BAYESIAN MODEL SELECTION IN SPATIAL LATTICE MODELS

Victor De Oliveira

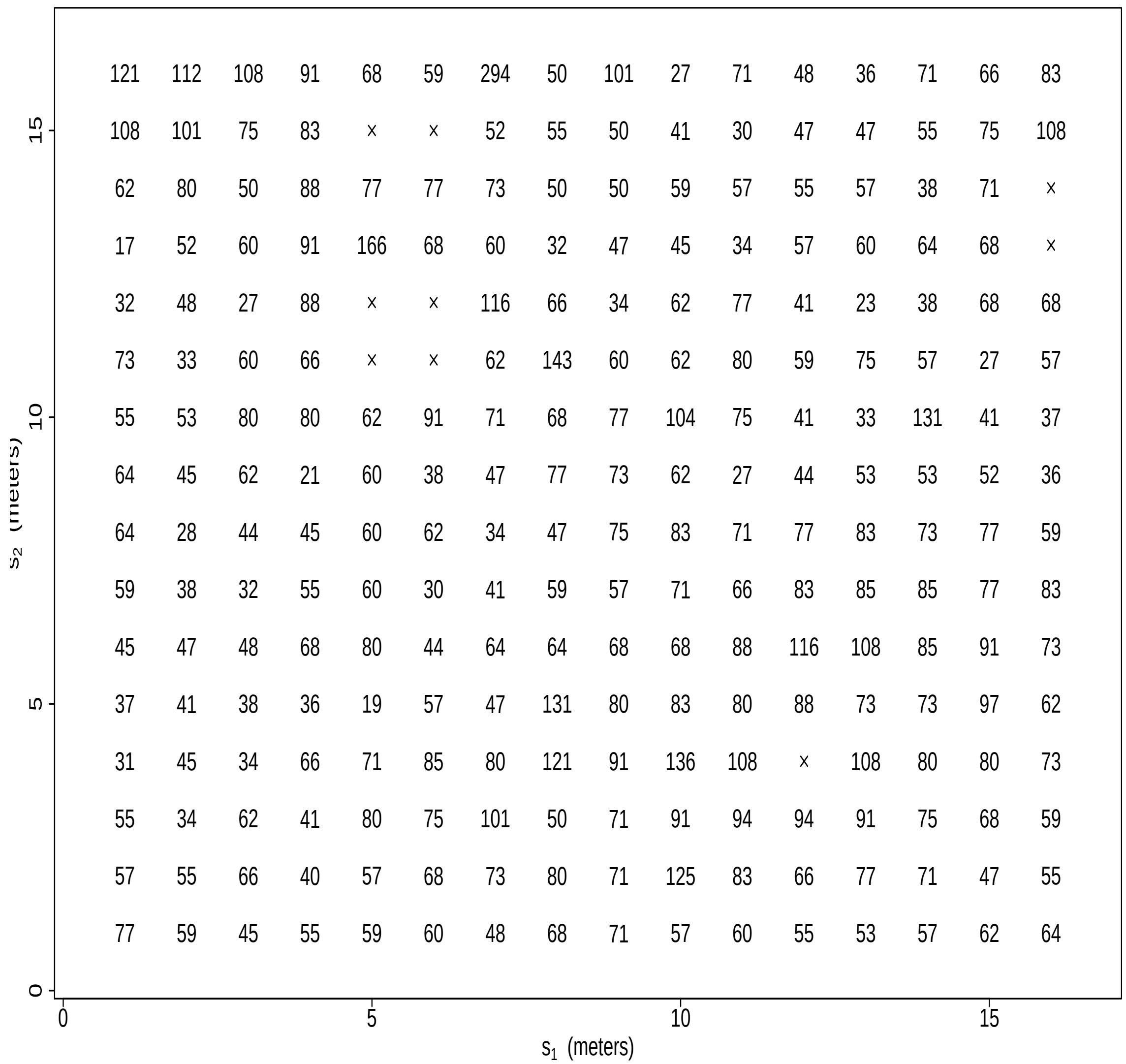
Department of Management Science and Statistics
The University of Texas at San Antonio
San Antonio, TX
USA
victor.deoliveira@utsa.edu
http://faculty.business.utsa.edu/vdeolive

