Statistical Inference for Spatial and Structured Population Epidemic Models

P.D. O'Neill
University of Nottingham

Some Apologies

- This review is far from comprehensive
- It reflects my own interests
- If I omit your work, it's not personal….
- …..but please mention it in discussions this week!

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

1. Space and Structure

What are space and structure…?

Do they matter for inference?

1. Space and Structure

“Real” space

- Botanical epidemics
- Animal diseases in the wild
- How to measure distance?

1. Space and Structure

Networks

- STD networks
- Transportation networks
- Long-range interactions

1. Space and Structure

Structured populations

- Household models
- Different levels of mixing
- Small-scale structure e.g. hospital wards

1. Space and Structure

Q: Are space/structure important for statistical inference?

1. Space and Structure

Q: Are space/structure important for statistical inference?

Answer 1:

Yes, because of confounding

- Mechanism of disease spread depends on both spatial components and infection rates

1. Space and Structure

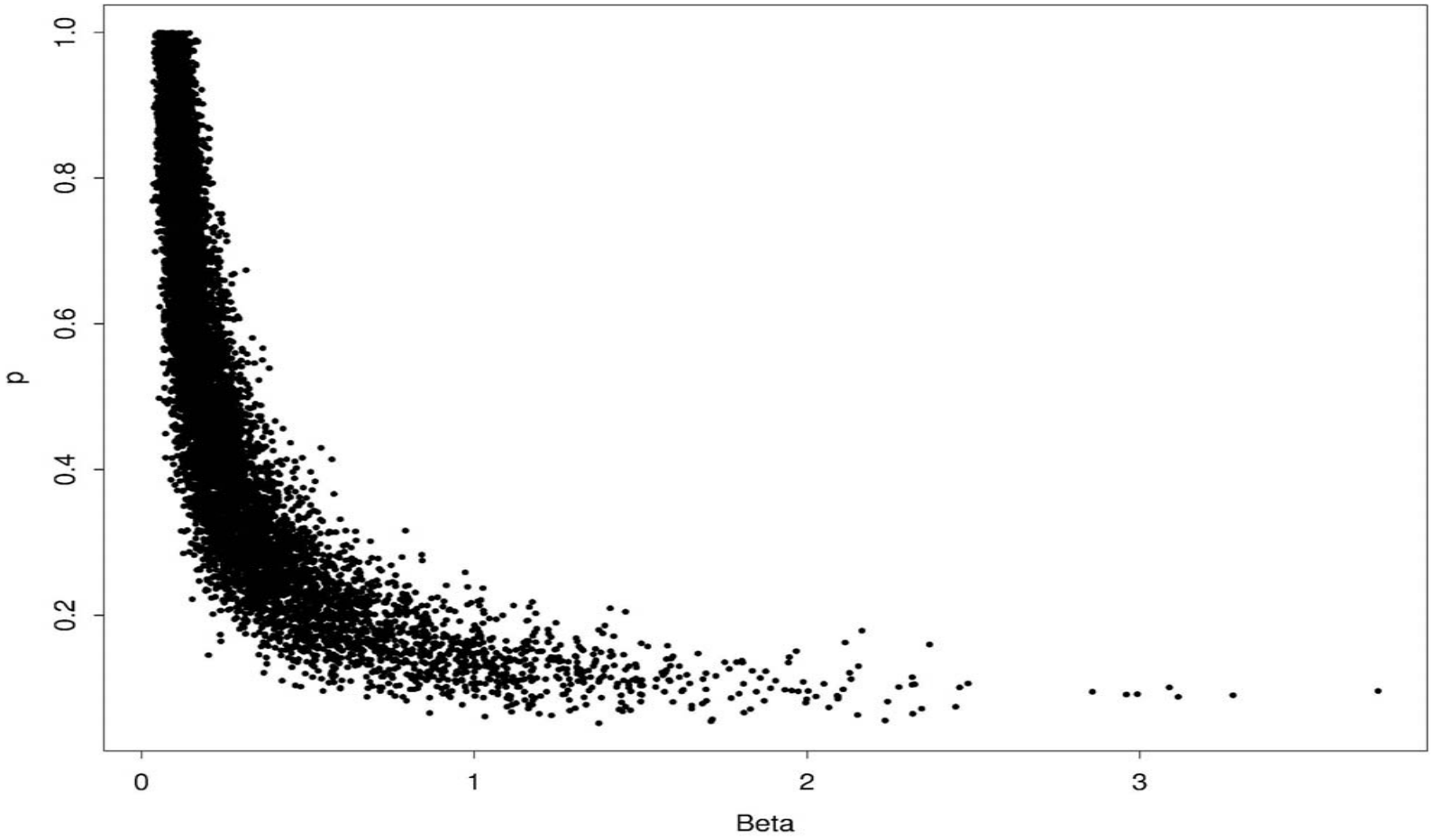
Example (Britton and O'Neill, 2002)

