LAWS: A Locally Adaptive Weighting and Screening Approach to Spatial Multiple Testing

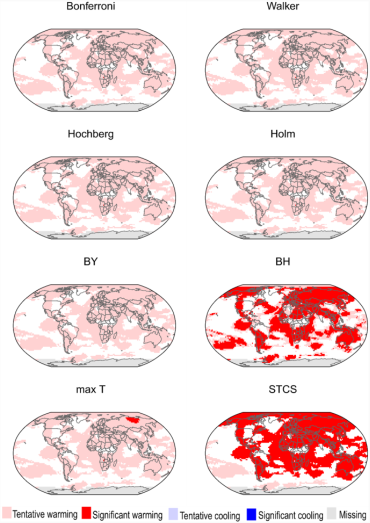
Jieun Shin

Table of Contents

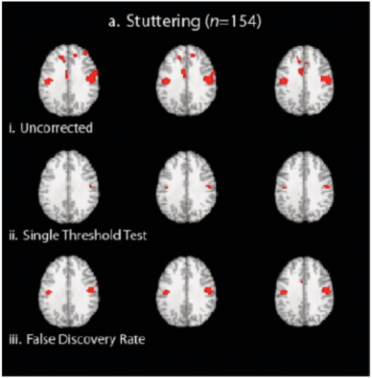
1. Introduction

1. Introduction

- Spatial multiple testing



(a) environmental study



(b) neuroimaging

1. Introduction

- Exploiting spatial structures can help identify signals more accurately and improve the interpretability of FDR analyses.
- The spatial structures and covariates are used to form new hypotheses.
- Most spatial multiple testing methods have assumed that clusters are known a priori, or the dependence structure can be estimated from data.

1. Introduction

Main idea

- To develop FDR methods for spatial analysis that are capable of adaptively learning the sparse structure.
- Data-driven procedure without prior knowledge on cluster, parametric assumption.
- Inference with auxiliary information, spatial location.

1. Introduction

A locally adaptive weighting and screening (LAWS)

1. Estimate the local sparsity structure using a screening approach.
2. Constructs spatially adaptive weights to reorder the p-values.
3. Chooses a threshold to adjust for multiplicity.

2. Model and Problem Formulation

2. Model and Problem Formulation

- Let $\mathcal{S} \subset R^d$ donate a d-dimensional spatial domain.
- Let $\mathbb{S} \subset \mathcal{S}$ define a finite, regular lattice where hypotheses are located on.
- Data are observed at every location $s \in \mathbb{S}$.
- In this paper, consider the infill-asymptotics framework and assume $\mathbb{S} \longrightarrow \mathcal{S}$

2. Model and Problem Formulation

- Let $\theta(s)$ be a binary variable, with $\theta(s) = 1$ and $\theta(s) = 0$, respectively, indicating the presence and absence of a signal of interest at location s.
- A multiple testing problem for identifying of spatial signal:

$$H_0(s) : \theta(s) = 0 \quad vs \quad H_1(s) : \theta(s) = 1, \quad s \in \mathbb{S}.$$

- Let $\{T(s) : s \in \mathbb{S}\}$ be the summary statistic at location s, $p(s)$ is a p-value corresponding to $T(s)$
- The conditional cumulative distribution functions (CDF) of the p-values are given by

$$\mathbb{P}\{p(s) \leq t|\theta(s)\} = \{1 - \theta(s)\}t + \theta(s)G_1(t|s),$$

where $t \in [0, 1]$ and $G_1(t|s)$ is the non-null p-value CDF at s.