Joint work with J.J. Song

The Fourth Erich L. Lehmann Symposium, May 9–12, 2011

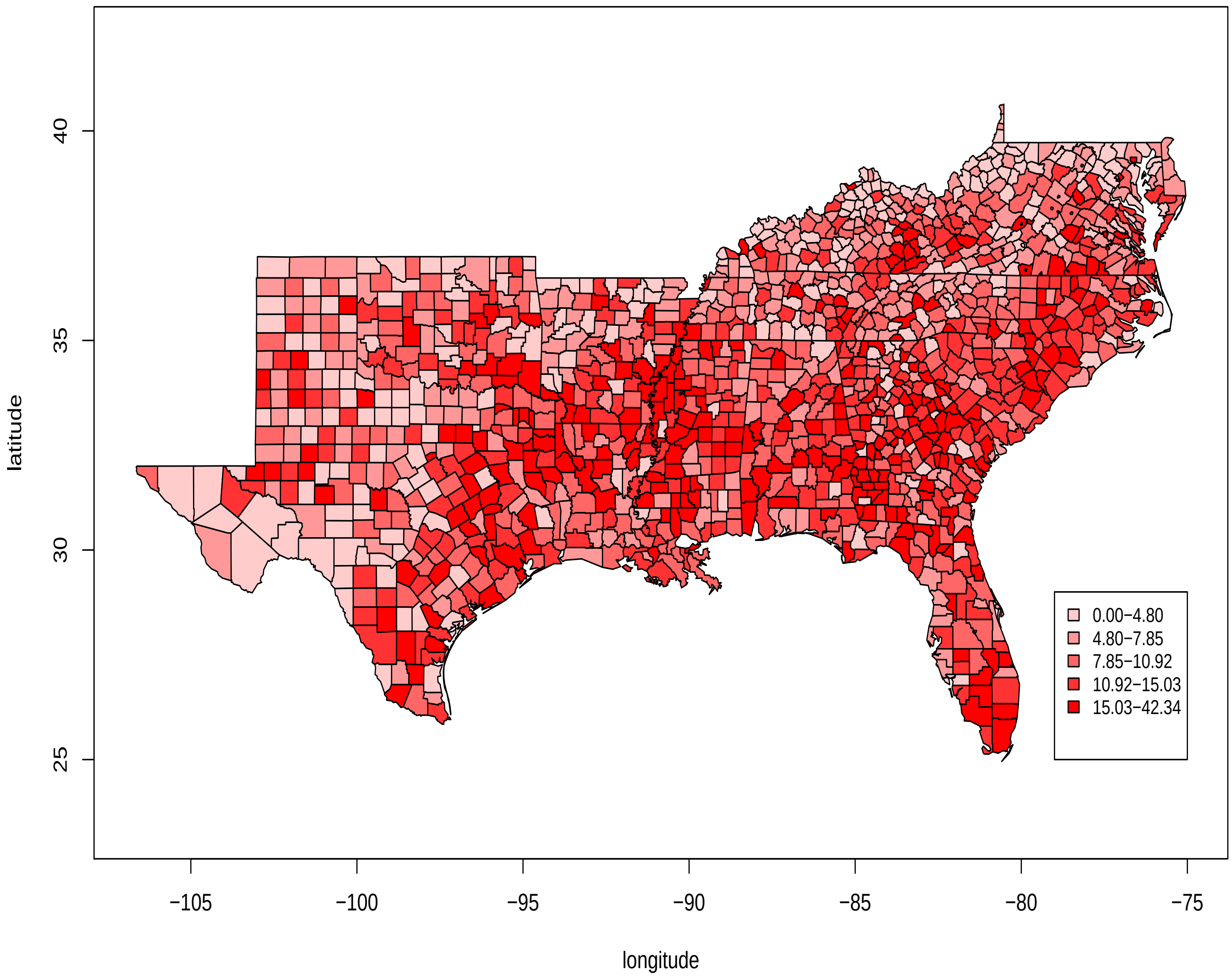
Example 1: Phosphate Data

Raw phosphate concentrations (in mg P/100 g of soil) collected over 16 by 16 regular lattice during several years in archaeological region of Greece



Example 2: Crime Data

Homicide rates per $100,000$ habitants for 1980 in the south of US, with $n = 1412$ counties



Models for Spatial Lattice Data

- Conditional Autoregressive (CAR) Models:
Mostly studied and applied in Statistical literature

