Bayesian Spatial Modeling

Lin Zhang

Department of Biostatistics
School of Public Health
University of Minnesota

Overview

- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:
  - geographically referenced, and often presented as maps, and
  - highly multivariate, with many important predictors and response variables, and
  - temporally correlated, as in longitudinal or other time series structures.

$\Rightarrow$ motivates hierarchical modeling and data analysis for complex spatial (and spatiotemporal) data sets.

Outline

- Types of Spatial Data
- Bayesian Spatial Modeling of Point-Referenced Data
- Bayesian Spatial Modeling of Areal Data
- Areal vs point-level models

Types of Spatial Data

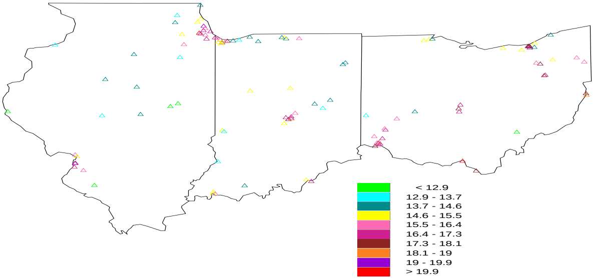
Consider a random variable $Y(s)$ observed in a spatial domain D. We can classify spatial data into three basic types, depending on the nature of the set D:

- *Point-reference data*, where D is a fixed subset of $\mathbb{R}^r$, and s varies continuously over D (geostatistical data)
  - Examples: climate data; air pollution data
  - Interest often lies in infering the entire spatial process and predicting outcomes at new locations.

- *Areal data*, where D is again fixed but partitioned into a finite number of areal units with well-defined boundaries (lattice data)
  - Examples: county-level or state-level epidemiological data
  - Interests often lies in smoothing.

- *Point pattern data*, where D is itself random
  - Examples: residence of persons suffering from a particular disease; locations of a certain species of tree in a forest
  - Interests often lies in clustering.

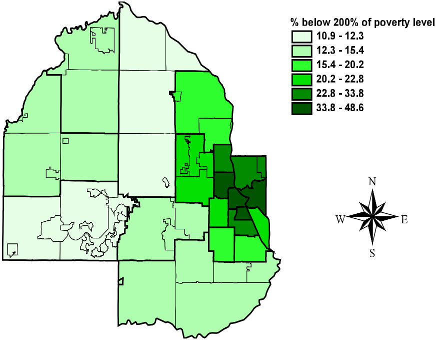
Example 1: Map of PM2.5 sampling sites



plotting color indicates range of average monitored PM2.5 level over the year 2001

Question of interest: How do the PM2.5 levels relate to regional industrialization, traffic density, and other covariates?

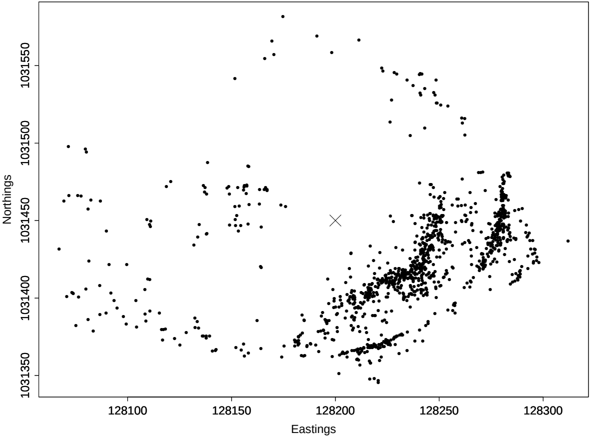
Example 2: Poverty map



`ArcView` map of percent of surveyed population with household income below 200% of the federal poverty limit for regional survey units in Hennepin County, MN

Question of interest: The geographic distribution of low-income family in Hennepin County after accounting for spatial correlation?

Example 3: Locations of Shrub in a forest



Question of interest: Is there a clustering pattern of the spatial distribution?

