MCMC 2: Lecture 4
Household models

Phil O'Neill Theo Kypraios
School of Mathematical Sciences
University of Nottingham

Contents

- 1. Household models
- 2. Longitudinal data
- 3. Final outcome data

Contents

- 1. Household models
- 2. Longitudinal data
- 3. Final outcome data

1. Household models

Many infectious disease studies incorporate households into the study design.

There are numerous reasons to do this, including

- convenience of collecting data
- stability of target population
- get data on within-household spread

1. Household models

Models

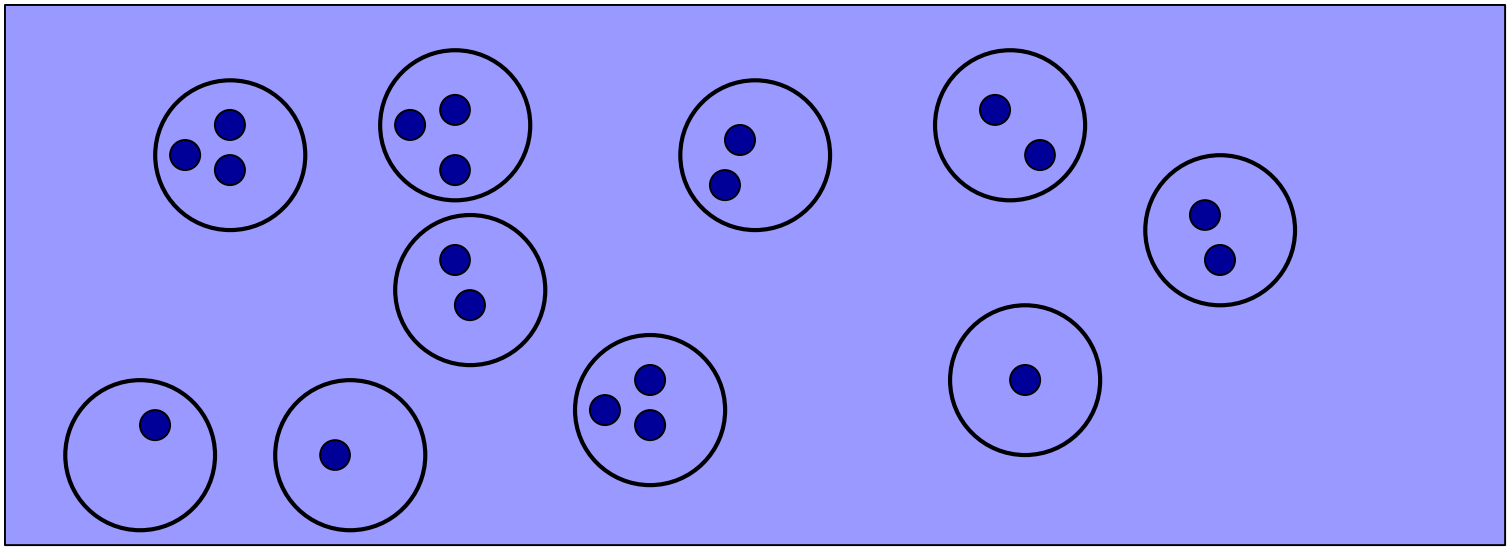
There are various models for disease transmission which incorporate households. Our focus here is on models in which households are independent, i.e. the fates of different households are independent of each other.

Models which relax this assumption will be mentioned in the next session.

1. Household models

Models

Henceforth we assume that a population of N individuals is partitioned into households, which need not all be the same size.



1. Household models

Models

We need two key ingredients:

- How the disease enters a household;
- What happens within a household.

1. Household models

Models

- How the disease enters a household

A typical assumption is that each individual in the population has, independently, a constant “risk” per unit time of becoming infected from the community.

1. Household models

Models

- How the disease enters a household

Formally: individual k becomes infected according to a Poisson process of rate b_k, say. Thus,

P(k avoids infection for T time units)

$$= \exp(- T\, b_k)$$

1. Household models

Models

- How the disease enters a household

The rate b_k is typically either the same for every individual ($b_k = b$ for all k), or else it may depend on the type of individual k.

Here, “type” might mean adult / child / vaccinated / unvaccinated / etc...

1. Household models

Models

- What happens within a household

A standard assumption is that the disease spreads according to an SIR or SEIR model within a household.

Again types might feature: both the infectivity and susceptibility of an individual might be type-dependent.

1. Household models

Models

- What happens within a household

A common assumption for households is that the infection rate β is not scaled by the household size.

This means that we assume that each infective has contacts with each susceptible at rate β.

