Bayesian Hierarchical Models

Gavin Shaddick,
Millie Green, Matthew Thomas
University of Bath

6th - 9th December 2016
APPLICATIONS OF BAYESIAN HIERARCHICAL MODELS
OUTLINE

Spatial epidemiology

Disease mapping

Statistical methods for smoothing risks
Spatial epidemiology
**What is Spatial Epidemiology?**

- Epidemiology is the study of the distribution of diseases in human populations.
- Disease risk depends on the classic epidemiological triad of person (genetics/behaviour), place and time.
- Spatial epidemiology focuses on the second of these.
- Place is a surrogate for exposures present at that location
  - environmental exposures in water/air/soil
  - lifestyle characteristics of those living in particular areas.
Growing interest in spatial epidemiology

- Public interest in effects of environmental hazards/pollution.
- Epidemiological interest in differences in disease rates across different areas.
- Data availability: collection of health data at different geographical scales.
- Increase in computing power and methods
  - Geographical Informations Systems (GIS).
- Development of statistical/epidemiological methods for investigating disease ‘clusters’. 
THE NEED FOR SPATIAL METHODS

- Many epidemiological studies are spatial
  - many are spatio-temporal!
- When do we need to ‘worry’?
  - acknowledge the spatial component
  - are we explicitly interested in the spatial pattern of disease incidence?
    - disease mapping
    - cluster detection.
  - is the clustering a nuisance quantity that we wish to acknowledge, but are not explicitly interested in?
    - spatial regression.
**Types of Spatial Data**

- **Point data**
  - ‘exact’ residential locations exist for cases and controls.

- **Count data**
  - aggregation
  - typically over administrative units
  - ecological, in that they are collected across groups
  - in spatial studies the groups are geographical areas.
Disease mapping
Overview of Disease Mapping

- The estimation and presentation of summary measures of health outcomes.
- The aims of disease mapping include:
  - simple description
  - hypothesis generation
  - allocation of health care resources, assessment of inequalities
  - estimation of background variability in underlying risk in order to place epidemiological studies in context.
- Unfortunately there are well-documented difficulties with the mapping of raw estimates since, for small areas and rare diseases in particular, these estimates will be dominated by sampling variability.
**Example:** Lip Cancer in Scotland


**Data**

- observed and ‘expected’ number of cases (based on the county age populations)
- this allows the calculation of the standardised morbidity ratio
  - ratio of the observed to the expected cases.
- a covariate measuring the proportion of the population engaged in agriculture, fishing, or forestry (AFF)
- the projections of the longitude and latitude of the area centroid, and the ‘position’ of each county expressed as a list of adjacent counties.
**Example: Lip Cancer in Scotland**