Bayesian regression of point-level data

- Basic Model:

$$Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s})$$

  The residual is partitioned into two pieces:

  - spatial error $w(\mathbf{s})$: a stationary Gaussian process, introducing the partial sill (σ^2) and range (ϕ) parameters.
  - non-spatial error $\epsilon(\mathbf{s})$: adding the nugget (τ^2) effect.

- Interpretations attached to $\epsilon(\mathbf{s})$:
  - pure error term: model is not perfectly spatial; σ^2 and τ^2 are spatial and nonspatial variance components
  - measurement error or replication variability, causing spatial discontinuity in $Y(\mathbf{s})$
  - microscale variability, present at distances smaller than the smallest inter-location distance

Stationarity and Isotropy

- Stationarity:
  - A spatial process is said to be strong stationary if for any set $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and any vector $\mathbf{h}$, the distribution of $(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))$ is the same as that of $(Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h}))$.
    $\Rightarrow$ *invariant joint distribution after spatial shift!*
  - A spatial process is called weak stationarity: if $\mu(\mathbf{s}) \equiv \mu$ and $Cov(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $\mathbf{h} \in \Re^r$.
    $\Rightarrow$ *invariant moments (mean & covariance) after spatial shift!*

- Isotropy:
  - Under weak stationarity, the covariance between any two locations can be summarized by a covariance function $C(\mathbf{h})$, which depends only on the separation vector $\mathbf{h}$.
  - If $C(\mathbf{h})$ depends upon the separation vector only through its length $||\mathbf{h}||$, i.e. $C(\mathbf{h})\,||\mathbf{h}||$ then we say that the process is *isotropic*.

Some common isotropic covariograms

Model	Covariance function, $C(d)$ $\quad (d = \|\|h\|\|)$
Spherical	$C(d) = \begin{cases} 0 & \text{if } d \geq 1/\phi \\ \sigma^2 \left[1 - \frac{3}{2}\phi d + \frac{1}{2}(\phi d)^3\right] & \text{if } 0 < d \leq 1/\phi \end{cases}$
Exponential	$C(d) = \begin{cases} \sigma^2 \exp(-\phi d) & \text{if } d > 0 \end{cases}$
Powered Exp	$C(d) = \begin{cases} \sigma^2 \exp(-\|\phi d\|^p) & \text{if } d > 0 \end{cases}$
Gaussian	$C(d) = \begin{cases} \sigma^2 \exp(-\phi^2 d^2) & \text{if } d > 0 \end{cases}$
Matérn	$C(d) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu} d\phi)^\nu K_\nu(2\sqrt{\nu} d\phi) & \text{if } d > 0 \end{cases}$

- Two parameters: spatial variance (partial sill) σ^2, and decay parameter ϕ.
- $\rho(d) = C(d)/\sigma^2$ is often called the correlation function.

Isotropic spatial models

- Suppose for the two error terms:

$$\begin{aligned} \mathbf{w} = [w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n)]^T &\sim N(\mathbf{0}, \sigma^2 H(\phi)) \\ \boldsymbol{\epsilon} = [\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n)]^T &\sim N(0, \tau^2 I) \end{aligned}$$

- $H_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \phi)$, where ρ is a valid (and typically isotropic) correlation function.

- Combining the two errors, we have the likelihood:

$$\mathbf{Y}|\boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 I),$$

  where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \phi)^T$.

- Gaussian kriging models are special cases of the general linear model, with a particular specification of the covariance matrix

$$\Sigma = \sigma^2 H(\phi) + \tau^2 I.$$

Prior specification

- In Bayesian framework, we require a prior $\pi(\boldsymbol{\theta})$, so the posterior is:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

- Typically, independent priors are chosen for the parameters:

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi)$$

  Useful candidates are multivariate normal for $\boldsymbol{\beta}$, and inverse gamma for σ^2 and τ^2. Specification of $\pi(\phi)$ depends upon choice of ρ function; a uniform or gamma prior is usually selected.

