Module 11: Lesson 4 Lecture Notes

Callum Arnold

20 July, 2021

Contents

Household models

- Many ID studies incorporate households into study design
    - Convenience of collecting data
    - Stability of target pop
    - Get data on within-household spread
- Here, focus on models in which the households are independent
    - May be reasonable if select households that are geographically separated
- Assume population of N individuals is partitioned into households
    - No requirement that they are the same size
- Household models require 2 key things
    - How the disease enters household
        * E.g. a community transmission risk
    - What happens within a household

How the disease enters a household

- Typically assume each individual in population has a constant risk per unit time of becoming infected by the community
    - Individually k becomes infected according to a Poisson process of rate b_k
        * Process that describes events happening randomly through time

* $P(k$ avoids infection for T time units$) = \exp(-Tb_k)$
  - The rate b_k is typically the same for all individuals
  - Can make rate b_k depend on the **type** of individual k
    * e.g. adult/child, vaccinated/unvaccinated etc
    * More types require more data

What happens within a household

- Spread via SEIR model within household
- Types may feature
  - Infectivity and susceptibility of an individual may be type-dependent
- Common assumption is that the infection rate β is not scaled by the size of the household
  - Each infective has contacts with each susceptible at rate β
  - Normally, we assume infection rate for each infective individual $\propto \beta/N$
    * At start of outbreak when $S \approx N$ therefore the overall rate $= \frac{\beta}{N}N = \beta$
    * Therefore, if look at one individual the overall rate of them having infectious contacts is β, and independent of the population size!
      · Frequency-dependent contact rate, e.g. don't expect to have many more contacts if you move from a small town to large city
  - Within households, need to move to density-dependent contact rate
    * Don't scale by N, so assume infection rate for each infective individual $\propto \beta N$

Longitudinal data

- We assume we have data through time e.g. daily observations
- Approach the inference problem in a similar manner to SIR data
- Have modelling process where we write down a likelihood and multiply by the priors of the parameters to get a density up to proportionality
- Assuming households are independent
  - If you can work out the likelihood of what happens in one household, the likelihood can be written as a product over all households
  - Still likely have to do data augmentation steps as before
- Now the parameter updates typically require M-H update steps within an MCM algorithm

Worked example

- Single household of $n = 4$ individuals in which we observe 2 removals
- Constant rate of outside infection λ
- Markov SIR model for within-household transmission
- Less constraint on when the infection times can happen

– i.e. Can be infected from outside the household, so infections no longer have to occur before everyone in the household has recovered

i$_2$ need not be before r$_1$
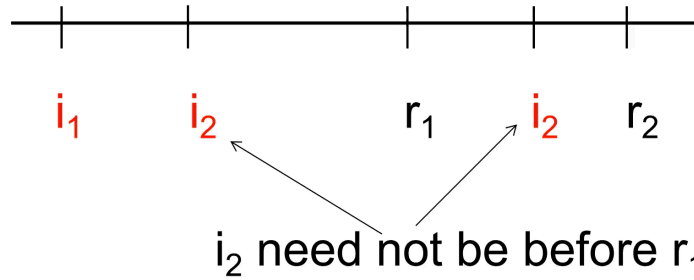
- Augmented likelihood for the household
  - Similar to before, but have updated the infection rate to include λ
  - Now neither β nor λ have Gamma distributed full conditional distributions, but γ still does

$$\pi(i_2, r_1, r_2 | i_1, \beta, \gamma, \lambda) = (\beta I(i_2-) + \lambda) \exp\left(-\int_{i_1}^{r_2} (\beta S(t) I(t) + \lambda) dt\right) \times \gamma^2 \exp\left(-\gamma(r_1 + r_2 - i_1 - i_2)\right)$$

- Augmented likelihood
  - likelihood is the product of the likelihoods in household k
- In constructing the posterior density, only need to include the prior density for the model parameters once (**not once per household**)

Final outcome data

- Suppose now the data consists only of **the final number of cases in each household**
- Without any temporal data, we can't estimate all of the parameters with respect to **real time**
- Therefore assume the infectious period is known and $E(T_I) = 1$
- As before, assume households are independent so calculate likelihood for 1 household and take product
- Consider a household with n individuals
- Now let T be the number who ever became infected $(0 \leq T \leq n)$
- Assume constant rate of outside infection λ and an SIR model for within-household contact with infection rate β
  - Not talking about infectious period any more as assume mean $= 1$, therefore no gamma term
- Wish to calculate $P(T = k), \quad k = 0, ..., n$

Removing time

- If we think about the final outcome in the correct way, it's possible to ignore the timing of events e.g. whether infections came first from outside or inside

Ignoring external infections

- To start, let's look at a single household and ignore external infections
- Let's look at individual A.
  - If A ever infected, they remain so for random time T_I, e.g. $T_I = \tau$
  - For a period of time τ, they have contacts with other household individuals according to $n - 1$ independent Poisson procesess, each with rate β
  - Can write a list of individuals A would infected if the others were susceptible
- Each individual in the population has a list
  - Who would you infect if you were infective and everyone else was susceptible?
- Can deduce who ends up getting infected!
- Let's imagine we have all the lists below:
  - $1 = \{4, 5\}, \quad 2 = \{4, 1, 5\}, \quad 3 = \{1\}, \quad 4 = \{2\}, \quad 5 = \{\}$
  - 1 is the initial infection
  - 1 infects 4 and 5
  - 4 infects 2
  - 2 only infects 5 as 4 and 1 aren't susceptible
  - 5 doesn't infect anyone
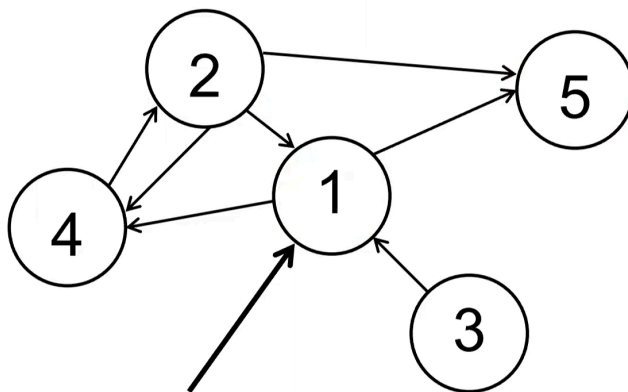
- The distribution of the final number infected in the epidemic is the same as the distribution of the number of individuals found on the final list (who ultimately got infected)
- All we care about is if the person gets infected, not who was the cause of each infection