- Simultaneously Autoregressive (SAR) Models:
Mostly studied and applied in Econometric/geography literature

All of these require specifying a neighborhood system

Neighborhood Systems

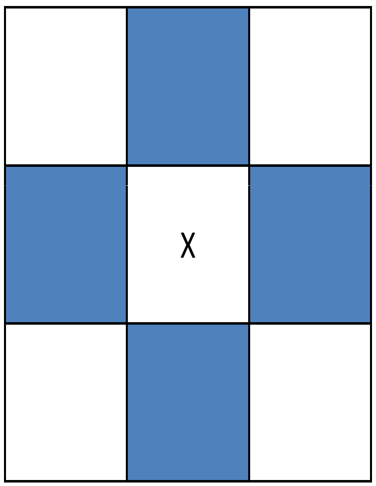
Sites $\{1, \ldots, n\}$ are endowed with neighborhood system, $\{N_i : i = 1, \ldots, n\}$, where $N_i =$ neighbors of site i.
Examples:

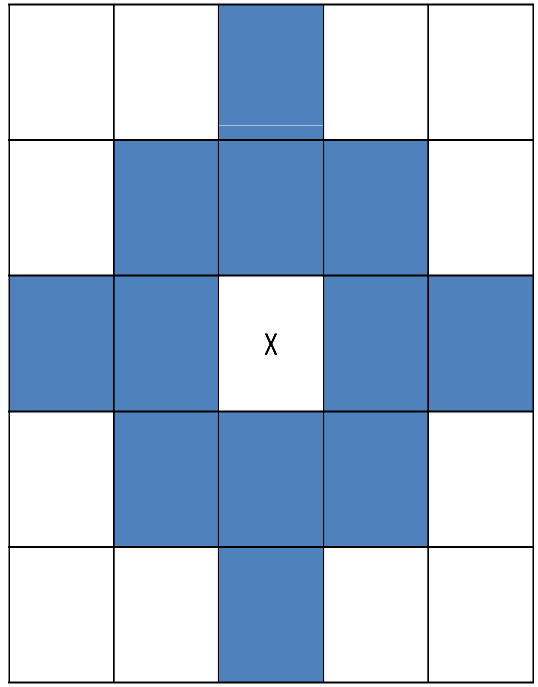
$$N_i = \{j : \text{site } j \text{ shares a boundary with site } i\}$$

$$N_i = \{j : 0 < d_{ij} < r\}$$

with $r > 0$ and d_{ij} the distance between sites i and j

First and second order neighborhood systems





Goal

Model selection for spatial lattice data using a default Bayesian approach, where the competing models:

- Have the same mean structure
- Have different covariance structures

CAR MODELS

Conditional Specification: For $i = 1, \ldots, n$

$$(Y_i \mid \mathbf{Y}_{(i)}) \sim \mathrm{N}(\mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} c_{ij}(Y_j - \mathbf{x}_j'\boldsymbol{\beta}), \tau_i^2)$$

- $\mathbf{Y}_{(i)} = \{Y_j, j \neq i\}$
- $\mathbf{x}_j' = (x_{j1}, \ldots, x_{jp})$
- $\boldsymbol{\beta} \in \mathbb{R}^p, \quad \tau_i > 0$
- $c_{ij} \geq 0$ and $c_{ij} > 0$ iff $i \sim j$

Let $M = \text{diag}(\tau_1^2, \ldots, \tau_n^2)$ and $C = (c_{ij})$ satisfy

- $M^{-1}C$ is symmetric, so $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$
- $M^{-1}(I_n - C)$ positive definite

Joint Specification:

$$\mathbf{Y} \sim \mathsf{N}_n(X\boldsymbol{\beta}, (I_n - C)^{-1}M)$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$

Parameterization

- $M = \sigma^2 G$, with $\sigma^2 > 0$ unknown and G diagonal (known)

- $C = \phi W$, with ϕ 'spatial parameter' and $W = (w_{ij})$ nonnegative "weight" known matrix (not necessarily symmetric), and $w_{ij} > 0$ iff $i \sim j$

Let $A = (a_{ij})$ [neighborhood matrix]:

$a_{ij} = 1$ if $i \sim j$, and $a_{ij} = 0$ otherwise

Classes of CAR Models

- Homogeneous CAR (HCAR):

$$G = I_n \quad , \quad W = A$$

- Weighted CAR (WCAR) (Besag et al. 1991):

$$G = \text{diag}(|N_1|^{-1}, \ldots, |N_n|^{-1}) \quad , \quad W = GA$$

with $|N_i| = \sum_{j=1}^{n} a_{ij}$

- Autocorrelation CAR (ACAR) (Cressie & Chang, 1989):

$$G = \text{diag}(|N_1|^{-1}, \ldots, |N_n|^{-1}) \quad , \quad W = G^{1/2} A G^{-1/2}$$

Facts Assume the above conditions hold and $G^{-1}M$ is symmetric. Then:

(a) $G^{-1/2}WG^{1/2}$ is symmetric

(b) $G^{-1/2}WG^{1/2}$ and W have the same nonzero eigenvalues, and all are real

(c) M and C determine a CAR model iff $\sigma^2 > 0$ and $\phi \in (\lambda_n^{-1}, \lambda_1^{-1})$, with $\lambda_1 \geq \ldots \geq \lambda_n$ ordered eigenvalues of $G^{-1/2}WG^{1/2}$

Parameter space: $\Omega = \mathbb{R}^p \times (0, \infty) \times (\lambda_n^{-1}, \lambda_1^{-1})$

SAR MODELS

Conditional Specification: For $i = 1, \ldots, n$

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} b_{ij}(Y_j - \mathbf{x}_j'\boldsymbol{\beta}) + \epsilon_i$$

- $\epsilon_i \sim \text{N}(0, \xi_i^2)$, independent
- $\boldsymbol{\beta} \in \mathbb{R}^p, \quad \xi_i > 0$
- $b_{ij} \geq 0$ and $b_{ij} > 0$ iff $i \sim j$

Let $M = \text{diag}(\xi_1^2, \ldots, \xi_n^2)$ and $B = (b_{ij})$ satisfy that $I_n - B$ is nonsingular. Then

Joint Specification:

$$\mathbf{Y} \sim \text{N}_n(X\boldsymbol{\beta}, (I_n - B)^{-1}M(I_n - B')^{-1})$$

Particular Model:

- $M = \sigma^2 I_n$
- $B = \phi A$

so

$$\mathbf{Y} \sim \mathsf{N}_n(X\boldsymbol{\beta}, \sigma^2((I_n - \phi A)^2)^{-1}$$

Parameter space: $\Omega = \mathbb{R}^p \times (0, \infty) \times (\lambda_n^{-1}, \lambda_1^{-1})$, with $\lambda_1 \geq \ldots \geq \lambda_n$ the ordered eigenvalues of A

MODEL SELECTION

Let $M_1, M_2, \ldots, M_k$ be the candidate models ($k \geq 2$)

M_j is either HCAR, WCAR, ACAR or SAR
parameterized by $\boldsymbol{\eta}_j = (\boldsymbol{\beta}, \sigma_j^2, \phi_j) \in \Omega_j$
with covariance depending on G_j and A_j

$\phi_j \in (1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$ with
$\lambda_1^{(j)} \geq \lambda_2^{(j)} \geq \ldots \geq \lambda_n^{(j)}$ eigenvalues of:

- A_j in case of HCAR, ACAR and SAR
- $G_j^{1/2} A_j G_j^{1/2}$ in case of WCAR

The approach proposed here assumes all models have the same mean structure

Likelihood for M_j

$$L_j(\boldsymbol{\eta}_j; \mathbf{y}) =$$

$$(2\pi\sigma_j^2)^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{\phi_j}^{-1}|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma_j^2}(\mathbf{y} - X\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\phi_j}^{-1} (\mathbf{y} - X\boldsymbol{\beta})\}$$

where

$$\boldsymbol{\Sigma}_{\phi_j}^{-1} = \begin{cases} I_n - \phi_j A_j & \text{for HCAR models} \\ G_j^{-1} - \phi_j A_j & \text{for WCAR models} \\ G_j^{-1} - \phi_j G_j^{-1/2} A_j G_j^{-1/2} & \text{for ACAR models} \\ (I_n - \phi_j A_j)^2 & \text{for SAR models} \end{cases}$$

Prior for M_j

$$\pi(\boldsymbol{\eta}_j \mid M_j) \propto \frac{\pi(\phi_j \mid M_j)}{\sigma_j^2} \mathbf{1}_{\Omega_j}(\boldsymbol{\eta}_j)$$