- Informativeness: $\pi(\boldsymbol{\beta})$ can be "flat" (improper), but ϕ and *at least* one of σ^2 and τ^2 require informative priors.

Hierarchical modeling

- We can recast the foregoing in a hierarchical setup by considering a conditional likelihood on the spatial random effects $\mathbf{w} = (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))$.
- First stage:
$$\mathbf{Y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim N(X\boldsymbol{\beta} + \mathbf{w}, \tau^2 I)$$
The $Y(\mathbf{s}_i)$ are *conditionally independent* given the $w(\mathbf{s}_i)$'s.
- Second stage:
$$\mathbf{w}|\sigma^2, \phi \sim N(\mathbf{0}, \sigma^2 H(\phi))$$
- Third stage:
$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\phi)$$

Marginal or Conditional Model?

- Choice: Fit the marginal model as $f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ or the conditional model as $f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{w})p(\mathbf{w}|\boldsymbol{\theta})p(\boldsymbol{\theta})$.
- Note that the posterior $p(\boldsymbol{\theta}|\mathbf{y})$ is the same for the original and hierarchical settings.
- Fitting the marginal model is computationally more stable:
  - lower dimensional sampler (no $\mathbf{w}$'s)
  - $\sigma^2 H(\phi) + \tau^2 I$ more stable than $\sigma^2 H(\phi)$
- BUT the conditional model allows conjugate full conditionals for σ^2, τ^2 (inverse gamma), $\boldsymbol{\beta}$, and $\mathbf{w}$ (Gaussian) – easy updates!
- Marginalized model will need Metropolis or slice sampling updates for σ^2, τ^2, and ϕ. But these usually work well and often converge faster than the full Gibbs updates.

Recover the spatial effects w's

- Interest often lies in the spatial surface $\mathbf{w}|\mathbf{y}$.
- Have we lost the $\mathbf{w}$'s with the marginalized sampling?
- No. Note that

$$p(\mathbf{w}|\mathbf{y}) = \int p(\mathbf{w}|\boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$$

  and,

$$p(\mathbf{w}|\boldsymbol{\theta}, \mathbf{y}) \propto f(\mathbf{y}|\mathbf{w}, \boldsymbol{\beta}, \tau^2) p(\mathbf{w}|\sigma^2, \phi)$$

  is a multivariate normal distribution.
- Thus we can generate posterior samples of $\mathbf{w}$ by
  - First generate $\boldsymbol{\theta}^{(g)} \sim p(\boldsymbol{\theta}|\mathbf{y})$ for $g = 1, \ldots, G$
  - Second generate $\mathbf{w}^{(g)} \sim p(\mathbf{w}|\boldsymbol{\theta}^{(g)}, \mathbf{y})$ given each $\boldsymbol{\theta}^{(g)}$

Spatial prediction (Bayesian kriging)

- Often we need to predict the response Y at a new site $\mathbf{s}_0$ with associated covariates $\mathbf{x}_0 \equiv \mathbf{x}(\mathbf{s}_0)$.
- It requires the predictive distribution:

$$\begin{aligned} p(y_0|\mathbf{y}, X, \mathbf{x}_0) &= \int p(y_0, \boldsymbol{\theta}|\mathbf{y}, X, \mathbf{x}_0) d\boldsymbol{\theta} \\ &= \int p(y_0|\mathbf{y}, \boldsymbol{\theta}, X, \mathbf{x}_0) p(\boldsymbol{\theta}|\mathbf{y}, X) d\boldsymbol{\theta} \end{aligned}$$

  Note that $p(y_0|\mathbf{y}, \boldsymbol{\theta}, X, \mathbf{x}_0)$ is normal since $p(y_0, \mathbf{y}|\boldsymbol{\theta}, X, \mathbf{x}_0)$ is multivariate normal!