SIR model on a Bernoulli random graph

Parameters include infection rate β, edge probability p

Inference using MCMC methods

1. Space and Structure



1. Space and Structure

Example: Different levels of mixing

- Should schools/workplaces be shut to prevent influenza spread?

Hard to answer without good estimates of infection rates at different levels

1. Space and Structure

Q: Are space/structure important for statistical inference?

Answer 2:

Yes, because that is the question…

1. Space and Structure

Example: Isolation for HCAIs in ICUs

- Does isolating patients help to prevent spread of nosocomial infections?
- Can address by statistical inference with model-choice methods.

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

2. Data

Examples of data types:

- Data in both space and time
- Snapshot data at a given time = t
- Final outcome data at time = T

(t is a fixed time. T is a stopping time.)

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

3. Methods of inference

Mechanistic models (deterministic or stochastic) used to describe disease spread dynamics.

Such models have parameters such as

- Infection rates….
- Spatial-spread parameters….
- Within-host parameters….. etc

3. Methods of inference

Objective of statistical inference is (usually) to provide

- estimates of the model parameters
- some measure of uncertainty of the estimates

using the available data.

3. Methods of inference

For deterministic models, this means trying to match the (single) model outcome to the observed data as closely as possible.

For example, minimise least-squares errors.

3. Methods of inference

For stochastic models, estimation requires the formulation of a likelihood, i.e. given data x and model parameters θ,

$$L(\theta) = f(x \mid \theta)$$

$$= P(\text{data given parameters})$$

(In many cases, f is actually a density)

3. Methods of inference

Given a likelihood, estimation can proceed along various lines, most common approaches being

- Maximum-likelihood estimation
- Bayesian estimation

3. Methods of inference

However: likelihood might be very hard (or impossible) to evaluate.

