2022 SISMID Module 9
Lecture 3: Disease Mapping for Count Data

Jon Wakefield and Lance Waller

Departments of Statistics and Biostatistics
University of Washington

Outline

Simple Summaries and Motivation for Smoothing Models

Bayesian Inference

Non-Spatial Hierarchical Models

Spatial Hierarchical Models

Discussion

Parameter Interpretation

Simple Summaries and Motivation for Smoothing Models

Mapping

Mapping has a long history in many fields, in particular epidemiology, and may be defined as the estimation and presentation of summary measures of outcomes of interest. Background reading, Wakefield *et al.* (2000).

The aims of mapping include

- simple description,
- hypothesis generation,
- allocation of resources, assessment of inequalities, and
- estimation of background variability in underlying risk in order to place studies in context.

In this lecture we consider Poisson models for count data aggregated over areas, but the methods we describe can be applied to normal and binomial data, amongst others.

The models we describe in this section, can also be used for regression modeling, and assessment of clustering.

Motivation and Context

We begin by noting a number of non-statistical issues, with an emphasis on epidemiology, public/global health, for more background see Chapters 12 and 13 of Elliott *et al.* (2000):

- In broad-scale studies (in particular international endeavors), data comparability is a major issue.
- Precise disease definition (via ICD codes) is also extremely important.
- Mortality data tend to be more reliable than incidence data, but the latter are in general of greater epidemiological interest, because incident cases are closer in time to exposure.
- Prevalence data are also more common than incidence data.

What's the Denominator?

In general for count data gathered in a particular geographical region and over some time period, the starting model will often be binomial or Poisson.

In either case, it will be important to determine an appropriate denominator.

If the response we are counting is on humans (e.g., disease counts, crimes, unemployed,...) then the denominator (on which to base a raw rate, for example), will often be the population of the area.

However, in the area of crime mapping, an appropriate denominator may not be totally obvious, and number of household may be more reasonable for residential burglaries, than population, and for car thefts, one would want the number of vehicles available to steal. See Ratcliffe (2010) for further discussion.

Motivation and Context

There is a trade-off when a geographical scale is chosen:

- Larger geographical areas providing more stable rates and less problems of migration, but relative risk summaries may be distorted due to the large aggregation of individuals.
- If the target shows marked variation within a particular area this information will be lost – if a particular subregion has a spike in the surface then this will be diluted under aggregation.
- Larger study regions are likely to offer greater contrasts (range of covariates *x*) in relative risks and exposures.

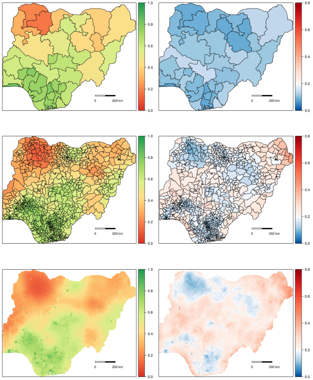

Figure 1: Vaccination coverage for: Admin-1 (top), Admin-2 (middle), 1 × 1 km (bottom). Left: posterior mean. Right: the width of a 95% credible interval.

Presentation

- Chloropleth (areas shaded) are the most popular kind of maps, but isopleth (contours) and cartograms (size of areas proportional to denominator) have also been used.
- Choice of color is important – multiple colors (hues) can be confusing, shading with a single color can work well.
- Two colors can work if 'intuitive', for example, for a relative risk, areas above 1 can be colored in red, with darker shades being used for values further from 1; areas below 1 can be colored in blue with darker shades being used for values further from 1.
- Cut-points should be chosen to be scientifically meaningful and convey as much information as possible.

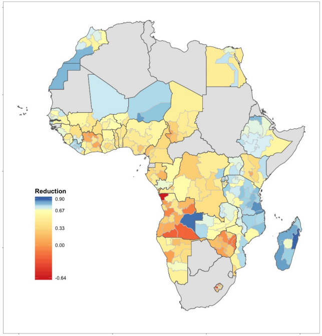

Figure 2: Reduction in under 5 mortality from 1990–2015.

Bayesian Inference

Bayesian Inference

Bayesian inference is a convenient framework within which to implement smoothing models.

- A Data Model (Likelihood) is probabilistically combined with
- A Penalization (Prior) that expresses beliefs about the parameters θ encoding the model.
- Combination occurs via **Bayes Theorem:**

$$\underbrace{p(\theta|y)}_{\textbf{Posterior}} \propto \underbrace{L(\theta)}_{\textbf{Likelihood}} \times \underbrace{\pi(\theta)}_{\textbf{Prior}}.$$

- On the log scale:

$$\underbrace{\log p(\theta|y)}_{\textbf{Updated Beliefs}} = \underbrace{\log L(\theta)}_{\textbf{Data Model}} + \underbrace{\log \pi(\theta)}_{\textbf{Penalization}}.$$

