Bayesian SAE using Complex Survey Data
Lecture 4A: Hierarchical Spatial Bayes Modeling

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington
Motivation

Spatial Hierarchical Models for Normal Data

Spatial Hierarchical Models for Binomial Data

Overview of Spatial Random Effects Models

Normal and Binomial Examples

Discussion

Technical Appendix: The Conditional Spatial Model
Motivation
Smoothing

In the last lecture, we considered hierarchical models that shrunk estimates towards a central value, with no consideration of the geography of the areas.

In general, we might expect unknown parameters of interest in areas that are “close” to be more similar than in areas that are not “close”.

We would like to encode this observation in a model, in order to smooth locally in space, in order to provide more reliable 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 parameters.

Unfortunately the modeling 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.
Spatial Hierarchical Models for Normal Data
Previously, we examined the non-spatial random effects model:

\[ Y_{ik} = \beta_0 + \delta_i + \epsilon_{ik}, \]

Mean of Area \( i \)

with \( \delta_i \sim iid \ N(0, \sigma_\delta^2) \) – these are the area-specific deviations (the random effects) from the overall level \( \beta_0 \) – and \( \epsilon_{ik} \sim iid \ N(0, \sigma_\epsilon^2) \), is the measurement error.

We extend this model to

\[ Y_{ik} = \beta_0 + \delta_i + S_i + \epsilon_{ik}, \]

Mean of Area \( i \)

where \( S_i \) are spatial random effects.

We are separating the residual variability into:

- Unstructured area-level variability \( \delta_i \).
- Spatial area-level variability \( S_i \).
- Measurement error \( \epsilon_{ik} \).
We will not go into detail on prior specification or computation for spatial models, in the accompanying R notes, we show how INLA provides a means for computing posterior summary measures, with sensible prior choices.

For more details on space-time modeling with INLA, see Blangiardo and Cameletti (2015).
Spatial Hierarchical Models for Binomial Data
We first consider the model

\[ Y_i | \theta_i \sim \text{ind Binomial}(n_i, \theta_i) \]  \hspace{1cm} (1)

with

\[ \log \left( \frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x_i \beta_1 + S_i + \delta_i, \]  \hspace{1cm} (2)

where

- the random effects $\delta_i | \sigma^2_\delta \sim \text{iid } N(0, \sigma^2_\delta)$ represent non-spatial overdispersion,
- $S_i$ are random effects with spatial structure.
- We describe two possible forms for the spatial random effects.
Overview of Spatial Random Effects Models
In general, there have been two approaches to modeling spatial dependence:

- **Local conditional** modeling: in our context, are usually used for **area** data.
- **Geostatistical** modeling: in our context, are usually used for **point** data.
The local approach, an early reference to which is Besag (1974), is based on conditional specifications $S_i | S_{-i}$, where

$$S_{-i} = (S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_n).$$

In general, the only variables in $S_{-i}$ that are relevant are the neighbors (suitably defined), which we write as $S_i | S_j, j \in \text{ne}(i)$.

In words, what is the distribution of $S_i$, given we know the values taken by the neighboring random variables $S_j, j \in \text{ne}(i)$ – known as a Markov Random Field (MRF) model.
The geostatistical approach, see for example Stein (1999), is based on the specification of the full multivariate distribution of

$$\mathbf{S} = (S_1, \ldots, S_n)$$

Kriging, which is used for prediction in many spatial contexts, may be derived from a multivariate normal geostatistical model.

For modeling area-level data, we will concentrate on conditionally specified spatial models\(^1\).

\(^1\)though return to the above model in the last lecture when we consider construction of a continuous surface
A Conditional Spatial Model

We need to specify a rule for determining the neighbors of each area.

In an epidemiological context the areas are not regular in shape.

This is in contrast to image processing applications in which the data are collected on a regular grid.

Hence, there is an arbitrariness in specification of the neighborhood structure.
To define *neighbors*, the most common approach is to take the neighborhood scheme to be such that two areas are treated as neighbors if they share a common boundary.