Key reasons for this are

- Model intractability
- Missing data e.g. do not observe infections

3. Methods of inference

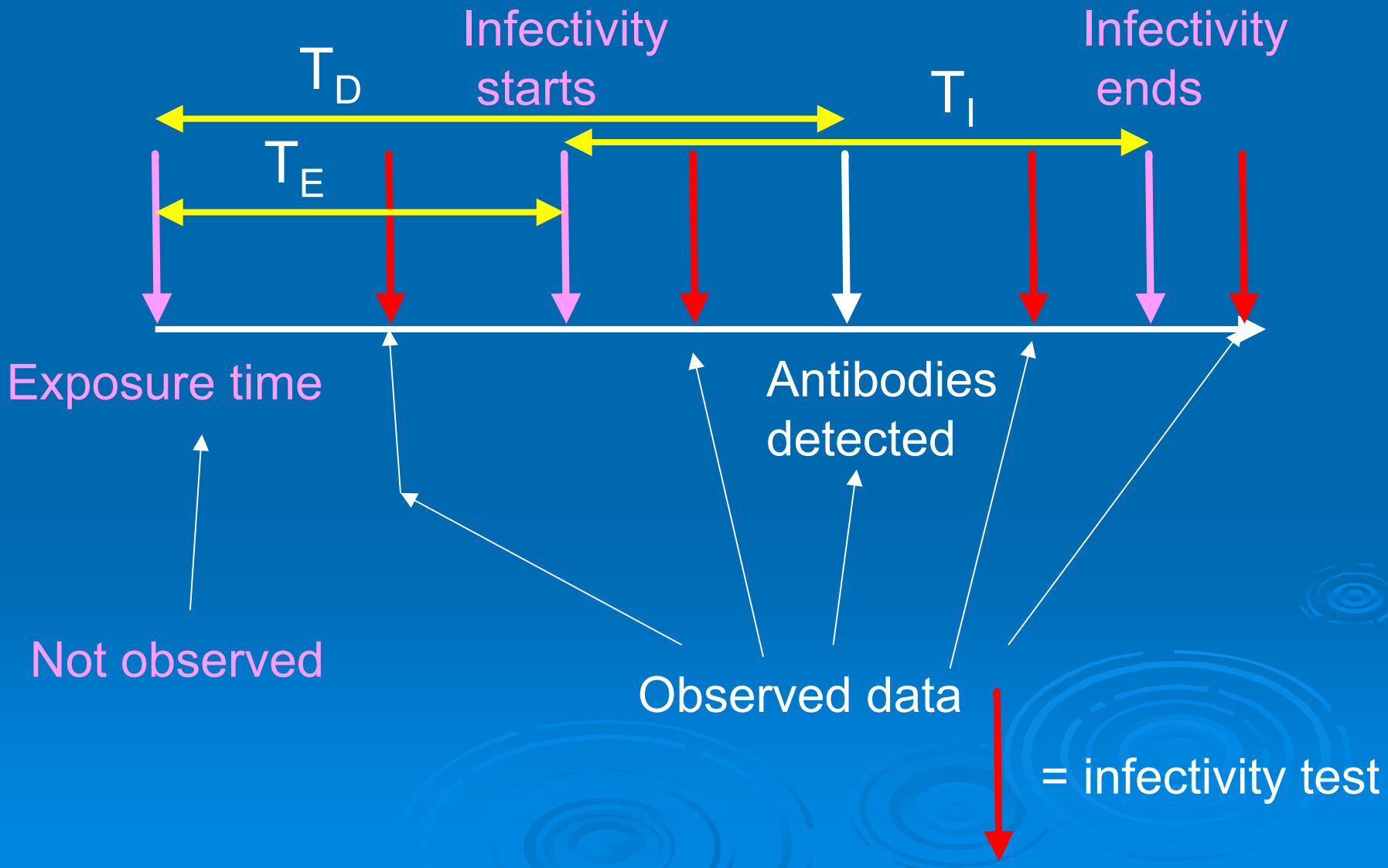
Example: Swine Fever in penned pigs (Höhle, Jørgensen, and O'Neill, 2005)

- Animal experiment
- Pigs kept in adjacent pens
- Regular tests to detect infected pigs
- Interest in infection rates and efficacy of control measures

3. Methods of inference

- SEIR (Susceptible-Exposed-Infective-Removed) spatial model with exposure/infection/removal times unobserved
- Data consist of antibody detection times and testing for infectivity
- Likelihood is intractable…
- …but augmented likelihood with event times is tractable

3. Methods of inference



3. Methods of inference

Model includes:

- Time from exposure to antibody detection (T_D)
- Time from exposure to infectivity (T_E)
- Period of infectivity (T_I)
- Spatially-dependent infection rate

3. Methods of inference

- If exposure, infection and removal times are imputed and fixed (“best guess”) then could e.g. proceed via maximum likelihood
- Obvious problem with this approach is that it requires additional assumptions

3. Methods of inference

- If exposure, infection and removal times are unknown then likelihood involves integrating over all such possible times – difficult because of inter-dependencies
- Thus exposure times are treated as additional model parameters (in MCMC framework in this case)

3. Methods of inference

MCMC (Markov Chain Monte Carlo)

- Target probability density of interest
  = f(parameters | data)
- MCMC works by constructing a Markov chain whose stationary distribution is f
- Run chain for a long time; samples from chain are (approx) samples from f

3. Methods of inference

MCMC (Markov Chain Monte Carlo)

- In practice, implementation requires finding ways of making the Markov chain move around easily
- E.g. here, how to update the unknown exposure/infection/removal times?

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

5. Botanical Epidemics

- Plant experiments can provide very rich data sets…
- …which naturally leads to issues of experimental design, e.g.
  - How many data are really necessary?
  - How often should data be collected?