Bayesian Inference

- In a Bayesian analysis the complete set of unknowns (parameters) is summarized via the **multivariate posterior distribution**.
- The marginal distribution for each parameter may be summarized via its mean, standard deviation, or quantiles.
- It is common to report the **posterior median** and a **90% or 95% posterior range** for parameters of interest.
- The range that is reported is known as a credible interval.
- The computations required for Bayesian inference (integrals) is often not trivial and many be carried out using a variety of analytic, numeric and simulation based techniques.
- We use the integrated nested Laplace approximation (INLA), introduced by Rue *et al.* (2009).
- Book-length treatments:
  - Blangiardo and Cameletti (2015) – space-time models.
  - Wang *et al.* (2018) – general models.
  - Krainski *et al.* (2018) – advanced space-time models.

Bayes Example

- Imagine the data model is normal with an unknown mean μ:

  $$\bar{y} \mid \mu \sim \mathrm{N}(\mu, \sigma^2/n),$$

  where σ^2/n is assumed known ($\sigma/\sqrt{n}$ is the standard error).
- We also imagine the prior is normal:

  $$\mu \sim \mathrm{N}(m, v),$$

  so that values of the mean μ that are (relatively) far from m are penalized.
- The log posterior is:

$$\underbrace{\log p(\mu \mid y}_{\text{Updated Beliefs}}) = -\underbrace{\frac{n}{2\sigma^2}(\bar{y} - \mu)^2}_{\text{Data Model}} - \underbrace{\frac{1}{2v}(\mu - m)^2}_{\text{Penalization}}.$$

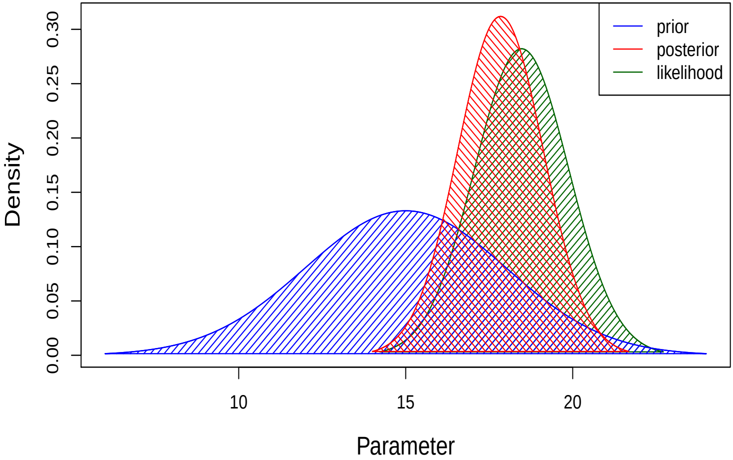

Figure 3: Normal data model with $n = 10$, $\overline{y} = 19.3$ and standard error 1.41. The prior for μ has mean $m =$15 and $v = 3^2$. The posterior for the parameter μ is a compromise between the two sources of information: the posterior mean is 18.5 and the posterior standard deviation is 1.28.

Non-Spatial Hierarchical Models

Instability of the Naive Relative Risk Estimate

We now consider the Poisson model, which is widely used in spatial epidemiology.

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

For the model $Y_i|\theta_i \sim \text{Poisson}(E_i\theta_i)$, the MLE is $\widehat{\theta}_i = \text{SMR}_i = \frac{Y_i}{E_i}$, with variance

$$\text{var}(\widehat{\theta}_i) = \frac{\theta_i}{E_i},$$

which is estimated by

$$\widehat{\text{var}}(\widehat{\theta}_i) = \frac{\widehat{\theta}_i}{E_i}$$

so that areas with small E_i have high associated variance.

In general, a confidence interval for θ_i is the exponentiated confidence interval for the log parameter (details shortly).

Surveillance

We imagine separate monthly surveillance for each of three areas over a 10-year period.

Data simulated from the model

$$Y_t | \theta \sim_{ind} \text{Poisson}(E\theta),$$

$t = 1, \ldots, 60$ months, where the relative risk is $\theta = 1$ in each case.

Recall that the MLE of the SMR in each time period is $\widehat{\theta}_t = Y_t / E$ with variance proportional to $1/E$ so that areas with small expected numbers have high variability.

The expected numbers differ in the three plots in Figure 4, and the resultant instability in the SMR is apparent.

For the $E = 0.2$ case there are a number of time periods with high estimates (and estimates of zero also!).