Two options for $\pi(\phi_j \mid M_j)$:

- Uniform:

$$\pi^U(\phi_j \mid M_j) = \mathbf{1}_{(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})}(\phi_j)$$

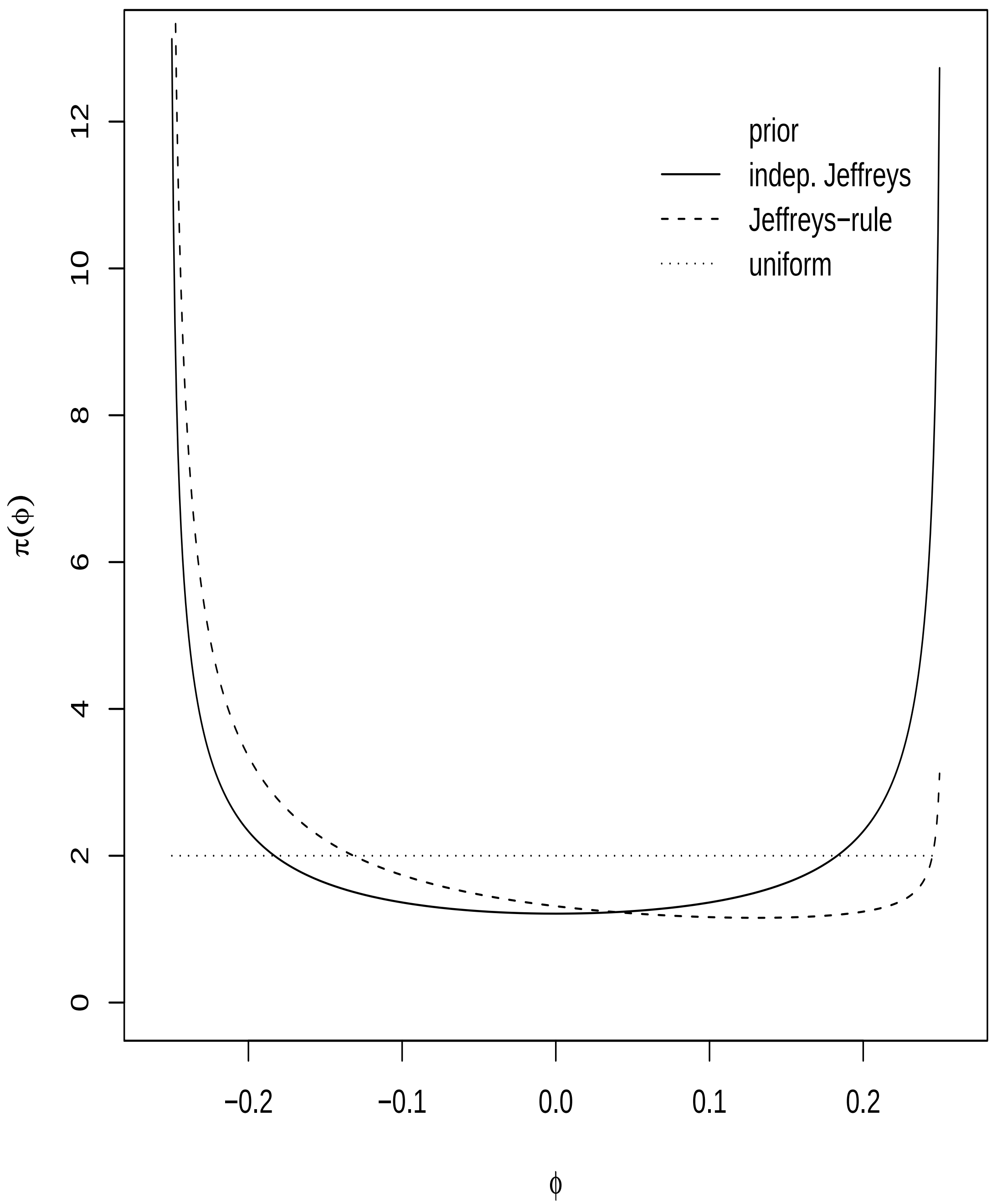
- Independence Jeffreys:

$$\pi^{J1}(\phi_j \mid M_j) =$$

$$\left\{ \sum_{i=1}^{n} \left(\frac{\lambda_i^{(j)}}{1 - \phi_j \lambda_i^{(j)}}\right)^2 - \frac{1}{n} \left[\sum_{i=1}^{n} \frac{\lambda_i^{(j)}}{1 - \phi_j \lambda_i^{(j)}}\right]^2 \right\}^{\frac{1}{2}} \mathbf{1}_{(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})}(\phi_j)$$