Contents

- 1. Household models
- 2. Longitudinal data
- 3. Final outcome data

2. Longitudinal data

Here we assume that we have data through time, e.g. daily observations.

We can approach the inference problem in a similar manner to that described previously for the SIR model.

2. Longitudinal data

Specifically, the likelihood can now be written as a product over all households, due to the assumption of independence.

However, parameter updates now typically require Metropolis-Hastings update steps within an MCMC algorithm.

2. Longitudinal data

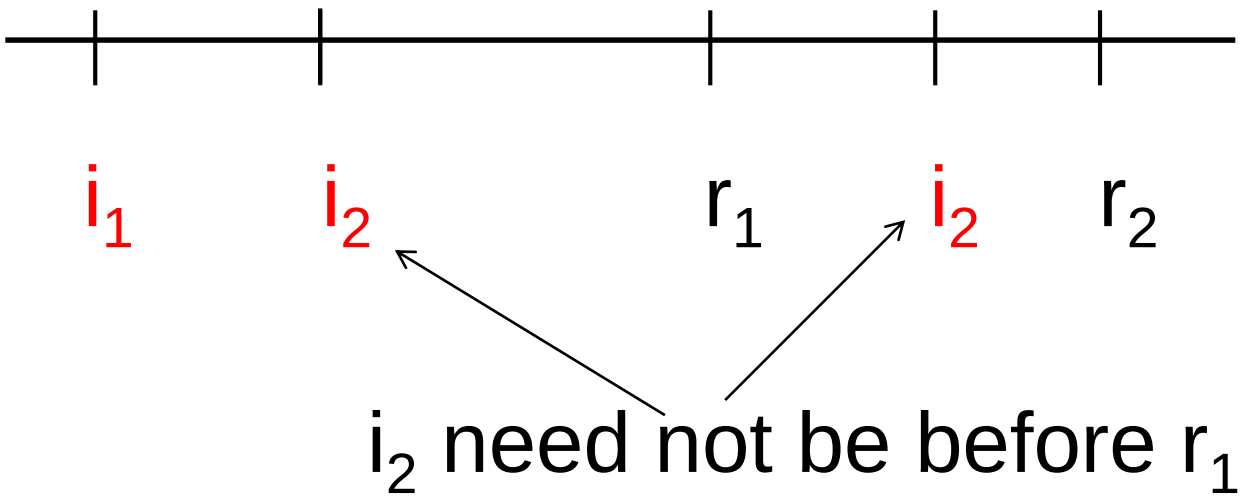
Example Consider a single household of n=4 individuals in which we observe two removals (= symptom-appearance times, say).

We assume

- a constant rate of outside infection λ
- a Markov SIR model for within-household transmission.

2. Longitudinal data

Example Let i_1 and i_2 denote the two infection times and $r_1 < r_2$ denote the two removal times. Data = $\{r_1 , r_2 \}$.



2. Longitudinal data

Example

$$\pi\,(i_2, r_1, r_2 \mid i_1, \beta, \gamma, \lambda)$$

$$= (\beta\, I(i_2-) + \lambda)\, \exp\left(- \int_{i_1}^{r_2} (\beta\, S(t) I(t) + \lambda)\, dt \right)$$

$$\times\, \gamma^{2} \exp(-\gamma(r_1 + r_2 - i_1 - i_2))$$

Note now that neither β nor λ has a Gamma-distributed full conditional distribution, although *γ* still does.

2. Longitudinal data

As mentioned above,

likelihood = Π_k likelihood in household k

Note that in constructing the posterior density, you only need to include the prior density for the model parameters once (i.e. not once per household).

Contents

- 1. Household models
- 2. Longitudinal data
- 3. Final outcome data

3. Final outcome data

Suppose now that the data consist only of the final number of cases in each household.

The first observation is that without temporal data we can no longer estimate all parameters with respect to “real time”.

We therefore assume that the infectious period distribution is known and $E(T_I) = 1$.

3. Final outcome data

Our objective is to estimate the remaining model parameters.

Since households are assumed independent, to evaluate the likelihood we need to find the likelihood for a given household.

3. Final outcome data

Specifically, consider a household with n individuals. Let T denote the number who ever become infected, $0 \leq T \leq n$.

Assume a constant rate λ of infection from outside, and an SIR model for within-household contact with infection rate β.

We wish to calculate $P(T = k)$, $k = 0,...,n$.

3. Final outcome data

Removing “time”

We first consider how the final outcome of the epidemic can be constructed without explicitly considering event times.

To begin with, focus on a single household, and ignore infections from outside the household.

3. Final outcome data

Removing “time”

Suppose that the household contains n individuals.