This is reasonable if all regions are (at least roughly) of similar size and arranged in a regular pattern (as is the case for pixels in image analysis where these models originated), but is not particularly attractive otherwise (but reasonable practical alternatives are not available).
Various other neighborhood/weighting schemes are possible:

- One can take the neighborhood structure to depend on the distance between area centroids and determine the extent of the spatial correlation (i.e. the distance within which regions are considered neighbors).
- One could also define neighbors in terms of cultural similarity.

In typical applications it is difficult to assess whether the spatial model chosen is appropriate, which argues for a simple form, and to assess the sensitivity of conclusions to different choices.
A Conditional Spatial Model

A common model, due to Besag *et al.* (1991), is to assign the spatial random effects an intrinsic conditional autoregressive (ICAR) prior.

Under this specification it is assumed that the spatial random effect is drawn from a normal distribution whose mean is the mean of the neighbors’ random effects, with variance proportional to one over the number of neighbors (so more neighbors, less variability).

Formally,

$$S_i | S_j, j \in ne(i) \sim N \left( \bar{S}_i, \frac{\sigma^2_s}{m_i} \right),$$

where $ne(i)$ is the set of neighbors of area $i$, $m_i$ is the number of neighbours, and

$$\bar{S}_i = \frac{1}{m_i} \sum_{j \in ne(i)} S_j$$

is the mean of the spatial random effects of these neighbors.
The parameter $\sigma^2_s$ is a conditionally \textit{variance} and its magnitude determines the amount of spatial variation.

Recall, we split the residual variability as

$$\delta_i + S_i.$$ 

The variance parameters $\sigma^2_\epsilon$ and $\sigma^2_s$ have different interpretations.

Both are defined on the same scale, but $\sigma_\epsilon$ has a \textit{marginal} interpretation while $\sigma_s$ has a \textit{conditional} interpretation.

Specifically, for area $i$, the variance of $S_i$ is conditional on $S_j, j \in \text{ne}(i)$.

Hence the variances are not directly comparable; the random effects $\epsilon_i$ and $S_i$ are comparable, however (so side-by-side maps of the contributions are useful).

Bottom line: Larger values of $\sigma^2_s$ are indicative of greater spatial dependence.
Normal and Binomial Examples
Figure 1: Comparison of area averages: Posterior medians from non-spatial model (described in Lecture 3) versus MLEs (left). Posterior medians from spatial model versus MLEs (right).

The shrinkage is less predictable with the spatial model, which is because of the local adaptation.
Motivating Example: Normal Data

Figure 2: Spatial (left) and non-spatial (left) random effects from the spatial+IID model.

The IID contribution is much smaller than the spatial contribution.
Motivating Example: Normal Data

Figure 3: Non-spatial random effects $\delta_i$ from the non-spatial model (left) and spatial random effects (right) random effects $S_i$.

The non-spatial model random effects are trying to pick up the spatial structure!
Motivating Example: Normal Data

Figure 4: Estimates of area averages of weight via MLE’s (left) and posterior medians from spatial model (right).

The extremes are attenuated under the spatial model.
Motivating Example: Normal Data

Figure 5: Posterior median estimates of area averages of weight via non-spatial hierarchical model with $\beta_0 + \delta_i$ (left) and spatial hierarchical model $\beta_0 + \delta_i + S_i$ (right); $\delta_i$ are iid and $S_i$ are spatial random effects.

Some differences between the estimates, but relatively minor.
Motivating Example: Binomial Data

Figure 6: Spatial (left) and non-spatial (left) random effects from the spatial+iid model with \( \logit(p_i) = \beta_0 + \delta_i + S_i; \delta_i \) are iid and \( S_i \) are spatial random effects.

The majority of the between-area variability is spatial.
Motivating Example: Binomial Data

Figure 7: Non-spatial random effects from the non-spatial model (left) and spatial random effects (right) random effects.