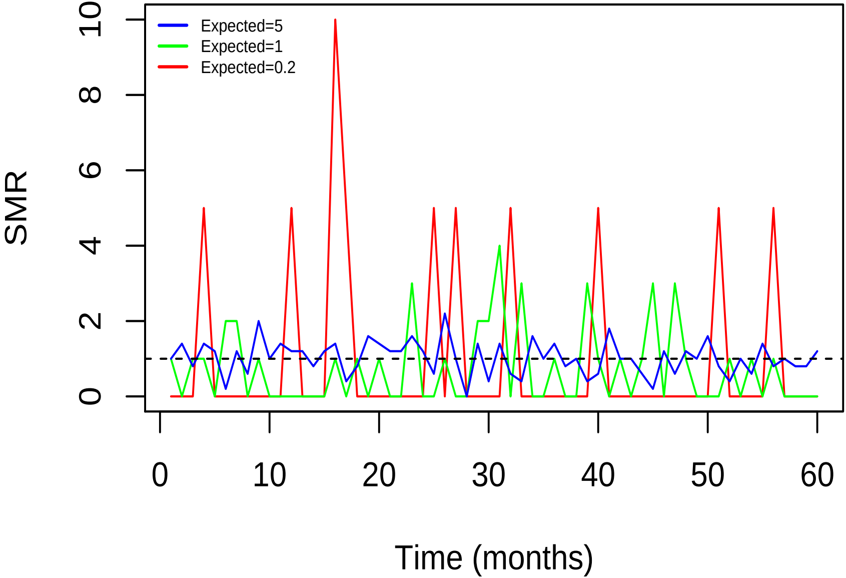

Figure 4: Simulations from the Poisson distribution under different expected numbers.

Scottish Lip Cancer Data

Incidence rates of lip cancer in males in 56 counties of Scotland, registered in 1975–1980. These data were originally reported in the mapping atlas of Kemp *et al.* (1985).

The Scottish lip cancer data have been widely analyzed, because they have been around a long time, and the SIRs display a lot of spatial variability.

The form of the data is:

- Observed and expected number of cases (based on the county age populations, details shortly) – allows the calculation of the standardized morbidity ratio, the 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.

Area i	Cases Y_i	Exp E_i	Prop AFF	SMR	Project N (km)	Project E (km)	Adjacent Counties
1	9	1.4	0.16	6.43	834.7	162.2	5,9,19
2	39	8.7	0.16	4.48	852.4	385.8	7,10
3	11	3.0	0.10	3.67	946.1	294.0	12
4	9	2.5	0.24	3.60	650.5	377.9	18,20,28
5	15	4.3	0.10	3.49	870.9	220.7	1,12,19
6	8	2.4	0.24	3.33	1015.2	340.2	Island
7	26	8.1	0.10	3.21	842.0	325.0	2,10,13,16,17
8	7	2.3	0.07	3.04	1168.9	442.2	Island
...							
48	3	9.3	0.01	0.32	654.7	282.0	24,44,47,49
49	28	88.7	0.00	0.32	666.7	267.8	38,41,44,47,48,52,53,5
50	6	19.6	0.01	0.31	736.5	342.2	21,29
51	1	3.4	0.01	0.29	678.9	274.9	34,38,42,54
52	1	3.6	0.00	0.28	683.7	257.8	34,40,49,54
53	1	5.7	0.01	0.18	646.6	265.6	41,46,47,49
54	1	7.0	0.01	0.14	682.3	267.9	34,38,49,51,52
55	0	4.2	0.16	0.00	640.1	321.5	18,24,30,33,45,56
56	0	1.8	0.10	0.00	589.9	322.2	18,20,24,27,55

Scottish Lip Cancer Data

- Figure 6 shows the SMRs for the Scottish lip cancer data, and indicates a large spread with an increasing trend in the south-north direction.
- The variance of the estimate is $\mathrm{var}(\mathrm{SMR}_i) = \mathrm{SMR}_i / E_i$, which will be large if E_i is small.

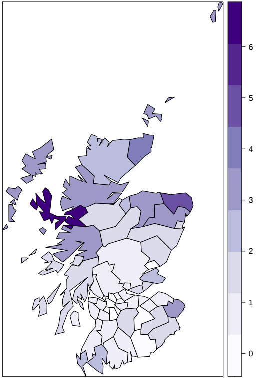

Figure 5: SMRs in 56 counties of Scotland.

Scottish Lip Cancer Data

- 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).

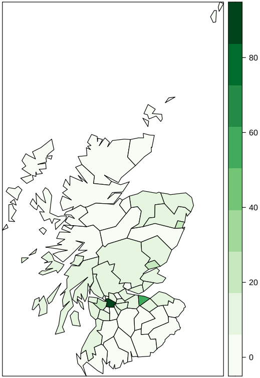

Figure 6: SMRs in 56 counties of Scotland.

Smoothing Models

The above considerations of instability again lead to models for smoothing the SMRs using hierarchical/random effects models that use the data from the totality of areas to provide more reliable estimates in each of the constituent areas.

Overview of Models:

- Basic Poisson Model: No smoothing.
- Random Effects Models:
  - Poisson-Lognormal: Non-spatial smoothing.
  - Poisson-Lognormal-Spatial: Spatial and non-spatial smoothing.
- Covariates may be added to each of these in order to smooth over covariate space.
- Computation for these models is a separate issue.