<table>
<thead>
<tr>
<th>County No.</th>
<th>Obs Cases $Y_i$</th>
<th>Exp Cases $E_i$</th>
<th>Prop AFF</th>
<th>SMR</th>
<th>Project N (km)</th>
<th>Project E (km)</th>
<th>Adjacent Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1.4</td>
<td>0.16</td>
<td>6.43</td>
<td>834.7</td>
<td>162.2</td>
<td>5,9,19</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>8.7</td>
<td>0.16</td>
<td>4.48</td>
<td>852.4</td>
<td>385.8</td>
<td>7,10</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>3.0</td>
<td>0.10</td>
<td>3.67</td>
<td>946.1</td>
<td>294.0</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>2.5</td>
<td>0.24</td>
<td>3.60</td>
<td>650.5</td>
<td>377.9</td>
<td>18,20,28</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>4.3</td>
<td>0.10</td>
<td>3.49</td>
<td>870.9</td>
<td>220.7</td>
<td>1,12,19</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>2.4</td>
<td>0.24</td>
<td>3.33</td>
<td>1015.2</td>
<td>340.2</td>
<td>Island</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>8.1</td>
<td>0.10</td>
<td>3.21</td>
<td>842.0</td>
<td>325.0</td>
<td>2,10,13,16,17</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>2.3</td>
<td>0.07</td>
<td>3.04</td>
<td>1168.9</td>
<td>442.2</td>
<td>Island</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>2.0</td>
<td>0.07</td>
<td>3.00</td>
<td>781.4</td>
<td>194.5</td>
<td>1,17,19,23,29</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>2</td>
<td>5.6</td>
<td>0.01</td>
<td>0.36</td>
<td>640.8</td>
<td>277.0</td>
<td>24,31,46,48,49,53</td>
</tr>
<tr>
<td>48</td>
<td>3</td>
<td>9.3</td>
<td>0.01</td>
<td>0.32</td>
<td>654.7</td>
<td>282.0</td>
<td>24,44,47,49</td>
</tr>
<tr>
<td>49</td>
<td>28</td>
<td>88.7</td>
<td>0.00</td>
<td>0.32</td>
<td>666.7</td>
<td>267.8</td>
<td>38,41,44,47,48,52,53,54</td>
</tr>
<tr>
<td>50</td>
<td>6</td>
<td>19.6</td>
<td>0.01</td>
<td>0.31</td>
<td>736.5</td>
<td>342.2</td>
<td>21,29</td>
</tr>
<tr>
<td>51</td>
<td>1</td>
<td>3.4</td>
<td>0.01</td>
<td>0.29</td>
<td>678.9</td>
<td>274.9</td>
<td>34,38,42,54</td>
</tr>
<tr>
<td>52</td>
<td>1</td>
<td>3.6</td>
<td>0.00</td>
<td>0.28</td>
<td>683.7</td>
<td>257.8</td>
<td>34,40,49,54</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>5.7</td>
<td>0.01</td>
<td>0.18</td>
<td>646.6</td>
<td>265.6</td>
<td>41,46,47,49</td>
</tr>
<tr>
<td>54</td>
<td>1</td>
<td>7.0</td>
<td>0.01</td>
<td>0.14</td>
<td>682.3</td>
<td>267.9</td>
<td>34,38,49,51,52</td>
</tr>
<tr>
<td>55</td>
<td>0</td>
<td>4.2</td>
<td>0.16</td>
<td>0.00</td>
<td>640.1</td>
<td>321.5</td>
<td>18,24,30,33,45,56</td>
</tr>
<tr>
<td>56</td>
<td>0</td>
<td>1.8</td>
<td>0.10</td>
<td>0.00</td>
<td>589.9</td>
<td>322.2</td>
<td>18,20,24,27,55</td>
</tr>
</tbody>
</table>
Example: Lip Cancer in Scotland

Figure: SMRs for male lip cancer in 56 counties of Scotland.
EXAMPLE: LUNG AND BRAIN CANCER IN NORTH-WEST ENGLAND

- Two tumors
  - one non-rare (lung)
  - and one rare (brain).
- Analysis performed at ward level (144 wards)
  - incidence data by postcode.
- Brain cancer
  - the median number of cases per ward over the 11 year period is 6
  - range of 0 to 17.
- Lung cancer
  - the median number is 20
  - range 0–60.
- ‘Expected counts’ were based on ward-level populations from the 1991 census, by 5-year age bands and sex.
EXAMPLE: LUNG AND BRAIN CANCER IN NORTH-WEST ENGLAND

Figure: SIRs for (a) lung cancer, and (b) brain cancer in the North-West of England.
Non-statistical issues

- There is a trade-off when a geographical scale is chosen
  - larger geographical areas provide more stable rates
  - relative risk summaries may be distorted due to the large aggregation of individuals.
- If the relative risk shows marked variation within a particular area this information will be lost
  - high relative risks will be diluted under aggregation.
- The size of the areas chosen also determines the sort of questions that can be posed
  - larger areas often mean greater contrasts in relative risks and exposures
  - localized effects can only be detected with data at a smaller level of aggregation.
**Statistical Issues**

▶ Unfortunately there are well-documented difficulties with the mapping of raw estimates since, for small areas and rare diseases in particular, these estimates will be dominated by sampling variability.

