Chapter 4.2 - Lattice processes
Lecture notes

Odd Kolbjørnsen Geir Storvik

March 6, 2017

Previously

- Lattice models
- MRF – Markov random field
- Neighborhood (undirected graph)
- Clique
- Negpotential function
- Besag's lemma (conditional vs joint distribution)
- Hammersley- Clifford theorem
- Auto spatial models
  - CAR model
  - Latent Gaussian process
  - Auto logistic model (Ising model)
  - auto Poisson model

Today

- Gibbs distribution
- How to construct a CAR from scratch
- Examples of Models in INLA (disease mapping, etc)
- Examples of models outside INLA (Potts model)

Gibbs distribution/ Gibbs measure

From statistical mechanics:

$$\Pr(\mathbf{y}) = \frac{1}{Z(\beta)} \exp\{-\beta E(y)\}$$

with $\beta = \frac{1}{T}$ inverse temperature; $E(y)$ potential function.
$Z(\beta)$ partition function

Recall: Hammersley - Clifford theorem. On the graph (G)
Graph = Nodes (=Lattice) + Edges (=Neighbors)

$$\Pr(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{c(\boldsymbol{\theta})} \exp\{Q(\mathbf{y};\boldsymbol{\theta})\} = \frac{1}{c(\boldsymbol{\theta})} \exp\left\{\sum_{c \in C_G} \psi_c(y_c, \boldsymbol{\theta})\right\}$$

with C_G being the set of all cliques, $Q(\mathbf{0}, \boldsymbol{\theta}) = 0$, and $c(\boldsymbol{\theta})$ is a normalizing constant.

Gaussian CAR models

$$\mathbf{Y} \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}) = \text{MVN}(\mathbf{0}, \mathbf{Q}^{-1})$$

- Conditional Gaussian distributions with

$$E[Y(s_i)|\mathbf{Y}_{-i}] = \sum_{j \in \mathcal{N}(\mathbf{s}_i)} c_{ij} Y(\mathbf{s}_j)$$
$$\text{var}[Y(s_i)|\mathbf{Y}_{-i}] = \tau_i^2$$

- If legal, $\mathbf{Q} = \mathbf{M}^{-1}[\mathbf{I} - \mathbf{C}]$ with $\mathbf{M} = \text{diag}\{\tau_i^2\}$
- If $\mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ is symmetric and positive definite, then
- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ is sparse if $\mathbf{C}$ is sparse!
- $\mathbf{M}$ always positive definite and symmetric
- $j \in \mathcal{N}(\mathbf{s}_i) \Leftrightarrow i \in \mathcal{N}(\mathbf{s}_j)$ imply $c_{ij} = 0 \Leftrightarrow c_{ji} = 0$
- Need $\tau_i^{-2} c_{ij} = \tau_j^{-2} c_{ji}$ for neighbors
- Also, the c_{ij} can not be too large for getting positive definiteness.

Parameterizations of CAR models

$$\mathbf{Y} \sim \text{MVN}(\mathbf{0}, (\mathbf{M}^{-1}[\mathbf{I}-\mathbf{C}])^{-1}) = \text{MVN}(\mathbf{0}, [\mathbf{I}-\mathbf{C}]^{-1}\mathbf{M})$$

- $\mathbf{M} = \text{diag}\{\tau_i^2\}, \tau_i > 0 \quad \forall i$
- Often
  - $\mathbf{C} = \phi\mathbf{H}$, $\mathbf{H}$ known and/or
  - $\mathbf{M} = \tau^2\mathbf{\Delta}$, $\mathbf{\Delta}$ known diagonal matrix.
- What requirements on (ϕ, τ)?
- Note: Need $\mathbf{M}^{-1}\mathbf{C}$ symmetric:
  - $\mathbf{Q}_y = \mathbf{M}^{-1}[\mathbf{I}-\mathbf{C}] = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{C}$, $\mathbf{M}^{-1}$ automatically symmetric
  - General: $\tau_i^{-2}C_{ij} = \tau_j^{-2}C_{ji}$
  - $\mathbf{M} = \tau^2\mathbf{\Delta}, \mathbf{C} = \phi\mathbf{H} \Rightarrow \Delta_{ii}^{-1}h_{ij} = \Delta_{jj}^{-1}h_{ji}$
- Need also $\mathbf{\Sigma}_Y = (\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}$ positive definite
- Equivalent to $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I}-\mathbf{C})$ positive definite

CAR: Requirements for positive definite

- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ positive definite?