Adding external infections

- Need to find out which individuals become infected from outside the household
  - Again, don't need to know when they become infected
- Treat individuals infected from outside as if they are initial infections within the household
  - Now have SIR model within the household with our calculated number of initial infections

- Probability of avoiding infection from outside the household during time t becomes the probability they avoid infection during the entire epidemic
    - $p = \exp(-\lambda t)$
    - If looking at seasonal data, effectively assuming epidemic is finished at end of season
- Assuming each individual in household avoids external infection independent of each other, it follows a Binomial distribution
    - $Y \sim \text{Binom}(n, 1-p)$
- We are interested in $P(T = k)$
    - $P(T = k) = \sum_{0 \leq y \leq n} P(T = k | Y = y) P(Y = y)$
        * Where $P(Y = y) = \frac{n!}{(n-y)!y!}(1-p)^y(p)^{n-y}$
            · Probability that y number of individuals in the household did get infected
        * Indicates that $k - y$ individuals have to get infected inside the household
        * Sum over all possible values
            · However, some probabilities can be 0, e.g. when $y > k$ as k is the total number of individuals infected
    - $P(T = k | Y = y)$ is the probability that $k - y$ susceptibles become infected in and SIR model with y initial infectives and $n - y$ susceptibles
        * We can evaluate with "triangular equations" for an SIR model

Triangular equations for an SIR model
- Consider SIR model with m initial susceptibles, a initial infectives, and α is the infection rate between two individuals
- Define $p(k)$ as the probability k of initial susceptibles ever become infected ($k = 0, 1, ..., m$)
- Let $f(x) = E[\exp(-xT_I)]$ be the moment generating function of the infectious period distribution
- For $0 \leq j \leq m$:
    - $\sum_{k=0}^{j} \binom{m-k}{j-k} \frac{p(k)}{[f(\alpha(m-j))]^{k+a}} = \binom{m}{j}$
    - Evaluate recursively by setting $j = 0, \quad j = 1, \quad ..., \quad j = m$ to get $p(0), p(1), ..., p(m)$
- For example
    - Set $T_I = 1$, then $f(x) = E[\exp(-x)]$
    - Setting $j = 0$ gives:
        * $p(0) = \exp(-3\alpha)$
    - Setting $j = 1$ gives:
        * $3p(0)\exp(2\alpha) + p(1)\exp(4\alpha) = 3$
        * $p(1) = 3\exp(-4\alpha)(1 - \exp(-\alpha))$
- Note that this requires a **recursive function**

Data and likelihood

- Suppose we have data consisting of the set of numbers $\vec{n} = \{n(j,k)\}$
    - $n(j,k)$ is the number of household in which j out of k initial susceptibles become infected
- The likelihood takes the form
    - $\pi(\vec{n}|p,\beta) = \prod_{j,k} q(j,k)^{n(j,k)}$
        * Remember that p is the probability of escaping infection
        * β is used in the triangular equations instead of α
        * $q(j,k) = P(T = j)$ for a household containing k susceptible individuals
            · It isn't explicitly written down but we can numerically calculate it
- The target density is:
    - $\pi(p,\beta,|\vec{n}) \propto \pi(\vec{n}|p,\beta)\pi(p,\beta)$
    - p is a probability and could be updated by Gaussian random walk ($p < 0$ and $p > 1$ must be rejected), or an independence sampler (e.g. U$[0,1]$), for example
        * $\beta^* \sim \mathcal{N}(\beta, \sigma^2)$

Fixed infectious periods

- Because we're thinking about final outcome data, there is no time information
- Let's look at the special case where $T_I = 1$
- The triangular equations yield expressions in terms of $f(x) = E[\exp(-xT_I)] = \exp(-x)$
    - The only difference between different choices of T_I is how different are the moment generating functions i.e., if not that different, don't see much difference in the inference for the parameters in the model e.g. p, β
    - $\sum_{k=0}^{j} \binom{m-k}{j-k} \frac{p(k)}{[f(\beta(m-j))]^{k+a}} = \binom{m}{j}$
        * Note we've reparamaterized this using β (our infection rate) instead of α
    - $[f(\beta(m-j))]^{k+a} = \exp(-\beta(m-j)(k+a)) = q^{(m-j)(k+a)}$
        * Where $q = \exp(-\beta)$
    - Now, we can put our parameters p, q in context (for individual A)
        * $p = P(A$ avoids external infection$)$
        * $q = P(A$ avoids infection from one infected household member$)$
            · This arises from the Poisson process assumption:
            · How likely is it that there are no points in a Poisson process that you observe for length one time period (because $T_I = 1$) where the rate of events is β
- The within-household model has the same final outcome distribution as the Reed-Frost model
- The whole model has the same final outcome distribution as the Longini-Koopman model

Discussion

Why use this household model vs Reed-Frost

- Reed-frost is discrete case, but has strong assumptions about long latent period and fixed length infection period

- Household model is more general and makes fewer assumptions

  - Can have different infectious period distributions

  - Continuous time so can incorporate other data with temporal data