Indian Buffet Epidemics

A non-parametric Bayesian Approach to Modelling Heterogeneity

Ashley Ford, Gareth Roberts

Department of Statistics, University of Warwick

Inference For Epidemic-related Risk, 2011

Outline

Motivation

- Need a model between
  - homogeneous mixing
  - over complex models with unknown parameters.
- Many have been proposed
  - household, spatial, multi-type
- Availability of data
  - contact surveys
    - RFID
    - POLYMOD
  - commuting data
- A non-parametric model for the heterogeneity that can represent a wide range of departures from homogeneity.

Motivation

- Need a model between
  - homogeneous mixing
  - over complex models with unknown parameters.
- Many have been proposed
  - household, spatial, multi-type
- Availability of data
  - contact surveys
    - RFID
    - POLYMOD
  - commuting data
- A non-parametric model for the heterogeneity that can represent a wide range of departures from homogeneity.

Motivation

- Need a model between
  - homogeneous mixing
  - over complex models with unknown parameters.
- Many have been proposed
  - household, spatial, multi-type
- Availability of data
  - contact surveys
    - RFID
    - POLYMOD
  - commuting data
- A non-parametric model for the heterogeneity that can represent a wide range of departures from homogeneity.

Motivation

- Need a model between
  - homogeneous mixing
  - over complex models with unknown parameters.
- Many have been proposed
  - household, spatial, multi-type
- Availability of data
  - contact surveys
    - RFID
    - POLYMOD
  - commuting data
- A non-parametric model for the heterogeneity that can represent a wide range of departures from homogeneity.

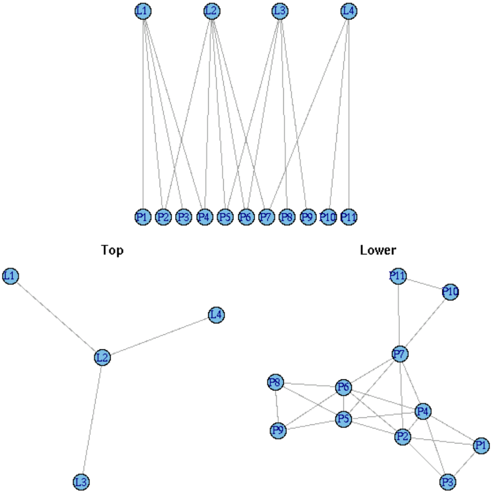
Places and People

- 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
- represented as
  - a bipartite graph
  - an $N \times K$ binary matrix Z

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
- represented as
  - a bipartite graph
  - an $N \times K$ binary matrix Z

Example bipartite graph



	L1	L2	L3	L4
1	1			
2	1	1		
3	1			
4	1	1		
5		1	1	
6		1	1	
7		1		1
8			1	
9			1	
10				1
11				1

Epidemics on bipartite graphs

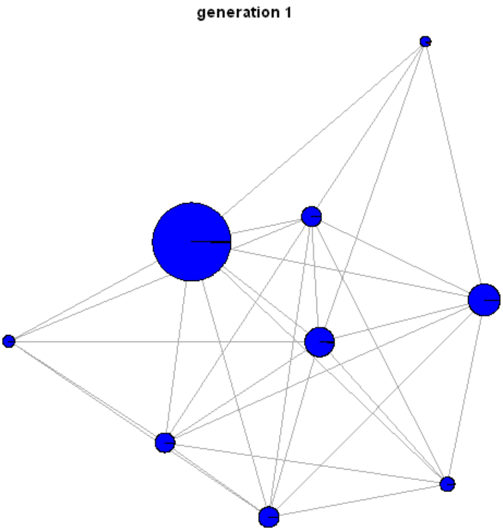
Extend homogeneous mixing epidemic models to bipartite graph

- e.g. SIR, SEIR, SIS, Reed-Frost

Approaches to defining the infection rate

- a single infection rate could apply to all pairs of individuals connected through one or more locations
- Each class has an associated infection rate λ_k
- Rate of infections on a susceptible individual j is $\sum z_{jk} \lambda_k N^I_{k,t}$
  - $N^I_{k,t}$ is the number that are in classs k and infective at time t.

Simulation



Household epidemic models

Global and within household infection rates $\lambda_g$and λ_h

household size m
number of housholds n_h
a bipartite graph representation with adjacency matrix $N \times (n_h + 1)$where $N = n_h m$
e.g. for 4 houses of sizes 2,3,3,4 the adjacency matrix is

λ_g	λ_h	λ_h	λ_h	λ_h
1	1			
1	1			
1		1		
1		1		
1		1		
1			1	
1			1	
1			1	
1				1
1				1
1				1
1				1

Household epidemic models

Global and within household infection rates λ_g and λ_h

household size m
number of housholds n_h
a bipartite graph representation with adjacency matrix $N \times (n_h + 1)$where $N = n_h m$
e.g. for 4 houses of sizes 2,3,3,4 the adjacency matrix is

λ_g	λ_h	λ_h	λ_h	λ_h
1	1			
1	1			
1		1		
1		1		
1		1		
1			1	
1			1	
1			1	
1				1
1				1
1				1
1				1

Multi-type model

A multi-type model with infection rate $\gamma_{i,j}$ between an infective in group i and a susceptible in group j .