$\Rightarrow$ same algorithm in previous slide:

- First, generate $\boldsymbol{\theta}^{(g)} \sim p(\boldsymbol{\theta}|\mathbf{y})$ for $g = 1, \ldots, G$
- Second, draw $Y_0^{(g)}$ from $p(y_0|\mathbf{y}, \boldsymbol{\theta}^{(g)}, X, \mathbf{x}_0)$ given each $\boldsymbol{\theta}^{(g)}$

Spatial Generalized Linear Models

- Some data sets preclude Gaussian modeling; $Y(\mathbf{s})$ may not even be continuous!
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - precipitation or deposition was measurable or not
  - price is high or low for home at location $\mathbf{s}$
  - number of insurance claims by residents of a single family home at $\mathbf{s}$
- In Bayesian framework, it's easy to extend spatial regression to GLM $\Rightarrow$ replace Gaussian likelihood by an appropriate exponential family member,
  - Binomial likelihood for binary data
  - Poisson (or over-dispersed Poisson) for count data

  and then introduce the spatial process in modeling the transformed mean response (with some link function).

Spatial GLM (cont'd)

Say, we have binary outcomes $Y(\mathbf{s}_i)$, $i = 1, \ldots, n$:

- First stage: we assume $Y(\mathbf{s}_i) \sim Bern(p(\mathbf{s}_i))$, with

$$logit(p(\mathbf{s}_i)) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i) \ .$$

- Second stage: Model $w(\mathbf{s})$ as a Gaussian process:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 H(\phi))$$

- Third stage: Priors and hyperpriors
- Note: Usually no pure error term $\epsilon(\mathbf{s})$, but possible computational advantage.

Bayesian spatial model in WinBUGS

- Data: Observations are home values (based on recent real estate sales) at 50 locations in Baton Rouge, Louisiana, USA.
- The response $Y(\mathbf{s})$ is a binary variable, with

$$Y(\mathbf{s}) = \begin{cases} 1 & \text{if price is ``high'' (above the median)} \\ 0 & \text{if price is ``low'' (below the median)} \end{cases}$$

- Observed covariates include the house's age and total living area.
- Model: We fit a generlaized linear model

$$\begin{aligned} Y(\mathbf{s}_i) &\sim Bernoulli(p(\mathbf{s}_i)), \\ logit(p(\mathbf{s}_i)) &= x^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i). \end{aligned}$$

Non-Gaussian point-level model in WinBUGS

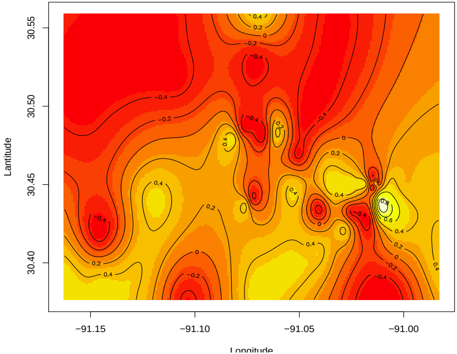
- Code for spatial GLM (logistic) model

```
model{
    for (i in 1:N) {
       Y[i] ~ dbern(p[i])
       logit(p[i]) <- w[i]
       mu[i] <- beta[1]+beta[2]*LivingArea[i]/1000+beta[3]*Age[i]
    }

    for (i in 1:3) {beta[i] ~ dnorm(0.0,0.001)}
    w[1:N] ~ spatial.exp(mu[], x[], y[], spat.prec, phi, 1)
    phi ~ dunif(0.1,10)

    spat.prec ~ dgamma(0.1, 0.1)
    sigmasq <- 1/spat.prec
}
```

- Summaries of posterior distributions

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta[1]	-1.19	1.01	0.08325	-3.857	-0.9445	0.1727	1001	5000
beta[2]	0.6922	0.5378	0.04714	0.02099	0.5597	2.123	1001	5000
beta[3]	-0.002336	0.02242	5.901E-4	-0.053	-8.878E-4	0.0382	1001	5000
phi	5.656	2.507	0.04839	1.281	5.637	9.756	1001	5000
sigmasq	1.663	1.678	0.1267	0.1903	1.149	6.384	1001	5000