2. Model and Problem Formulation

- Define sparsity level at location s

$$\pi(s) = \mathbb{P}\{\theta(s) = 1\}.$$

- smoothness: $\pi(s)$ varies smoothly as a continuous function of s.
  - provide the key structural information.
  - can be exploited to integrate information from nearby locations

2. Model and Problem Formulation

- The decision at location s, testing unit, is represented by a binary variable $\delta(s)$, where

$$\delta(s) = \begin{cases} 1, & \text{if } \mathrm{H}_0(s) \text{ is rejected,} \\ 0, & \text{o.w.} \end{cases}$$

- The BH FDR is defined as

$$FDR = \mathbb{E}\left[\frac{\sum_{s\in\mathbb{S}}\{1-\theta(s)\}\delta(s)}{\max\{\sum_{s\in\mathbb{S}}\delta(s), 1\}}\right].$$

- The power of FDR procedure $\boldsymbol{\delta} = \{\delta(s) : s \in \mathbb{S}\}$ can be evaluated using the expected number of true positive:

$$ETP(\boldsymbol{\delta}) = \mathbb{E}\left[\sum_{s\in\mathbb{S}}\theta(s)\delta(s)\right]$$

3. Structure-Adaptive Weighted p-value

3. Structure-Adaptive Weighted p-value

- how to improve the ranking by exploiting the spatial pattern and constructing structure-adaptive weights to adjust the p-values.
- covariate-adjusted mixture model:

$$X(s) \overset{iid}{\sim} f(x|s) = \{(1-\pi(s)\}f_0(x|s) + \pi(s)f_1(x|s),$$

  where the covariate s encodes useful side information, $f_0(x|s)$ and $f_1(x|s)$ are the null and non-null densities, $\pi(s)$ is the sparsity level and $f(x|s)$ is the mixture density.
- Define the conditional local FDR

$$CLfdr(x|s) = \mathbb{P}(\theta(s) = 0|x, s) = \frac{\{1-\pi(s)\}f_0(x|s)}{f(x|s)}$$

  its thresholding rule is optimal in the sense that it maximizes the ETP subject to the constraint on FDR.

3. Structure-Adaptive Weighted p-value

- However, CLfdr cannot handle dependent tests.
- Let
$$\Lambda(x|s) = \frac{1-\pi(s)}{\pi(s)} \frac{f_0(x|s)}{f_1(x|s)}.$$
Then $\text{CLfdr} = \Lambda/(\Lambda+1)$ is monotone in Λ.
- the first term means the information of the sparsity structure that reflects how frequently signals appear in the neighborhood.
- the second term means the information exhibited by the data itself that indicates the strength of evidence against the null.
- the second term is replaced by the *p*-value because it is difficult to calculate.

3. Structure-Adaptive Weighted p-value

- Then, define the weighted p-values:

$$p^w(s) = \min\left\{\frac{1-\pi(s)}{\pi(s)}p(s), 1\right\} = \min\left\{\frac{p(s)}{w(s)}, 1\right\}, \quad s \in \mathbb{S},$$

where $w(s) = \pi(s)/(1-\pi(s))$.

4. Spatial Multiple Testing by LAWS

4-1. Sparsity Estimation via Screening

- estimate (i) the sparsity level $\pi(s)$ and (ii) threshold t_w
- Since the direct estimation of $\pi(s)$ is very difficult, we instead introduce an intermediate quantity to approximate $\pi(s)$:

$$\pi^{\tau}(s) = 1 - \frac{\mathbb{P}(p(s) > \tau)}{1 - \tau}, \quad 0 < \tau < 1.$$

4-1. Sparsity Estimation via Screening

- In the smoothing step, we exploit the structural assumption that $\pi^{\tau}(s)$ varies as a smooth function of spatial location s.
- A kernel function K to assign weights to observation according to their distances to s.
- For any given grid $\mathbb{S}$ on $\mathcal{S} \subset \mathbb{R}^d$, let $K : \mathbb{R}^d \to \mathbb{R}$ be a positive, bounded and symmetric kernel function satisfying

$$\int_{\mathbb{R}^d} K(t)dt = 1, \quad \int_{\mathbb{R}^d} tK(t)dt = 0, \quad \int_{\mathbb{R}^d} t^{\mathsf{T}} tK(t)dt < \infty.$$

- Denote by $K_h(t) = h^{-1}K(t/h)$, where h is the bandwidth.

