Spatial Bayesian Variable Selection Models on Functional Magnetic Resonance Imaging Time-Series Data

by: KJ Lee, GL Jones, BS Caffo & SS Bassett

Bayesian Analysis 9(3), 699-732

Warwick Statistics

October 16, 2014

Outline

Introduction

- BOLD signal modelling for task-realted fMRI
- Keystone: model complexity/computational efficiency
- Main contributions:
  - Spatio-temporal correlations
  - Variable selection

Data

- Part of longitudinal AD study
- Sample consists of older, well-educated, right handed controls
- Investigation of the Stroop paradigm
  - Automatic behaviour vs decision rule
  - Several brain regions involved
  - In this study: **WORD**, **BLUE** , **BLUE**
- Experimental design:
  - Block design
  - 465 total time points, scanning time 2sec
  - Standard preprocessing...
  - 79 $\times$ 95 $\times$ 68 template, 2mm^3 voxels

BOLD modelling

- For voxel $v = 1, \ldots, N$ and time $i = 1, \ldots T_v$ assume:

$$\mathbf{y}_v = \mathbf{X}_v \boldsymbol{\beta}_v + \boldsymbol{\epsilon}_v, \quad \boldsymbol{\epsilon}_v \sim \mathcal{N}_{T_v}(0, \sigma_v^2 \boldsymbol{\Lambda}_v)$$

  with $\mathbf{y}_v = [y_{v,1}, \ldots, y_{v,N_v}]^\top$, etc...
- Variable selection introduces as:

$$\mathbf{X}_v(\gamma_v)\boldsymbol{\beta}_v(\gamma_v)$$

  where γ_v has 0,1

Prior distributions (1/2)

- $\boldsymbol{\beta}_v(\boldsymbol{\gamma}_v)$ have Zellner's g-prior, with mean estimated from data
- σ_v^2 independent:

$$\pi(\sigma_v^2) \propto \frac{1}{\sigma_v^2}$$

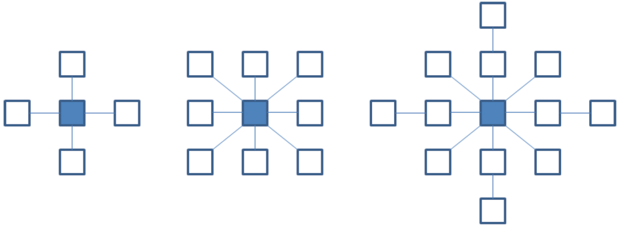
- Several possibilities for $\boldsymbol{\Lambda}_v$:
  - $\mathbf{I}$
  - $\boldsymbol{\Lambda}_v(i,j) = \rho_v^{|i-j|}$: AR(1) structure
    - $\rho_v \stackrel{iid}{\sim} \mathrm{Unif}(-1,1)$
    - EB approach, $\hat{\rho}_v$ as the MLE

Prior distributions (2/2)

- γ have binary spatial Ising priors

$$\pi(\gamma \mid \theta) \propto \exp\left\{\sum_{v=1}^{N} \alpha_v \gamma_v + \theta \sum_{v \sim k} \omega_{v,k} I(\gamma_v = \gamma_k)\right\}$$

where $\alpha_v = \log \frac{P(\gamma_v=1)}{1-P(\gamma_v=1)}$ and $\theta \sim \mathrm{Unif}(0, \theta_{\max})$



Posterior inferences (1/3)

- Full posterior computationally prohibitive
- However, is it really needed?
- Focus on the following quantities:
  - Activation probabilities:
    $$\pi\left(\gamma_{v,j} = 1 \mid \mathbf{y}\right)$$
  - Effect magnitudes:
    $$\mathbb{E}\left[\boldsymbol{\beta}_v \mid \mathbf{y}\right]$$
- The rest are mere details...