If $\gamma_{i,j} = \gamma_{ji}$ and $\sum_{j \neq i} \gamma_{i,j} \leq \gamma_{i,i} \, \forall i$ it can be represented as a bipartite model.

These conditions will usualy apply if the groups are geographically separate,

but may not if the groups are split by ages or varying susceptibility and infectiousness e.g. $\gamma_{i,j} = \alpha_i \beta_j$.

With m types the bipartite representation has m columns for the within type infections and $m(m-1)/2$ for the between type infections.

Multi-type model

A multi-type model with infection rate $\gamma_{i,j}$ between an infective in group i and a susceptible in group j .

If $\gamma_{i,j} = \gamma_{ji}$ and $\sum_{j \neq i} \gamma_{i,j} \leq \gamma_{i,i} \; \forall \, i$ it can be represented as a bipartite model.

These conditions will usualy apply if the groups are geographically separate,

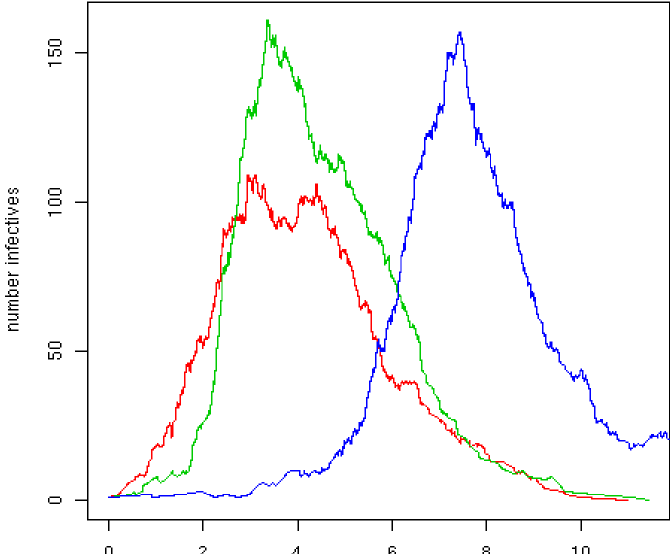
but may not if the groups are split by ages or varying susceptibility and infectiousness e.g. $\gamma_{i,j} = \alpha_i \beta_j$.

With m types the bipartite representation has m columns for the within type infections and $m(m-1)/2$ for the between type infections.

Multi-type model

$\gamma_{1,1} - \gamma_{1,2} - \gamma_{1,3}$	$\gamma_{2,2} - \gamma_{1,2} - \gamma_{2,3}$	$\gamma_{3,3} - \gamma_{1,3} - \gamma_{2,3}$	$\gamma_{1,2}$	$\gamma_{1,3}$	$\gamma_{2,3}$
1			1	1	
1			1	1	
1			1	1	
	1		1		1
	1		1		1
		1		1	1
		1		1	1
		1		1	1
		1		1	1

Examples



Other Models

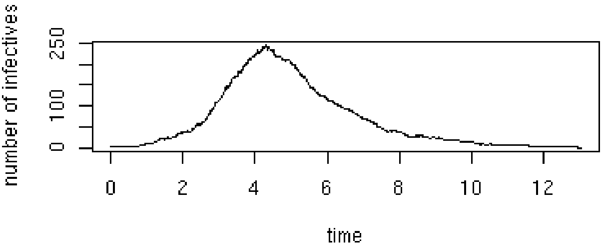
- three level
- commuting
- known labels

MCMC inference for known contact matrix

log likelihood when observed on $[0, T_{\max}]$
with n infections at T_j^I and removals at T_j^R

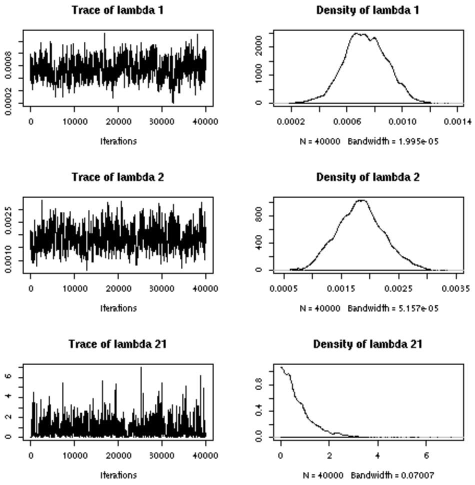
$$\sum \log \eta_j(T_j^I) \qquad - \int_0^{T_{\max}} \sum_j \eta_j(t)dt \qquad +$$
$$\sum_j \log g(T_j^R - T_j^I) \quad + \sum_j \log\{1 - G(T_{\max} - T_j^I)\}$$

where $\eta_j(t) = \sum_k z_{jk} \lambda_k N^I_{k,t-}$ is the instantaneous rate of infections on individual j, g and G are the pdf and cdf of time to recovery.
A simulation with Z 1000×21 and $\lambda_k = .5/n_k$ has 626 infections