4-1. Sparsity Estimation via Screening

- At location s, define

$$v_h(s, s') = \frac{K_h(s - s')}{K_h(0)},$$

for all $s' \in \mathbb{S}$

- $K_h(s - s')$ is computed as a function of the Euclidean distance $||s - s'||$, $h > 0$ is a scalar.
- The quantity $m_s = \sum_{s' \in \mathbb{S}} v_h(s, s')$ as the "total mass" or "total number observations" at location s.
- Thus, m_s is calculated by borrowing strength from points close to s while placing little weight on points far apart from s.

4-1. Sparsity Estimation via Screening

- Next in screening step, we first apply a screening procedure with threshold τ to obtain a subset $\mathcal{T}(\tau) = \{s \in \mathbb{S} : p(s) > \tau\}$
- The empirical count is given by

$$\sum_{s' \in \mathcal{T}_\tau} v_h(s, s').$$

- And the expected count can be calculated theoretically as

$$\{\sum_{s' \in \mathbb{S}} v_h(s, s')\}\{1 - \pi^\tau(s)\}(1 - \tau).$$

- Setting Equations equal, we obtain the following estimate

$$\hat{\pi}^\tau(s) = 1 - \frac{\sum_{s' \in \mathcal{T}_\tau} v_h(s, s')}{\sum_{s' \in \mathbb{S}} v_h(s, s')(1 - \tau)}$$

4-2. Data-Driven Procedure

- Define the locally adaptive weights

$$\hat{w}(s) = \frac{\hat{\pi}(s)}{1 - \hat{\pi}(s)}, \quad s \in \mathbb{S}$$

  where $\hat{\pi}(s)$ is estimated by the screening appreoach.
- For the stability of the algorithm, we take $\hat{\pi}(s) = (1 - \nu)$ if $\hat{\pi}(s) > 1 - \nu$ and take $\hat{\pi}(s) = \nu$ if $\hat{\pi}(s) < \nu$ with $\nu = 10^{-5}$.
- Next we order the weighted p-values from the smallest to largest.

4-2. Data-Driven Procedure

- If $\pi(s)$ is known and the threshold is given by t_w, then the expected number of false positives (EFP) can be calculated as

$$\mathrm{EFP} = \sum_{s\in\mathbb{S}} \mathbb{P}\{p^w(s) \leq t_w|\theta(s) = 0\}\mathbb{P}\{\theta(s) = 0\} = \sum_{s\in\mathbb{S}} \pi(s)t_w.$$

- If $t_w = p^{\hat{w}}_{(j)}$, then $\sum_{s\in\mathbb{S}} \hat{\pi}(s)p^{\hat{w}}_{(j)}$ rejections are likely to be false positive.
- And $j^{-1}\sum_{s\in\mathbb{S}} \hat{\pi}(s)p^{\hat{w}}_{(j)}$ provides a good esimate of the false discovery proportion (FDP).

4-2. Data-Driven Procedure

The LAWS procedure

1. Order the weighted p-values from the smallest to largest $p^{\hat{w}}_{(1)}, \ldots, p^{\hat{w}}_{(m)}$ and denote corresponding null hypotheses $H_{(1)}, \ldots, H_{(m)}$.
2. Let $k^{\hat{w}} = \max\{j : j^{-1} \sum_{s \in \mathbb{S}} \hat{\pi}(s) p^{\hat{w}}_{(j)} \leq \alpha\}$.
3. Reject $H_{(1)}, \ldots, H_{(k^{\hat{w}})}$.

5. Simulation

5. Simulation

- To create the screening subset $\mathcal{T}$, choose τ as the p-value threshold of the BH procedure at $\alpha = 0.9$
- Generate $|\mathbb{S}|$ hypotheses from the following normal mixture model:

$$X(s) \overset{iid}{\sim} \{(1-\theta(s)\}N(0,1) + \theta(s)N(\mu,1), \quad \theta(s) \sim Bernoulli(\pi(s))$$

- One dimensional Setting: linear block, triangle block pattern
- Two and three dimensional setting with same patterns

5-1. One Dimensional Setting with Piece-Wise Constants

- Setup: $m = 5000$, and the signals appear

$$\pi(s) = 0.9 \text{ for } s \in [1001, 1200] \cup [2001, 2200];$$
$$\pi(s) = 0.6 \text{ for } s \in [3001, 3200] \cup [4001, 4200],$$

$\pi(s) = 0.01$ for rest of the locations.

