Spatial CUSUM for Signal Region Detection

Xin Zhang
*Iowa State University*, xinzhang@iastate.edu

Zhengyuan Zhu
*Iowa State University*, zhuz@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/creativecomponents

Part of the *Applied Statistics Commons*, and the *Statistical Methodology Commons*

**Recommended Citation**

Zhang, Xin and Zhu, Zhengyuan, "Spatial CUSUM for Signal Region Detection" (2018). *Creative Components*. 120.
https://lib.dr.iastate.edu/creativecomponents/120

This Creative Component is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Creative Components by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Spatial CUSUM for Signal Region Detection

Xin Zhang
Department of Statistics, Iowa State University

Zhengyuan Zhu
Department of Statistics, Iowa State University

Abstract
Detecting weak clustered signal in spatial data is important but challenging in applications such as medical image and epidemiology. A more efficient detection algorithm can provide more precise early warning, and effectively reduce the decision risk and cost. To date, many methods have been developed to detect signals with spatial structures. However, most of the existing methods are either too conservative for weak signals or computationally too intensive. In this paper, we consider a novel method named Spatial CUSUM (SCUSUM), which employs the idea of the CUSUM procedure and false discovery rate controlling. We develop theoretical properties of the method which indicates that asymptotically SCUSUM can reach high classification accuracy. In the simulation study, we demonstrate that SCUSUM is sensitive to weak spatial signals. This new method is applied to a real fMRI dataset as illustration, and more irregular weak spatial signals are detected in the images compared to some existing methods, including the conventional FDR, FDR_L and scan statistics.

Keywords: Spatial signal detection, CUSUM, FDR, weak dependence, fMRI.

1. Introduction
Spatial signal detection is an important topic in many fields, including astrophysics (Abazajian and Kaplinghat (2012); Gladders and Yee (2000)), brain imaging analysis (Craddock et al. (2012); Zhang et. al. (2011); Blumensath et al. (2013); Shen et al. (2013)), epidemiology (Kulldorff and Nagarwalla (1995); Tango (2000); Wheefer (2007)), meteorology (Sun et al. (2015)) etc. Typically, given a spatial domain \( \mathcal{D} \), e.g. an brain image or a geographical map, if there is no spatial signal, all the observations could be regarded to follow the same distribution. While with the existing of spatial signals, the responses within a unknown sub-region are from a different distribution. Locating signal regions with low signal-noise ratio is meaningful in the early detection and warning systems: In the early stage of abnormality, the spatial signal is very weak compared with the measurement noise; however, an accurate early warning could effectively reduce the decision risk and avoid unnecessary but lethal cost. Such warning systems have been studied and applied in many practical cases, e.g. disease and weather monitoring (Thomson and Connor (2001); Grover-Kopec et al. (2005); Breed (2011)). Therefore, there will be a huge breakthrough if weak spatial signals can be efficiently identified.

Thus far, many methods and algorithms have been developed for spatial signal detection. One class of approaches to identify spatial clusters is the spatial scan statistics (Glaz et al. (2009); Glaz et al. (2001); Priebe et al. (2005); Glaz and Balakrishnan (2012) etc.) Scan statistics, also known as window statistics, was first proposed in Naus (1965). The idea is to perform likelihood ratio tests on all the scan windows of different sizes and locations and identify the significant windows.
as clusters. This method was designed to find unusual clusters of randomly positioned points. Naus (1982) developed the asymptotic distribution for the scan statistics and proposed the method to find the maximum cluster of points on a line or circle, the length of the longest success run in Bernoulli trials, and the generalized birthday problem. Kulldorff (1999) extended the framework of the conventional scan statistics to multidimensional scenario, including two-dimensional scan statistics on the plane or on a sphere and three-dimensional scan statistics in space or in spacetime. However, if the shape of the true cluster is not circle or ellipsoid, the power of the traditional scan statistics will significantly reduce. Additionally, without p-value adjustment, the detection result from scan statistics might be too aggressive (Zhang et al. (2010)).

Another branch of the detection methods is based on multiple testing and false discovery rate (FDR) controlling (Benjamini and Hochberg (1995); Benjamini and Yekutieli (2001); Genovese et al. (2002); Miller et al. (2001); Zhang et al. (2011); Tango (2000); Sun et al. (2015)). Multiple hypothesis testing is concerned with testing several statistical hypotheses simultaneously, and false discovery rate is a criterion designed to control the expected proportion of rejected null hypotheses that are incorrect rejections:

\[
\text{FDR} = \mathbb{E}\left[ \frac{\# \text{incorrect rejections}}{\# \text{rejected null hypotheses}} \right].
\]

In spatial signal detection, the statistical hypotheses are about whether locations belong to signal region or not. Genovese et al. (2002) applied multiple testing to functional neuroimaging data and used FDR to find a threshold for signal classification. Their experiments showed that FDR worked more conservatively when the correlations between hypotheses are high. Miller et al. (2001) applied FDR procedure to astrophysical data, and showed that FDR had a similar rate of correct detections and significantly less false detections compared with certain standard testing procedures. In Tango (2000), multiple testing was used to detecting spatial disease clusters. To improve detecting effectiveness, Zhang et al. (2011) proposed a testing procedure named FDR\(_L\) with the consideration of the spatial structures. By aggregating the local p-values, FDR\(_L\) could avoid the lack of identification phenomenon and improve the detection sensitivity. Sun et al. (2015) developed an oracle procedure, which optimally controlled the false discovery rate, false discovery exceedance and false cluster rate, for multiple testing of spatial signals. The tropospheric ozone data in eastern USA were analyzed with their method to show the detection effectiveness. Although FDR and its variants have good statistical interpretation and can be easily implemented, the prior knowledge about the null distribution is required and many signals can be missed when signal-noise ratio is small.

In this paper, we will introduce a novel detecting method named Spatial CUSUM (SCUSUM) to identify spatial signal regions. We assume the expected value of the signal regions is different (usually higher) from that of the indifference regions, and noise processes are zero-mean and independent. SCUSUM has two steps: First applying moving window and CUSUM cut-off to estimate signal weight for each location; then determining a threshold with FDR controlling. Moving window method has been broadly used to analyze temporal and spatial data (Páez et al. (2008); Haas (1990)), where the neighboring data is utilized to capture local feature. Similarly, in our work, the spatial data is projected into an array, which can be analyzed by the CUSUM procedure. The CUSUM procedure (or cumulative summation) is well-known to locate changepoints in time series (Horváth and Hušková (2012); Cho et al. (2016); Gromenko et al. (2017); Wang and Samworth (2018); Aue et al. (2009)). With the CUSUM transformation, the testing statistics can be compared with the standard Brownian Bridge to test the existence of the changepoint. The location of
changepoint is where CUSUM reaches the maximum. In our work, by repeating CUSUM cut-off with moving window, the detection frequency can be calculated for each location. We define this frequency as signal weight. Then, the null density and alternative density can be approximated by bounded density estimation method. A proper threshold could be found based on FDR to identify the singal region. Our theoretical results show that SCUSUM could asymptotically reduce the misclassification rate to zero with probability 1. The experiment section support that our method could detect more weak spatial signals, compared with the existing methods, including FDR, FDR$_L$ and scan statistics.

The rest of the paper is organized as follows. The spatial signal detection problem is formulated mathematically in Section 2. The details of our proposed method are introduced in Section 3. Section 4 presents simulation comparisons between SCUSUM and FDR$_L$ under different signal strengths and noise dependence structures. An application of four methods (SCUSUM, scan statistics, FDR and FDR$_L$), to a real fMRI data is given in Section 5. We give the conclusions in Section 6. The proofs are shown in Appendix.