MCMC inference for known contact matrix

RW - Metropolis-Hastings for 21 parameters



true values 0.00058, 0.00150, 0.17000

non-parametric inference

- Parametric distributions
  - Gaussian, Cauchy
  - exponential, gamma, log-normal, Weibull
- Non-parametric
  - histograms
  - kernel density estimate

Indian Buffet Process- a culinary metaphor

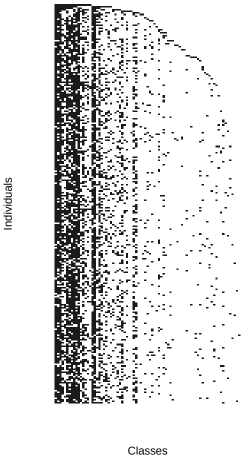
Introduced by Griffiths and Ghahramani (2005) .

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

A distribution over all binary matrices with N rows

- Expected number in each row is α
- Expected number of non zero columns $\alpha \sum_j^N 1/j$

Indian Buffet Process - example



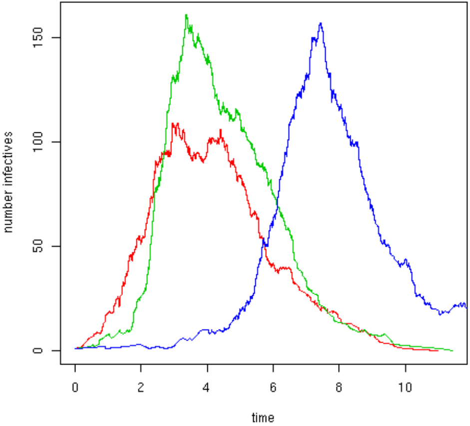
IBP Z generated with $N = 260$, $\alpha = 15$

Limit of Finite K

- ψ_k is the probability that an individual is in class k
- $\psi_k \sim Beta(\alpha/K, 1)$ or from stick breaking
- The model for Z is: $z_{ik}|\psi_k \sim Bernoulli(\psi_k)$ independently
- The Indian Buffet process is obtained as $K \to \infty$

Indian Buffet Epidemic

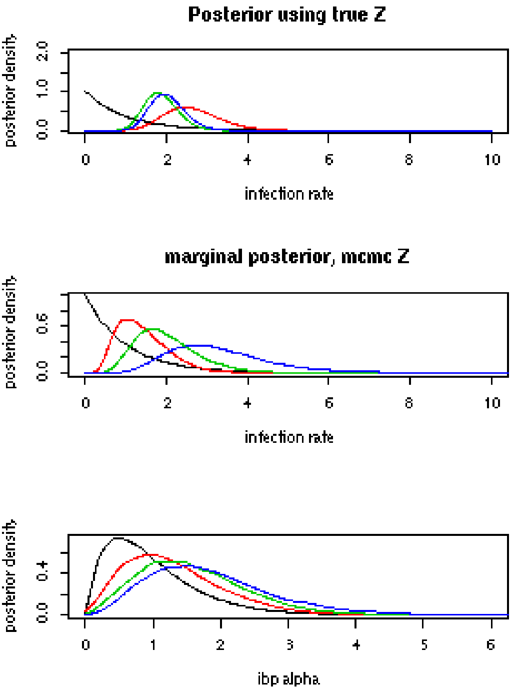
Combine the bipartite graph model with the Indian Buffet Process as a prior for the contact graph.



$\alpha = 6,\ N = 600,\ K_+ = 31$, different initial infectives in 8,4,1

MCMC for Indian Buffet Epidemic

- A challenging MCMC problem
  - Very high dimensional
  - multi-modal
- Proposals
  - non centered, using fixed K
  - independence, using sequential IBP


Posterior using true Z
posterior density
0.0
1.0
2.0
0
2
4
6
8
10
infection rate
marginal posterior, mcmc Z
posterior density
0.0
0.6
0
2
4
6
8
10
infection rate
posterior density
0.0
0.4
0
1
2
3
4
5
6
ibp alpha

Conclusions

- Bi-partite graph epidemics provide a generic formulation for modeling and inference.
- The Indian Buffet Epidemic provides a non-parametric model for heterogeneity in contact processes.
- MCMC inference is possible on small epidemics
  - Work continues on extending the size which can be handled

For Further Reading I

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.