Estimated Spatial Effects



- Export the posterior samples of the spatial effects w_i into .txt file.
- Read the .txt file into R, and plot image on w_i posterior medians.
- negative residuals (i.e. lower prices) in the north; positive residuals (i.e. higher prices) in the south

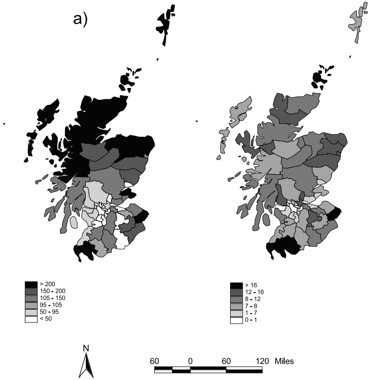
Spatial GLM: comments

- Our use of spatial random effects in the (transformed) mean encourages the means of spatial variables at proximate locations to be similar to each other
- But the observed $Y(\mathbf{s})$ and $Y(\mathbf{s}')$ need not be similar.
- Thus second stage spatial modeling is attractive for spatial explanation in the *mean*, while direct (first stage) spatial modeling is better for encouraging proximate observations to be similar.
- For computation, unlike the Gaussian case, an MCMC algorithm will have to update $\mathbf{w}$ as well as $\boldsymbol{\beta}, \sigma^2, \phi$, and γ.

Spatial modeling of areal data

- Suppose we have observations $\mathbf{y} = (y_1, \ldots, y_n)$ observed from areal units $i = 1, \ldots, n$ in a region D of interest.
- Key interest: To obtain a joint distributional model of $\{Y_i, i = 1, \ldots, n\}$.
- Difficulty: When the number of areal units is very large (say, a fine-resolution image or a large geographic region with small units), it's hard to write down the joint distribution.
- Solutions: Works with full conditionals of the Y_i's, which will yield a valid joint distribution!

Example: Scottish lip cancer data



Panel (a): Standardized mortality rate, $SMR_i = 100Y_i/E_i$ for lip cancer in $n = 56$ districts, 1975-1980.

Panel (b): One covariate, x_i = percentage of the population engaged in agriculture, fishing or forestry (AFF)

Lip Cancer Example – Likelihood

- Consider the Poisson model for disease count in each district:

$$
\begin{aligned}
Y_i|\eta_i &\overset{ind}{\sim} Pois(E_i \eta_i)\ , \text{ where} \\
Y_i &= \text{observed disease count from district } i \\
E_i &= \text{expected count (known) from district } i \\
\eta_i &: \text{relative risk of the disease in district } i.
\end{aligned}
$$

- We model log-relative risk as

$$\psi_i = log(\eta_i) = \beta_0 + x_i'\beta + \phi_i + \theta_i,$$

where x_i is the explanatory spatial covariate, AFF, with coefficient β.

- Note the mean structure also contains two sets of random effects!
  - ϕ_i: spatial clustering random effects
  - θ_i: unstructural heterogeneity random effects

- Interest: find the estimated posterior of the AFF effect β_1, and map the fitted relative risks $E(e_i^{\psi}|\mathbf{y})$.

Conditional autoregressive (CAR) model

- Assuming Y_i's are Gaussian-distributed, the conditional autoregressive (CAR) model specifies the full conditional distribution for each Y_i

$$p(y_i | y_j, j \neq i) = N\left(\sum_{j \sim i} b_{ij} y_j, \tau_i^2\right)$$

  where $j \sim i$ indicates that units i and j are neighbors.

- The coefficients b_{ij} introduces spatial smoothing between neighboring pairs, and τ_i^2 gives the conditional variance of unit i.

- Question: Can these full conditionals specification lead to a valid joint distribution of $\mathbf{Y}$?

- Answer: Yes!

