating h but this Markov chain has stationary distribution $\pi(\theta \mid \mathbf{Y})$. Proof is a simple reversibility argument (see Marjoram, Molitor, Plagnol, Tavare, 2003).

Likelihood-free MCMC [cont'd]

- As before, $\mathbf{Y}^* = \mathbf{Y}$ is very unlikely (or has zero probability) in most cases.
- Approximate likelihood-free MCMC:
  - If current state is θ, propose a move to θ^* according to a transition kernel $q(\cdot \mid \theta)$.
  - Generate $\mathbf{Y}^* \sim f(\cdot \mid \theta^*)$.
  - If $(\rho(S(\mathbf{Y}^*), S(\mathbf{Y})) > \epsilon)$, reject θ^* (stay at θ). If $(\rho(S(\mathbf{Y}^*), S(\mathbf{Y})) < \epsilon$ calculate acceptance probability, $\alpha(\theta, \theta^*) = \min\left(1, \frac{p(\theta^*)}{p(\theta)} \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\right)$. Accept θ^* as the next state with probability $\alpha(\theta, \theta^*)$.
- Avoided evaluating h. The idea: if ϵ is small this Markov chain has *approximately* the right stationary distribution $\pi(\theta \mid \mathbf{Y})$. No really sound theoretical basis for convergence of estimates based on this algorithm.

Sequential Monte Carlo without likelihoods

- For many models (especially state-space models), rejection sampling or Markov chain Monte Carlo can be highly inefficient and accordingly require far more iterations than may be practical to implement.
- Sequential Monte Carlo methods (Doucet et al, 2001) may be much more efficient/practical in such situations. However, these algorithms also require evaluation of the likelihood.
- Sisson et al. (2007) develop a likelihood-free version of sequential Monte Carlo.

ABC

- ABC has been used for inference for the temporal SEIR model (McKinley, Cook, Deardon, 2009).
- Likelihood-free MCMC appears to be harder to tune and less efficient to use than likelihood-free sequential Monte Carlo. (McKinley et al., 2009). Also useful to consider Beaumont et al. alternative to ABC with MCMC and SMC.
- ABC approaches require choosing appropriate summary statistics and tolerance levels and accuracy of inference is sensitive to these choices.

Inference for parameters in deterministic models

- To learn about mechanisms and processes, scientists build complicated models that are usually numerical solutions of complex mathematical models.
- Translated into computer code so that they can study simulations of their physical processes under different conditions.
- When the model is deterministic and complex, it may not be obvious how to learn about the model parameters.
- For stochastic models where simulation is expensive, it may not be possible to use likelihood-free inference.
- Learning about parameters in a complex deterministic or stochastic model using real ('field') data can be thought of as a 'computer model emulation/calibration' problem.

Complex stochastic models and likelihood inference

- Consider the general case that the probability model for Z depends on some parameter θ.
- If the likelihood function for this probability model is explicit, we have $\mathcal{L}(Z \mid \theta)$ and we can perform likelihood-based inference, either finding the maximum likelihood estimator of θ or the posterior distribution for $\theta \mid Z$ after specifying a prior for θ.
- If the (assumed) mechanism/process to simulate Z is provided, but no likelihood corresponding to it is available, the likelihood is *implicit* and hence likelihood-based inference may be challenging. Advantage of this: scientists build models that correspond to their scientific interests and goals.

Computer model emulation via Gaussian processes

- An emulator (or 'meta-model') is an approximation of a complex computer model.
- An emulator is constructed by fitting a model to a training set of runs from the complex computer model.
- The emulator serves as a surrogate for the computer model, and is much faster/simpler. Hence, it is possible to simulate output from the emulator very quickly.
- The advantage of doing it in a probabilistic framework:
  - Uncertainties associated with interpolation (predictions), for example greater uncertain where there is less training data information.
  - Probability model: useful for statistical inference.
  - "Without any quantification of uncertainty, it is easy to dismiss [computer] models." (A.O'Hagan)

Computer model calibration

- Statistical problem: given (i) model (simulation) output at several parameter settings, and (ii) observations of the real process being modeled by the computer code, what is the value of the parameter that best 'fits' the observations?
- Notation:
  - Computer model output $\mathbf{Y} = (Y(\theta_1), \ldots, Y(\theta_n))$.
  - Observation Z, assumed to be a realization of computer model at 'true' θ + discrepancy + measurement error.
  - Want to perform inference for 'true' θ.
- Ideally done in a Bayesian setting:
  - There is often real prior information about θ.
  - The likelihood surface for θ may often be highly multimodal; useful to have access to the full posterior distribution.
  - If θ is multivariate, may be important to look at bivariate and marginal distributions (easier w/ sample-based approach).

Computer model calibration [cont'd]

- Field data = computer model + model discrepancy (structural error, biases) + measurement error

$$Z(x) = Y(x, \theta) + \delta(x, \theta) + \epsilon(x).$$

x: controllable (e.g. location/time), θ is unknown parameter.

- It is important to model $\delta(x, \theta)$ (not appropriate to assume i.i.d. error), as this may result in over-fitting/biased θ as it tries to compensate for model inadequacy.
  - GP model for $Y(x, \theta)$ since it is an unknown function.
  - GP model for $\delta(x, \theta)$. It is also an unknown function.
  - $\epsilon(x) \overset{iid}{\sim} N(0, \psi), \psi > 0$.
- Place prior on θ, inference based on $\pi(\theta \mid \mathbf{Y}, \mathbf{Z})$.
- Obvious that there are a lot of identifiability issues.

Some of the many topics not discussed

- *Filtering* approaches. New approach: iterated filtering (Ionides et al., 2006) — uses a Bayesian formulation to obtain maximum likelihood estimates.
- *Network models* recognize that each individual has a finite set of contacts to whom he can pass infection. Learning about the network allows individual-level dynamics. This is an active area of research, and has seen the development of statistical approaches. For e.g. Keeling and Eames (2005).
- *Multitype epidemic models* with varying levels of mixing: the population is divided into different groups with different transmission rates within and between the groups. For e.g. Demiris and O'Neill (2005).

Summary

- Modeling disease dynamics is a very active and exciting research area with a huge number of interesting applications.
- Historically, disease modeling has been dominated by deterministic, differential equations-based models.
- While there are many researchers working on statistical methods, sound inferential techniques are still in their infancy relative to the forward models developed.
- Modeling dynamics in space is a particularly interesting and challenging area for research.

Acknowledgments and references

- Most of this lecture follows the book: Keeling and Rohani (2007) “Modeling Infectious Diseases in Humans and Animals.”
- He, Ionides, King (2009) “Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.”
- Ottar Bjørnstad’s notes.
- Bjørnstad, Finkenstad, Grenfell (2002) “Dynamics of measles epidemics: Estimating scaling of transmission rates using a Time series SIR model.”
- Comments from D.Welch.

Acknowledgments and references

- M.A.Beaumont, W.Zhang and D.J.Balding (2002, Genetics) “Approximate Bayesian Computing.”
- Online lecture notes by Ed Ionides: `http://www.stat.lsa.umich.edu/~ionides/`
- D.J.Wilkinson (2007) “Stochastic Modeling for Systems Biology.”
- Saltelli, Chan, Scott (2000; 2008) "Sensitivity Analysis".