The Big Picture

Consider the situation in which I want to estimate the relative risk θ in an area with a disease count y and expected number of E.

A starting sampling model is $Y|\theta \sim \text{Poisson}(E\theta)$ – the MLE (SMR) is $\widehat{\theta} = y/E$.

Suppose I tell you that based on the larger region in which the area is contained, a prior for θ is the LogNormal distribution with known parameters μ and σ^2 – the prior median is $\exp(\mu)$.

A Bayesian analysis can be performed with the two-stage model:

$$\begin{aligned} Y|\theta &\sim \text{Poisson}(E\theta) \\ \theta &\sim \text{LogNormal}(\mu, \sigma^2) \end{aligned}$$

The posterior distribution for θ will reflect the information in the data, as quantified through the Poisson sampling model, and the prior – the resultant inference will typically produce narrower intervals than a confidence interval, though the estimate will be moved from the SMR, towards $\exp(\mu)$ – shrinkage.

The Big Picture

This is all wishful thinking, as how do we come up with μ and σ^2?

Suppose I have data (y_i, E_i) from $i = 1, \ldots, n$ areas – we can use all of the data to estimate the parameters of the LogNormal distribution, μ and σ^2.

This gives the classic three stage hierarchical model (see next slide).

Hierarchical models have a long history and are designed to produce improved estimates of the collection $\theta_1, \ldots, \theta_n$ – but the shrinkage means each area's estimate is biased (in a frequentist sense).

However, the variance of the estimate in each area is reduced, so we get improved mean squared error (MSE)[1].

But the shrinkage means that hierarchical models are not good for hot spot detection.

---

[1] MSE = Var + Bias^2

Poisson-Lognormal Models

A Poisson-lognormal non-spatial random effect model is given by:

$$\begin{array}{rcl} Y_i|\theta_i & \sim_{ind} & \text{Poisson}(E_i\theta_i) \\ \log\theta_i & = & \beta_0 + e_i \\ e_i|\sigma_e^2 & \sim_{iid} & \text{N}(0,\sigma_e^2) \end{array}$$

where e_i are area-specific random effects that capture the residual or unexplained (log) relative risk of the event in area i, $i = 1, \ldots, n$.

Here we have

$$\lambda_i = \text{e}^{e_i} \sim_{iid} \text{LogNormal}(0, \sigma_e^2).$$

We have a single parameter controlling the spread of the random effects, σ_e^2.

We use the INLA computational approach (Rue *et al.*, 2009) – see Riebler *et al.* (2016) for details on how a variety of spatial models may be analyzed with INLA.

INLA

- Relatively recently an approach has emerged that combines Laplace approximations and numerical integration in a very efficient manner, see Rue *et al.* (2009) for more detail.
- The method is designed for latent Gaussian models, which for our purposes, means modeling with latent normal random effects, that have independent, spatial, or space-time structure.
- Suppose the outcomes y_i, have density $p(y_i|\boldsymbol{x}_i, \alpha^\star)$, $i = 1, \dots, n$, where $\boldsymbol{x}|\theta \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1})$ and $\alpha^\star$ are dispersion parameters (e.g. the measurement error variance for a normal sampling model).
- The $\boldsymbol{x}$ represent the random effects, along with regression coefficients.
- We also have priors for $\alpha^\star$ and $\boldsymbol{\theta}$ — these priors may be non-normal – penalized complexity (PC) priors are recommended (Simpson *et al.*, 2017).
- Let $\boldsymbol{\alpha} = [\boldsymbol{\theta}, \boldsymbol{\alpha}^\star]$ denote the non-Gaussian parameters.

INLA

The posterior is,

$$\begin{aligned}\pi(\boldsymbol{x}, \boldsymbol{\alpha} \mid \boldsymbol{y}) &\propto \pi(\boldsymbol{\alpha})\pi(\boldsymbol{x} \mid \boldsymbol{\alpha}) \prod_{i=1}^{n} p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\alpha}) \\ &\propto \pi(\boldsymbol{\alpha})|Q(\boldsymbol{\theta})|^{p/2} \exp\left\{-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i=1}^{n} \log p(y_i|\boldsymbol{x}_i, \boldsymbol{\alpha})\right\}\end{aligned}$$

The random effects are always added on the linear predictor scale:

- linear for a normal sampling model
- log-linear for a Poisson sampling model
- logistic for a sampling binomial model

IID Model Estimates

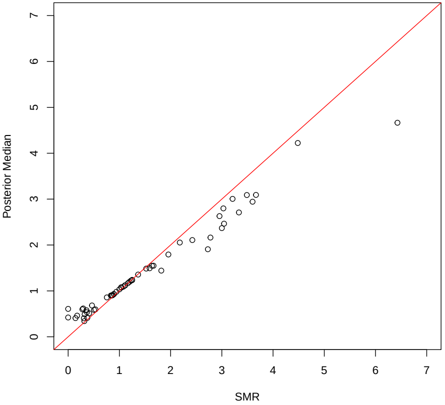

Figure 7: Comparison of posterior medians from IID Lognormal model and SMRs. The shrinkage of the Bayes estimates is evident.