2. Problem Formulation

Let $\mathcal{D}$ be the entire spatial domain, $s$ present the location belonging to $\mathcal{D}$ and $x(s)$ be the observed data at location $s$. Consider $\mathcal{D}_d$ be the signal region in $\mathcal{D}$, and its complement set $\mathcal{D}_d^c$ be the indifference region. We assume that under $H_0$, there is no signal region (i.e. $\mathcal{D}_d = \emptyset$) and $x(s)$ has the same mean process $\mu$; while under $H_1$, $x(s)$ has mean $\mu_1$ if $s \in \mathcal{D}_d$ and $\mu_0$ if $s \in \mathcal{D}_d^c$.

Hence, the following additive model for the random variables $X = \{x(s), s \in \mathcal{D}\}$ is considered:

$$x(s) = \mu_0 \mathbb{I}(s \in \mathcal{D}_d^c) + \mu_1 \mathbb{I}(s \in \mathcal{D}_d) + \epsilon(s),$$

(2)

where both $\mu_0$ and $\mu_1$ are the unobserved mean (w.l.o.g, assume $\mu_1 \geq \mu_0$,) $\epsilon(s)$ is the independent noise and $\mathbb{I}(\cdot)$ is the indicator function. Note here, we don’t assign any distribution model to the noise. The only requirement for noise is that it has zero mean and i.i.d. Our goal is to identify the signal region $\mathcal{D}_d$.

3. Proposed method

In this section, we will give the details of our proposed method named Spatial CUSUM (SCUSUM), which has two steps: 1) For each location, we first estimate the signal weight, which is expected to be large in the signal region $\mathcal{D}_d$, while small in the indifference region $\mathcal{D}_d^c$. 2) Given a significant level $\alpha$, a threshold is determined based on FDR idea. In section 3.1, we describe the way of using the moving window idea to project a spatial domain into a sequence and then estimating the signal weight for each location with the CUSUM cut-off. In section 3.2, we introduce how to estimate the null distribution $f_{H_0}$ and alternative distribution $f_{H_1}$ with estimated signal weights, following by the step to determine the detection threshold. In section 3.3, we briefly discuss neighbor size selection for moving window in the first step.

3.1. The first step: signal weight estimation

Our signal weight estimation method is inspired by the CUSUM procedure for changepoint detection in time series. However, for spatial signal detection, the conventional CUSUM is impractical,
mainly due to lack of natural order for spatial observations, which are located in $\mathbb{R}^2$ or $\mathbb{R}^3$ (in our work, we focus on $\mathbb{R}^2$.) Hence, we consider to use the moving window technique to construct proper sequences.

Given a square spatial domain $\mathcal{D}$ and a neighbor size $k$, we can select one point $(x, y)$ from $\{(p + \frac{1}{2}, q + \frac{1}{2}), \ p = 1 : k, \ q = 1 : k\}$. Then we can divide $\mathcal{D}$ into $b$ non-overlapping blocks $\{B_i\}_{i=1}^b$ of size $k \times k$, with $(x, y)$ as the initial grid point (see Figure 1.) So we have $n = \sum_{i=1}^b n_i$, where $n$ is the total number of observations and $n_i$ is the observations in the block $B_i$.

According to model 2, if block $B_i$ is inside the signal region, i.e. $B_i \in \mathcal{D}_s$, then all the observations in it follow $x(s) = \mu_1 + \epsilon(s), \ \forall x \in B_i$; if $B_i$ is inside the indifference region, i.e. $B_i \in \mathcal{D}_c$, then $x(s) = \mu_0 + \epsilon(s), \ \forall x \in B_i$; if $B_i$ is at the boundary of $\mathcal{D}_s$ and $\mathcal{D}_c$, then the observations $x(s)$ follows a mixture model: $x(s) = \mu_1 + \epsilon(s)$ with probability $p_i$ and $x(s) = \mu_0 + \epsilon(s)$ with $1 - p_i$, where $p_i$ is the ratio of signal points inside $B_i$ (see Figure 2.)

Next, based on the above division, we construct two sequences to capture the feature of these spatial observations. The first sequence is sample sequence: A random sample, denoted as $\gamma_i$, is drawn from block $B_i$. It could be regarded as the ‘representative’ for this block. Meanwhile, we construct the second sequence by computing the block mean without the ‘representative’, $\tilde{\mu}_i = \frac{\sum_{x \in B_i} x - \gamma_i}{\sum_{x \in \mathcal{D}} 1(\{x \in B_i\}) - 1}$. As the number of observations $n_i$ in $B_i$ increases, the pseudo block mean gets closer to the true block mean, i.e. $\tilde{\mu}_i \rightarrow \frac{\sum_{x \in B_i} x}{\sum_{x \in \mathcal{D}} 1(\{x \in B_i\})}$. Hence, the pseudo block mean $\tilde{\mu}_i$ could present the local block mean. Based on the analysis in the last paragraph, we
could easily derive the following results for \( \{\gamma_i\}_{i=1}^b \) and \( \{\tilde{\mu}_i\}_{i=1}^b \):

\[
\gamma_i = \begin{cases} 
\mu_1 + \epsilon, & \text{if } B_i \in \mathcal{D}_{af} \\
\mu_0 + \epsilon, & \text{if } B_i \in \mathcal{D}_{cf} \\
\mu_1 z + \mu_0 (1-z) + \epsilon, & \text{if } B_i \text{ at boundary}
\end{cases}
\]  

(3)

\[
\mathbb{E}[\tilde{\mu}_i] = \begin{cases} 
\mu_1, & \text{if } B_i \in \mathcal{D}_{af} \\
\mu_0, & \text{if } B_i \in \mathcal{D}_{cf} \\
\approx p_i \mu_1 + (1-p_i) \mu_0, & \text{if } B_i \text{ at boundary}
\end{cases}
\]  

(4)

where \( z \sim \text{Ber}(p_i) \). (3) and (4) show that even though \( \{\gamma_i\}_{i=1}^b \) and \( \{\tilde{\mu}_i\}_{i=1}^b \) are independent (see Lemma 1), they have similar patterns: the closer \( B_i \) is to \( \mathcal{D}_{af} \), the more likely it has large \( \gamma_i \) and \( \tilde{\mu}_i \), and vice versa. Hence, we could consider to rearrange \( \{\gamma_i\}_{i=1}^b \) according to the decreasing order of \( \{\tilde{\mu}_i\}_{i=1}^b \), denoted as \( \{\gamma_i^*\}_{i=1}^b \). Intuitively, if there is no signal region, then \( \{\gamma_i^*\}_{i=1}^b \) should be around \( \mu_0 \); otherwise, \( \{\gamma_i^*\}_{i=1}^b \) should have three parts: the first part presenting signal blocks is around \( \mu_1 \), the second part is the interim from \( \mu_1 \) to \( \mu_0 \) and the third part is indifference blocks around \( \mu_0 \) (see Figure 3.)

Lemma 1 Based on model 2, \( \{\gamma_i\} \) and \( \{\tilde{\mu}_i\} \) are independent. As the number of observations in each block \( n_i \) goes to infinity, i.e. \( \min n_i \to \infty \), we have the following: under the null hypothesis \( H_0 \): there is no signal region, then \( \{\gamma_i\}_{i=1}^b \) is an i.i.d sequence; under the alternative hypothesis \( H_1 \): signal region exists, then there exists \( l_1 \) and \( l_2 \) with \( 0 \leq l_1 < l_2 \leq b 
\]

\[
\mathbb{E}[\gamma_i^*] = \begin{cases} 
\mu_1, & \text{if } 0 \leq i < l_1 \\
(\mu_0, \mu_1) & \text{if } l_1 \leq i < l_2 \\
\mu_0, & \text{if } l_2 \leq i \leq n,
\end{cases}
\]  

(5)

and \( l_1 \) is the number of blocks inside \( \mathcal{D}_{af} \), and \( (b-l_2) \) is the number of blocks inside \( \mathcal{D}_{af} \).