The non-spatial model random effects are trying to pick up the spatial structure!
Motivating Example: Binomial Data

Figure 8: MLEs of area diabetes risk (left) and posterior medians from the spatial hierarchical model (right).
Motivating Example: Binomial Data

Figure 9: Posterior median estimates of area diabetes risk via non-spatial hierarchical model (left) and spatial hierarchical model (right).

Estimates are very similar!
Motivating Example: Binary Outcome

Figure 10: Comparison of area averages. Posterior standard deviation versus standard errors of MLEs on the probability scale, for the non-spatial hierarchical model (left), and the spatial hierarchical model (right).

The problem of standard errors being estimated as zero is clearly alleviated, and the two sets of posterior standard deviations are quite similar.
Motivating Example: Binary Outcome

Figure 11: Bias of MLEs, with confidence intervals (left). Bias of posterior medians, with credible intervals (right).

If we calculate,

$$\frac{1}{n} \sum_{i=1}^{n} |\hat{p}_i - p_i|,$$

we get 0.026 (MLE) and 0.018 (Bayes).
Discussion
If the data are sparse in an area, averages and totals are unstable because of the small denominators.

More reliable estimates can be obtained by using the totality of data to inform on the distribution, both locally and globally, of the averages across the study region.

A GLMM can include spatial dependence relatively easily, with the ICAR model being particularly popular.
Discussion

Four levels of understanding for hierarchical models, in descending order of importance:

- The intuition on global and local smoothing.
- The models to achieve this.
- How to specify prior distributions.
- The computation behind the modeling.

Overall Strategy

- First, calculate empirical means and map them. Also look at map of standard errors and/or confidence intervals.
- Fit non-spatial random effects models.
- Fit the ICAR+IID spatial model.
- Add in covariates if available.
References


Technical Appendix: The Conditional Spatial Model
This is a little counterintuitive but stems from spatial models having two aspects, the strength of dependence and the magnitude of spatial dependence, and in the ICAR model there is only a single parameter which controls both aspects.

In the joint model (with covariance $\sigma_s^2 \rho^{d_{ij}}$ for example) the strength is determined by $\rho$ and the total amount by $\sigma_s^2$.

A non-spatial random effect should always be included along with the ICAR random effect since this model cannot take a limiting form that allows non-spatial variability.

In the joint model with $S_i$ only, this is achieved as $\rho \to 0$.

If the majority of the variability is non-spatial, inference for this model might incorrectly suggest that spatial dependence was present.

Prior specification is difficult for the conditional variance is difficult because it has a conditional rather than a marginal interpretation.

See Fong et al. (2010) for details on how to handle this problem.
Let $Q/\sigma_s^2$ denote the precision matrix of the ICAR model.

For simplicity, suppose all areas are connected to at least one other area.

The elements $Q_{ij} = 0$ if $S_i$ and $S_j$ are conditionally independent, i.e., not neighbors.

The elements $Q_{ij} = -1$ if $S_i$ and $S_j$ are conditionally dependent, i.e., neighbors.

The elements $Q_{ii} = m_i$, where $m_i$ is the number of neighbors of area $i$.

Hence, most of the elements of $Q$ are zero (so the matrix is sparse) and this aids greatly in computation, see Rue and Held (2005) for details.
The form of the joint ‘density’ is

\[
p(s|Q, \sigma_s^2) = (2\pi)^{-1/2} |Q|^{1/2} \sigma_s^{-(n-1)/2} \exp \left( -\frac{1}{2\sigma_s^2} s^\top Q s \right)
\]

\[
= (2\pi)^{-1/2} |Q|^{1/2} \sigma_s^{-(n-1)/2} \exp \left( -\frac{1}{2\sigma_s^2} \sum_{i \sim j} (s_i - s_j)^2 \right)
\]

where \( i \sim j \) means \( i \) and \( j \) are neighbors.

This is not a true density since it is not proper; \( Q \) is singular and has rank \( n - 1 \).

The ICAR model is an example of a **Gaussian Markov Random Field**.

Note the contrast with the multivariate model in which \( \Sigma_{ij} = 0 \) if the marginal covariance between \( S_i \) and \( S_j \) is zero.