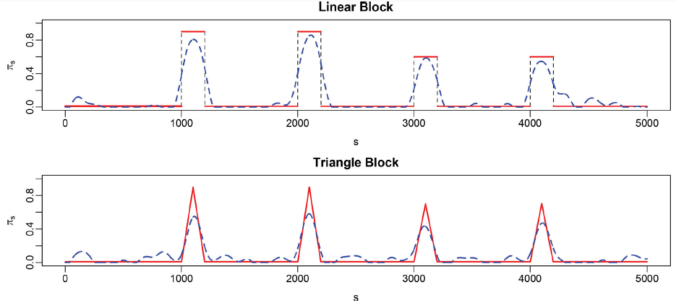


Figure 2: True $\pi(s)$(solid line) vs estimated $\pi(s)$(dashed line) in one-dim.

5-1. One Dimensional Setting with Piece-Wise Constants

- the top row: vary μ from 2 to 4.
- the bottom row: fix $\mu = 2.5$, $\pi(s)$ very 0.3 to 0.9 in signal and $\pi(s)$ fix 0.01 the rest.
- The empirical FDR and power are computed over 200 replications with nominal level $\alpha = 0.05$

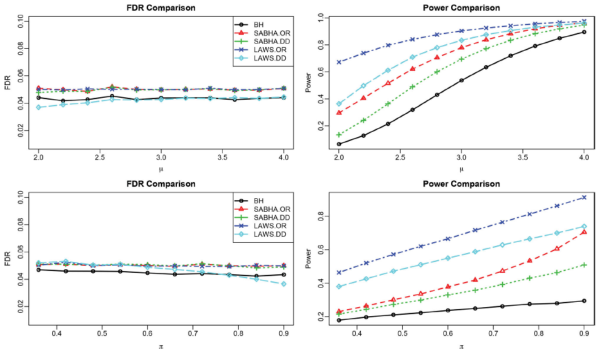


Figure 3: FDR and power comparisons in the linear block pattern.

5-2. One Dimensional Setting with Triangular Patterns

- The setting is same with linear block

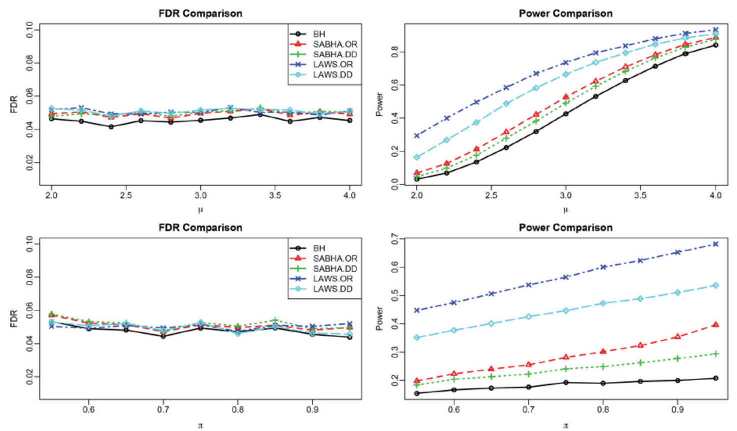


Figure 4: FDR and power comparisons in the triangular pattern.

5-3. Two-Dimensional Setting

- Generate data on a 200×200 lattice where the signals located on a double triangle or a rectangle shape.
- Let $\pi(s) = 0.9$ for the triangle and left half of the rectangle,
- $\pi(s) = 0.6$ for the right partition,
- $\pi(s) = 0.01$ for the rest.