Lemma 1 shows that under \( H_1 \), the projected sequence \( \{\gamma_i^*\}_{i=1}^b \) has a changepoint in \([l_1, l_2]\) and the conventional CUSUM could help locate a cut-off index near or inside the interval. First, we compute the CUSUM statistics for \( \{\gamma_i\}_{i=1}^b \) at each location:

\[
\tilde{\gamma}_r = |\sum_{i=1}^{r} \gamma_i^* - \frac{r}{b} \sum_{i=1}^{b} \gamma_i^*|.
\]  

(6)

Then the cut-off index is \( t = \arg \max_i \tilde{\gamma}_i \). The following theorem guarantees the accuracy of the cut-off index.

Theorem 1 Under the alternative hypothesis: if the signal region exists, as the number of block \( b \) and the number of observations in each block \( n_i \) go to infinity, then the cut-off index \( t \) based on CUSUM procedure will fall into the interval \([l_1, l_2]\) with probability 1, i.e. \( \mathbb{P}(l_1 \leq t \leq l_2) = 1 \) as \( b \to \infty \) and \( \min n_i \to \infty \).
Theorem 1 ensures that the cut-off procedure could asymptotically separate the signal region and indifference (see Figure 3 (b)). Also, with the given spatial domain, as the block size becomes finer (equivalent to $b \to \infty$), the number of the blocks at the boundary is decreasing. Hence we have $(l_2 - l_1)/b \to 0$. Combining the results from Theorem 1, the number of misclassified locations goes to zero.

Above theoretical results require $b \to \infty$ and min $n_i \to \infty$. In practice, with limited observations, the detected result might be affected by the initial point selection, especially when the signal region is not regular. Thus, we could eliminate the effect of initial point by going through all the possible initial points (see Figure 1.) We summarize our method in Algorithm 1. Also, we could sufficiently extract the local information and eliminate the effect of randomly sampled ‘representatives’ by repeated Algorithm 1 more than once. With these steps, we could estimate the signal weights $\{w(s)\}$ (or $\{\tilde{w}(s)\}$) by computing the detected frequency of each location. The larger signal weight means the location is more likely to belong to the signal region.

3.2. The second step: Threshold estimation with FDR

With the estimated signal weight and a given significant level $\alpha$, we could identify the signal region with FDR. Firstly, the weights are in $[0, 1]$, and could be considered as the possibilities that the locations have signals. Thus we could use density estimation with the boundary correction method to estimate the distribution of the signal weights, $f(x)$. Many density estimations have been studied in previous works (Chen (1999); Cowling and Hall (1996); Jones and Foster (1996); Cattaneo et al. (2017).) In our work, we use the local polynomial density estimation method from Cattaneo et al. (2017).

In the following, we analyze the characteristic of $f(x)$, which could help us estimate the threshold. Under the null hypothesis $H_0$, with Lemma 1, we know that $\{\gamma_i^{*}\}_{i=1}^b$ are i.i.d. and the corresponding blocks are random indexed. Hence with CUSUM cut-off procedure, the distribution for signal weight $f(x)$ is symmetric and has lower value with $x = 0$ and $x = 1$ (see Figure 4 (a).)
Algorithm 1 Moving Window detecting method for signal weight

Require: observed data \( \{x(s)\} \) on grid \( \{(p,q), p = 1 : n, q = 1 : n\} \), neighbor size \( k \), repeat times \( m \);
Ensure: corresponding signal weights \( \{w(s)\} \) or \( \{\tilde{w}(s)\} \);

1: for \( (x,y) \) in \( \{(p + \frac{1}{2}, q + \frac{1}{2}), p = 1 : k, q = 1 : k\} \) do
2: Divide \( \mathcal{D} \) into blocks \( \{B_i\}_{i=1}^b \) of size \( k \times k \) based on \( (x,y) \);
3: Sample one observation from each block \( \gamma_i \);
4: Estimate the block mean \( \hat{\mu}_i \);
5: Reorder \( \{\gamma_i\}_{i=1}^b \) according to \( \{\hat{\mu}_i\}_{i=1}^b \) decreasingly as \( \{\gamma_i\}_{i=1}^b \);
6: Conduct CUSUM transformation on \( \{\gamma_i\}_{i=1}^b \) as \( \{\tilde{\gamma}_i\}_{i=1}^b \);
7: Find the location \( t \) where \( \{\tilde{\gamma}_i\}_{i=1}^b \) reaches maximum;
8: Define the blocks corresponding to the first \( t \) elements in \( \{\gamma_i\}_{i=1}^b \) as signal block, and so do the observations in these blocks;
9: end for
10: Compute corresponding signal weight \( w(s) = \frac{\text{detected times for } x(s)}{k^2} \);
11: Repeat above produce \( m \) times and obtain \( \{w^i(s)\}_{i=1}^m \);
12: Compute the average signal weights at each location \( \{\tilde{w}^i(s)\} : \tilde{w}^i(s) = \sum_{i=1}^m w^i(s)/m \);

Lemma 2 Under the null hypothesis \( H_0 \), as the number of block \( b \) and the number of observations in each block \( n_i \) go to infinity, then the density for signal weights \( f(x) \) is symmetric.

Under the alternative hypothesis \( H_1 \), the distribution \( f(x) \) should be composited by the null part \( f_{H_0}(x) \) and alternative part \( f_{H_1}(x) : f(x) = f_{H_0}(x) + f_{H_1}(x) \). The observations inside the signal region are more likely to be detected, i.e. the corresponding signal weight gets close to 1 and vice versa for the observations inside the indifference region. With finer block division, the fraction of the observations in the blocks at the boundary goes to 0. Therefore, \( f_{H_1}(x) \) has the ‘peak’ near 1 and \( f_{H_0}(x) \) has the ‘peak’ around 0, which implies that \( f(x) \) has two ‘peaks’ near the boundaries separately and a ‘valley’ in the middle of \([0,1]\). Then we can use the line search to locate the ‘valley’, say \((t^*, f(t^*))\), and conduct linear interpolation between the two points \((t^*, f(t^*))\) and \((1, 0)\). Obviously, the null density \( f_{H_0}(x) \) is controlled by

\[
\tilde{f}_{H_0}(x) = \begin{cases} f(x), & \text{if } 0 \leq x \leq t^* \\ f(t^*)(1 - \frac{x - t^*}{1 - t^*}), & \text{if } t^* < x \leq 1 \end{cases}
\]

(7)

which could be used as estimated null density (see Figure 4 (b).) Recall the definition of the marginal false discovery rate (mFDR) (Genovese and Wasserman (2002); Sun et al. (2015)):

\[
mFDR = \frac{\mathbb{E}[\#\text{false positive}]}{\mathbb{E}[\#\text{rejected}]}.
\]

(8)

Thus, we could control mFDR with given significant level \( \alpha \) by finding a threshold \( c \) so that

\[
c = \arg \min_x \left( \frac{\tilde{f}_{H_0}(x)}{f(x)} \leq \alpha \right).
\]

(9)
Figure 4: The idea for the second step: (a) Under $H_0$, there is no signal (first column.) The histogram of signal weights are shown in the second column and it’s symmetric with ‘peak’ around 0.5. The estimated density is shown in the third column; (b) Under $H_1$, the histogram (the second column) is composited with two parts and has higher values around boundaries 0 and 1, lower values at the middle part. The estimated densities are shown in third column: the black curve is $f(x)$, green one is estimated null density $\tilde{f}_{H_0}(x)$ and red one is estimated alternative density $\tilde{f}_{H_1}(x) = f(x) - \tilde{f}_{H_0}(x)$.