▶ For the model

\[ Y_i \sim \text{Poisson}(E_i \theta_i) \]

statistically, the best estimate of the rate \( \hat{\theta}_i \) will be

\[ \hat{\theta}_i = \text{SMR}_i = \frac{Y_i}{E_i} \]

with variance

\[ \text{var}(\hat{\theta}_i) = \frac{\theta_i}{E_i} \]

so that areas with small \( E_i \) have high associated variance.
EXAMPLE: SCOTTISH LIP CANCER

- The variance of the estimate is $\text{var}(\text{SMR}_i) = \frac{\text{SMR}_i}{E_i}$.
- This will be large if $E_i$ is small.
- For the Scottish data the expected numbers are highly variable, with range 1.1–88.7.
- This variability suggests that there is a good chance that the extreme SMRs are based on small expected numbers
  - many of the large, sparsely populated, rural areas in the north have high SMRs.
Figure: Estimates versus standard errors for 56 counties of Scotland.
Smoothing Models

- The above considerations led to methods being developed to smooth the SMRs.
- Use hierarchical/random effects models that use the data from all of the areas to provide more reliable estimates in each of the constituent areas.
Statistical methods for smoothing risks
**Spatial Models**

- In general we might expect relative risks in areas that are ‘close’ to be more similar than in areas that are not ‘close’.
- We would like to exploit this information in order to provide more reliable relative risk estimates in each area.
- This is analogous to the use of a covariate $x$, in that areas with similar $x$ values are likely to have similar relative risks.
- Unfortunately the modelling of spatial dependence is much more difficult since spatial location is acting as a surrogate for unobserved covariates.
- We need to choose an appropriate spatial model, but do not directly observe the covariates whose effect we are trying to mimic.
Consider the model

\[ Y_i|\beta, \gamma, U_i, V_i \sim \text{Poisson}(E_i \mu_i e^{U_i+V_i}) \]

\[ \log \mu_i = g(S_i, \gamma) + f(x_i, \beta), \]

- \( S_i = (S_{i1}, S_{i2}) \) denotes spatial location, the centroid of area \( i \),
- \( f(x_i, \beta) \) is a regression model,
- \( g(S_i, \gamma) \) is an expression that we may include to capture large-scale spatial trend, e.g. long-term spatial trend

\[ f(S_i) = \gamma_1 S_{i1} + \gamma_2 S_{i2}, \]

- The random effects \( V_i \sim \text{iid } N(0, \sigma_v^2) \) represent non-spatial overdispersion,
- \( U_i \) are random effects with spatial structure.
One approach is to specify the distribution of the random effect for a particular area, $U_i$, as if we knew the values of the spatial random effects, $U_j$, in ‘neighbouring areas’.

We therefore need to specify a rule for determining the ‘neighbours’ of each area.

Commonly areas $i$ and $j$ are taken to be neighbours if they share a common boundary.
- Various other neighbourhood/weighting schemes are possible.
- The neighbourhood structure could depend on the distance between area centroids.
- Determine the distance within which regions are considered neighbours.

Figure: Close-up of a region of Birmingham.
The ICAR model

- A common model is to assign the spatial random effects an intrinsic conditional autoregressive (ICAR) prior.
- Under this specification it is assumed that

\[ U_i | U_j, j \in \partial_i \sim N \left( \bar{U}_i, \frac{\omega_u^2}{m_i} \right), \]

where \( \partial_i \) is the set of neighbours of area \( i \), \( m_i \) is the number of neighbours, and \( \bar{U}_i \) is the mean of the spatial random effects of these neighbours.
- The parameter \( \omega_u^2 \) is a conditional variance and its magnitude determines the amount of spatial variation.
- The variance parameters \( \sigma_v^2 \) and \( \omega_u^2 \) are on different scales, \( \sigma_v \) is on the log odds scale while \( \omega_u \) is on the log odds scale, conditional on \( U_j, j \in \partial_i \); hence they are not comparable.
Notice that if $\omega_u^2$ is ‘small’ then although the residual is strongly dependent on the neighboring value the overall contribution to the residual relative risk is small.

