A Bayesian Approach for Spatial Analysis of Lung Cancer Mortality Rates in Ohio

Hu Wei

June 14, 2005
A Bayesian Approach for Spatial Analysis of Lung Cancer Mortality Rates in Ohio

Hu Wei

1. Introduction

The analysis of spatial dispersion of the risk of occurrence of an event, e.g., disease, is usually done via maps of incidence rates, where a set of areas is shaded according to the values of a variable of interest. The objective of disease mapping is to infer the geographic distribution of the rates thus highlighting geographic areas with high or low incidence or mortality rates of a specific disease, such as lung cancer, and the variability of such rates over a state or country. Disease maps can also be used to identify spatial clusters which may be due to common environmental, demographic or cultural effects shared by neighboring regions.

For a map divided into \( J \) contiguous areas, let \( y = (y_1, \ldots, y_J) \) be the observed number of occurrences of a certain disease in area \( j \) during the period of investigation, and \( e = (e_1, \ldots, e_J) \) be the expected number of occurrences. For a disease like cancer mortality a Poisson model is commonly adopted for each \( y_j \) (Cressie, 1993). If the disease is noncontagious the numbers of cases are mutually independent. According to the Poisson model, each \( y_j \) has mean \( \psi_j e_j \) where \( \Psi = (\psi_1, \ldots, \psi_J) \) are the relative risks. Relative risk is a measure of how much a particular risk factor (e.g., cigarette smoking) influences the risk of a specified outcome (e.g., lung cancer). \( \Psi = (\psi_1, \ldots, \psi_J) \) are specific of each area and these are the parameters of interest. So, the adopted model is given by

\[
Y_j | e_j, \psi_j \sim \text{Poisson}(e_j \psi_j), \quad j = 1, \ldots, J.
\]

The expected count \( e_j \) is assumed to be a known quantity which value is based on known risk factors. In the analysis in this project it is assumed that there are no confounding factors and the expected values are computed as \( n_i \sum_j y_j / \sum_j n_j \) where \( n_i \) is the population in area \( i \). In this work these values are close to the values computed by applying statewide age-sex adjusted rates to age-sex population in each county since age-sex distribution in each area is close to overall distribution.
2. Maximum Likelihood Estimate

In the classical approach the maximum likelihood estimate of the relative risks is given by \( \hat{\psi}_j = \frac{y_j}{e_j} \), with estimated standard error \( s_j = \sqrt{\frac{y_j}{e_j}} \), and is usually called the standardized mortality rate (SMR). Several problems with this approach have already been identified in the literature. First, more extreme values of the estimates of the relative risks (which usually dominate the map) may be based on a few cases only in areas with small population. The second problem is that rare events in small areas can lead to extra-Poisson variation, i.e., there is more heterogeneity in the population than is assumed by the Poisson model. Another characteristic commonly found in this context and that is not taken into account by the classical approach is the possibility of spatial correlation in the relative risks. Such correlation may be due for example to spatially correlated covariates and not included in the model.

2.1 Estimating relative risks using MLE approach

Information was collected concerning the number of lung cancer mortality in 1994 at each county of Ohio. Figure 1 shows the population distribution in counties of Ohio. In the map in Figure 2, maximum likelihood estimates (SMR) were calculated and the counties received different tones on the map according to the SMR value. Note that from the dispersion in Figure 3 that there is a large variation in the mortality rates for small populations. As we can see, counties with larger populations received a positive evaluation relative to those with small populations, which can lead to wrong decisions concerning the mortality rates by county. Figure 4 shows the probability of relative risk greater than 1 with cutoff points at 0.025 and 0.975. It seems as if the mortality rates are being underestimated in some counties and it is also not easy to identify any sort of spatial pattern on the map.
Figure 1: Population distribution in counties of Ohio

Figure 2: Relative risks obtained via maximum likelihood estimates (SMR) for lung cancer mortality numbers in counties of Ohio
3. Hierarchical Models for Relative Risks

The problems with MLE can be taken into account by allowing the relative risk to vary within each area and in this case the Bayesian approach is appropriate specifying prior distributions to the parameters $\Psi$.