IID Model Estimates

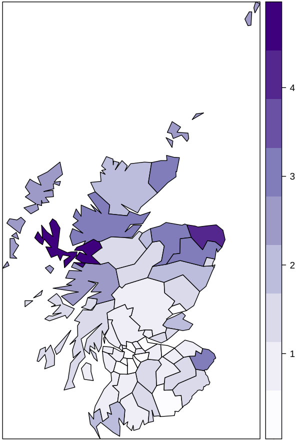

Figure 8: Map of posterior medians of relative risk estimates – note the narrowing of the range.

AFF Covariate

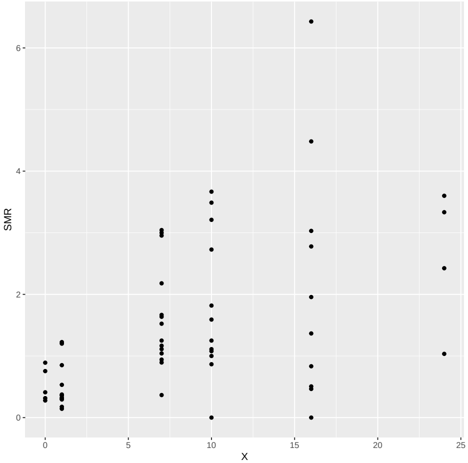

Figure 9: SMR versus percentage of population in agriculture, fishing and farming, in Scottish counties.

Covariates

We now add the AFF covariate x to the model:

$$\begin{aligned} Y_i|\theta_i &\sim_{ind} \text{Poisson}(E_i\theta_i) \\ \log\theta_i &= \beta_0 + \beta_1 x_i + e_i \\ e_i|\sigma_e^2 &\sim_{iid} \text{N}(0, \sigma_e^2) \end{aligned}$$

This results is a RR posterior median estimate of $\exp(\widehat{\beta}_1) = 1.07$ so that an area with 1% higher in AFF has a 7% increase in risk of male lip cancer.

This is an ecological association, and should not be assumed to hold for the men in the areas – this runs the risk of the ecological fallacy (Wakefield, 2008).

Parameter Interpretation

The posterior median of σ_e is 0.582 and a 95% interval is

$$[1/\sqrt{5.13}, 1/\sqrt{1.70}] = [0.44, 0.77].$$

A more interpretable quantity is an interval on the residual relative risk (RRR).

The RRRs follow a lognormal distribution LogNormal$(0, \sigma_e^2)$ so a 95% interval is $\exp(\pm 1.96 \times \sigma_e)$.

A posterior median of a 95% RRR interval is

$$\begin{aligned}[\exp(-1.96 \times \text{median}(\sigma_e)), \exp(1.96 \times \text{median}(\sigma_e))] \\ = [\exp(-1.96 \times 0.582), \exp(1.96 \times 0.582)] \\ = [0.320, 3.13]\end{aligned}$$

which is quite wide.

Spatial Hierarchical Models

Poisson Spatial Model

The previous model assumed that deviations from the overall level arose from the same distribution, with no spatial pattern – we might expect local fluctuations.

A natural spatial model is:

$$\begin{aligned} Y_i|\theta_i &\sim_{ind} \text{Poisson}(E_i\theta_i) && (1) \\ \log\theta_i &= \beta_0 + x_i\beta_1 + \delta_i && (2) \end{aligned}$$

where δ_i have a spatial component.

The BYM Model

In the popular BYM model Besag, York and Mollié (1991):

$$\delta_i = S_i + e_i,$$

where

- $e_i \sim_{ind} \mathrm{N}(0, \sigma_e^2)$.
- The spatial effects S_i are modeled conditional on the neighbors:

$$S_i | \underbrace{\{S_j = s_j, j \sim i\}}_{\text{Neighbors of } i}, \sigma_s^2 \sim \mathrm{N}\left(\overline{s}_i, \frac{\sigma_s^2}{m_i}\right),$$

  where $\overline{s}_i = \frac{1}{m_i} \sum_{j \sim i} s_j$ is the mean of the neighbors of area i and m_i is the number of such neighbors.
- σ_s^2 is a smoothing parameter: large values indicate large spatial variability.
- This form is known as an intrinsic conditional autoregression (ICAR).

Example of a Neighborhood Scheme

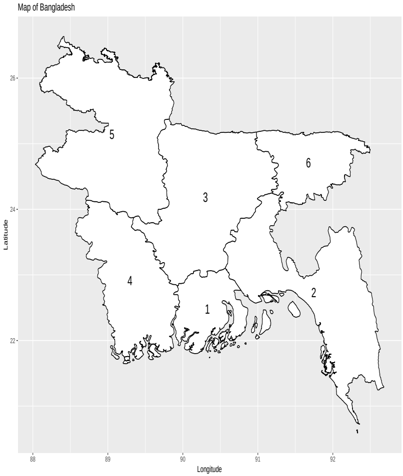

Figure 10: Common boundary neighbor scheme for Bangladesh divisions.