Posterior inferences (2/3)

- We know that:

$$\mathbb{E}\left[\boldsymbol{\beta}_v \mid \mathbf{y}\right] = \sum_{\gamma_v} \mathbb{E}\left[\boldsymbol{\beta}_v \mid \gamma_v, \mathbf{y}\right] \pi\left(\gamma_v \mid \mathbf{y}\right)$$

- Also:

$$\pi\left(\gamma_{v,j} = 1 \mid \mathbf{y}\right) = \int \pi\left(\gamma_{v,j} = 1 \mid \rho_v, \gamma_{-(v,j)}, \mathbf{y}\right) \times \\ \times \pi\left(\rho_v \mid \mathbf{y}\right) \pi\left(\gamma_{-(v,j)} \mid \mathbf{y}\right) d\rho_v d\gamma_{-(v,j)}$$

- Thus we only need to know $\pi\left(\gamma, \boldsymbol{\rho} \mid \mathbf{y}\right)$

Posterior inferences (3/3)

- Now we can approximate:

$$\mathbb{E}\left[\boldsymbol{\beta}_v \mid \mathbf{y}\right] \approx \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_v \left(\boldsymbol{\gamma}_v^{[k]}\right)$$

and:

$$\pi\left(\gamma_{v,j} = 1 \mid \mathbf{y}\right) \approx \frac{1}{K} \sum_{k=1}^{K} \pi\left(\gamma_{v,j} = 1 \mid \rho_v^{[k]}, \gamma_{-(v,j)}^{[k]}, \mathbf{y}\right)$$

- $\boldsymbol{\gamma}^{[k]}, \boldsymbol{\rho}^{[k]}$ via MCMC (see paper)
- Activation probability threshold: 0.8722

Simulation study

- 30×30 2D image
- 100 times points
- Signal simulated from:

$$\mathbf{y}_v = \mathbf{X}_v(\gamma_v)\boldsymbol{\beta}_v(\gamma_v) + \boldsymbol{\epsilon}_v, \quad \epsilon_v \sim \mathcal{N}_{100}\left(0, \sigma_v^2 \boldsymbol{\Lambda}_v\right)$$

- One regressor, 5% signal; rest parameters fixed/simulated from priors
- $B = 10$ runs in total

Sensitivity analyses (1/2)

- ρ on $\theta = 0.7$

Prior for ρ	Estimate $\hat{\theta}$	MCSE
Uniform$(-1, 1)$	0.73	0.0017
EB	0.74	0.0015
$\Lambda_v = I_{100}$	0.78	0.0015

- ρ on accuracy

Prior for ρ	$\Lambda_v = I_{100}$	Uniform$(-1, 1)$	EB
Accuracy (%)	91.38	97.38	97.16
False Positive Rate (%)	13.59	0.045	0.04

- ω on accuracy

Weight	1/2	1	2
Accuracy (%)	97.11	97.28	97.25
False Positive (%)	1.20	1.36	1.34

Sensitivity analyses (2/2)

- Activation probability threshold on accuracy

Critical value	.7946	.8722	.9650
Accuracy (%)	97.44	97.28	95.93
False Positive (%)	2.30	1.36	0.50

- EB AR(1) under assumption violations

Models	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1, 1)
Acc (%)	97.38 (0.97)	96.40 (0.97)	95.98 (0.67)	97.88 (1.75)	96.01 (0.65)
FP (%)	0.68 (0.041)	2.23 (0.077)	0.97 (0.041)	0.98 (0.067)	0.97 (0.043)

- Identity correlation under assumption violations

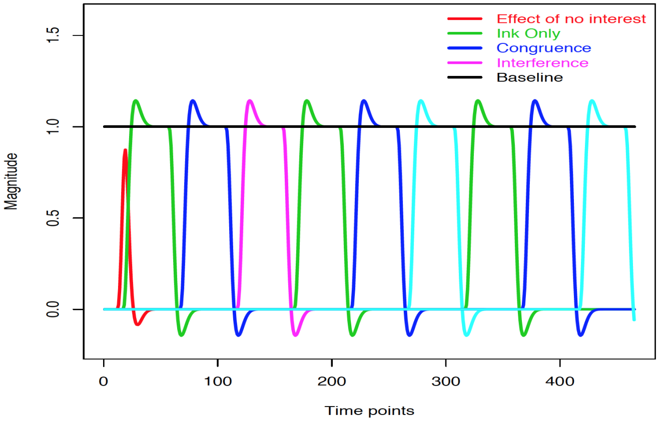
Models	AR(1)	AR(2)	MA(1)	MA(2)	ARMA(1, 1)
Acc (%)	95.11 (1.26)	94.43 (1.47)	98.56 (0.07)	96.88 (0.08)	93.67 (1.65)
FP (%)	8.74 (0.41)	10.22 (0.87)	0.10 (0.004)	6.00 (0.64)	9.79 (1.43)