CAR: Requirements for positive definite

- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ positive definite?
- Since $\mathbf{M}^{1/2}$ is positive definite ($\tau_i > 0$),
- equivalent to that $\mathbf{M}^{-1/2}(\mathbf{I} - \mathbf{C})\mathbf{M}^{1/2} = \mathbf{I} - \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$ is positive definite

CAR: Requirements for positive definite

- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ positive definite?
- Since $\mathbf{M}^{1/2}$ is positive definite ($\tau_i > 0$),
- equivalent to that $\mathbf{M}^{-1/2}(\mathbf{I} - \mathbf{C})\mathbf{M}^{1/2} = \mathbf{I} - \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$ is positive definite
- $\mathbf{A}$ symmetric ($\mathbf{A} = \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$)
  - spectral decomposition: $\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T$ where $\mathbf{T}\mathbf{T}^T = \mathbf{I}$ and $\boldsymbol{\Lambda}$ contains eigenvalues.
  - $\mathbf{A}$ is positive definite if all eigenvalues are positive.

CAR: Requirements for positive definite

- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ positive definite?
- Since $\mathbf{M}^{1/2}$ is positive definite ($\tau_i > 0$),
- equivalent to that $\mathbf{M}^{-1/2}(\mathbf{I} - \mathbf{C})\mathbf{M}^{1/2} = \mathbf{I} - \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$ is positive definite
- $\mathbf{A}$ symmetric ($\mathbf{A} = \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$)
  - spectral decomposition: $\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T$ where $\mathbf{T}\mathbf{T}^T = \mathbf{I}$ and $\boldsymbol{\Lambda}$ contains eigenvalues.
  - $\mathbf{A}$ is positive definite if all eigenvalues are positive.
  - $\mathbf{I} - \mathbf{A} = \mathbf{I} - \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T = \mathbf{T}\mathbf{T}^T - \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T = \mathbf{T}[\mathbf{I} - \boldsymbol{\Lambda}]\mathbf{T}^T \Rightarrow \mathbf{I} - \boldsymbol{\Lambda}$ contain eigenvalues of $\mathbf{I} - \mathbf{A}$.

CAR: Requirements for positive definite

- $\mathbf{Q} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{C})$ positive definite?
- Since $\mathbf{M}^{1/2}$ is positive definite ($\tau_i > 0$),
- equivalent to that $\mathbf{M}^{-1/2}(\mathbf{I} - \mathbf{C})\mathbf{M}^{1/2} = \mathbf{I} - \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$ is positive definite
- $\mathbf{A}$ symmetric ($\mathbf{A} = \mathbf{M}^{1/2}\mathbf{C}\mathbf{M}^{-1/2}$)
  - spectral decomposition: $\mathbf{A} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T$ where $\mathbf{T}\mathbf{T}^T = \mathbf{I}$ and $\boldsymbol{\Lambda}$ contains eigenvalues.
  - $\mathbf{A}$ is positive definite if all eigenvalues are positive.
  - $\mathbf{I} - \mathbf{A} = \mathbf{I} - \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T = \mathbf{T}\mathbf{T}^T - \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}^T = \mathbf{T}[\mathbf{I} - \boldsymbol{\Lambda}]\mathbf{T}^T \Rightarrow \mathbf{I} - \boldsymbol{\Lambda}$ contain eigenvalues of $\mathbf{I} - \mathbf{A}$.
- If λ eigenvalue of $\mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2}$, then $1 - \lambda$ eigenvalue of $\mathbf{I} - \mathbf{M}^{-1/2}\mathbf{C}\mathbf{M}^{1/2}$
- Positive definite if $1 - \lambda > 0$, or $\lambda < 1$.