Consider a single individual A. If A ever became infected they remain so for a random time T_I ; suppose $T_I = \tau$, say.

3. Final outcome data

Removing "time"

If infected, then for a period of time τ, individual A has contacts with each other individual in the household according to n-1 independent Poisson processes, each of rate β.

It follows that we can construct a list of individuals that A would infect if A was infected, and the others susceptible.

3. Final outcome data

Removing “time”

Now suppose that each individual in the population had such a “list”. It follows that we can easily deduce who actually gets infected, once the initial infectives are chosen.

3. Final outcome data

Removing “time”: Example

Suppose n = 5 and the lists are

1 = { 4, 5 }

2 = { 4, 1, 5 }

3 = { 1 }

4 = { 2 }

5 = { }

and suppose 1 is initially infective.

3. Final outcome data

Removing “time”: Example

1 = {4, 5}, 2 = {4, 1, 5}, 3 = {1}, 4={2}, 5={ }

1 infects 4 and 5

4 infects 2

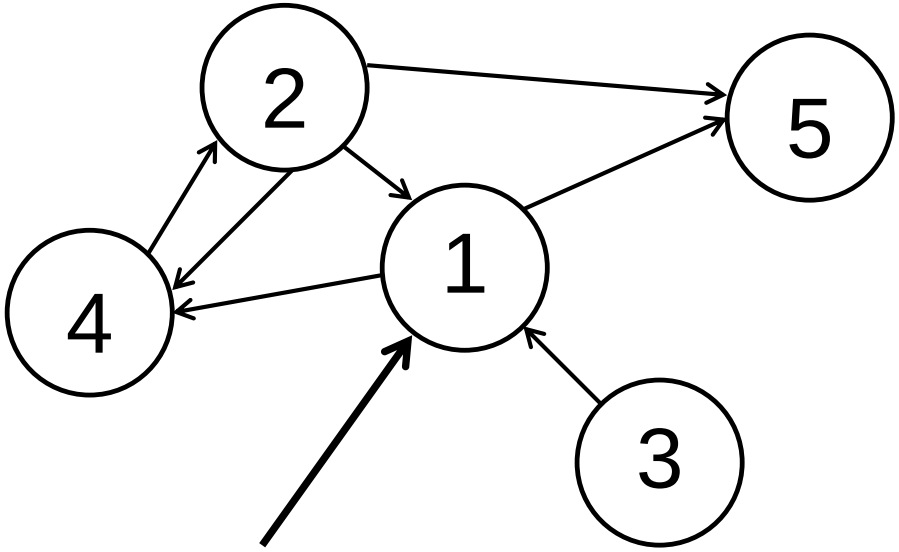
2 infects 4,1,5 (ignore all since already infected)

5 infects no-one

3. Final outcome data

Removing “time”: Example

1 = {4, 5}, 2 = {4, 1, 5}, 3 = {1}, 4={2}, 5={ }



3. Final outcome data

Removing “time”

It follows that the distribution of the final number infected in the epidemic is the same as the distribution of the number of individuals found on the final list in this construction.

3. Final outcome data

Removing “time”

Similar reasoning can be applied to infection from outside the household: all that is required is knowledge of which individuals ever became infected from outside, rather than when they became infected.

3. Final outcome data

Removing “time”

We can therefore proceed by assuming any individuals infected from outside the household are infected first, and so we can treat them as the “initial infectives” in the household.

3. Final outcome data

Removing “time”

Recall that the probability that an individual avoids infection from outside the household for a period of time t is

$$\exp(-\lambda t)$$

Since we cannot estimate t from the data, instead we simply define

$$p = \exp(-\lambda t)$$

3. Final outcome data

Removing “time”

Now, since each individual in the household avoids infection from outside independently, it follows that the number infected from outside, Y say, has a Binomial distribution with parameters

$$Y \sim \text{Bin}(n, 1-p).$$

3. Final outcome data

Back to P(T=k)

It follows that

$$P(T = k) = \sum_{0 \le y \le n} P(T=k \mid Y=y)\, P(Y=y)$$

where $P(Y=y) = (n! / y!(n-y)!)\, (1-p)^{y}\, (p)^{n-y}$.

probability mass function of Binomial(n,1-p)

3. Final outcome data

Back to P(T=k)

Further, P(T=k | Y=y) is simply the probability that k-y susceptibles become infected in an SIR model with y initial infectives and n-y initial susceptibles.

Note that this probability is zero for $k < y$.

3. Final outcome data

Back to P(T=k)

To evaluate P(T=k | Y=y) we can use the "triangular equations" for an SIR model, defined as follows.