ICAR prior for spatial random effects $\boldsymbol{s}$ cab be written as,

$$p(\boldsymbol{s}|\sigma_s^2) \propto \exp\left(-\frac{1}{2}\boldsymbol{s}^\top \boldsymbol{Q}\boldsymbol{s}\right).$$

Precision matrix, $\boldsymbol{Q} = \boldsymbol{R}/\sigma_s^2$, $\boldsymbol{R}$:

$$\begin{bmatrix} 3 & -1 & -1 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & -1 \\ -1 & -1 & 5 & -1 & -1 & -1 \\ -1 & 0 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 & 0 \\ 0 & -1 & -1 & 0 & 0 & 2 \end{bmatrix}$$

BYM2 Prior

We have random effects contributions $S_i + e_i$, and in the BYM2 model the original BYM model is reparameterized (Riebler *et al.*, 2016; Simpson *et al.*, 2017).

The vector of random effects $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_n]^\top$ is written as,

$$\boldsymbol{\delta} = \sigma_\delta \left[\sqrt{1-\phi}\,\boldsymbol{e} + \sqrt{\phi}\,\boldsymbol{S}^\star \right]$$

where $\boldsymbol{e}$ are iid and $\boldsymbol{S}^\star$ are ICAR random effects rescaled to have a comparable (generalized) variance to the IID terms.

Then,

$$\text{var}(\boldsymbol{\delta}) = \sigma_\delta^2 \left[(1-\phi)\boldsymbol{I} + \phi \boldsymbol{Q}^{\star -} \right],$$

so

- σ_δ^2 is the total variance and
- ϕ is the proportion of the variance that is spatial.

The BYM2 model is very popular, because of its relative simplicity, ease of implementation, and the extensive experience gathered on its use.

BYM2 Hyperpriors

Recall the BYM2 model with random effects,

$$\boldsymbol{\delta} = \sigma_\delta[\sqrt{1-\phi}\boldsymbol{e} + \sqrt{\phi}\boldsymbol{S}^\star]$$

where $\boldsymbol{x}^\star$ have been rescaled to have a comparable (generalized) variance to the IID terms. Then,

$$\text{var}(\boldsymbol{\delta}) = \sigma_\delta^2[(1-\phi)\boldsymbol{I} + \phi\boldsymbol{Q}^{\star -}].$$

For a prior on σ_δ^2, we specify two values (U_1, α_1) such that $\Pr(\sigma_\delta < U_1) = \alpha_1$, e.g., $U_1 = 0.5, \alpha_1 = 0.95$.

In terms of the mixing parameter, $\phi = 0$ corresponds to the base model, i.e., no spatial variation.

We again specify (U_2, α_2) set through $\Pr(\phi < U_2) = \alpha_2$, e.g., $U_2 = 0.5, \alpha_2 = 0.5$.

Neighborhood Models

- 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).

A Conditional Spatial Model

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.

Spatial Model Estimates

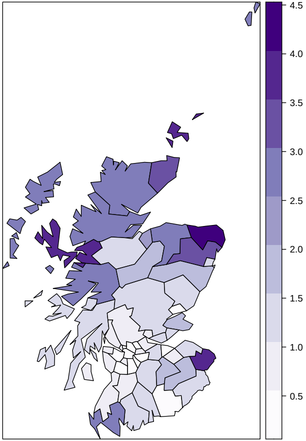

Figure 11: Map of posterior medians of relative risk estimates from the spatial model with covariate.

Spatial model fit

The posterior median of the total standard deviation (on the log relative risk scale) is $1/\sqrt{4.45} = 0.47$.

The posterior median for the proportion of the residual variation that is spatial is 0.96.

Note that the posterior mean estimate of β_1, log relative risk associated with AFF is reduced from 0.068 to 0.026 when moving from the non-spatial to spatial model – this is known as confounding by location.

The model attributes spatial variability in risk to either the covariate (which has a strong spatial pattern) or to the spatial random effects – in the non-spatial model, all of the spatial variation goes into the covariate term.

Spatial Model Estimates

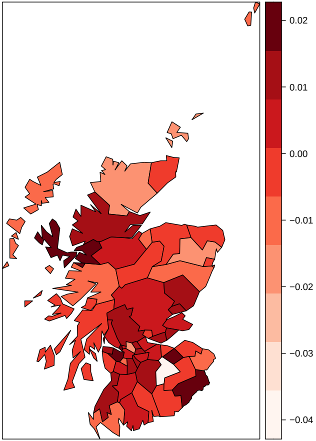

Figure 12: Map of posterior medians of non-spatial component of the log residual relative risk (e_i) from the spatial model with covariate.