CAR: Requirements (cont)

- Assume $\mathbf{C} = \phi\mathbf{H}$, so $\mathbf{A} = \phi\mathbf{M}^{1/2}\mathbf{H}\mathbf{M}^{-1/2}$
- $\tilde{\lambda}$ eigenvalue of $\mathbf{M}^{1/2}\mathbf{H}\mathbf{M}^{-1/2}$ gives $\lambda = \phi\tilde{\lambda}$ eigenvalue of $\mathbf{A}$.
- Need $\lambda < 1$ or $\phi\tilde{\lambda} < 1$.
- For $\tilde{\lambda}$ positive, need $\phi < \lambda^{-1}$
- For $\tilde{\lambda}$ negative, need $\phi > \lambda^{-1}$
- Let $\tilde{\lambda}_{(1)} \leq \tilde{\lambda}_{(2)} \leq \cdots \leq \tilde{\lambda}_{(n)}$ be ordered eigenvalues
- We have $\tilde{\lambda}_{(1)} < 0$, $\tilde{\lambda}_{(n)} > 0$ (not obvious!)
- Requirement: $\tilde{\lambda}_{(1)}^{-1} < \phi < \tilde{\lambda}_{(n)}^{-1}$

CAR: Requirements (cont)

- Assume $\mathbf{C} = \phi\mathbf{H}$, so $\mathbf{A} = \phi\mathbf{M}^{1/2}\mathbf{H}\mathbf{M}^{-1/2}$
- $\tilde{\lambda}$ eigenvalue of $\mathbf{M}^{1/2}\mathbf{H}\mathbf{M}^{-1/2}$ gives $\lambda = \phi\tilde{\lambda}$ eigenvalue of $\mathbf{A}$.
- Need $\lambda < 1$ or $\phi\tilde{\lambda} < 1$.
- For $\tilde{\lambda}$ positive, need $\phi < \lambda^{-1}$
- For $\tilde{\lambda}$ negative, need $\phi > \lambda^{-1}$
- Let $\tilde{\lambda}_{(1)} \leq \tilde{\lambda}_{(2)} \leq \cdots \leq \tilde{\lambda}_{(n)}$ be ordered eigenvalues
- We have $\tilde{\lambda}_{(1)} < 0$, $\tilde{\lambda}_{(n)} > 0$ (not obvious!)
- Requirement: $\tilde{\lambda}_{(1)}^{-1} < \phi < \tilde{\lambda}_{(n)}^{-1}$
- If $\mathbf{M} = \tau^2\mathbf{\Delta}$,, τ just a scaling factor, same requirements

Inhomogeneous processes

- $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\delta}$, $\boldsymbol{\delta} \sim \text{Gau}(\mathbf{0}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M})$
- For $\mathbf{M} = \tau^2\boldsymbol{\Delta}$, with different diagonal elements, called *inhomogeneous process*.
- Define $\widetilde{\mathbf{Y}} = \boldsymbol{\Delta}^{-1/2}\mathbf{Y}$, $\widetilde{\mathbf{X}} = \boldsymbol{\Delta}^{-1/2}\mathbf{X}$, $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\delta}$,
- Then $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\beta + \widetilde{\boldsymbol{\delta}}$, $\widetilde{\boldsymbol{\delta}} \sim \text{Gau}(\mathbf{0}, \tau^2(\mathbf{I} - \widetilde{\mathbf{C}})^{-1})$
  $\widetilde{\mathbf{C}} = \boldsymbol{\Delta}^{-1/2}\mathbf{C}\boldsymbol{\Delta}^{1/2}$
- Note: $c_{ij} = 0 \Leftrightarrow \tilde{c}_{ij} = 0$
- For $\mathbf{C} = \phi\mathbf{H}$: $\widetilde{\mathbf{C}} = \phi\boldsymbol{\Delta}^{-1/2}\mathbf{H}\boldsymbol{\Delta}^{1/2}$

How to model **C** (or **H**)?