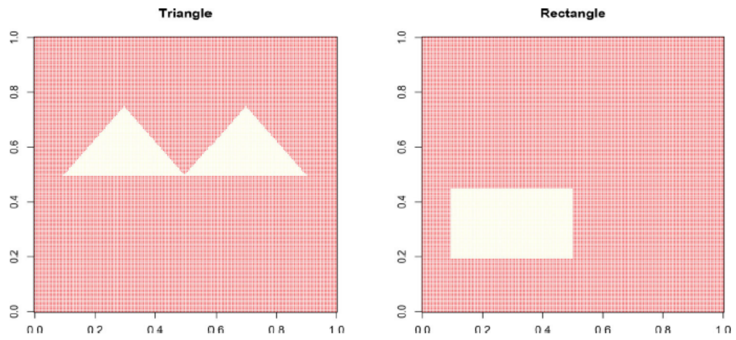


Figure 5: Two-dimensional triangle and rectangle pattern.

5-3. Two-Dimensional Setting

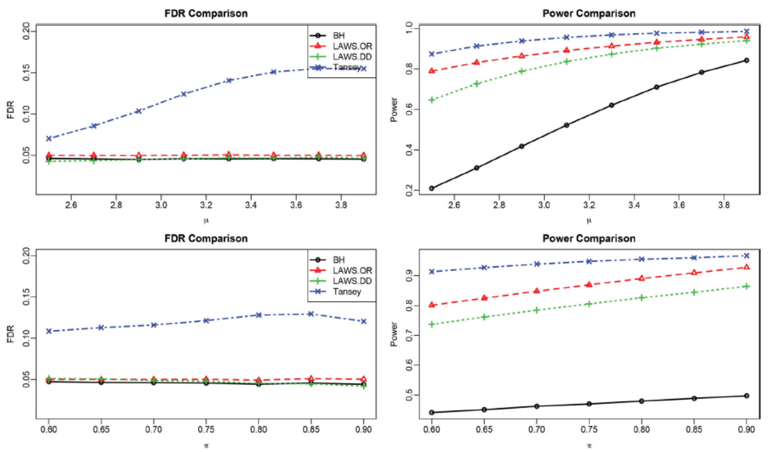


Figure 6: FDR and power comparisons in two-dimensional triangle pattern.

5-3. Two-Dimensional Setting

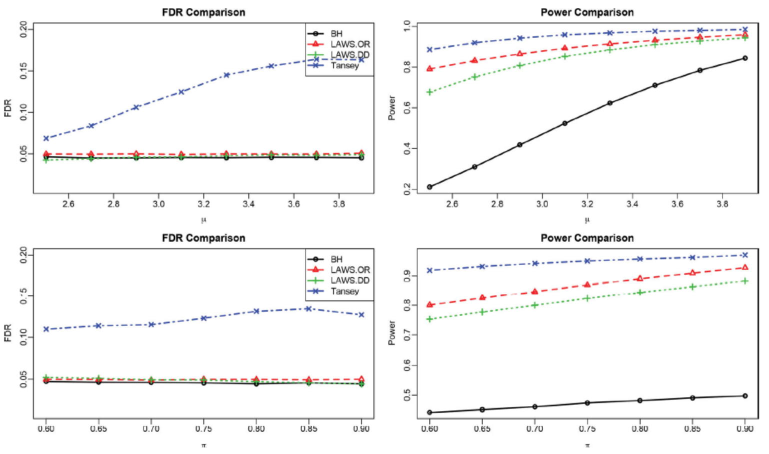


Figure 7: FDR and power comparisons in two-dimensional rectangle pattern.

5-4. Three-Dimensional Setting

- Generate the data on a $20 \times 25 \times 30$ lattice, where the signals are located on a cubic with $10 \times 10 \times 15$.