Algorithm 2 The signal region detection method

Require: Signal weights $\{w(s)\}$ or $\{\tilde{w}(s)\}$, significant level $\alpha$;

Ensure: corresponding detected result;

1: Estimate the density curve $f(x)$ based on signal weights, $x \in [0, 1]$;
2: Estimate null density $f_{H_0}(x)$ and alternative density $f_{H_1}(x)$ with (7);
3: Compute mFDR and find the threshold $c$ with (9);
4: Obtain the detected result with the threshold $c$;

The observations with signal weight larger than $c$ are the detected signals. This step is summarized in Algorithm 2.
3.3. Neighbor size selection

In this part, we consider the selection of neighbor size $k$, and this mainly affects the accuracy in Section 3.1. Intuitively, the larger $k$ means the larger block and tends to over-smooth; while the smaller $k$ implying the small block might lose spatial information.

In order to make the "right" cut-off, we need to ensure two points: 1) the variance of the pseudo block mean $\{\hat{\mu}_i\}$ should be as small as possible, so that we could reasonably rearrange the 'representatives' $\{\gamma_i\}$; 2) the length of the rearranged 'representative' sequence $\{\gamma^*_i\}$ should be as long as possible, which could ensure the cut-off location $t$ fall into $[l_1, l_2]$ with probability 1. For the first point, we need to make the number of observations in each block $n_i$ go to infinity; for the second point, the length of $\{\gamma^*_i\}$ is the number of blocks $b$. And the relationship between $b$, $k$ and $n_i$ could be approximated as:

$$\begin{cases} 
\min n_i \approx k^2 \\
\frac{b}{k^2} \approx \frac{n}{k^2}
\end{cases} \quad (10)$$

Hence, we could get the following trade-off problem:

$$k_{opt} = \arg \min_k k^2 + C_1 \frac{n}{k^2} = \sqrt{C_1 n}, \quad (11)$$

where $C_1$ is a given weight to reflect which part we want to emphasize and $n$ is the total number of observations.

Of course, the above analysis is based on the theoretical result. In practice, the neighbor size selection depends on the specific problem and application. Related discussions on neighbor size selection could be found in existing works (Wang et al. (2006); Hall et al. (1995); Sun et al. (2015).)

4. Simulation Study

In this section, we will use simulation to show the effectiveness of our proposed method. We compare SCUSUM with FDR$_L$, because the two methods are designed to detect irregular signals with false discovery rate controlling. All the examples are simulated in the image with $100 \times 100$ pixels. Although in model 2 we didn’t specify the distribution for noise process, we consider independent standard normal distribution $N(0, 1)$ for noise term and generate the data according to the model:

$$x(i, j) = \mu(i, j) + \epsilon(i, j), \quad i, j = 1, \ldots, 100, \quad (12)$$

where $\mu(i, j) = 0$ for $(i, j) \in D_c^c$, and $\mu(i, j) \neq 0$ for $(i, j) \in D_c$, the ‘L’ shape and ‘H’ shape shown in Figure 5: the black region is signal region $D_c$ and white region is the indifference part. The total number of signal pixels is 1288. Here we mainly concern about the accuracy of classification, both false positive and false negative. We set the repeated time $m$ in Algorithm 1 as 50. Additional, for FDR$_L$, we do a standard normal test on each pixels and then apply the algorithm on the corresponding $p$-values.

Figure 5: Ground truth of our simulation setting
Table 1: Detected accuracy comparison between SCUSUM and FDR$_L$: the signal-noise ratio is changed from 0.8 to 2 and neighbor size is chose from {3, 5, 10}. In the two algorithms, significant level $\alpha$ is 0.05.

| neighbor size | Signal $\mu$ | SCUSUM | FDR | FDR$_L$ |
|---------------|--------------|--------|-----|--------|
|               | false negative | false positive | FDR | false negative | false positive | FDR |
| k=3           | 0.8          | 0.6338 | 0.0010 | 0.0186 | 0.9831 | 0.0003 | 0.0917 |
|               | 1            | 0.4286 | 0.0010 | 0.0115 | 0.9380 | 0.0006 | 0.0484 |
|               | 1.5          | 0.1953 | 0.0005 | 0.0044 | 0.5158 | 0.0039 | 0.0508 |
|               | 2            | 0.1271 | 0.0006 | 0.0042 | 0.1478 | 0.0072 | 0.0535 |
| k=5           | 0.8          | 0.3750 | 0.0009 | 0.0094 | 0.8034 | 0.0020 | 0.0586 |
|               | 1            | 0.2750 | 0.0009 | 0.0082 | 0.5351 | 0.0047 | 0.0612 |
|               | 1.5          | 0.1599 | 0.0019 | 0.0147 | 0.0971 | 0.0112 | 0.0770 |
|               | 2            | 0.1014 | 0.0027 | 0.0194 | 0.0159 | 0.0150 | 0.0928 |
| k=10          | 0.8          | 0.3240 | 0.0066 | 0.0588 | 0.1433 | 0.0286 | 0.1793 |
|               | 1            | 0.2526 | 0.0084 | 0.0692 | 0.0569 | 0.0394 | 0.2163 |
|               | 1.5          | 0.1642 | 0.0133 | 0.0968 | 0.0076 | 0.0651 | 0.3051 |
|               | 2            | 0.1288 | 0.0149 | 0.1034 | 0.0027 | 0.0751 | 0.3358 |

In Table 1, the simulation results for SCUSUM and FDR$_L$ are shown. Under each setting, we repeat simulation 100 times. For the two algorithms, we preset the significant level $\alpha$ as 0.05. In the signal region, $\mu$ ranges from 0.8 to 2. Also to show the effect of neighbor size selection, we choose $k$ from {3, 5, 10}. We can see that when our method could control FDR under the presetted significant level $\alpha = 0.05$ with small neighbor size $k = 3, 5$; while FDR$_L$ would allow FDR a little bit higher than 0.05. When the neighbor size $k$ is small ($k = 3$), SCUSUM outperforms both in false negative and false positive: though the false negative for SCUSUM is 0.6338 when $k = 3$ and $\mu = 0.8$, the corresponding false positive is 0.001, a very small proportion and the false negative for FDR$_L$ under the same setting is 0.9831, almost 1. In addition, when we allow the neighbor size $k$ to be 5, imply a little larger block, the false negative for SCUSUM would decrease to be less than 0.50 while the false positive is controlled less than 0.003. Compared with FDR$_L$, with neighbor size as 5, the false negative is over 0.50 with low signal strength 0.8 and 1.0, while the false positive is almost higher than 0.003. These experiment data implies that our proposed method, SCUSUM, performs better than FDR$_L$.

In Figure 6, we show the probability maps of the pixels being detected by the two methods with give $\alpha = 0.05$. Here we set the neighbor size $k = 5$, which we think has performance from Table 1 (relatively lower false positive and false negative). We range the signal strength from 0.5 to 2. In the probability maps, the darker the color is, the higher probability the corresponding point is signal. Comparing the probability maps with the ground truth (see Figure 5), it could be easily see that with high signal strength $\mu \geq 1.5$, the two methods have almost the same performance; while the signal is much too low (e.g. $\mu = 0.5$,) SCUSUM has a better detected result than FDR$_L$. Also we can see that in the results of FDR$_L$ there are some shadows outside of 'L' and 'H' signal region, which are false positive; while for SCUSUM, the detections for indifference region are more 'white' (no
Spatial CUSUM for Signal Region Detection

Figure 6: Comparison of the detection probability for SCUSUM and FDR$_L$ under different signal strength: the darker the color is, the higher probability the corresponding point is signal.

shadows.) To some degree, these probability maps are consistent with the experiment data in Table 1.