CAR model (cont'd)

- The Brook's Lemma ensures that we can retrieve the joint distribution based on the given the full conditionals specified by CAR model.
- Based on Brook's Lemma, we can derive the joint distribution as

$$p(y_1, y_2, ...y_n) \propto \exp\left\{-\frac{1}{2}\mathbf{y}'D^{-1}(I - B)\mathbf{y}\right\},$$

  where $B = \{b_{ij}\}$ and D is diagonal with $D_{ii} = \tau_i^2$.
- This suggests a multivariate normal distribution of $\mathbf{Y}$ with

$$\mu_y = 0 \text{ and } \Sigma_Y = (I - B)^{-1}D$$

- To ensure the validity of the joint distribution, the covariance matrix must be symmetric and positive definite!

CAR model (cont'd)

- $D^{-1}(I - B)$ being symmetric requires $\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}$ for all i, j.
- Let W be some user-defined proximity matrix, which is a $p \times p$ matrix with off-diagonal entries w_{ij} indicating spatial connection between i and j units, and diagonal entries $w_{ii} = 0$.
- Assuming W is symmetric, choose $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, then

$$p(y_1, y_2, \ldots y_n) \propto \exp\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\},$$

  where D_w is diagonal with $(D_w)_{ii} = w_{i+}$. This is called Intrinsic autoregressive (IAR) model!
- The joint distribution is equivalent to

$$p(y_1, y_2, \ldots y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{i\sim j} w_{ij}(y_i - y_j)^2\right\},$$

  a function only on the pairwise difference between neighboring units.

Proximity matrices

- w_{ij} can be viewed as 'weights' that determines the spatial connection between i and j.
  - Larger w_{ij} if j is closer to i.

- Choices for w_{ij}:
  - $w_{ij} = 1$ if i, j share a common boundary (possibly a common vertex)
  - w_{ij} is an *inverse* distance between units
  - $w_{ij} = 1$ if distance between units is $\leq K$
  - $w_{ij} = 1$ for m nearest neighbors.

  W is typically symmetric, but need not be.

- $\widetilde{W}$: standardized entries $\tilde{w}_{ij} = w_{ij}/w_{i+}$, where i by $w_{i+} = \sum_j$
  - $\widetilde{W}$ is row stochastic, i.e. $\widetilde{W}\mathbf{1} = 1$, but need not be symmetric.

CAR model (cont'd)

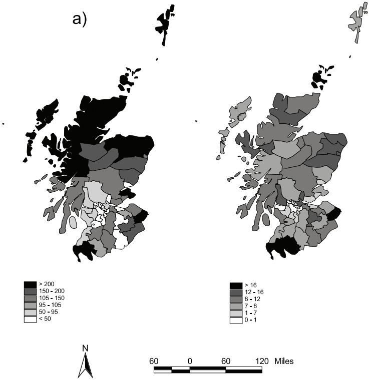
- Question: Is the IAR model a valid joint distribution of $\mathbf{Y}$?

$$p(y_1, y_2, ...y_n) \propto \exp\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\},$$

- No! The joint distribution is improper since $(D_w - W)\mathbf{1} = \mathbf{0}$;
  $\Rightarrow$ the precision matrix of the distribution is singular!

- The IAR model is NOT a valid likelihood distribution, but can be used as prior for the random effects in a regression model!

Example: Scottish lip cancer data



Panel (a): Standardized mortality rate, $SMR_i = 100Y_i/E_i$ for lip cancer in $n = 56$ districts, 1975-1980.

Panel (b): One covariate, x_i = percentage of the population engaged in agriculture, fishing or forestry (AFF)

Lip Cancer Example – Likelihood

- Consider the Poisson model for disease count in each district:

$$\begin{aligned} Y_i|\eta_i &\overset{ind}{\sim} Pois(E_i\eta_i)\ ,\ \text{where} \\ Y_i &= \text{observed disease count from district } i \\ E_i &= \text{expected count (known) from district } i \\ \eta_i &: \text{relative risk of the disease in district } i. \end{aligned}$$

- We model log-relative risk as

$$\psi_i = log(\eta_i) = \beta_0 + x_i'\beta + \phi_i + \theta_i,$$

  where x_i is the explanatory spatial covariate, AFF, with coefficient β.