In the Bayesian approach, further to information in the data, i.e. the number of occurrences in each area, we need to specify a prior distribution $p(\Psi)$ for the relative
risks which provides information concerning their variability along the map. Bayesian inference is then based on the combination of these two sources of information via the posterior distribution of the relative risks $p(\Psi \mid y)$ obtained via Bayes theorem. The prior distribution usually depends on hyperparameters $\gamma$ so that the marginal posterior of $\Psi$ is given by

$$p(\Psi \mid y) = \int p(\Psi, \gamma \mid y) d\gamma$$

Point estimates of the relative risks can be obtained via location measures of the distribution (1) while scale measures provide information on the uncertainty of these estimates. In general the integrals involved in the computation of these measures cannot be obtained analytically or even by numerical integration and approximation methods are necessary.

### 3.1 Prior Specification

The prior distribution of the relative risks should be structured so as to accommodate the effect of factors measured at the level of each area as well as the possibility of spatial variation. One form of combining unstructured priors with information concerning the spatial structure proposed in the literature consists in modeling the logarithm of the relative risk as the sum of two independent components: the nonspatial random effect $\theta_j$ and the spatial random effect $\phi_j$. In the literature, the nonspatial components are usually modeled independently as $\theta_j \sim N(0, 1/\tau_\theta)$, describing the unstructured heterogeneity, and the spatial components so as to indicate that geographically close areas tend to present similar risks. One way of expressing this spatial structure is conditional specification via Markov random fields models where the distribution of each $\phi_j$ given all the other elements $\{\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_J\}$ depends only on its neighborhood (Cressie, 1993). A commonly used model is the Gaussian intrinsic conditional autoregression where the conditional distribution of each $\phi_j$ is given by

$$\phi_j \mid \phi_{-j} \sim N\left( \frac{\sum_{i \in \delta_j} w_{ij} \phi_i}{\sum_{i \in \delta_j} w_{ij}}, \frac{1}{\tau_\phi \sum_{i \in \delta_j} w_{ij}} \right)$$

where $\phi_j$ represents the set of areas which are neighbor of area $j$. One important restriction in this specification is that the matrix of weights $W$ must be symmetric. It
should be noted that the specification of this CAR structure leads to a prior joint
distribution for the relative risks given by
\[
\phi_j \mid \phi_{-j} \sim N \left( \frac{\sum_{i \in \delta_j} w_{ij} \phi_i}{\sum_{i \in \delta_j} w_{ij}}, \frac{1}{\tau^2 \sum_{i \in \delta_j} w_{ij}} \right)
\]

This prior is improper as it is based on paired differences between the \( \phi_j \)'s. In
practice, a sum to zero constraint is imposed to these random effects in order to
guarantee identifiability (GeoBUGS manual).

Although other possibilities exist, the simplest and most commonly used
neighborhood structure is defined by the existence of a common border of any length
between the areas. In this case, the weights \( w_{ij} \) are specified as \( w_{ij} = 1 \) if \( i \in \delta_j \) and \( w_{ij} = 0 \) otherwise so that \( \sum_{i \in \delta_j} w_{ij} \) is simply the number of neighbors of area \( j \). So, the
conditional prior mean of \( \phi_j \) is given by the arithmetic average of the spatial effects
from its neighbors and the conditional prior variance is proportional to the number of
neighbors. This will also be the structure adopted in this work.

Another way of expressing the spatial structure is joint specification. In this study, a
powered exponential function is utilized for the elements of the correlation matrix
(GeoBUGS manual):
\[
f(d_{ij}; \phi, k) = \exp[(-\phi d_{ij})^k]
\]
where \( d_{ij} = \) distance between areas \( i \) and \( j \), the parameter \( \phi \) controls the rate of
decline of correlation with distance, and \( k \) is a smoothing constant between 0 and 2. In
this study, \( k \) is set to 1 in all joint specification models.

A set of covariates is also related to the relative risks. The vector \( \beta \) of covariate
coefficients will be specified from a multivariate normal prior distribution with mean
zero and variance-covariance matrix \( \Sigma = I \sigma_\beta^2 \), i.e. the \( \beta_j \)'s are assumed independent a
priori.