Though in model 2 we assume the noise to be independence, now we try to apply SCUSUM to dependence spatial observations, and compare with FDR$_L$. Here we consider to use the Exponential Covariance Model (Gelfand et al. (2010)) to generate dependence data. The covariance matrix is

$$C(s_i, s_j) = \exp(-\frac{\|s_i - s_j\|}{r}),$$

(13)

where $s_i$ and $s_j$ are two locations, $\|s_i - s_j\|$ is the distance between the two location, and $r$ is the dependence scale. The larger $r$ means the stronger dependence. Then the data are generate from multivariate normal distribution with above covariance matrix and corresponding mean, $\mu_0$ for $\mathcal{D}_{df^c}$ and $\mu_1$ for $\mathcal{D}_{df}$. We range scale $r$ from $\{0.1, 0.3, 0.5\}$, and the corresponding covariances for unit distance are $\{0.00004, 0.03567, 0.13533\}$. We set the neighbor size $k = 5$ for both two algorithms and $\alpha = 0.05$ to control marginal FDR. The results are shown in Figure 7 and Table 2. It could be seen that with the weak dependence, SCUSUM could still detect the signal region efficiently while control the false positive. However for FDR$_L$, the false negative increases significantly with stronger dependence (see Table 2). Also Table 2 shows that larger dependence scale leads to larger false positive, false negative and FDR for SCUSUM. This gives us a hint that for larger scale dependence noise we need to choose relatively larger blocks. Nevertheless, we can see that SCUSUM could recognize the signal region with higher probability than FDR$_L$, when the noise dependence is weak.

5. Real Data Experiment

In this section, we apply four methods, SCUSUM, scan statistics, the conventional FDR and FDR$_L$, to a real fMRI data to illustrate their differences in real data application. The fMRI data has been analyzed in some previous works (Maitra (2009); Zhang and Zhu (2012) etc.)
Figure 7: Comparison of the detection probability for SCUSUM and FDR\textsubscript{L} on data with different dependence scales: the darker the color is, the higher probability the corresponding point is signal. Here we set the neighbor size \( k = 5 \) and signal \( \mu = 1 \).

Table 2: Summary for detection probabilities on dependence data. The dependence scale is changed from 0.1 to 0.5, neighbor size is chose as 5 and signal strength is 1.

| Dependence Scale | \( r = 0.1 \) | \( r = 0.3 \) | \( r = 0.5 \) |
|------------------|--------------|--------------|--------------|
| false negative   | SCUSUM 0.2721 | 0.2820       | 0.3108       |
|                  | FDR\textsubscript{L} 0.5213 | 0.5879       | 0.7317       |
| false positive   | SCUSUM 0.00095 | 0.00154      | 0.00418      |
|                  | FDR\textsubscript{L} 0.00499 | 0.00411      | 0.00280      |
| FDR              | SCUSUM 0.0086 | 0.0141       | 0.0386       |
|                  | FDR\textsubscript{L} 0.0622 | 0.0604       | 0.0590       |

Figure 8 (a) shows six slices of the fMRI images in a total of 22 slices. Each individual image has 128 \( \times \) 128 pixels. All these images show activities in different regions by heat maps. The pixels’ values are the transformations of \( p \)-values from a previous study, which should follow a standard normal distribution, and we only care about detection of the regions with positive values.

As for the conventional FDR, the lack of identification phenomenon happened, e.g. all the signals of the fifth slice in Figure 8 (b) are missed. Also without considering spatial correlation, for example in the first slice of Figure 8 (b), some detected 'signals' are scattered around, which means
that some of them might be false positive. Although FDR$_L$ could make full use of the neighbor information of spatially structured data and improve the detection efficiency, many weak signals are missed, e.g. in fifth slice and sixth slice of Figure 8 (c), many active regions are not detected. For scan statistics, even though almost all the ‘hot’ pixels are detected, the ‘signal’ regions are too large, which is doubtful.

Similar to the conclusion in Section 4, SCUSUM is more likely to detect weak signals compared with FDR methods. In all the six slices, the detected regions are larger and cover the regions detected by FDR methods. Meanwhile, within each slice, SCUSUM could identify several irregularly shaped clusters. We can see that the detected regions form natural clusters, and they are the spatially grouped ‘hot’ parts in the raw images. These results show that our proposed method might be more suitable for detection of irregular shaped and weak spatial signals.

6. Conclusion

In this work, we proposed a spatial signal detection method, SCUSUM, which could accommodate the local spatial information. SCUSUM consists of two steps: firstly signal weights are estimated by moving window projecting and CUSUM cut-off; then a threshold is determined with given significant level $\alpha$ to control marginal false discovery rate. Our simulation study shows that our method has a better performance compared to FDR$_L$ method. Empirically, SCUSUM tends to detect spatially grouped and weak signals, which are missed by the other two methods. Finally, our method is applied to a real fMRI data to illustrate its detection effectiveness.

In model 2, though our method doesn’t need to specify the distribution for noise process, the noise processes are assumed to be independent, and this is a strong assumption in spatial statistics. In the future work, it could be possible to consider a dependent noise process. In our simulation, the result shows that with weak spatial dependent noise process, SCUSUM could still maintain it effectiveness. Another key issue is to consider how to combine multi-source image data. So far, we only consider the observations are scalar and use moving window idea to project spatial observations to a sequence. However, this projecting method could be problematic when the observation of a spatial location is high-dimensional data or functional data.
Figure 8: Detected signal regions for the real fMRI dataset. The first column is the raw data. Conventional FDR approach (second column), FDR\textsubscript{L} approach (third column), Scan statistics (fourth column) and our proposed SCUSUM method (fifth column) are shown. The black part is the detected signal region. Here, significant level is $\alpha = 0.0001$. 
References

Kevork N Abazajian and Manoj Kaplinghat. Detection of a gamma-ray source in the galactic center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission. *Physical Review D*, 86(8):083511, 2012.

Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 2009.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

Thomas Blumensath, Saad Jbabdi, Matthew F Glasser, David C Van Essen, Kamil Ugurbil, Timothy EJ Behrens, and Stephen M Smith. Spatially constrained hierarchical parcellation of the brain with resting-state fmri. *Neuroimage*, 76:313–324, 2013.

David S Breed. Weather monitoring techniques, November 15 2011. US Patent 8,060,308.

Matias D Cattaneo, Michael Jansson, and Xinwei Ma. lpdensity: Local polynomial density estimation and inference. 2017.

Song Xi Chen. Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2):131–145, 1999.

Haeran Cho et al. Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038, 2016.

Ann Cowling and Peter Hall. On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 551–563, 1996.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

Christopher Genovese and Larry Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):499–517, 2002.

Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.
Michael D Gladders and HKC Yee. A new method for galaxy cluster detection. i. the algorithm. *The Astronomical Journal*, 120(4):2148, 2000.

Joseph Glaz and Narayanaswamy Balakrishnan. *Scan statistics and applications*. Springer Science & Business Media, 2012.

Joseph Glaz, Joseph I Naus, Sylvan Wallenstein, Sylvan Wallenstein, and Joseph I Naus. *Scan statistics*. Springer, 2001.

Joseph Glaz, Vladimir Pozdnyakov, and Sylvan Wallenstein. *Scan statistics: methods and applications*. Springer Science & Business Media, 2009.

Oleksandr Gromenko, Piotr Kokoszka, and Matthew Reimherr. Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):29–50, 2017.

Emily Grover-Kopec, Mika Kawano, Robert W Klaver, Benno Blumenthal, Pietro Ceccato, and Stephen J Connor. An online operational rainfall-monitoring resource for epidemic malaria early warning systems in africa. *Malaria Journal*, 4(1):6, 2005.

Timothy C Haas. Kriging and automated variogram modeling within a moving window. *Atmospheric Environment. Part A. General Topics*, 24(7):1759–1769, 1990.

Peter Hall, Joel L Horowitz, and Bing-Yi Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995.

Lajos Horváth and Marie Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648, 2012.

MC Jones and PJ Foster. A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, pages 1005–1013, 1996.

Martin Kulldorff. Spatial scan statistics: models, calculations, and applications. In *Scan statistics and applications*, pages 303–322. Springer, 1999.

Martin Kulldorff and Neville Nagarwalla. Spatial disease clusters: detection and inference. *Statistics in medicine*, 14(8):799–810, 1995.

