Identifying the support of rectangular signals in Gaussian noise

Jiyao Kou
Department of Statistics, Stanford University, Stanford, CA, USA

ABSTRACT
We consider the problem of identifying the support of the block signal in a sequence when both the length and the location of the block signal are unknown. The multivariate version of this problem is also considered, in which we try to identify the support of the rectangular signal in a hyper-rectangle. In this article, we greatly generalize the requirement in Jeng, Cai, and Li and in particular, we allow the length of the block signal to grow polynomially with the length of the sequence. A statistical boundary above which the identification is possible is presented and an asymptotically optimal and computationally efficient procedure is proposed under the assumption of Gaussian white noise. The problem of block signal identification is shown to have the same statistical difficulty as the corresponding problem of detection, in the sense that whenever we can detect the signal, we can identify the support of the signal. Some generalizations are also considered here: (1) the case of multiple block signals, (2) the robust identification problem where the noise distribution is unspecified, (3) the block signal identification problem under the exponential family setting.

ARTICLE HISTORY
Received 21 December 2020
Accepted 15 August 2021

KEYWORDS
Block signal; rectangular signal; support identification; multidimensional; penalized scan

MATHEMATICS SUBJECT CLASSIFICATION
62: Statistics

1. Introduction

Block signal detection and identification in a long one-dimensional sequence is a challenging and important problem and arises in many applications, see, for example, Gangnon and Clayton (2001) and Neill (2009) in epidemiology; and Jeng, Cai, and Li (2010) and Stranger et al. (2007) in Copy Number Variation. Block signal detection decides whether there exists any block signal in the sequence while block signal identification further identifies the support of the block signal. There has been a large body of work on signal detection, see, e.g., Glaz and Balakrishnan (2012) and Dumbgen and Spokoiny (2001) on the scan statistic; Arias-Castro, Donoho, and Huo (2005, Arias-Castro, Candès, and Durand 2011) for geometric objects and cluster detection; Donoho and Jin (2004), Delaigle and Hall (2009) and Donoho and Jin (2015) for sparse signals detection and identification, in which signal is not assumed to have block structure; Dümbgen and Walther (2008) and Rivera and Walther (2013) about density inference; and Walther (2010), Rufibach and Walther (2010), Chan (2009) and Chan and Walther (2013) for more recent results on block signal detection using the penalized scan and...
average likelihood ratio. For the block signal identification problem in the univariate case, in Jeng, Cai, and Li (2010), the authors characterized the identifiable region under the assumption of Gaussian white noise and \( \log |I_n^*| = o(\log n) \), where \( n \) is the length of the sequence and \( |I_n^*| \) is the length of the block signal. They also proposed the Likelihood Ratio Selector (LRS) procedure and established its optimality under the above assumptions. However, their result excludes the common and important situation where \( |I_n^*| = n^{1-\beta} \) for \( 0 < \beta < 1 \), and in fact LRS is not optimal in this situation. Moreover, LRS procedure needs to pre-specify a parameter \( L \) which is some number greater than \( |I_n^*| \). Such \( L \) is not always easy to pre-specify and the misspecification may cause misidentification.

In this article, we establish the block signal identification theory under a more general assumption. A computational efficient procedure is proposed and its optimality is established under the assumption of Gaussian white noise. We note that in our procedure, there is no unknown parameters to be pre-specified. Moreover, our results show that the block signal detection and block signal identification have the same statistical difficulty, although the latter seems to be more challenging than the former.

In addition, we consider in this article several generalizations of the block signal identification problem. Firstly, we generalize our results to multi-dimensional case, where we show our procedure remains optimal in identifying rectangular signals in the hyper-rectangle. Secondly, we consider the extension to multiple block signals, where under certain assumptions, our procedure remains optimal in identifying all block signals. Last but not the least, we consider the robust identification problem where the noise distribution is unspecified and discuss about the block signal identification under the exponential family setting in the discussion section.

The rest of the article is organized as follows. In Section 2, our identification procedure is introduced in the univariate case and its optimality is established under the assumption of Gaussian white noise. In Section 3, we extend our theorem to the multi-dimensional case. In Section 4, we consider the case of multiple block signals. In Section 5, a simulation study is carried out to illustrate our previous results. In Section 6, we discuss about some generalizations and future research topics.