Spatial Model Estimates

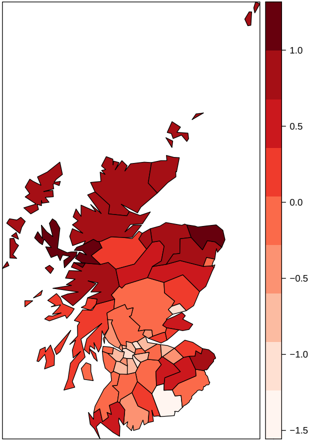

Figure 13: Map of posterior medians of spatial component of the log residual relative risk (S_i) from the spatial model with covariate. Note the very different scale from the previous map.

Discussion

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.
- We have described a popular "BYM2" smoothing model in the context of Poisson data – the relative risk was modeled as

  $$\eta_i = \beta_0 + \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + \delta_i,$$

  with random effects δ_i that have independent and spatial components.
- This same η_i can be used with other types of data, including normal and binomial – in these cases η_i is on the linear and logistic scales.

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 averages (e.g., the SMRs) and map them. Also look at map of standard errors and/or confidence intervals.
- Fit non-spatial random effects models.
- Fit the BYM2 spatial model.
- Covariates can be added, if available.

References

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*, **43**, 1–59.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley and Sons.

Elliott, P., Wakefield, J. C., Best, N. G., and Briggs, D. J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press, Oxford.

Kemp, I., Boyle, P., Smans, M., and Muir, C. (1985). *Atlas of Cancer in Scotland, 1975–1980: Incidence and Epidemiologic Persepctive*. IARC Scientific Publication 72, Lyon, France, International Agency for Research on Cancer.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Ratcliffe, J. (2010). Crime mapping: spatial and temporal challenges. In *Handbook of Quantitative Criminology*, pages 5–24. Springer.

Riebler, A., Sørbye, S., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, **25**, 1145–1165.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.

Simpson, D., Rue, H., Riebler, A., Martins, T., and Sørbye, S. (2017). Penalising model component complexity: A principled, practical approach to constructing priors (with discussion). *Statistical Science*, **32**, 1–28.

Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health*, **29**, 75–90.

Wakefield, J. C., Best, N. G., and Waller, L. A. (2000). Bayesian approaches to disease mapping. In P. Elliott, J. C. Wakefield, N. G. Best, and D. Briggs, editors, *Spatial Epidemiology: Methods and Applications*, pages 104–27. Oxford University Press, Oxford.

Wang, X., Yue, Y., and Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC.

Appendix: Parameter Interpretation

Parameter interpretation

We consider the model used for count data $Y_1, \dots, Y_n$ from a rare outcome, with mean $\mathrm{E}[Y_i] = E_i\theta_i$, where θ_i is the relative risk (RR) associated with area i.

It is important to emphasize that the RRs are associated with areas (which constitute the level of the analysis), the RRs should not be interpreted as the RRs associated with the individual within the areas.

Recall the “relative” is with respect to the reference risks that were used to calculate the expected numbers.

In the following we state the model on the loglinear scale (which makes sense statistically/computationally), and on the natural scale (which is more interpretable).

The RR is not the empirical RR (this is the SMR, i.e., Y_i/E_i) but the super-population RR — i.e., the RR associated with a hypothetical infinite population living in area i.

Parameter interpretation

Constant Relative Risk Model:

$$\mathrm{E}[Y_i] = E_i e^{\beta_0} = E_i \theta_0, \qquad i = 1, \ldots, n$$

- RRs are the same in all areas.
- θ_0 is the relative risk in each area.
- To obtain a credible interval for the relative risk in each we would need to account for parameter uncertainty.
- Under this (usually unrealistic) model, we obtain the posterior distribution $p(\theta_0|\boldsymbol{y})$, from which we can obtain the usual summaries – posterior mean, median, standard deviation, variance, quantiles (to obtain credible intervals).
- The posterior does not correspond to uncertainty about the empirical RR, but for the superpopulation RR θ_0.

Parameter interpretation

Saturated Model:

$$\mathrm{E}[Y_i] = E_i e^{\beta_i} = E_i \theta_i, \qquad i = 1, \ldots, n$$

- RRs are unique in each area with no modeled similarity.
- θ_i is the relative risk associated with area *i*.

Parameter interpretation

Covariate Model:

$$\mathrm{E}[Y_i] = E_i e^{\beta_0+\beta_1 x_i} = E_i \theta_0 \theta_1^{x_i}, \qquad i = 1, \ldots, n$$

- RRs differ across areas, due to the association with a covariate x_i.
- $\theta_0 = \exp(\beta_0)$ is the RR of an area with $x = 0$.
- $\theta_1 = \exp(\beta_1)$ is the multiplicative change in the areal RR associated with a 1-unit increase in x.
- The only between-area variability in the areal RR is due to the covariate.