Ranjan Maitra. Assessing certainty of activation or inactivation in test–retest fmri studies. *Neuroimage*, 47(1):88–97, 2009.

Christopher J Miller, Christopher Genovese, Robert C Nichol, Larry Wasserman, Andrew Connolly, Daniel Reichart, Andrew Hopkins, Jeff Schneider, and Andrew Moore. Controlling the false-discovery rate in astrophysical data analysis. *The Astronomical Journal*, 122(6):3492, 2001.

JL Naus. Clustering of random points in two dimensions. *Biometrika*, 52(1-2):263–266, 1965.

Joseph I Naus. Approximations for distributions of scan statistics. *Journal of the American Statistical Association*, 77(377):177–183, 1982.
Antonio Páez, Fei Long, and Steven Farber. Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8):1565–1581, 2008.

Carey E Priebe, John M Conroy, David J Marchette, and Youngser Park. Scan statistics on enron graphs. *Computational & Mathematical Organization Theory*, 11(3):229–247, 2005.

Xilin Shen, Fuyuze Tokoglu, Xenios Papademetris, and R Todd Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415, 2013.

Wenguang Sun, Brian J Reich, T Tony Cai, Michele Guindani, and Armin Schwartzman. False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):59–83, 2015.

Toshiro Tango. A test for spatial disease clustering adjusted for multiple testing. *Statistics in medicine*, 19(2):191–204, 2000.

Madeleine C Thomson and Stephen J Connor. The development of malaria early warning systems for africa. *Trends in parasitology*, 17(9):438–445, 2001.

Jigang Wang, Predrag Neskovic, and Leon N Cooper. Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence. *Pattern Recognition*, 39(3):417–423, 2006.

Tengyao Wang and Richard J Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83, 2018.

David C Wheeler. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in ohio, 1996–2003. *International Journal of Health Geographics*, 6(1):13, 2007.

Chunming Zhang, Jianqing Fan, and Tao Yu. Multiple testing via fdrl for large scale imaging data. *Annals of statistics*, 39(1):613, 2011.

Lingsong Zhang and Zhengyuan Zhu. Spatial multiresolution cluster detection method. *arXiv preprint arXiv:1205.2106*, 2012.

Zhenkui Zhang, Renato Assunção, and Martin Kulldorff. Spatial scan statistics adjusted for multiple clusters. *Journal of Probability and Statistics*, 2010, 2010.
Appendix A. Proof for Lemma 1

Proof In assumed model 2, the noise processes are independent. Thus, in the $i$th block, the sampled representative $\gamma_i$ is independent with the rest observations, which implies $\gamma_i$ and $\bar{\mu}_i$ are independent, i.e. $\gamma_i \perp \bar{\mu}_i$. Also the representaties and pseudo block means between blocks are independent. These lead that $\{\gamma_i\}$ and $\{\bar{\mu}_i\}$ are independant.

Under the null hypothesis $H_0$, i.e. there is no signal and $\mathbb{E}[\bar{\mu}_i] = \mu_0$, pseudo block mean sequence $\{\bar{\mu}_i\}$ are i.i.d, as well as representative sequence $\{\gamma_i\}$. For the distribution of $\{\gamma_i\}$, we have following:

$$f(\gamma_1 \leq y_1, \ldots, \gamma_b \leq y_b) = \sum_{(1) \ldots (b) \in S_b} f(\gamma_1 \leq y_1, \ldots, \gamma_b \leq y_b | \bar{\mu}_1 \geq \ldots \geq \bar{\mu}_b) f(\bar{\mu}_1 > \ldots > \bar{\mu}_b)$$

$$= \sum_{(1) \ldots (b) \in S_b} f(\gamma_1 \leq y_1, \ldots, \gamma_b \leq y_b | \bar{\mu}_1 \geq \ldots \geq \bar{\mu}_b) \frac{1}{b!}$$

$$= \sum_{(1) \ldots (b) \in S_b} f(\gamma_1 \leq y_1, \ldots, \gamma_b \leq y_b) \frac{1}{b!}$$

$$= f(\gamma_1 \leq y_1, \ldots, \gamma_b \leq y_b).$$

Here $(1) \ldots (b)$ is the one possible decreasing order for $\{\bar{\mu}_i\}$, $S_b$ presents the set of all the possible orders. The first equation is according to bayesian formula; the second equation is because of $f(\bar{\mu}_1 > \ldots > \bar{\mu}_b) = \int_{\bar{\mu}_1 > \ldots > \bar{\mu}_b} \text{d}n(\bar{\mu}_1, \ldots, \bar{\mu}_b) = 1/b!$ under the independence of $\{\bar{\mu}_i\}$; the third equation is due to independence between $\{\gamma_i\}$ and $\{\bar{\mu}_i\}$; the fourth equation is because $\{\gamma_i\}$ are independent; the fifth equation is due to the cardinality of $S_b$ is $1/b!$. From above result, we reach that $\{\gamma_i\}$ are also i.i.d, having the same distribution with $\{\gamma_i\}$.

Under the alternative hypothesis $H_1$, as the number of observations in each block $n_i$ goes to infinity, the weak law of larger number supports that $\bar{\mu}_i \xrightarrow{p} \mathbb{E}[\bar{\mu}_i]$. Thus, with (3) and (4), (5) holds. And $l_1$ is the number of block inside the signal region, $(b - l_2)$ is the number of block inside the indifference region and $(l_2 - l_1)$ is the number of block at the boundary.

Appendix B. Proof for Theorem 1

Proof The proof idea is similar with Aue et al. (2009). W.l.o.g, here we consider the variance of noise processes is 1, $\mu_0 = 0$ and $\mu_1 = \Delta$. Define the ratio of signal region to the entire spatial domain is $\theta$, hence, as blocks become finer and finer ($b \to \infty$), the ratio of the blocks with signal representative to the total blocks is getting closer to $\theta$.

Under $H_1$, define the following events:

$$\begin{align*}
A_1 &= \{\mu_i \leq \mu_j, B_i \in \mathcal{D}_d, B_j \in \mathcal{D}_d^c, \forall i, j\}, \\
A_2 &= \{\mu_k \leq \mu_j, B_i \in \mathcal{D}_d, B_k \text{ at boundary, } \forall i, k\}, \\
A_3 &= \{\mu_j \leq \mu_k, B_j \in \mathcal{D}_d^c, B_k \text{ at boundary, } \forall j, k\}.
\end{align*}$$

Event $(A_1 \cup A_2 \cup A_3)^c$ presents the scenario that the order of $\{\gamma_i\}$ is from signal blocks $[1, l_1)$ to interim (boundary) blocks $[l_1, l_2]$ and then to indifferent blocks $(l_2, b]$ (see Figure 3 (b)). And with Lemma 1, $\mathbb{P}((A_1 \cup A_2 \cup A_3)^c) = 1$, as $\min n_i \to \infty$. 