(De Oliveira & Song, 2008; De Oliveira, 2011)

(a)



prior
indep. Jeffreys
Jeffreys−rule
uniform
π(ϕ)
ϕ
0
2
4
6
8
10
12
−0.2
−0.1
0.0
0.1
0.2

Bayes Factors & Posterior Model Probabilities

$$\begin{aligned}\frac{\pi(M_i \mid \mathbf{y})}{\pi(M_j \mid \mathbf{y})} &= \frac{m(\mathbf{y} \mid M_i)\pi(M_i)}{m(\mathbf{y} \mid M_j)\pi(M_j)} \\ &= B_{ij} \times \text{prior odds}_{ij}\end{aligned}$$

where

$$m(\mathbf{y} \mid M_j) = \int_{\Omega_j} L_j(\boldsymbol{\eta}_j \mid \mathbf{y})\pi(\boldsymbol{\eta}_j \mid M_j)d\boldsymbol{\eta}_j,$$

and

$$B_{ij} = \frac{m(\mathbf{y} \mid M_i)}{m(\mathbf{y} \mid M_j)}$$

Hence

$$\begin{aligned}\pi(M_j \mid \mathbf{y}) &= \left(\sum_{l=1}^{k} \frac{\pi(M_l)}{\pi(M_j)} B_{lj}\right)^{-1}, \quad j = 1, \ldots, k \\ &= \frac{m(\mathbf{y} \mid M_j)}{\sum_{l=1}^{k} m(\mathbf{y} \mid M_l)}, \quad \text{when } \pi(M_j) = \frac{1}{k}\end{aligned}$$

Remarks

- Bayes factors and posterior model probabilities are, in general, undetermined when improper priors are used

- Important exception occurs when competing models have same invariance structure, up to individual model parameters that have proper priors (Berger et al., 1998)

- CAR and SAR models fit this exception when all the competing models have the same mean structure and $\pi(\phi_j \mid M_j)$ is proper

Fact As $\phi_j \to 1/\lambda_i^{(j)}$; $i = 1$ or n

$$\pi^{J1}(\phi_j \mid M_j) = O((1 - \phi_j \lambda_i^{(j)})^{-1})$$

so $\pi^{J1}(\phi_j \mid M_j)$ is not integrable
(De Oliveira & Song, 2008).

Instead we use $(\pi^{J1}(\phi_j \mid M_j))^r$, with $r < 1$, which is proper and has the same "shape".

For $j = 1, \ldots, k$:

$$m(\mathbf{y} \mid M_j) = K c_j \int_{1/\lambda_n^{(j)}}^{1/\lambda_1^{(j)}} h(\phi_j, M_j, \mathbf{y}) d\phi_j$$

where

$$h(\phi_j, M_j, \mathbf{y}) =$$

$$|\boldsymbol{\Sigma}_{\phi_j}^{-1}|^{1/2} |X' \boldsymbol{\Sigma}_{\phi_j}^{-1} X|^{-1/2} (S_{\phi_j}^2)^{-(n-p)/2} \pi(\phi_j \mid M_j)$$

$$S_{\phi_j}^2 = (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\phi_j})' \boldsymbol{\Sigma}_{\phi_j}^{-1} (\mathbf{y} - X\widehat{\boldsymbol{\beta}}_{\phi_j})$$

$$\widehat{\boldsymbol{\beta}}_{\phi_j} = (X' \boldsymbol{\Sigma}_{\phi_j}^{-1} X)^{-1} X' \boldsymbol{\Sigma}_{\phi_j}^{-1} \mathbf{y}$$

$$K = \frac{\Gamma(\frac{n-p}{2})}{\pi^{\frac{n-p}{2}}} \quad , \quad c_j = \left(\int_{1/\lambda_n^{(j)}}^{1/\lambda_1^{(j)}} \pi(\phi_j \mid M_j) d\phi_j \right)^{-1}$$