This class of models has been used in most recent works on disease mapping (see for
eexample Waller et al., 1997; Sun et al., 2000). The use of information from other
areas in the region under study should reduce the effect of random fluctuations not associated to the relative risk. Taking spatial correlation among neighboring areas into account is expected to produce smoother and more informative maps.

3.2 Fully Bayesian Estimation

In the fully Bayesian approach the idea is to add another level in the model hierarchy by specifying a prior distribution for the hyperparameters $\gamma$ (Gelman et al., 2003). In this case the inference on the relative risks will be based on the marginal posterior distribution (1) which often cannot be obtained analytically. Analytical or numerical approximations are needed. In particular, Markov chain Monte Carlo methods (MCMC) will be employed to obtain a sample from the joint posterior distribution of $(\Psi, \gamma)$, automatically generating samples from the marginal posteriors of $\Psi$ and $\gamma$.

Defining the parameter vectors $\theta = (\theta_1, \cdots, \theta_J)$ and $\phi = (\phi_1, \cdots, \phi_J)$, the joint posterior distribution of all parameters is expressed as

$$p(\theta, \phi, \tau_\theta, \tau_\phi \mid y) \propto p(y \mid \theta, \phi)p(\theta \mid \tau_\theta)p(\phi \mid \tau_\phi)p(\tau_\theta)p(\tau_\phi)$$

This joint posterior distribution takes into account a conditional independence structure. From a sample from this posterior we can obtain estimates of the relative risks via $\psi_j = \exp(\theta_j + \phi_j)$.

In the highest level of the hierarchy prior distributions are specified to the prior precisions $\tau_\theta$ and $\tau_\phi$. The Gamma family of prior distributions is conditionally conjugate, i.e. the full posterior conditional distribution is also Gamma. This conditional conjugacy allows that $\tau_\theta$ and $\tau_\phi$ be easily updated in the algorithm used here. A common choice in the literature is the non-informative (proper) prior Gamma($\epsilon, \epsilon$) with small values for $\epsilon$. However, this specification attributes low prior probability to small values of the standard deviation and consequently a spatial structure for example might be imposed a priori. Kelsall and Wakefield (2002) verified that the estimation of relative risks can be highly dependent of the choice of prior parameters and within a class of Gamma priors they suggest a Gamma(0.5;0.0005) distribution as a sensible choice.
The posterior distribution is clearly analytically intractable. One method to obtain values from the joint posterior is via simulation of a sufficiently large number of dependent observations of the parameter vector as an ergodic Markov chain. In particular, the algorithm known as Gibbs sampler is useful in the context of Markov random fields where the joint posterior distribution is complicated but the full conditional posterior distributions have simple forms (Gelman et al., 2003). This method is implemented here with the software WinBUGS, freely available on the Internet in the address http://www.mrcbsu.cam.ac.uk/bugs. The example WinBUGS code used in this work is provided in the appendix.

**Estimating relative risks using Bayesian approach**

The dataset concerning lung cancer mortality rates in 1994 by county in Ohio was again analyzed. The Bayesian hierarchical model as described above was estimated with a conditionally conjugate Gamma prior for the hyperparameters $\tau_\theta$ and $\tau_\phi$. The variation in the relative risks along the counties was modeled according to the spatial prior (conditional and joint specifications), plus the effects of selected covariates. The covariates used in this study are described below.

- **Log.emissions**: Logarithm of the amount of air pollutants by county of Ohio in 1994 (Source: EPA).
- **Poverty**: Estimated percentage of poverty population in 1994 (Source: Census).

![Figure 5: Emissions of air pollutants by county in 1994.](image-url)
The maps in Figures 5 and 6 were constructed with the values of the two covariates. The WinBUGS software was then used to perform 1200 simulations from the full conditional posterior distributions, from which the first 300 were discarded as burn-in. So, all the results here are based on a sample of 900 values. From this sample several characteristics of the posterior distribution may be estimated, and the main interest here is the spatial variation of the relative risks.