In the end of this section, we make some notations. For two series \( a_n \) and \( b_n \), we define \( a_n \ll b_n \) if \( a_n = o(b_n) \), or equivalently, \( \frac{a_n}{b_n} \to 0 \). We may use this notation and the small-o notation interchangeably. For a set of random variables \( X_n \) and a corresponding set of constant \( a_n \), we define \( X_n = o_p(a_n) \) if \( X_n / a_n \) converges to 0 in probability. Similarly, we define \( X_n = O_p(a_n) \) if for any \( \epsilon > 0 \), there exists a finite \( M \) such that \( P(|X_n/a_n| > M) < \epsilon \) for all \( n \).

### 2. Block signal identification under Gaussian white noise

Let’s first consider block signal identification in the univariate setting. Suppose we observe that

\[
Y_i = \mu 1_{I_n}(i) + Z_i, \quad i = 1, \ldots, n
\]

where \( Z_i \) are i.i.d standard normal random variables and unknown interval \( I_n^* = (j_n, k_n), 0 \leq j_n < k_n < n \) and
\[
I_n^*(i) = \begin{cases} 
1 & i \in I_n^* \\
0 & i \notin I_n^* 
\end{cases}
\]

\(\mu\) is an unknown number and for simplicity we assume that \(\mu\) is non-negative. If \(\mu\) is non-positive, we can replace \(Y_i\) by \(-Y_i\). Our goal is to estimate the support \(I_n^*\) in Model (1) where \(I_n^* = \emptyset\) means \(\mu = 0\).

For block signal detection, which is a testing problem, we want to maximize the power of the test while controlling the type I error. Similarly, in the corresponding block signal identification problem, we want to approximately find the start and the end point of the block signal (when it exists) with high probability while control the type I error.

Let \(H_0\) denote the null case that there exists no signal in the sequence and \(H_1\) denote the case where there exists a block signal \(I_n^*\): Define the distance between two intervals \(I_1\) and \(I_2\) as

\[
D(I_1, I_2) = \frac{1}{\sqrt{\|I_1\|\|I_2\|}} \left(\frac{1}{\|I_1\|} + \frac{1}{\|I_2\|}\right). 
\]

This metric is first introduced in Arias-Castro, Donoho, and Huo (2005) and is also used in Jeng, Cai, and Li (2010). In addition, we define

\[
D(I_1, \emptyset) = 1 \quad \text{for} \quad I_1 \neq \emptyset, \quad \text{and} \quad D(\emptyset, I_2) = 1 \quad \text{for} \quad I_2 \neq \emptyset.
\]

The definition of consistency for the block signal identification problem is given below.

**Definition 1.** We call a procedure \(\mathcal{P}\) consistent if its estimated interval \(\hat{I}_n\) satisfies

\[
P_{H_0}(\hat{I}_n \neq \emptyset) \leq \alpha \quad (2)
\]

and

\[
P_{H_1}(D(\hat{I}_n, I_n^*) > \delta_n) \to 0 \quad (3)
\]

for some \(\delta_n = o(1)\), where \(\emptyset\) denotes the empty set and \(\alpha\) denotes the significance level.

In this section, we focus on those \(I_n^*\) satisfying the following property: there exists a \(\kappa > 0\), such that \(|I_n^*| \ll n^{1-\kappa}\). This mild assumption includes all intervals with length \(n^{1-\beta}\) for \(0 < \beta \leq 1\), but not those with length \(n / \log n\). The set of intervals considered here greatly extends those in Jeng, Cai, and Li (2010), in which the authors require \(\log(|I_n^*|) \ll \log n\).