Note

- For posterior model probabilities to be well defined and calibrated, the proportionality constants in the likelihoods and priors of all competing models should be retained

- Computation of $m(\mathbf{y} \mid M_j)$ involves one-dimensional integration over a bounded interval

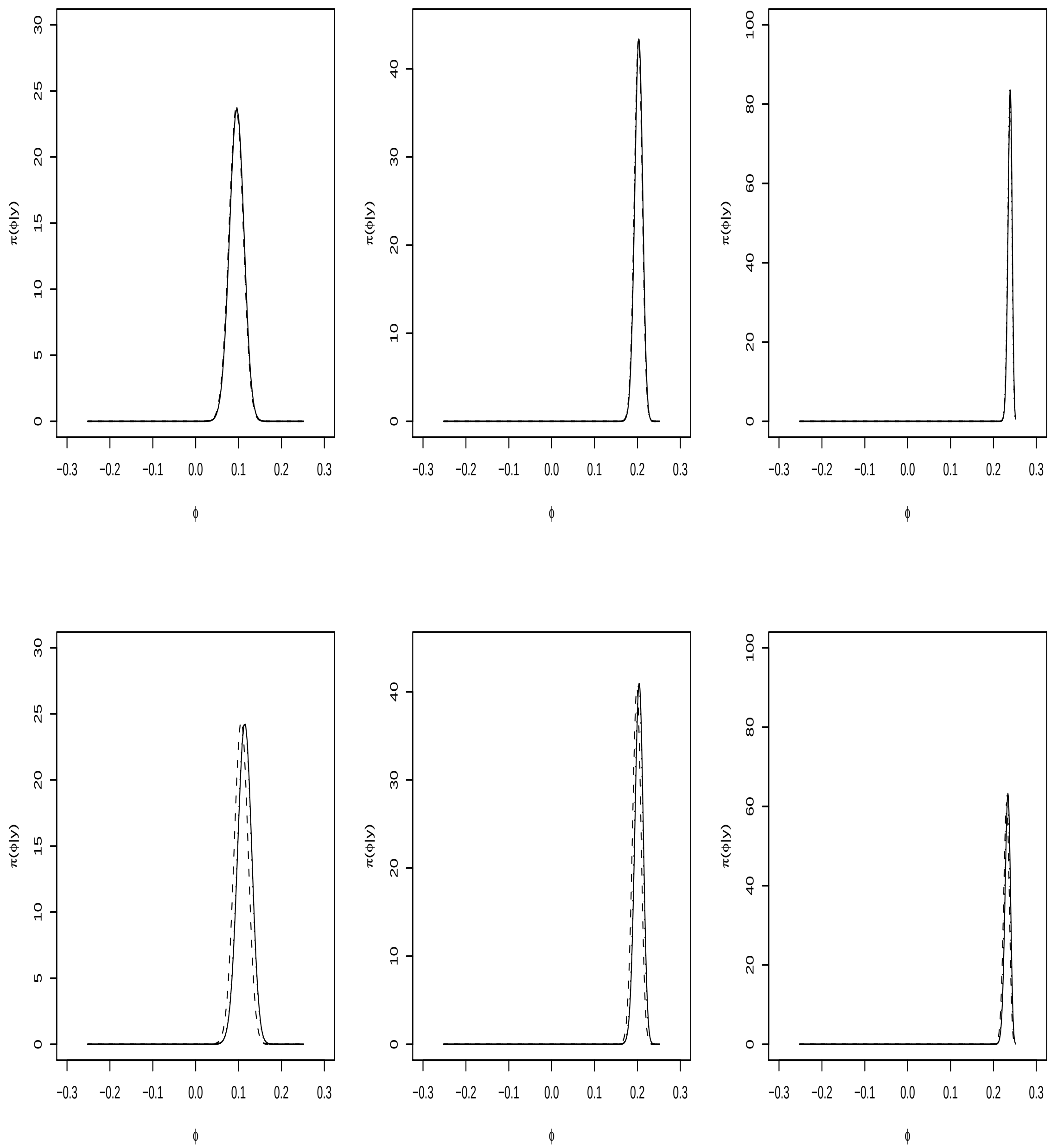
Computation

- Computation of $\widehat{c}_j$ straightforward: numerical quadrature or Monte Carlo

$$\widehat{c}_j = \left((\frac{1}{\lambda_1^{(j)}} - \frac{1}{\lambda_n^{(j)}}) \frac{1}{m} \sum_{l=1}^{m} (\pi^{J1}(\phi_j^{(l)} \mid M_j))^{1/2} \right)^{-1}$$

with $\phi_j^{(1)}, \ldots, \phi_j^{(m)} \overset{\text{iid}}{\sim} \text{unif}(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$