18
With event \((A_1 \cup A_2 \cup A_3)^c\), following we show that the probability of cut-off location \(t = \arg \max_i \tilde{\gamma}_i\) falling into \([l_1, l_2]\) would converge to 1. To proof that, we define statistics

\[
Q(r) = \tilde{\gamma}_r^2 = \left(\sum_{i=1}^{r} \gamma_i^* - \frac{r}{b} \sum_{i=1}^{b} \gamma_i^* \right)^2.
\] (15)

First consider the probability of event \(B_1(N) = \{t \geq (l_2 + N)\}\), with \((l_2 + N) \leq b\), \(N\) is a fixed constant. Define \(R(r; l_2) = Q(r) - Q(l_2)\), and note that \(Q(l_2)\) is a constant.

\[
R(r; l_2) = Q(r) - Q(l_2)
= \left(\sum_{i=1}^{r} \gamma_i^* - \frac{r}{b} \sum_{i=1}^{b} \gamma_i^* \right)^2 - \left(\sum_{i=1}^{l_2} \gamma_i^* - \frac{l_2}{b} \sum_{i=1}^{b} \gamma_i^* \right)^2
= \left[\sum_{l_2+1}^{r} \gamma_i^* - (r - l_2) \tilde{\gamma}_r^* \right] \left[\sum_{i=1}^{l_2} \gamma_i + \sum_{j=1}^{l_2} \gamma_j - (r + l_2) \tilde{\gamma}_r^* \right]
\]

where \(\tilde{\gamma}_r^* = \frac{1}{b} \sum_{i=1}^{b} \gamma_i^*\). And with equation (5), following equations hold:

\[
(\text{I}) = \sum_{i=l_2+1}^{r} \epsilon_i - (r - l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i - (r - l_2) \theta \Delta,
\] (16)

\[
(\text{II}) = \sum_{i=1}^{r} \epsilon_i + \theta b \Delta + \sum_{j=1}^{l_2} \epsilon_j + \theta b \Delta - (r + l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i - (r + l_2) \theta \Delta
= \sum_{i=1}^{r} \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j - (r + l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i + (2 \theta b - (r + l_2) \theta) \Delta.
\] (17)

Define the following statistics:

\[
\begin{align*}
E^1(r; l_2) &:= \sum_{i=l_2+1}^{r} \epsilon_i - (r - l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i, \\
E^2(r; l_2) &:= \sum_{i=1}^{r} \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j - (r + l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i, \\
D^1(r; l_2) &:= - (r - l_2) \theta \Delta, \\
D^2(r; l_2) &:= (2 \theta b - (r + l_2) \theta) \Delta.
\end{align*}
\] (18)

So \((\text{I}) = E^1(r; l_2) + D^1(r; l_2)\) and \((\text{II}) = E^2(r; l_2) + D^2(r; l_2)\).

As \(b \rightarrow \infty\), we have

\[
\max_{(l_2+N) \leq r \leq b} D^1(r; l_2)D^2(r; l_2) = \max_{(l_2+N) \leq r \leq b} \left[ - (r - l_2) \theta \Delta \right] \left[ (2 \theta b - (r + l_2) \theta) \Delta \right]
= \max_{(l_2+N) \leq r \leq b} - \theta^2 \Delta^2 (r - l_2)(2 - \frac{r + l_2}{b})b
= - \theta^2 \Delta^2 N (2 - \frac{2l_2 + N}{b})b,
\]
the last equation is due to the $D^1(r; l_2)D^2(r; l_2)$ reaches the maximum with $r = (l_2 + N)$. Also $\forall \epsilon \geq 0$, we have

$$
\limsup_{b \to \infty} \mathbb{P}(\max_{(l_2 + N) \leq r \leq b} D^1(r; l_2)D^2(r; l_2) > -\epsilon) = \limsup_{b \to \infty} \mathbb{P}(-\theta^2 \Delta^2 N(2 - \frac{2l_2 + N}{b})b > -\epsilon) = \limsup_{b \to \infty} \mathbb{P}(\theta^2 \Delta^2 N(2 - \frac{2l_2 + N}{b})b \leq \epsilon) = 0
$$

If we could prove $D^1D^2(r; l_2)$ is the leading term in $R(r; l_2)$, then

$$
\limsup_{b \to \infty} \mathbb{P}(B_1(N)) = \limsup_{b \to \infty} \mathbb{P}(t \geq l_2 + N) = \limsup_{b \to \infty} \mathbb{P}(\max_{(l_2 + N) \leq r \leq b} R(r; l_2) > 0) = \limsup_{b \to \infty} \mathbb{P}(\theta^2 \Delta^2 N(2 - \frac{2l_2 + N}{b})b + O(1) \leq \epsilon) = 0
$$

Hence $\mathbb{P}(t \geq l_2) = \bigcup_{N=0}^{b-l_2} \mathbb{P}(B_1(N)) = 0$.

Following lemmas support that $D^1(r; l_2)D^2(r; l_2)$ is the leading term in $R(r; l_2)$, with $r \in [l_2 + N, b]$.

Lemma 3  With the assumptions of Theorem 1, given $N$, $\forall \epsilon > 0$,

$$
\limsup_{b \to \infty} \mathbb{P}(\max_{(l_2 + N) \leq r \leq b} \frac{|E^1(r; l_2)E^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \geq \epsilon) = 0.
$$

Proof  With the brief derivation, we have,

$$
\max_{(l_2 + N) \leq r \leq b} \frac{|E^1(r; l_2)E^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \leq \max_{(l_2 + N) \leq r \leq b} \frac{|E^1(r; l_2)||E^2(r; l_2)|}{\theta \Delta^2(r - l_2)(2b \theta - (l_2 + r))} = O(1) \max_{(l_2 + N) \leq r \leq b} \frac{|E^1(r; l_2)||E^2(r; l_2)|}{(r - l_2)(l_2 + r)} = O(1) \max_{(l_2 + N) \leq r \leq b} \left[ \frac{|E^1(r; l_2)|}{(r - l_2)} \frac{|E^2(r; l_2)|}{(l_2 + r)} \right]^{(III)}
$$

the last equation is because of $l_2 > \theta \Delta$.

For (III),

$$
\frac{|E^1(r; l_2)|}{(r - l_2)} \leq \frac{\sum_{i=l_2+1}^{r} \epsilon_i}{(r - l_2)} + \frac{1}{b} \sum_{i=1}^{b} \epsilon_i |^{P} \to 0,
$$

where $P$ is a probability. The probability $P$ tends to $0$ as $b$ goes to infinity. The notation $|^{P} \to 0$ means that the probability of the event $P$ tends to $0$ as $b$ goes to infinity.
the last equation is due to the law of iterated logarithm and the weak law of large number. Similarly, for (IV),

\[ |E^2(r; l_2)| \leq \frac{\sum_{i=1}^{r} \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j + |(r + l_2) \frac{1}{b} \sum_{i=1}^{b} \epsilon_i|}{(l_2 + r)} \xrightarrow{p} 0. \]

Hence, with continuous mapping theorem, we have the result.

\[ \square \]

**Lemma 4** With the assumptions of Theorem 1, given \( N, \forall \epsilon > 0 \),

\[ \lim_{b \to \infty} \sup_{b \to \infty} \mathbb{P}\left( \max_{(t_2+N) \leq r \leq b} \frac{|E^1(r; l_2)D^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \geq \epsilon \right) = 0. \] (20)

**Proof** Similarly, we have

\[ \max_{(t_2+N) \leq r \leq b} \frac{|E^1(r; l_2)D^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \leq O(1) \max_{(t_2+N) \leq r \leq b} \frac{|E^1(r; l_2)|}{(r - l_2)} \frac{|D^2(r; l_2)|}{(r + l_2)}. \]

From Lemma 3, we know that \( \frac{|E^1(r; l_2)|}{(r - l_2)} \xrightarrow{P} 0 \). With directly derivation, \( \frac{|D^2(r; l_2)|}{(r + l_2)} = O(1) \). Hence, \( \frac{\max_{(t_2+N) \leq r \leq b} |E^1(r; l_2)D^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \xrightarrow{P} 0. \)

\[ \square \]

**Lemma 5** With the assumptions of Theorem 1, given \( N, \forall \epsilon > 0 \),

\[ \lim_{b \to \infty} \sup_{b \to \infty} \mathbb{P}\left( \max_{(t_2+N) \leq r \leq b} \frac{|D^1(r; l_2)E^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \geq \epsilon \right) = 0. \] (21)

**Proof** The idea is the same with Lemma 4:

\[ \max_{(t_2+N) \leq r \leq b} \frac{|D^1(r; l_2)E^2(r; l_2)|}{D^1(r; l_2)D^2(r; l_2)} \leq O(1) \max_{(t_2+N) \leq r \leq b} \frac{|D^1(r; l_2)|}{(r - l_2)} \frac{|E^2(r; l_2)|}{(r + l_2)} \xrightarrow{P} 0, \] (22)

with \( \frac{|D^1(r; l_2)|}{(r - l_2)} = O(1) \) and \( \frac{|E^2(r; l_2)|}{(r + l_2)} \xrightarrow{P} 0. \)

\[ \square \]

The above three lemmas support that \( D^1(r; l_2)D^2(r; l_2) \) is the leading term in \( R(r; l_2) \), with \( r \in [l_2 + N, b] \). Hence, the probability of event \( \{ t > l_2 \} \cup (A_1 \cup A_2 \cup A_3) \to 0, \) as \( b \to \infty \) and \( \min n_i \to \infty \).