5. Botanical Epidemics

- Experimental setting also reduces extent to which data are missing, e.g. may know when plants are infected
- This in turn makes likelihood evaluation simpler

5. Botanical Epidemics

- Models often use spatial (or dispersal) kernel, i.e. a way of modelling how likely infection is to occur at a given distance
- Choice of kernel? – model choice and goodness-of-fit issues

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

6. Two-level mixing

Model (Ball, Mollison, Scalia-Tomba 1997)

- Population size N, divided into households
- SIR (or SEIR) model
- Infectious period distribution assumed known
- Local infection rate λ_L, global infection rate λ_G

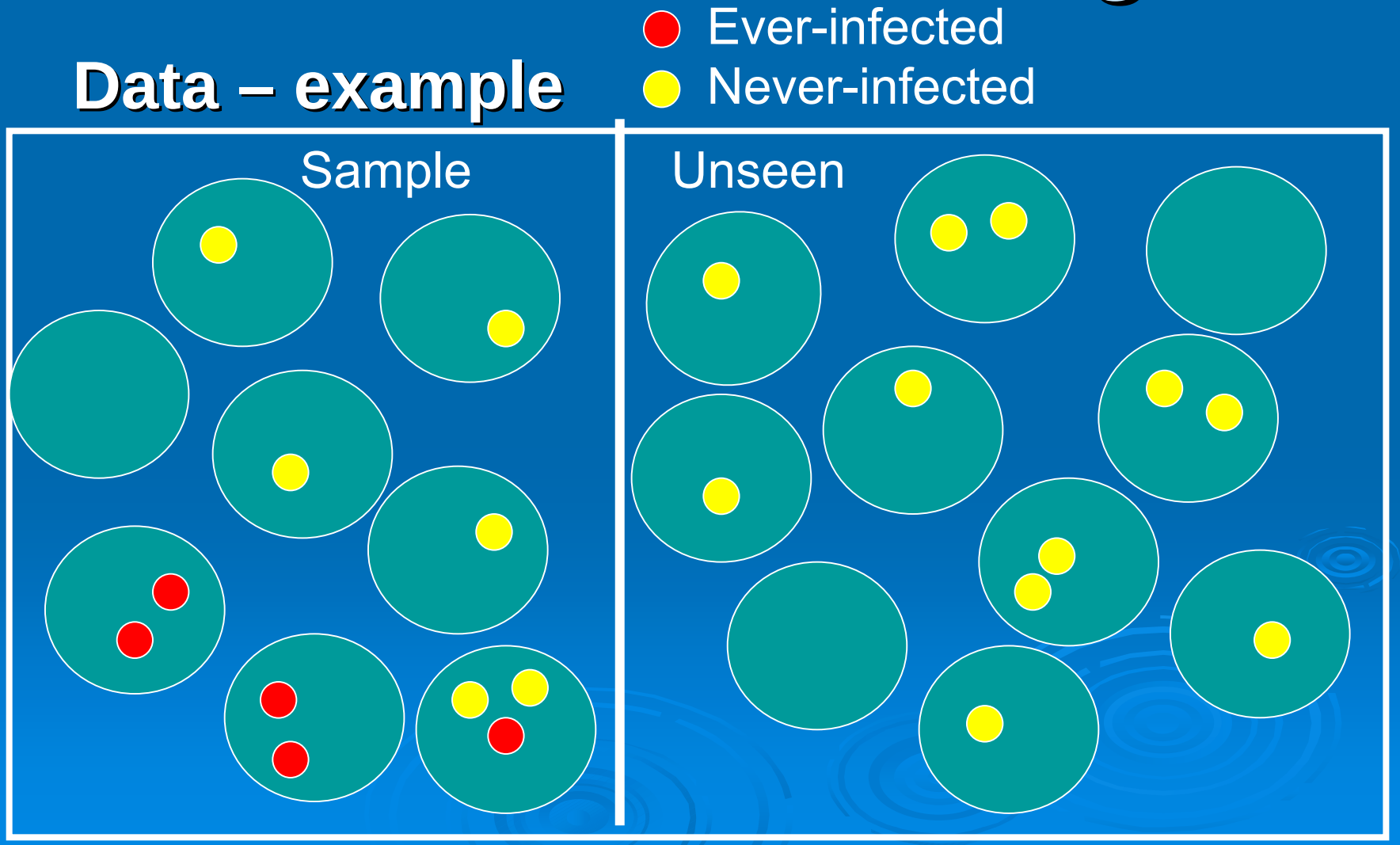
6. Two-level mixing

Data

- Sample from the population containing:
  - Precise household structure
  - Numbers initially susceptible
  - Numbers ever-infected during epidemic

6. Two-level mixing

Data – example



6. Two-level mixing

Likelihood: f(λ_L , λ_G | data)

- Intractable as it stands

Possible solutions:

- Use a simpler model with independent households (B, M, S-T 1997; etc)
- Do some kind of data imputation

6. Two-level mixing

Data imputation methods

- Impute final severity via approximation (Demiris and O'Neill, 2005a)
- Random graph methods

6. Two-level mixing

Random Graph method 1
(Demiris and O'Neill, 2005b)

For each pair of individuals (i,j):

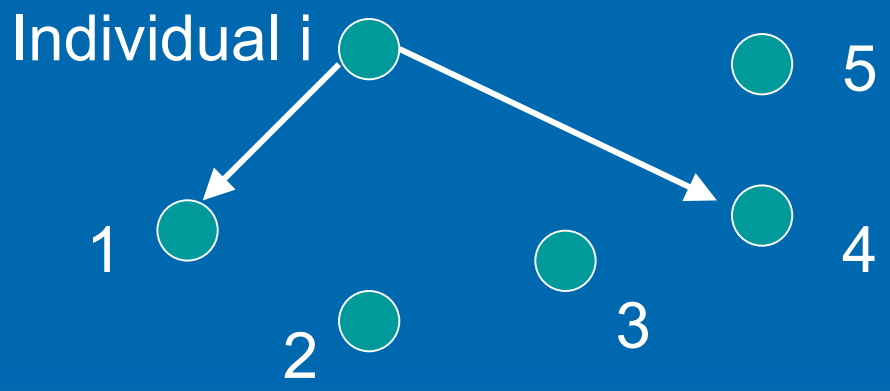
X(i,j) = Indicator{"i tries to infect j"}

P(X(i,j) = 1) easily evaluated

Knowledge of X's gives likelihood

6. Two-level mixing

Random Graph method 1



$X(i,1) = X(i,4) = 1 \qquad X(i,2) = X(i,3) = X(i,5) = 0$

$\text{Likelihood} = (1-\exp(-\lambda I))^2 \, (\exp(-\lambda I))^3$

6. Two-level mixing

Random Graph method 2

For each individual i:

X(i) = number of contacts i has

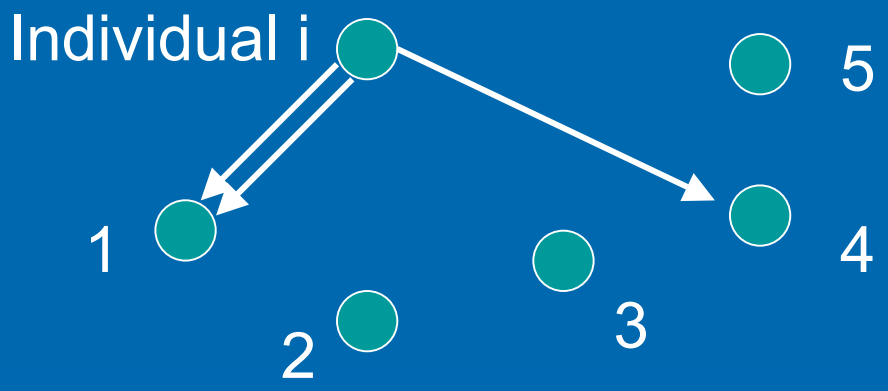
X(i) is Poisson (λ)

