Indian Buffet Epidemics

Ashley Ford, Gareth Roberts

EEE 2010

Outline

SIR Models

$$S \xrightarrow{\beta} I \xrightarrow{\gamma} R$$

$$N_t^S \qquad N_t^I \qquad N_t^R$$

$$N = N_t^S + N_t^I + N_t^R$$

$$P(N_{t+dt}^S = j-1 | N_t^S = j) = \beta N_t^S N_t^I dt$$

$$P(N_{t+dt}^I = j+1 | N_t^I = j) = \beta N_t^S N_t^I dt$$

$$P(N_{t+dt}^I = j-1 | N_t^I = j) = \gamma N_t^I dt$$

Inference

- What data is available ?
    - Epidemic complete ?
    - Infection times ?
- MLE well known with full data
    - see Andersson and Britton (2000)
- Martingale estimator Becker and Hasofer (1997)
- MCMC estimates O'Neill and Roberts (1999)

Inference

- What data is available ?
  - Epidemic complete ?
  - Infection times ?
- MLE well known with full data
  - see Andersson and Britton (2000)
- Martingale estimator Becker and Hasofer (1997)
- MCMC estimates O'Neill and Roberts (1999)

Inference

- What data is available ?
  - Epidemic complete ?
  - Infection times ?
- MLE well known with full data
  - see Andersson and Britton (2000)
- Martingale estimator Becker and Hasofer (1997)
- MCMC estimates O'Neill and Roberts (1999)

Inference

- What data is available ?
    - Epidemic complete ?
    - Infection times ?
- MLE well known with full data
    - see Andersson and Britton (2000)
- Martingale estimator Becker and Hasofer (1997)
- MCMC estimates O'Neill and Roberts (1999)

Indian Buffet Epidemics

- Need a model between homogeneous mixing and over complex models.
- Aim to fit the heterogeneity with two or three parameters that measure the departure from homogeneity.

Places and People

- Model heterogeneity in an epidemic amongst N people
- Each person belongs to 1 or more of many classes
  - e.g. households, schools, clubs, buses etcetera
- The classes are not specified
- A prior is put on class membership
  - represented as an $N \times K$ binary matrix Z
- An Indian Buffet Process

Places and People

- Model heterogeneity in an epidemic amongst N people
- Each person belongs to 1 or more of many classes
    - e.g. households, schools, clubs, buses etcetera
- The classes are not specified
- A prior is put on class membership
    - represented as an $N \times K$ binary matrix Z
- An Indian Buffet Process

Places and People

- Model heterogeneity in an epidemic amongst N people
- Each person belongs to 1 or more of many classes
  - e.g. households, schools, clubs, buses etcetera
- The classes are not specified
- A prior is put on class membership
  - represented as an $N \times K$ binary matrix Z
- An Indian Buffet Process

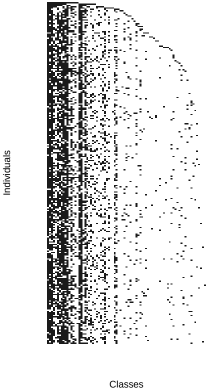
Indian Buffet Process

- Introduced by Griffiths and Ghahramani (2005)
- For each k $\psi_k$is the probability that an individual is in class k
- $\psi_k \sim Beta(\alpha/K, 1)$with α being the strength parameter of the IBP.
- The model for Z is: $z_{ik}|\psi_k \sim Bernoulli(\psi_k)$ independently
- The process is obtained as $K \to \infty$

A culinary metaphor

- N customers enter a restaurant one after another.
- The jth customer selects each dish with probability m_k/j
  - where m_k is the number of previous customers who have chosen a dish.
- He then tries $\text{Poisson}(\alpha/j)$ new dishes.

Indian Buffet Process example



IBP Z generated with $N = 260$, $K = 260$, $\alpha = 15$

Indian Buffet Epidemic

- The state of individual j is at time t is $x_{j,t} \in \{S,I,R\}$.
- Given Z , infections are independent with transition rates given by
  - $P(x_{j,t+dt} = I \mid x_{j,t} = S, Z) = \sum z_{jk} \lambda_k N^I_{k,t}\, dt$
  - where $N^I_{k,t}$is the number that are in classs k and infective at time t and $\lambda_k$is the infection rate within group k.
- $N^I_{k,t} = \sum_j z_{jk} \mathbf{1}(x_{j,t} = I)$
- A basic model has λ_k the same for all k.
- Intuitively it is reasonable to assume a greater per person infection rate in a small group such as a household
  - so take $\lambda_k = \lambda N_k^{-\nu}$