- Note the mean structure also contains two sets of random effects!
  - ϕ_i: spatial clustering random effects
  - θ_i: unstructural heterogeneity random effects

- Interest: find the estimated posterior of the AFF effect β_1, and map the fitted relative risks $E(e_i^{\psi}|\mathbf{y})$.

Priors

In Bayesian framework, we need to assign priors for $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\phi})$.

- We assume a flat prior for the fixed effect $\boldsymbol{\beta} = (\beta_0, \beta_1)$
- θ_i captures unstrunctural heterogeneity via the prior

$$\theta_i \overset{ind}{\sim} N(0, \tau_h^{-1}).$$

- ϕ_i captures spatial clustering via the CAR prior,

$$\phi_i | \phi_{j \neq i} \sim N\left(\frac{1}{m_i} \sum_{j \sim i} \phi_j, (m_i \tau_c)^{-1}\right),$$

  where m_i is the number of "neighbors" of region i.

- The CAR prior is translation invariant, so requires a constraint by centering $\sum_{i=1}^{n} \phi_i = 0$ (imposed numerically after each MCMC iteration).

Identifiability

- Note Y_i cannot inform about θ_i or ϕ_i, but only about their sum $\epsilon_i = \theta_i + \phi_i$.
- Though unidentified, the θ_i and ϕ are interesting in their own right, as is

$$\alpha = \frac{sd(\boldsymbol{\phi})}{sd(\boldsymbol{\theta}) + sd(\boldsymbol{\phi})},$$

  where $sd(\cdot)$ is the empirical marginal standard deviation function.
- We want to specify values τ_h and τ_c such that the priors
  - are proper that leads to acceptable convergence behavior, and
  - yet vague that still allows Bayesian learning.
- In addition, we also want a "fair" prior that balances between the unstructured heterogeneity and the spatial clustering (e.g. leads to $\alpha \approx 1/2$).

WinBUGS code to fit this model

The following WinBUGS code uses vague priors for τ_h and τ_c as suggested by Best et al. (1999).

```
model{
   for (i in 1 : regions) {
      O[i] ~ dpois(mu[i])
      log(mu[i]) <- log(E[i]) + beta0 + beta1*aff[i]/10 + phi[i] + theta[i]
      theta[i] ~ dnorm(0.0,tau.h)
      psi[i] <- theta[i] + phi[i]
      SMRhat[i] <- 100 * mu[i] / E[i]
      SMRraw[i] <- 100* O[i] / E[i]
   }
   phi[1:regions] ~ car.normal(adj[], weights[], num[], tau.c)

   beta0 ~ dnorm(0.0, 1.0E-5)  # vague prior on grand intercept
   beta1 ~ dnorm(0.0, 1.0E-5)  # vague prior on covariate effect

   tau.h ~ dgamma(1.0E-3,1.0E-3)  #  ``fair'' prior from Best et al.
   tau.c ~ dgamma(1.0E-1,1.0E-1)  #   (1999, Bayesian Statistics 6)

   sd.h <- sd(theta[]) # marginal SD of heterogeneity effects
   sd.c <- sd(phi[])   # marginal SD of clustering (spatial) effects
   alpha <- sd.c / (sd.h + sd.c)
}
```

Lip Cancer Results

priors for τ_c, τ_h	posterior for α mean	sd	l1 acf	posterior for β_1 mean	sd	l1 acf
$G(1.0, 1.0)$, $G(3.2761, 1.81)$	0.57	0.056	0.80	0.43	0.16	0.94
$G(.1, .1)$, $G(.32761, .181)$	0.65	0.071	0.89	0.40	0.14	0.93
$G(.1, .1)$, $G(.001, .001)$	0.83	0.099	0.97	0.37	0.13	0.92

priors for τ_c, τ_h	posterior for ϵ_1 mean	sd	l1 acf	posterior for ϵ_{56} mean	sd	l1 acf
$G(1.0, 1.0)$, $G(3.2761, 1.81)$	1.23	0.34	0.05	-0.68	0.50	0.02
$G(.1, .1)$, $G(.32761, .181)$	1.16	0.31	0.08	-0.54	0.39	0.09
$G(.1, .1)$, $G(.001, .001)$	1.15	0.30	0.12	-0.48	0.33	0.11