For the other side, consider event \( B_2(N) = \{ t \leq (l_1 - N) \} \), with a given \( N \) and \( l_1 \geq N \). Similarly define \( R(r; l_1) = Q(r) - Q(l_1) \), and following we will show \( R(r; l_1) < 0, \forall r < l_1 \)
with probability 1.

\[
R(r; l_1) = Q(r) - Q(l_1)
\]

\[
= \left( \sum_{i=1}^{r} \gamma_i^* - \frac{r}{b} \sum_{i=1}^{b} \gamma_i^* \right)^2 - \left( \sum_{i=1}^{l_1} \gamma_i^* - \frac{l_1}{b} \sum_{i=1}^{b} \gamma_i^* \right)^2
\]

\[
= \left[ - \sum_{i=r+1}^{l_1} \gamma_i^* - (r - l_1)\tilde{\gamma}^* \right] \left[ \sum_{i=1}^{r} \gamma_i + \sum_{j=1}^{l_1} \gamma_j - (r + l_1)\tilde{\gamma}^* \right]
\]

Define the following statistics:

\[
\begin{align*}
E^1(r; l_1) &:= - \sum_{i=r+1}^{l_1} \epsilon_i - (r - l_1)\tilde{\gamma}^*, \\
E^2(r; l_1) &:= \sum_{i=1}^{r} \epsilon_i + \sum_{j=1}^{l_1} \epsilon_j - (r + l_1)\tilde{\gamma}^*, \\
D^1(r; l_1) &:= -(l_1 - r)(1 - \theta)\Delta, \\
D^2(r; l_1) &:= (r + l_1)(1 - \theta)\Delta.
\end{align*}
\]

Similarly, we have \((V) = E^1(r; l_1) + D^1(r; l_1), (VI) = E^1(r; l_2) + D^1(r; l_2)\). Also, as \(b \to \infty\), \(l_1 = \theta b \to \infty\), so we have

\[
\max_{1 \leq r \leq (l_1 - N)} D^1(r; l_1)D^2(r; l_1) = [-N(r + l_1)(1 - \theta)^2\Delta^2]
\]

and

\[
\lim_{b \to \infty} \sup \mathbb{P} \left( \max_{1 \leq r \leq (l_1 - N)} D^1(r; l_1)D^2(r; l_1) > -\epsilon \right) = 0
\]

Following lemma shows that \(D^1(r; l_1)D^2(r; l_1)\) is the leading term in \(R(r; l_1)\), with \(r \in [1, l_1 - N]\).

**Lemma 6**  With the assumptions of Theorem 1, given \(N, \forall \epsilon > 0\), we have

\[
\begin{align*}
\lim_{b \to \infty} \sup \mathbb{P} \left( \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)E^2(r; l_1)|}{D^1(r; l_1)D^2(r; l_1)} \geq \epsilon \right) &= 0, \\
\lim_{b \to \infty} \sup \mathbb{P} \left( \max_{1 \leq r \leq (l_1 - N)} \frac{|D^1(r; l_1)E^2(r; l_1)|}{D^1(r; l_1)D^2(r; l_1)} \geq \epsilon \right) &= 0, \\
\lim_{b \to \infty} \sup \mathbb{P} \left( \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)D^2(r; l_1)|}{D^1(r; l_1)D^2(r; l_1)} \geq \epsilon \right) &= 0.
\end{align*}
\]
Proof Similarly with Lemma 3.5, we have following:

$$\left\{ \begin{array}{l}
\frac{|E_1(r; l_1)|}{(l_1 - r)} \leq \left| \sum_{i=r+1}^{l_1} \gamma_i \right| \frac{1}{l_1 - r} + \frac{1}{b} \sum_{i=1}^{b} \gamma_i \to 0;

\frac{|E_2(r; l_1)|}{(l_1 + r)} \leq \left| \sum_{i=1}^{l_1} \gamma_i + \sum_{i=1}^{l_1} \gamma_i \right| + \frac{1}{b} \sum_{i=1}^{b} \gamma_i \to 0;

\frac{|D_1(r; l_1)|}{(l_1 - r)} = \frac{|D_2(r; l_1)|}{(l_1 + r)} = O(1).
\end{array} \right.$$

Hence, we have the results

$$\max_{1 \leq r \leq (l_1 - N)} \frac{|E_1(r; l_1)| |E_2(r; l_1)|}{|D_1(r; l_1)| |D_2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|E_1(r; l_1)| |E_2(r; l_1)|}{(l_1 - r)(l_1 + r)} \to 0;$$

$$\max_{1 \leq r \leq (l_1 - N)} \frac{|D_1(r; l_1)| |E_2(r; l_1)|}{|D_1(r; l_1)| |D_2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|D_1(r; l_1)| |E_2(r; l_1)|}{(l_1 - r)(l_1 + r)} \to 0;$$

$$\max_{1 \leq r \leq (l_1 - N)} \frac{|E_1(r; l_1)| |D_2(r; l_1)|}{|D_1(r; l_1)| |D_2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|E_1(r; l_1)| |D_2(r; l_1)|}{(l_1 - r)(l_1 + r)} \to 0.$$

With above conclusion, we have

$$\lim_{b \to \infty} \sup \mathbb{P}(B_2(N)) = \lim_{b \to \infty} \sup \mathbb{P}(t \leq l_1 - N) = \lim_{b \to \infty} \mathbb{P}(\max_{1 \leq r \leq (l_1 - N)} R(r; l_1) > 0) = \lim_{b \to \infty} \mathbb{P}(N(r + \theta b)(1 - \theta)^2 \Delta^2 + O(1) \leq \epsilon) = 0$$

which implies \(\mathbb{P}(t \leq l_1) = \bigcup_{N=1}^{1} \mathbb{P}(B_2(N)) = 0\).

With these results, we have

$$\lim_{b \to \infty} \mathbb{P}(\{t \in [l_1, l_2]\}) \leq 1 - \lim_{b \to \infty} \sup \mathbb{P}(\{t < l_1\} \cup \{t > l_2\} \cup (A_1 \cup A_2 \cup A_3)) \to 1,$$

as \(b \to \infty\) and \(\min n_i \to \infty\).

Appendix C. Proof for Lemma 2

Proof This lemma is easy to prove: from Lemma 1, as \(\min n_i \to \infty\), \(\{\gamma_i^*\}\) are i.i.d. Consider sequence \(\{y_1, \ldots, y_b\}\), then \(\mathbb{P}(\gamma_1^* = y_1, \ldots, \gamma_b^* = y_b) = \mathbb{P}(\gamma_1^* = y_b, \ldots, \gamma_b^* = y_1)\). Note that in the left side \(\{\gamma_i^*\}\) equal to the reversed sequence \(\{y_b, \ldots, y_1\}\). Via applying CUSUM cut-off on \(\{y_1, \ldots, y_b\}\)
and \( \{y_0, ..., y_1\} \), the detection result is opposite. Hence the distribution for signal weights under \( H_0 \) is symmetric under the asymptotical setting.

\[ \square \]