C(i) = list of who is contacted

Knowledge of X's and C's gives likelihood

6. Two-level mixing

Random Graph method 2



X(i) = 3 C(i) = {1, 4, 1}

Likelihood = $\exp(-\lambda)\ \lambda^3\ (1/3!)\ (1/5)^3$

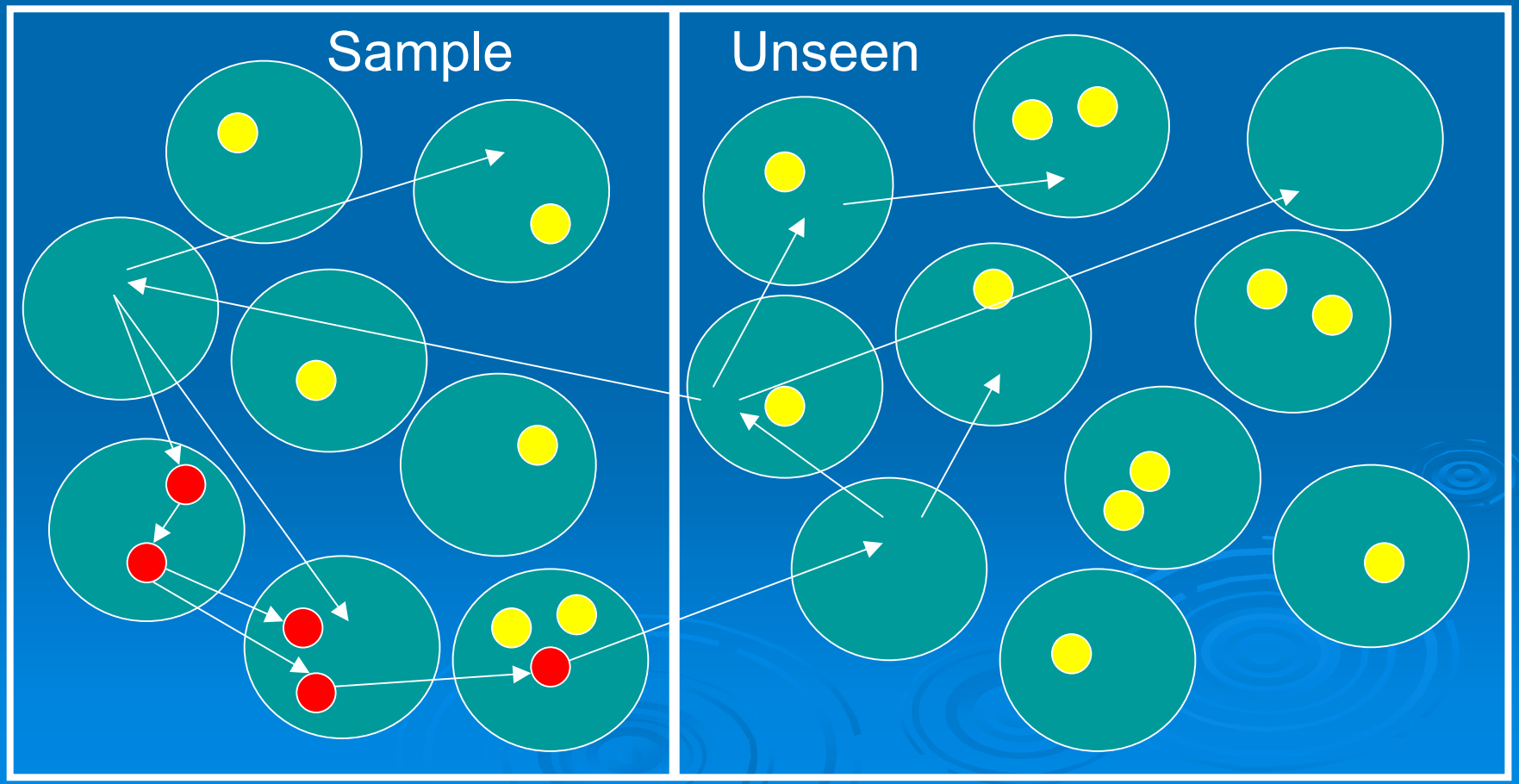
6. Two-level mixing

Random Graph methods

- Method 2 allows for Gibbs updates of parameters and is faster
- Method 1 easier to code
- Both methods struggle if unobserved population is large: correlations between infection rates and contacts

6. Two-level mixing

- Ever-infected
- Never-infected



6. Two-level mixing

Random Graph method 2 again

- X(i) = no. contacts ~ Poisson(λ)
- So X, λ strongly correlated
- Makes it hard to update either individually within the MCMC code

6. Two-level mixing

Random Graph method 2 again

- Non-centered parameterisation:
- Set d(i) ~ U(0,1)
- $X(i) = F^{-1}(d(i))$ F = cdf Poisson(λ)
- Here, d(i) and λ are independent
- MCMC mixing much improved

6. Two-level mixing

Random Graph methods

- Easily generalised to Multi-type setting
- Easily generalised to other structures (e.g. 3 levels of mixing; specific spatial structures)

Outline

1. Space and Structure
2. Data
3. Methods of inference
4. Example 1: UK Foot & Mouth
5. Example 2: Botanical epidemics
6. Example 3: Two-level mixing
7. Concluding remarks

7. Concluding remarks

- Wide variety of inference methods have been implemented
- MCMC methods appear powerful but are non-trivial to implement
- Goodness-of-fit and model choice methods need further work