Consider an SIR model with m initial susceptibles, a initial infectives, and infection rate α (between two individuals).

3. Final outcome data

Back to P(T=k)

Define p(k) as the probability that k of the initial susceptibles ever become infected,

k = 0, 1, ..., m.

Let $f(x) = E[\exp(- x T_I)]$ be the moment generating function of the infectious period distribution.

3. Final outcome data

Back to P(T=k)

Then for $0 \le j \le m$,

$$\sum_{k=0}^{j} \binom{m-k}{j-k} p(k) / [\, f(\alpha(m-j))]^{k+a} = \binom{m}{j}$$

From this formula we can recursively evaluate p(0), p(1), ..., p(m) by setting j=0, j=1, ..., j=m.

3. Final outcome data

Back to P(T=k)

Then for $0 \leq j \leq m$,

$$\sum_{k=0}^{j} \binom{m-k}{j-k} p(k) / [\, f(\alpha(m-j))]^{k+a} = \binom{m}{j}$$

These equations are often called “the triangular equations for the final size distribution”.

3. Final outcome data

Example: m=3, a=1, fixed infectious periods

Set $T_I = 1$. Then $f(x) = E[\exp(- x T_I)]$

$= \exp(- x)$.

Setting j=0 in formula gives

$p(0) = \exp(- 3 \alpha)$.

Setting j=1 gives

$3p(0) \exp(2 \alpha) + p(1) \exp(4 \alpha) = 3,$

so $p(1) = 3 \exp(-4 \alpha) (1 - \exp(- \alpha))$.

3. Final outcome data

Summary: P(T=k)

$$P(T = k) = \sum_{0 \le y \le n} P(T=k \mid Y=y)\, P(Y=y)$$

where $P(Y=y) = (n! / y!(n-y)!)\, (1-p)^{y}\, (p)^{n-y}$,

and $P(T=k \mid Y=y)$ can be evaluated using the triangular equations. Note that this requires a recursive function (in R, say).

3. Final outcome data

Data and likelihood

Suppose that the data consist of the set of numbers **n** = { n(j,k) }, where

n(j,k) = number of households in which j out of k initial suseptibles become infected.

3. Final outcome data

Data and likelihood

The likelihood takes the form

$$\pi\,(\mathbf{n} \mid p, \beta\,) = \Pi_{j,k}\;\; q(j,k)^{n(j,k)}$$

where q(j,k) = P(T=j) for a household containing k susceptible individuals.

3. Final outcome data

Bayesian inference and MCMC

The target density is

$$\pi(p, \beta \mid \mathbf{n}) \propto \pi(\mathbf{n} \mid p, \beta)\, \pi(p, \beta)$$

p is a probability and could be updated by e.g. Gaussian random walk ($p < 0$ and $p > 1$ must be rejected), or an independence sampler (e.g. U(0,1) proposal).

3. Final outcome data

Bayesian inference and MCMC

The parameter β could be updated by Gaussian random walk, for example.

3. Final outcome data

Fixed infectious periods

Special case: if the infectious period T_I is constant, $T_I = 1$, then the triangular equations yield expressions in terms of

$$f(x) = E\,[\exp(-x\,T_I)] = \exp(-x).$$

3. Final outcome data

Fixed infectious periods

$$\sum_{k=0}^{j} \binom{m-k}{j-k} p(k) / [\, f(\alpha(m-j))]^{k+a} = \binom{m}{j}$$

Specifically, we have

$[f(\alpha(m-j))]^{k+a} = \exp[\, -\alpha(m-j)(k+a)\,]$

$= q^{(m-j)(k+a)}$,

where $q = \exp(-\alpha)$.

3. Final outcome data

Fixed infectious periods

It therefore makes sense to reparameterise the model by defining

$$q = \exp (- \beta).$$

Note that $0 \leq q \leq 1$.

3. Final outcome data

Fixed infectious periods

With this parameterisation (p,q) we have that, for a susceptible A say,

p = P(A avoids outside infection)

q = P(A avoids infection from one infected household member)

3. Final outcome data

Fixed infectious periods

The within-household model has the same final outcome distribution as the so-called "Reed-Frost" epidemic model.

The whole model has the same final outcome distribution as the "Longini-Koopman" model.

References

- O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M. and Mollison, D. (2000) Analyses of infectious disease data from household outbreaks by Markov Chain Monte Carlo methods. *Applied Statistics* 49, 517-542.
- Andersson, H. and Britton, T. (2000) *Stochastic Epidemic Models and their Statistical Analysis*. Lecture notes in Statistics 151, Springer.