This is a little counterintuitive but stems from spatial models having two aspects, strength of dependence and total amount of spatial dependence, and in the ICAR model there is only a single parameter which controls both aspects.
Recall the model

\[ Y_i | \beta, \gamma, U_i, V_i \sim_{ind} \text{Poisson}(E_i \mu_i e^{U_i + V_i}) \]

with

\[ \log \mu_i = g(S_i, \gamma) + f(x_i, \beta), \]  

(1)

where

- \( S_i = (S_{i1}, S_{i2}) \) denotes spatial location, the centroid of area \( i \),
- \( f(x_i, \beta) \) is a regression model,
- \( g(S_i, \gamma) \) is an expression that we may include to capture large-scale spatial trend – the form

\[ f(S_i) = \gamma_1 S_{i1} + \gamma_2 S_{i2}, \]

is a simple way of accommodating long-term spatial trend.
- The random effects \( V_i \sim_{iid} N(0, \sigma_v^2) \) represent non-spatial overdispersion,
- \( U_i \) are random effects with spatial structure which we now consider ‘jointly’ rather than considering neighbours.
A Joint Model

- Assume that $\mathbf{U} = (U_1, ..., U_n)$ arise from a zero mean multivariate normal distribution with variances $\text{var}(U_i) = \sigma^2_u$ and correlations $\text{corr}(U_i, U_j) = \exp(-\phi d_{ij}) = \rho^{d_{ij}}$ where $d_{ij}$ is the distance between the centroids of areas $i$ and $j$, and $\rho > 0$ is a parameter that determines the extent of the correlation.

- This model is isotropic since it assumes that the correlation is the same in all spatial directions. We refer to this as the joint model, since we have specified the joint distribution for $\mathbf{U}$.

- More generally the correlations can be modeled as $\text{corr}(U_i, U_j) = \exp(-(\phi d_{ij})^\kappa)$. 
**Bayesian Analysis**

- Inference for models with a spatial component is often not straightforward using likelihood based approaches, and so Bayesian methods are commonly used.
- Unfortunately, most Bayesian models are not conducive to analytical analysis, and so are not available in standard software packages.
- Markov chain Monte Carlo (MCMC)
  - Win/OpenBUGS is a package that allows very general Bayesian modeling
  - GeoBUGS module contains a number of useful spatial models, and mapping facilities
  - Packages in R, for example CARBayes.
- Approximate Bayesian inference, for fast computation on big datasets
  - R-INLA.
**Example: Lip Cancer in Scotland**

*Figure:* (a) Unsmoothed SMRs and (b) Smoothed SMRs for lung cancer in the North-West of England.
**Example: Lung and Brain Cancer in North-West England**

![Figure: (a) Unsmoothed SIRs and (b) Smoothed SIRs for lung cancer in the North-West of England.](image-url)
EXAMPLE: LUNG AND BRAIN CANCER IN NORTH-WEST ENGLAND

Figure: (a) Unsmoothed SIRs and (b) Smoothed SIRs for brain cancer in the North-West of England.
SUMMARY OF SMOOTHING IN DISEASE MAPPING

- The aim is to reduce the instability inherent in SMRs based on small expected numbers.
- This is achieved by fitting a random effects model which assumes that area-level deviations from the regression model arise from a probability distribution.
- Comparing SMRs with smoothed estimates is important
  - if there are big changes, are they appropriate?
  - were the expected numbers small?
  - did the regression model fit this area well?