- Can show $[\text{cor}(Y_i, Y_j|Y_k, k \neq i, j)]^2 = c_{ij}c_{ji}$
- Need $0 \leq c_{ij}c_{ji} \leq 1$
- In general: c_{ij} can depend on distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. (distance between center points)
- Regular lattice: $c_{ij} = 0$ for
  - $||s_i - s_j|| > 1$ (1. order)
  - $||s_i - s_j|| > \sqrt{2}$ (2. order)
  - $||s_i - s_j|| > 2$ (3. order)
- Irregular lattice: $c_{ij} = 0$ for
  - i and j do not share border
  - $||s_i - s_j|| >$ threshold
- Size of c_{ij}: Depending on distance, number of neighbors

Special cases

Consider now $\mathbf{M} = \tau^2 \mathbf{\Delta}, \mathbf{C} = \phi \mathbf{H}$
Define $\mathbf{A}$ such that $a_{ij} = I(i$ and j are neighbors$)$.

- *Homogeneous CAR* (HCAR)
  $\mathbf{\Delta} = \mathbf{I}, \mathbf{H} = \mathbf{A}$
  Gives $\text{cor}(Y_i, Y_j | Y_k, k \neq i, j) = \phi$
  Need (at least) $\phi \in (0, 1)$
- *Weighted CAR* (WCAR)
  $\mathbf{\Delta} = \text{diag}\{|N(\mathbf{s}_i)|^{-1}\}$, $h_{ij} = a_{ij}|N(\mathbf{s}_i)|^{-1}$
  Gives $\text{cor}(Y_i, Y_j | Y_k, k \neq i, j) = \phi |N(\mathbf{s}_i)|^{-1/2} |N(\mathbf{s}_i)|^{-1/2}$
  Need (at least) $\phi \in (0, \min_{ij} |N(\mathbf{s}_i)|^{-1/2} |N(\mathbf{s}_i)|^{-1/2})$
- *Autocorrelated CAR* (ACAR)
  $\mathbf{\Delta} = \text{diag}\{|N(\mathbf{s}_i)|^{-1}\}$, $h_{ij} = a_{ij}|N(\mathbf{s}_i)|^{-1/2} |N(\mathbf{s}_j)|^{1/2}$
  Gives $\text{cor}(Y_i, Y_j | Y_k, k \neq i, j) = \phi$
  Need (at least) $\phi \in (0, 1)$

CAR models vs geostatistical models

- $\mathbf{Y} \sim \text{Gau}(\boldsymbol{\mu}, (\mathbf{I}-\mathbf{C})^{-1}\mathbf{M})$, $\text{Diag}(\mathbf{C}) = \mathbf{0}$
- Most important MRF model, often $\boldsymbol{\mu} = \mathbf{X}^T\beta$
- Building block for more complex models
- Geostatistical model: $\mathbf{Y} \sim \text{Gau}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$
  $C_Y(\mathbf{s}_j - \mathbf{s}_i) \approx 0 (= 0)$, for $||\mathbf{s}_j - \mathbf{s}_i||$ "large"
  $\boldsymbol{\Sigma}_\mathbf{Y}$ "sparse"
- MRF: $c_{ij} \approx 0(= 0)$ for $||\mathbf{s}_j - \mathbf{s}_i||$ "large"
  $\boldsymbol{\Sigma}_\mathbf{Y}^{-1}$ "sparse"
- MRF to geostat: $\boldsymbol{\Sigma}_Y = (\mathbf{I}-\mathbf{C})^{-1}\mathbf{M}$
- geostat to MRF: $\mathbf{M} = \text{Diag}(\boldsymbol{\Sigma}_Y^{-1}), \mathbf{C} = \mathbf{I} - \mathbf{M}\boldsymbol{\Sigma}_Y^{-1}$
- Differ in "sparsity"
- Differ in how to define dependence, distance versus neighborhood
- Which operations are simple? Building model vs Conditioning model.

Building model vs conditioning model

Model building: X,Y independent : $Z = X + Y$

$$C_Z = C_X + C_Y$$
$$Q_Z = Q_X - Q_X(Q_X + Q_Y)^{-1}Q_X$$
$$= Q_Y - Q_Y(Q_X + Q_Y)^{-1}Q_Y$$

Conditioning: X,Y independent : $Z = X + Y$

$$C_{Y|Z} = C_Y - C_Y(C_x + C_Y)^{-1}C_Y$$
$$Q_{Y|Z} = Q_Y + Q_X$$
$$Q_{X|Z} = Q_X + Q_Y$$

Example - Doctor-prescription

Data: Aaverage doctor-prescription amounts per consultation in cantons of the Midi-Pyrenees Department in southwest France.