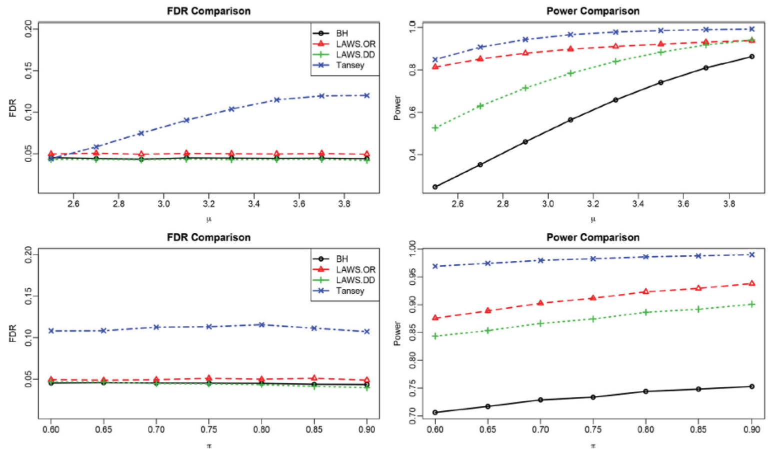


Figure 8: FDR and power comparisons in three-dimensional cubic pattern.

6. Application

6. Application

1. to identify two-dimensional spatial clusters.
2. to identify signal in three-dimensional image data.

6-1. Two-Dimensional Setting with Spatial Clusters

- Simulate data on a 200×200 lattice.
- Form two spatial clusters, donut and square shapes, which are signals.

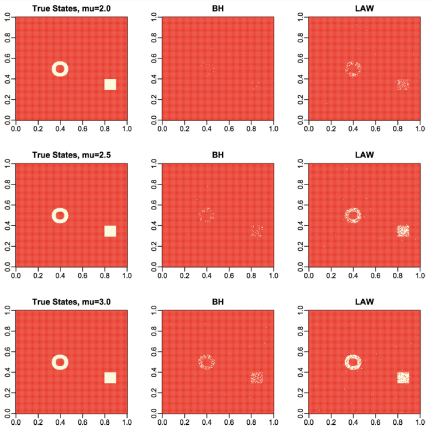


Figure 9: Spatial FDR analysis in two-dimensional setting.

6-2. Three-Dimensional Setting: fMRI Data

- A MRI data for a study of ADHA.
- Reduce the resolution of images from $256 \times 198 \times 256$ to $30 \times 36 \times 30$
- 931 subjects (356 are with ADHD and 575 are normal).
- Test statistics are computed by two-sample t-test.
- p-values are obtained by normal approximation.

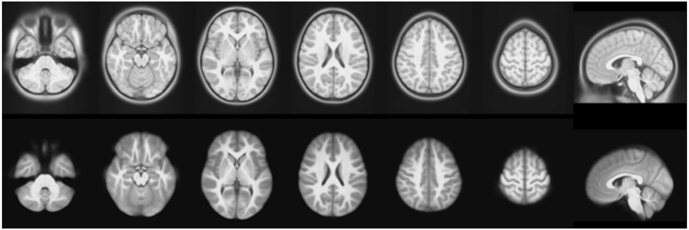


Figure 10: MRI image from Neuro Bureau

6-2. Three-Dimensional Setting: fMRI Data

- The LAWS identifies 538 regions, while BH recovers 349.
- LAWS has superior power performance over BH in identifying spatial singals.

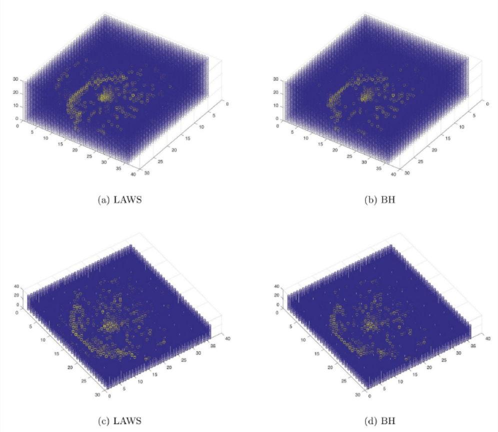


Figure 11: Significant brain regions (yellow), LAWS (left), BH (right).