Parameter interpretation

Random Effects Model:

$$\mathrm{E}[Y_i] = E_i \mathrm{e}^{\beta_0 + e_i} = E_i \theta_0 \delta_i, \qquad i = 1, \dots, n.$$

- RRs are assumed to be “tied together” via the random effects $e_i \sim \mathrm{N}(0, \sigma^2)$ – this is equivalent to $\delta_i \sim \mathrm{LogNormal}(0, \sigma^2)$.
- The “overall level” θ_0 is the RR associated with a “typical” area, i.e., an area with $\delta = 1$.
- The distribution of the RRs is $\mathrm{LogNormal}(\beta_0, \sigma^2)$. An alternative interpretation is that, across the study region, θ_0 is the median of the distribution of the RRs (the mean of the distribution of RRs is $\theta_0 \exp(\sigma^2/2)$).
- The $\mathrm{LogNormal}(0, \sigma^2)$ is the distribution of residual RRs across the map/study region. Residual here is relative to the overall level θ_0.
- The δ_i are the area-specifc adjustments to the overall level of the RR (which is θ_0).
- σ is the standard deviation of the log residual RRs.
- $\theta_0 \exp(\pm 1.96 \times \sigma)$ is an interval within which 95% of the area RRs lie.
- $\exp(\pm 1.96 \times \sigma)$ is an interval within which 95% of the area residual RRs lie

Parameter interpretation

Covariate and Random Effects Model:

$$\mathrm{E}[Y_i] = E_i e^{\beta_0+\beta_1 x_i+e_i} = E_i \theta_0 \theta_1^{x_i} \delta_i, \qquad i = 1, \ldots, n.$$

- RRs are assumed to be "tied together" across areas via the random effects $e_i \sim \mathrm{N}(0, \sigma^2)$.
- This is equivalent to $\delta_i \sim \mathrm{LogNormal}(0, \sigma^2)$ – these are the area-specifc adjustments to the covariate-adjusted level.
- RRs differ across areas, due to the association with a covariate x_i, and other independent adjustments δ_i.
- The $\mathrm{LogNormal}(0, \sigma^2)$ is the distribution of residual RRs across the map/study region. Residual here is relative to the covariate adjusted level $\theta_0 \theta_1^{x_i}$
- θ_0 is the RR of an area with $x = 0$ and $\delta = 1$.
- θ_1 is the multiplicative change in the areal RR associated with a 1-unit increase in x, for two areas with the same δ (or in a typical area with $\delta = 1$).

Parameter interpretation

BYM Random Effects Model:

$$\mathrm{E}[Y_i] = E_i \mathrm{e}^{\beta_0 + e_i + S_i} = E_i \theta_0 \delta_i \eta_i, \qquad i = 1, \dots, n.$$

- RRs are assumed to be “tied together” via global and local contributions, δ_i and η_i, respectively.
- The global (often referred to as unstructured) random effects have distribution $e_i \sim \mathrm{N}(0, \sigma^2)$ – which is equivalent to $\delta_i \sim \mathrm{LogNormal}(0, \sigma^2)$.
- The local (often referred to as structured) random effects have distribution $S_i | S_j, j \in \mathrm{ne}(i) \sim \mathrm{ICAR}(\sigma_s^2)$ – which is equivalent to $\eta_i | \eta_j, j \in \mathrm{ne}(i) \sim \mathrm{LogNormal}(0, \sigma_s^2)$.
- The “overall level” θ_0 is the RR associated with a “typical” area, i.e., an area with $\delta = \eta = 1$.
- The distribution of the RRs (and the residual RRs) is undefined, since the ICAR model is not proper.

Parameter interpretation

BYM Random Effects Model:

$$\mathrm{E}[Y_i] = E_i e^{\beta_0 + e_i + S_i} = E_i \theta_0 \delta_i \eta_i, \qquad i = 1, \dots, n.$$

- The LogNormal$(0, \sigma^2)$ is the distribution of the unstructured contribution to the residual RRs across the map/study region.
- The δ_i are the area-specifc unstructured adjustments to the overall level of the RR (which is θ_0), and η_i are the area-specifc structured adjustments to the overall level of the RR (which is θ_0).
- The σ_s parameter is very difficult to interpret.
- Under the BYM2 reparameterized version of BYM, we have

  $$\boldsymbol{b} = \sigma_b^2[\sqrt{1-\phi}\boldsymbol{I} + \sqrt{\phi}\boldsymbol{Q}^{\star}],$$

  and so work with the total variance (on the log scale) σ_b^2 and the proportion of the variability that is spatial ϕ.
- σ is the standard deviation of the log residual RRs.
- $\theta_0 \exp(\pm 1.96 \times \sigma_b)$ is an interval within which 95% of the area RRs lie.
- $\exp(\pm 1.96 \times \sigma_b)$ is an interval within which 95% of the areas residual RRs lie.