- 268 cantons (32 "missing cantons" with no data)
- Response
  - Z: *Average* prescription amount per consultation in 1999
- Several possible covariates
  - X: X-coordinate of the centroid (in meters according to NTF)
  - Y: Y-coordinate of the centroid (in meters according to NTF)
  - $X2$: percentage of patients 70 or older
  - $X1$: per-capita income
  - E: number of consultations in 1999 (1270-1784977)

INLA

Syntax for WCAR (Besag) model:

```
formula = log(Z) ~ X2+Y+f(NO,model="besag",graph=Canton.graph)

res = inla(formula,family="gaussian",data=dat)
```

Formula:

- Specifies `log(Z)` as response and `X2` and `Y` as covariates
- `f()` specifies a *random effect* (δ in our spatial model).
  Can have different models. Here the Besag (WCAR) model is specified. Requires a neighborhood structure, given in the `graph` option
- The function `inla` requires
  - The formulae
  - A model for the response, given by `family`
  - The data
  - Several other options possible, default choices imply
    - a Bayesian approach
    - Default priors on *hyperparameters*

Spatial rates model

$$\Delta_{ii} = M_i^{-1}$$
$$h_{ij} = \begin{cases} (M_i/M_j)^{1/2}, & j \in N(\mathbf{s}_i) \\ 0, & \text{elsewhere} \end{cases}$$
$$\mathbf{Q} = \frac{1}{\tau^2}\boldsymbol{\Delta}^{-1}[\mathbf{I} - \phi\mathbf{H}]$$

Gives

$$\text{var}(Y_i|Y_k, k \neq i) = M_i$$
$$\text{cor}(Y_i, Y_j|Y_k, k \neq i, j) = \phi \qquad \text{similar to ACAR}$$

In application: $M_i = E(i)$ (number of consultations)
Not directly available in INLA, but possible through transformation:
$\widetilde{\boldsymbol{\delta}} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\delta} \sim \text{Gau}(\mathbf{0}, \tau^2(\mathbf{I} - \phi\widetilde{\mathbf{H}})^{-1})$

$$\tilde{h}_{ij} = \begin{cases} M_i/M_j, & j \in N(\mathbf{s}_i) \\ 0, & \text{elsewhere} \end{cases}$$

Spatial rates model in INLA

INLA `generic1` model: $\mathbf{Q} = \xi(\mathbf{I} - \frac{\phi}{\lambda_{max}}\widetilde{\mathbf{H}})$,
$\phi \in (0, 1)$, λ_{max} maximum eigenvalue of $\widetilde{\mathbf{H}}$.
Our model:

$$\begin{aligned}\mathbf{Z} =& \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}\\ =& \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}^{1/2}\widetilde{\boldsymbol{\delta}} + \boldsymbol{\varepsilon}\end{aligned}$$

Possible in INLA by

- specifying `generic1` model for $\widetilde{\boldsymbol{\delta}}$ and
- including $\boldsymbol{\Delta}^{1/2}$ as weights

`Canton.R` script

Model comparison

Model selection tools for Bayesian approaches:

- Marginal likelihood:

$$p(\mathbf{z}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{y}} p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})p(\theta)d\mathbf{y}d\boldsymbol{\theta}$$

  Want it large!
  Can be sensitive to $p(\boldsymbol{\theta})$.
  In general difficult to compute, "easy" in INLA
- DIC: Bayesian alternative to AIC
  Want it large!
  Much used, but can give strange results
- Many other alternatives in the literature

Model selection - Doctors prescription

Marginal likelihood

```
> cbind(res.ind$mlik,res.besag.b$mlik,res.sr$mlik)
                                      [,1]       [,2]      [,3]
log marg-likel (integration)      111.5341 -73.78425 112.4315
log marginal-likel (Gaussian)     112.4530 -73.99076 112.3778
```

