Bayesian Modular & Multiscale Regression

Michele Peruzzi
with David Dunson

JSM 2018

Università Bocconi and Duke University

Multiscale regression

A single-resolution regression model is the usual

$$y = X\beta + \varepsilon$$

With multiple resolutions available, we may use

$$y = X_1\theta_1 + \cdots + X_K\theta_K + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

2 points of view

$$y = X_1\theta_1 + \cdots + X_K\theta_K + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

Two *multiscale points of view*:

1. multiple scales $X_1, \ldots, X_K$ are **pre-determined**.
   Goal = use them together in the same model
2. multiscale **interpretation** of single-scale data: we **create** scales $X_1, \ldots, X_K$, based on original data X, for interpretation purposes.

Compare with Wavelets:

1. cannot use prespecified scales
2. specific inflexible orthogonal decomposition of the data

But...

$$y = X_1\theta_1 + \cdots + X_K\theta_K + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

- a basic multiscale structure takes low resolutions as coarsening of high resolutions
$$X_j = X_{j+1}L_j$$
- this results in $\theta_1, \ldots, \theta_K$ being **not identifiable**

Bayesian modularization:

- break dependence, "cut"
- split the problem into smaller subproblems

BM&Ms

$$y = X_1\theta_1 + \cdots + X_K\theta_K + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

Module j $\in \{1, K\}$:

- $e_0 = y$
- sample $\tilde{\theta}_j \sim p(\theta_j | e_{j-1}, X_j)$ using model $e_{j-1} = X_j\theta_j + \varepsilon_j$
- Calculate $e_j = e_{j-1} - X_j\tilde{\theta}_j$

Modular posterior = collection of modules' posteriors

$$p_M(\theta_1, \ldots, \theta_K | y, X_{1:K}) = \pi_1(\theta_1 | y, X_1) \cdots \pi_K(\theta_K | e_{K-1}, X_K)$$

In **large samples and with K=2 modules**, BM&Ms posterior mean is

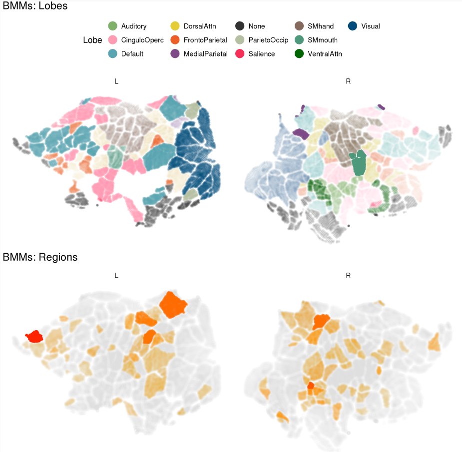
$$\bar{\mu}_{1:2} \approx \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 - L_1\hat{\beta}_1 \end{bmatrix}$$

where $\hat{\beta}_j = (X_j'X_j)^{-1}X_j'y$ and L_1 stretches $\hat{\beta}_1$ to the dimension of $\hat{\beta}_2$.
"Regression on residuals" + propagates uncertainty

Gender classification: (1) Predetermined scales

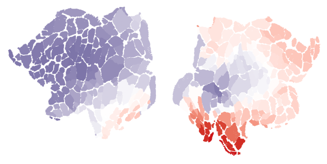
- X_1 corresponding to 26 lobes and X_2 to 333 regions for $n = 100$ subjects
- Spike-slab modules (variable selection prior for each scale)
- Using lobes, **fewer regions selected**. Accuracy $\approx 79\%$

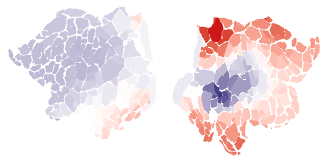
Posterior selection probability of lobes (top) and regions (bottom)

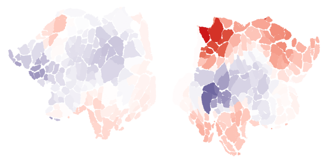


Gender classification: (2) Multiscale interpretation

- Use 333 regions and their centroid location in flattened 2D brain
- adaptively aggregate adjacent regions to create $X_1, \ldots, X_{K=5}$
- at most 30 groups (macroregions) necessary for best performance
- Below, model averaged results for $j = 1, 3, 5$







Out-of-sample acc. = 81.5%