- Computation of $m(\mathbf{y} \mid M_j)$ requires more care: $h(\phi_j, M_j, \mathbf{y})$ is highly peaked and concentrated near the right boundary for moderate or large sample sizes. Hence almost constant and very close to zero over most of the integration region, and common numerical quadrature or Monte Carlo estimates are often zero.


π(ϕ|y)
ϕ
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0
5
10
15
20
25
30
40
60
80
100

A Solution (Importance Sampling)

Let $\tilde{\phi}_j$ value that maximizes $h(\phi_j, M_j, \mathbf{y})$, $t \in [3, 4]$ and
$\omega_j = (1/\lambda_1^{(j)} - \tilde{\phi}_j)/t$.
Then

$$\widehat{m}(\mathbf{y} \mid M_j) =$$

$$\left(\Phi(t) - \Phi\left(t\frac{1/\lambda_n^{(j)} - \tilde{\phi}_j)}{1/\lambda_1^{(j)} - \tilde{\phi}_j)}\right)\right)\frac{\sqrt{2\pi}Kc_j\omega_j}{m}\sum_{l=1}^{m}\left(\frac{h(\phi_j^{(l)}, M_j, \mathbf{y})}{\exp\{-(\phi_j^{(l)} - \tilde{\phi}_j)^2/2\omega_j^2\}}\right)$$

where $\phi_j^{(1)}, \ldots, \phi_j^{(m)} \stackrel{\text{iid}}{\sim} N(\tilde{\phi}_j, \omega_j^2)$ truncated to $(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$

Example 1: Phosphate Data

- Data were transformed to become closer to Gaussian

- HCAR, WCAR, ACAR and SAR models as competing models

- First and second order neighborhood systems were entertained

- $E\{\tilde{Y}_i\}$ is β_1 $(p = 1)$ or $\beta_1 + \beta_2 s_{i1} + \beta_3 s_{i2}$ $(p = 3)$

- All models equally likely a priori

- Both default priors were considered

Results

models	HCAR-1	HCAR-2	WCAR-1	WCAR-2	ACAR-1	ACAR-2	SAR-1	SAR-2
				modified independence Jeffreys prior				
$p = 1$	0.099	2.2×10^{-8}	0.321	4.0×10^{-8}	0.443	5.1×10^{-8}	0.136	1.3×10^{-5}
$p = 3$	0.130	7.6×10^{-8}	0.249	9.2×10^{-8}	0.488	1.2×10^{-7}	0.132	1.9×10^{-5}
				uniform prior				
$p = 1$	0.085	4.3×10^{-7}	0.295	6.6×10^{-7}	0.416	6.6×10^{-7}	0.203	1.5×10^{-5}
$p = 3$	0.148	6.3×10^{-7}	0.221	1.6×10^{-9}	0.443	8.7×10^{-7}	0.186	2.1×10^{-5}

Example 2: Crime Data

- Significant explanatory variables:
an index of resource deprivation, an index of population structure, median age, divorce rate and unemployment rate

- HCAR, WCAR, ACAR and SAR models as competing models

- Consider the adjacency neighborhood system (AC), and two distance-based neighborhood systems with $r = 70$ miles (D70) and $r = 100$ miles (D100)

- All models equally likely a priori

- Both default priors were considered

Results

models	HCAR	WCAR	ACAR	SAR
	modified independence Jeffreys prior			
AC	4.2×10^{-6}	0	0	0
D70	0.857	0	0	0.065
D100	3.0×10^{-3}	0	0	0.074
	uniform prior			
AC	3.6×10^{-6}	0	0	0
D70	0.822	0	0	0.074
D100	3.4×10^{-3}	0	0	0.100

Conclusions

⊕ Method does not require nested competing models

⊕ Method provides interpretable measures of how strongly the data support each competing model

⊕ Method does not require assessing subjective priors for model parameters

⊖ Method requires all competing models to have the same mean structure