DIC

```
> cbind(res.ind$dic,res.besag.b$dic,res.sr$dic)
              [,1]        [,2]        [,3]
dic           -298.1838   -317.3547   -297.5507
p.eff         3.878958    43.89627    4.195396
mean.deviance -302.0628   -361.2509   -301.7461
deviance.mean -305.9418   -405.1472   -305.9415
```

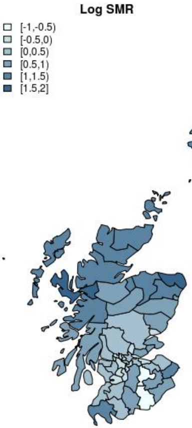
Spatial disease mapping

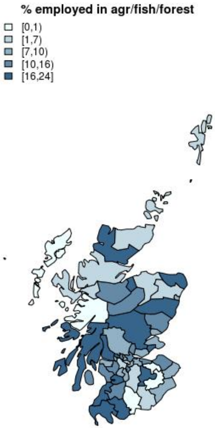
$$
\begin{aligned}
Z_i|Y_i \overset{ind}{\sim} & \text{Poisson}(E_i \exp(Y_i)) \\
Z_i =& \text{Observed disease count} \\
E_i =& \text{Expected count (known), and} \\
Y_i =& \mathbf{x}_i^T \boldsymbol{\beta} + \delta_i + \varepsilon_i \\
\delta_i|\delta_{j\neq i} \sim & N(|N_i|^{-1} \sum_{j \in N_i} \delta_j, 1/(\tau_c|N_i|)) \qquad \text{WCAR/Besag model} \\
\varepsilon_i \overset{ind}{\sim} & N(0, 1/\tau_\varepsilon)
\end{aligned}
$$

Usually include $\sum_i \delta_i = 0$ to make model identifiable.
Note: Often $Z_i|Y_i$ is Binomial$(N_i, p_i(Y_i))$ but large N_i and small p_i make Poisson distribution more convenient to use.
Often: Considering *standardized mortality ratio (SMR)*:

$$SMR_i \equiv Z_i/E_i$$

Scottish lip cancer data





$$Z_i|Y_i \sim \text{Poisson}(E_i \exp(Y_i))$$
$$Y_i = \beta_0 + \beta_1 x_i/10 + \delta_i + \varepsilon_i$$
$$x_i = \text{Percentage of population enganged in agriculture/fishing/forestry}$$

Disease mapping in INLA

```
library(INLA)
data(Scotland)
g = system.file("demodata/scotland.graph", package="INLA")
graph = inla.read.graph(g)
Scotland$Region2 = Scotland$Region
formula = Counts~ I(X/10) + f(Region,model="besag", graph=graph,
                            f(Region2,model="iid")
mod.scotland = inla(formula,family="poisson",E=E,data=Scotland)
```

Script Scottish.R

Computation when INLA not is possible

Inla code:

- INLA: Possible for latent processes being linear and Gaussian
- Nonlinearity/non-Gaussian: Monte Carlo metods

Computation when INLA not is possible

Inla code:

- INLA: Possible for latent processes being linear and Gaussian
- Nonlinearity/non-Gaussian: Monte Carlo metods

Monte Carlo methods:

- Assume interest in $p(\mathbf{y}|\mathbf{z})$
- Assume possible to simulate $\mathbf{y}^1, ..., \mathbf{y}^S$ from $p(\mathbf{y}|\mathbf{z})$
- Can approximate $E[g(\mathbf{y})|\mathbf{Z} = \mathbf{z}]$ by $S^{-1}\sum_{s=1}^{S} g(\mathbf{y}^s)$
- Problem: Difficult to simulate from $p(\mathbf{y}|\mathbf{z})$ directly

Computation when INLA not is possible

Inla code:

- INLA: Possible for latent processes being linear and Gaussian
- Nonlinearity/non-Gaussian: Monte Carlo metods

Monte Carlo methods:

- Assume interest in $p(\mathbf{y}|\mathbf{z})$
- Assume possible to simulate $\mathbf{y}^1, ..., \mathbf{y}^S$ from $p(\mathbf{y}|\mathbf{z})$
- Can approximate $E[g(\mathbf{y})|\mathbf{Z} = \mathbf{z}]$ by $S^{-1}\sum_{s=1}^{S} g(\mathbf{y}^s)$
- Problem: Difficult to simulate from $p(\mathbf{y}|\mathbf{z})$ directly

Markov chain Monte Carlo:

- $\mathbf{y}^{(s)}$ is generated only depending on $\mathbf{y}^{(s-1)}$ (Markov chain).
- As s increases, the distribution of $\mathbf{y}^{(s)}$ converges towards $p(\mathbf{y}|\mathbf{z})$
- Under some additional requirements, we also have that $(S-b)^{-1}\sum_{s=b+1}^{S} g(\mathbf{y}^{(s)})$ converges towards $E[g(\mathbf{Y})|\mathbf{Z} = \mathbf{z}]$.

Gibbs sampler

One of many many MCMC algorithms
Easy to implement, not always very efficient in spatial settings

Assume $\mathbf{y} = (y_1, ..., y_n)$

- For $s = 1, 2, 3, ..., B$
  - simulate $y_1^s \sim p(y_1 | y_2^{s-1}, ..., y_n^{s-1})$
  - simulate $y_2^s \sim p(y_2 | y_1^s, y_3^{s-1}, ..., y_n^{s-1})$
  - $\vdots$
  - simulate $y_n^s \sim p(y_n | y_1^s, y_2^s, ..., y_{n-1}^s)$

Note

- Often use a permutation of the ordering in the updates
- Only univariate updates
- Only need conditional distributions, will typically not require the global normalization constant.

Potts model

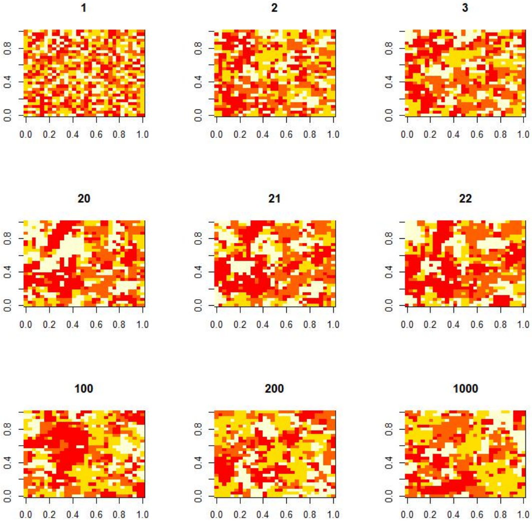
Model defined on a discrete set of values $Y_i \in \{1, ..., K\}$.
Defined trough the *Gibbs distribution* :

$$\Pr(\mathbf{Y} = \mathbf{y}) \propto \exp\{\sum_{i=1}^{n} \alpha_{i,y_i} + \tfrac{1}{2}\beta \sum_{i} \sum_{j \in N_i} I(y_i = y_j)\}$$

Conditional distribution:

$$\Pr(Y_i = k | Y_j = y_j, j \neq i) = \frac{\exp\{\alpha_{i,k} + \beta \sum_{j \in N_i} I(y_j = k)\}}{\sum_{l=1}^{K} \exp\{\alpha_{i,l} + \beta \sum_{j \in N_i} I(y_j = l)\}}$$

Simulations from Potts model, $\alpha_{i,k} = 0$, $\beta = 1$



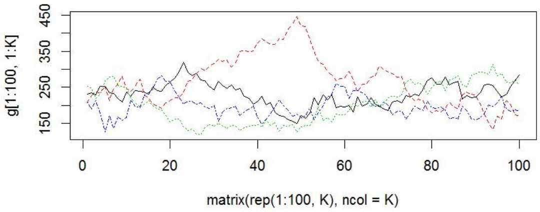
MCMC: How many iterations

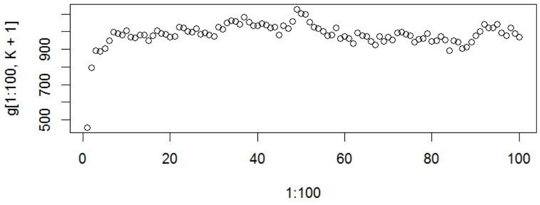
- Convergence often performed by eye (formal tests exist)
- Difficult to look at the whole process
- Usually considering summary statistics $g(\mathbf{y})$
- Potts model:

$$g_k(\mathbf{y}) = \sum_i I(y_i = k) \qquad k = 1, ..., K$$

$$g_{K+1}(\mathbf{y}) = \sum_{i \sim j} I(y_i = y_j) \qquad i \sim j \text{ means } i, j \text{ are neighboors}$$

g-functions for Potts model, per iteration 1-100

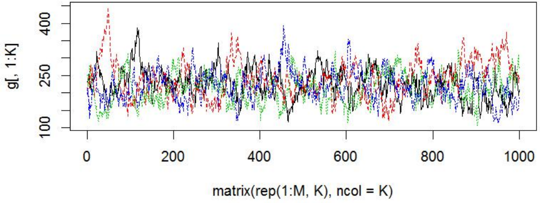


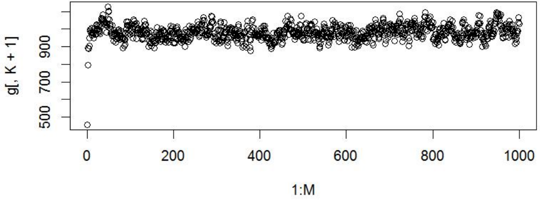


Note: Want

- Convergence, curve stabilize
- Small auto-correlation

g-functions for Potts model, per iteration

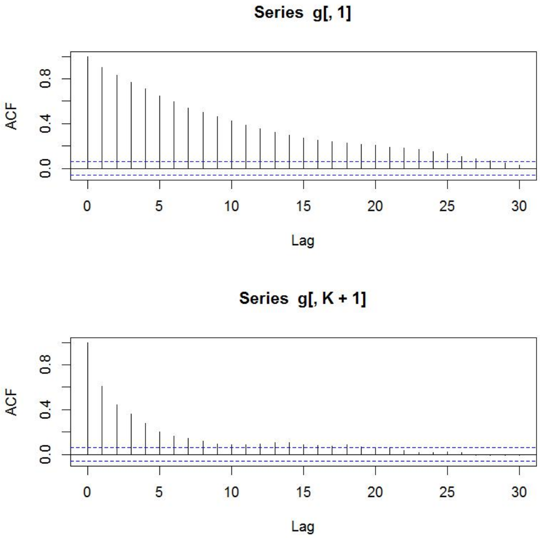




Note: Want

- Convergence, curve stabilize
- Small auto-correlation

ACF of g-functions for Potts model



Variance of MCMC

Assume $\mathbf{y}^{(s)} \approx p(\mathbf{y}|\mathbf{z})$

$$\begin{aligned}
\hat{\theta} =& S^{-1}\sum_{s=1}^{S} g(\mathbf{y}^{(s)})\\
\text{Var}[\hat{\theta}] =& S^{-2}[\sum_{s=1}^{S}\text{Var}[g(\mathbf{y}^{(s)})] + 2\sum_{h=1}^{S-1}\sum_{s=1}^{S-h}\text{Cov}[g(\mathbf{y}^{(s)}), g(\mathbf{y}^{(s+h)})]\\
=& S^{-2}[S\text{Var}[g(\mathbf{y}^{(s)})] + 2\sum_{h=1}^{S-1}(S-h)\text{Cov}[g(\mathbf{y}^{(s)}), g(\mathbf{y}^{(s+h)})]\\
=& S^{-1}\text{Var}[g(\mathbf{y}^{(s)})][1 + 2\sum_{s=1}^{S-1}(1-\tfrac{h}{S})\text{Cor}[g(\mathbf{y}^{(s)}), g(\mathbf{y}^{(s+h)})]]
\end{aligned}$$

Note: Need $\sum_{s=1}^{S-1}(1-\frac{h}{S})\text{Cor}[g(\mathbf{y}^{(s)}), g(\mathbf{y}^{(s+h)})] \stackrel{S\to\infty}{\to}$ Const