- AFF covariate is significantly $\neq 0$ under all 3 priors
- Convergence is slow for α, but rapid for ϵ_i.
- Excess variability in the data is mostly due to clustering ($E(\psi|\mathbf{y}) > .50$), but the posterior distribution for α is not robust to changes in the prior.

Proper Car Model

- "Proper CAR": replace $(D_w - W)$ by $(D_w - \rho W)$, such that

$$p(y_1, y_2, ...y_n) \propto \exp\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - \rho W)\mathbf{y}\},$$

  and choose ρ such that $\Sigma_y = \tau^2(D_w - \rho W)^{-1}$ exists.

- This in turn implies the full conditional

$$p(y_i|y_j, j \neq i) = N\left(\rho \sum_{j \neq i} \frac{w_{ij}}{w_{i+}} y_j, \frac{\tau^2}{w_{i+}}\right).$$

  $\Rightarrow$ Using proper CAR as a prior, ρ determines the spatial smoothing (shrinkage) of y_i toward the average of its neighbors!

- Choice of ρ: Σ_y exists if $\rho \in \left(1/\lambda_{(1)}, 1\right)$, where $\lambda_{(1)}$ is the smallest eigenvalue of $D_w^{-1/2} W D_w^{-1/2}$. In practice, the bound $\rho \in (0, 1)$ is often preferred.

To ρ or not to ρ?

- Advantages:
  - makes distribution proper
  - adds parametric flexibility
  - $\rho = 0$ interpretable as independence
- However, does ρ give sensible interpretation of "strength of spatial association"?
  - calibration of ρ as a correlation, e.g.,

$$\begin{aligned} \rho &= 0.80 \quad \text{yields } 0.1 \leq I \leq 0.15, \\ \rho &= 0.90 \quad \text{yields } 0.2 \leq I \leq 0.25, \\ \rho &= 0.99 \quad \text{yields } I \leq 0.5 \end{aligned}$$

  - So, used with random effects, scope of spatial pattern limited.
  - In a Bayesian framework, a prior on ρ that encourages a consequential amount of spatial association would place most of its mass near 1.

Comments on CAR models

- We specify Σ_y^{-1}, not Σ_Y, so does not directly model association.
- If $(\Sigma_y^{-1})_{ii} = 1/\tau_i^2$; $(\Sigma_y^{-1})_{ij} = 0 \Leftrightarrow$ conditional independence.
- Prediction at new sites is ad hoc, in that if

$$p(y_0|y_1, y_2, ...y_n) = N\left(\sum_j \frac{w_{0j}}{w_{0+}} y_j, \frac{\tau^2}{w_{0+}}\right)$$

  then $p(y_0, y_1, ...y_n)$ well-defined but not equivalent to the CAR model arising from the full conditionals of $Y_0, Y_1, \ldots, Y_n$.
- Non-Gaussian case: for binary data, the autologistic model

$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \psi \sum_{j \sim i} w_{ij} y_i y_j.$$

Comparing point-referenced and areal data models

- Spatial process vs. a single *n*-dimensional distribution
- Gaussian process vs. CAR model
- Modeling $\Sigma_{\mathbf{Y}}$ (Gaussian process) vs. $\Sigma_{\mathbf{Y}}^{-1}$ (CAR)
- Prediction vs. smoothing
- "big *N* problem" (with matrix inversion for likelihood evaluation) vs. NO "big *N* problem" (readily available full conditionals for Gibbs sampling)