Data analysis

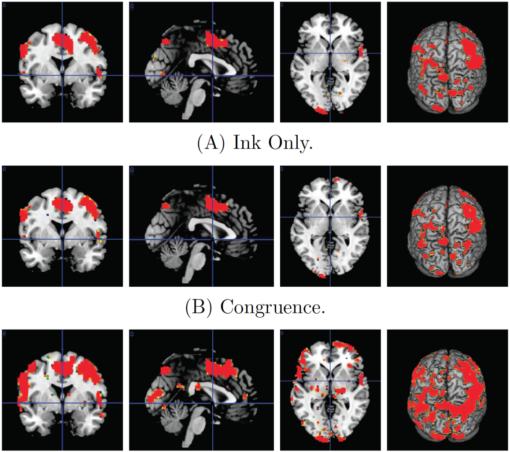
- Analysis of the dataset described earlier
- EB for correlation parameters
- Activation probability threshold 0.8772
- Weights: reciprocal of Euclidian distance
- 2 models:
    - I) Activation patterns constant (focus: regions)
    - II) Activation patterns change (focus: changes over time)

Design matrix

- Design matrix, convolved with HRF function:



Model I

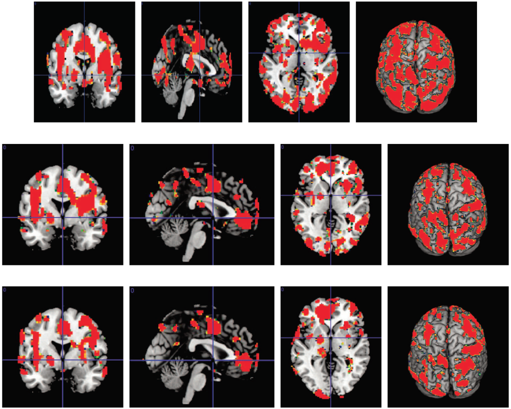


(A) Ink Only.

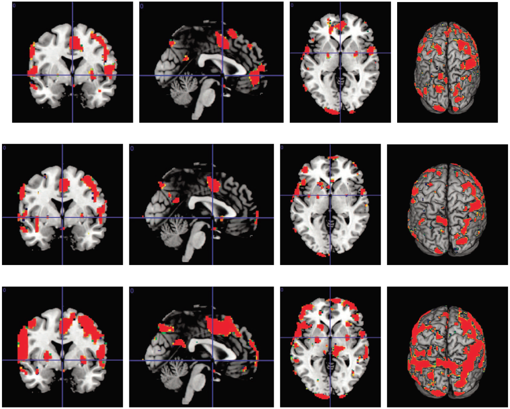
(B) Congruence.

(C) Interference.

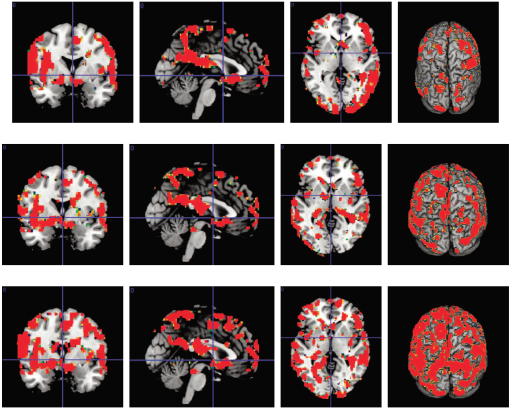
Model II: trial 1



Model II: trial 2



Model II: trial 3



Discussion

- Current model can:
  - Introduce both spatial and temporal correlations
  - Facilitates variable selection in fMRI regression
- But:
  - Interpretability not what practitioners used to
  - Cannot be applied to group-analyses
- **MORE AND MORE BAYESIAN MODELS APPLIED IN NEUROIMAGING!!!!!!**

THANK YOU!!!