Indian Buffet Epidemic

- The state of individual j is at time t is $x_{j,t} \in$ {S,I,R} .
- Given Z , infections are independent with transition rates given by
  - $P(x_{j,t+dt} = I \,|\, x_{j,t} = S,\, Z) = \sum z_{jk} \lambda_k N^I_{k,t}\, dt$
  - where $N^I_{k,t}$is the number that are in classs k and infective at time t and $\lambda_k$is the infection rate within group k.
- $N^I_{k,t} = \sum_j z_{jk} \mathbf{1}(x_{j,t} = I)$
- A basic model has λ_k the same for all k.
- Intuitively it is reasonable to assume a greater per person infection rate in a small group such as a household
  - so take $\lambda_k = \lambda N_k^{-\nu}$

MCMC Inference

- Augmented data
- Parameterisation
- Proposal

Summary

- A new model for epidemics incorporating heterogenity has been introduced.
- Initial steps towards inference taken.
- Planned developments
  - Develop MCMC algorithms
  - Apply to real data
- Questions

MCMC Inference

The parameters $\theta = (\lambda, \alpha, \gamma, \nu, K)$.
The log liklihood when the epidemic is observed on $[0, T_{\max}]$

$$\log f(T^I, T^R | z, \theta) = \sum \log \eta_j(T_j^I) - \int_0^{T_{\max}} \sum \eta_j(t) dt + \sum \log g(T_j^R - T_j^I) + \sum \log 1 - G(T_{\max} - T_j^I) \quad (1)$$

where η_j is the instantaneous rate of infections on individual j
g and G are the pdf and cdf of time to recovery

$$\eta_j(t) = \sum_k z_{jk} \lambda_k N_{k,t-}^I \quad (2)$$

$$\eta_j(t) = \lambda \sum_k z_{jk} N_{k,t-}^I / N_k^\nu \quad (3)$$

Random Walk Metropilis MCMC

Given complete data, i.e. observed infection and recovery times, the likelihood factorises so γ can be independently estimated. The steps in the algorithm are:

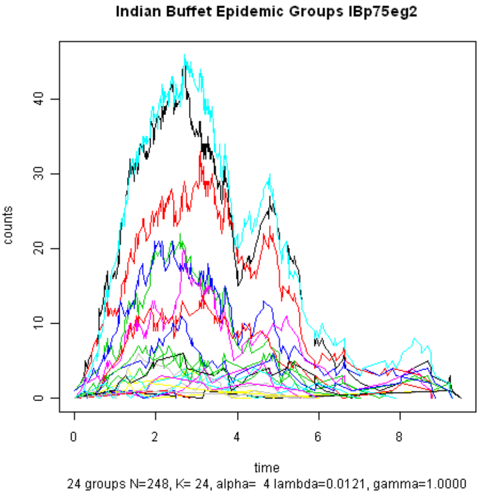
1. λ ∼MH using a random walk with Gaussian steps, folding at 0
2. α ∼MH using a random walk with Gaussian steps, folding at 0
3. Z ∼MH on Z, proposal .
    1. At each step K i.i.d. column flip probabilities ψ_k are sampled from a beta distribution with parameters K and $0.8/K$.
    2. Within each column, each bit is flipped independently with probability ψ_k.

These parameters where chosen so that the expected number of flips is close to 1 but there is a small chance of a large number of flips.

Diffusion Models

Defining $x_t = N_t^S/N$and $y_t = N_t^I/N$ we can approximate the process as an SDE
$dx = -\beta xydt + \sqrt{\beta xy/N}dB_1(t)$
$dy = (\beta xy - \gamma y)dt - \sqrt{\beta xy/N}dB_1(t) + \sqrt{\gamma y/N}dB_2(t)$
where dB_{1} and dB_{2} are independent Brownian motions.

Infectives in each group



Martingale estimation

A significant result is that of Becker and Hasofer (1997) An epidemic process has two obvious martingales

$$dM_1(t) = dN_t^S + \beta N_t^S N_t^I \quad (4)$$

$$dM_2(t) = dN_t^R - \gamma N_t^I \quad (5)$$

setting $\theta = \beta/\gamma$ two less obvious martingales

$$dM_3(t) = dM_1(t) + \theta N_t^S dM_2(t) \quad (6)$$

$$M_4(t) = \delta M_2(t) + \int_0^t H(\theta, \tau) dM_3 \quad (7)$$

For Further Reading I

H. Andersson and T. Britton.
Stochastic Epidemic Models and Their Statistical Analysis.
Springer, 2000.

T.L. Griffiths and Z. Ghahramani.
Infinite latent feature models and the Indian buffet process
(tech. rep. no. 2005-001).
Gatsby Computational Neuroscience Unit, 2005.

N.G. Becker and A.M. Hasofer.
Estimation in epidemics with incomplete observations.
Journal of the Royal Statistical Society. Series B (Methodological), 59(2):415–429, 1997.