In order to compare and select the more appropriate model among those considered here I use the Deviance information criterion (DIC), where lower values indicate a good model fit relative to the number of parameters in the model (Gelman et al., 2003):

\[ \text{DIC} = \hat{D}_{\text{avg}}(y) = 2 \hat{D}_{\text{avg}}(y) - D_{\theta}(y) \]

One advantage of the DIC is that it is easily computed during the simulation of the Markov chains. In Table 1 the computed values for the DIC are presented, from which it can be seen that the best model is the one that incorporates spatial and nonspatial random effects plus the covariate Poverty. The results also show that powered exponential specification was better than CAR prior distribution in terms of dependent variable fit. This is due to the fact that for between-area correlation the exponential model takes all areas into account while the CAR model only considers adjacent neighbors for each area.
Table 1: DIC values for each model obtained from WinBugs based on 1200 simulations (lowest value in bold).

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>No spatial structured variance</td>
<td>952.3</td>
</tr>
<tr>
<td>Structured &amp; unstructured CAR</td>
<td>926.5</td>
</tr>
<tr>
<td>Exp</td>
<td>916.3</td>
</tr>
<tr>
<td>log.emissions</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>923.6</td>
</tr>
<tr>
<td>Exp</td>
<td>916.7</td>
</tr>
<tr>
<td>poverty</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>924.3</td>
</tr>
<tr>
<td>Exp</td>
<td><strong>916.2</strong></td>
</tr>
<tr>
<td>log.emissions and poverty</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>921.7</td>
</tr>
<tr>
<td>Exp</td>
<td>917.4</td>
</tr>
</tbody>
</table>

The relative risks were estimated using one covariate Poverty. The map in Figure 7 was constructed with the posterior means of the $\psi$, as point estimates of the relative risks. Note that there were no big changes from the SMRs in areas with large populations, as we expected as these were good estimates of the relative risks. As shown in Figure 8, there are few extreme values than using SMR approach. On the other hand, areas with small populations benefited from the information coming from neighboring areas. Also, there is a risk pattern that concentrates in the south and southeast areas (mountain areas), then reduces in northwest directions. All in all, it is easier to identify counties in the state with similar relative risks.
Figure 7: Relative risks estimates in the Bayesian hierarchical model with covariate Poverty included.

Figure 8: The probability of relative risk greater than 1 (Bayesian)

The history graphs for selected posterior means in Figure 9 indicate that the model converges well after several hundred iterations.

Figure 9: History graphs
4. Discussion

In this work I adopted a Bayesian approach to estimate relative risks of a rare event occurrence in small areas. The problem of overdispersion found in the usual classical estimation was tackled via specification of suitable priors. The method was illustrated with a real data example. Estimates of the posterior distribution were obtained via MCMC methods where inference is based on an approximate sample from the posterior distribution.

The Bayesian hierarchical model adopted is intrinsically spatial thus incorporating a component that captures the large scale smooth variation of the risk in the whole region under study. For the real data example, a set of covariates which might be of potential influence on the relative risks were included and tested via DIC.

There are various interesting extensions of the model adopted in this work. In the CAR prior, other weight structures describing the neighborhood may be adopted. For example, the matrix of weights of neighborhood structure might be defined in terms of the length of the border between two areas. In this case, rather than a binary structure the weight $w_{ij}$ is equal to the length of the border between areas $i$ and $j$ so that the spatial influence from neighboring areas increases or decreases (a priori) according to the length of the border. Trying to include other covariates, time structures and studying space-time interactions will also be the subject of future work.
References


Appendix

WindBUGS code

model {
  # Likelihood
  for (i in 1 : N) {
    obs.m[i]  ~ dpois(mu[i])
    theta[i]  ~ dnorm(0, tau.theta)
    log(mu[i]) <- log(e.m[i]) + beta0 + beta2*pov[i] + phi[i] + theta[i]
    RR[i] <- exp(beta0 + beta2*pov[i] + phi[i] + theta[i])
  }
  # CAR prior distribution for relative risk:
  phi[1:N] ~ car.normal(adj[], weights[], num[], tau.phi)
  for(k in 1:sumNumNeigh) {
    weights[k] <- 1
  }
  # Other priors:
  beta0 ~ dflat()
  #beta1 ~ dnorm(0.0, 1.0E-5)
  beta2 ~ dnorm(0.0, 1.0E-5)
  tau.phi ~ dgamma(0.5, 0.0005)
  tau.theta ~ dgamma(0.5, 0.0005)
  sigma.phi <- sqrt(1 / tau.phi)
}

Model diagram