Before introducing our identification procedure, we first introduce the concept of the approximation set which is introduced in Rivera and Walther (2013); Walther (2010); Chan and Walther (2013). The idea of the approximation set is that we only need to consider intervals with endpoints on a grid as long as we can approximate each interval relatively well. In this section, we define our approximation set as below:

\[
\mathcal{I}_{\text{app}} = \bigcup_{\ell=1}^{\ell_{\text{max}}} \mathcal{I}_{\text{app}}(\ell) \bigcup \mathcal{I}_{\text{small}}, \quad \text{where} \quad \ell_{\text{max}} = \left\lfloor \log_2 \frac{n}{\log n} \right\rfloor
\]

in which,

\[
\mathcal{I}_{\text{app}}(\ell) = \{ (j, k) : j, k \in \{ id_{\ell}, i = 0, 1, \ldots \} \text{ and } m_{\ell} < k - j \leq 2m_{\ell} \},
\]

\[
\mathcal{I}_{\text{small}} = \{ (j, k) : k - j \leq m_{\ell_{\text{max}}} \}
\]

where \(m_{\ell} = n^{2^{-\ell}}, d_{\ell} = \lceil \frac{m_{\ell}}{6\sqrt{\ell}} \rceil\). A simple counting argument shows that \(|\mathcal{I}_{\text{app}}(\ell)| \leq
\[
\left(\frac{n}{d} + 1\right)\left(\frac{m_i}{d} + 1\right) \leq 144\ell^2d^\ell \quad \text{for} \quad \ell = 1, \ldots, \ell_{\text{max}} \quad \text{and} \quad |\mathcal{I}_{\text{small}}| \leq 2n \log n.
\]
Thus \(|\mathcal{I}_{\text{app}}| \leq \sum_{\ell=1}^{\ell_{\text{max}}} 144\ell^2d^\ell + 2n \log n = O(n \log n)

**Remark 1.** In fact we can let \(d_i = \lceil \frac{m_i}{c\ell^d} \rceil\) for some \(c > 0\) and \(\zeta \geq 0.5\). \(c\) and \(\zeta\) control the precision of the approximation set and the choice is a tradeoff between computational efficiency and approximation error: the larger the \(c\) and \(\zeta\), the better the approximation while the heavier the computation.

Define \(Y(I) = \sum_{i \in I} Y_i / \sqrt{|I|} \). Our identification procedure, denoted by \(\mathcal{P}_n\), works as follows:

If \(\max_{I \in \mathcal{I}_{\text{app}}} (Y(I) - \sqrt{2 \log \frac{|I|}{n}}) < \gamma_n(x)\), where \(\gamma_n(x)\) is the \((1 - x)\) quantile of the null distribution of \(\max_{I \in \mathcal{I}_{\text{app}}} (Y(I) - \sqrt{2 \log \frac{|I|}{n}})\), we claim there exists no signal, i.e., \(\hat{I}_n = \emptyset\).

Otherwise, our estimated interval is

\[
\hat{I}_n = \arg\max_{I \in \mathcal{I}_{\text{app}}} \left( Y(I) - \sqrt{2 \log \frac{en}{|I|}} \right). \quad (4)
\]

By Boole’s inequality, we can show that under the null distribution, \(P_n := \max_{I \in \mathcal{I}_{\text{app}}} (Y(I) - \sqrt{2 \log \frac{|I|}{n}}) = O_p(1)\), so \(\limsup_{n \to \infty} \gamma_n(x) < \infty\), see Rivera and Walther (2013).

It can be shown that \(\mathcal{P}_n\) is optimal for block signal identification. In fact, the procedure \(\mathcal{P}_n\) is consistent in identifying the support \(I_n^*\) whenever the signal is in the detectable region. We summarize this fact in the following theorem.

**Theorem 1.** Assume Model (1) and there exists a \(\kappa > 0\) such that \(|I_n^*| \ll n^{1-\kappa}\). If \(\mu \geq \left(\sqrt{2 \log \frac{en}{|I_n^*|}} + b_n\right) / \sqrt{|I_n^*|}\) with \(b_n \to +\infty\), then our identification procedure \(\mathcal{P}_n\) is consistent with any \(1 \gg \delta_n \gg \sqrt{\log \log n / \sqrt{\log n}}\). In addition, this procedure can be computed in \(O(n \log n)\) time.

From Section 2 of Chan and Walther (2013), \(\mu \geq \left(\sqrt{2 \log \frac{en}{|I_n^*|}} + b_n\right) / \sqrt{|I_n^*|}\) with \(b_n \to +\infty\) is necessary for any test to be consistent in detecting the signal \(I_n^*\). Since the identification problem is more challenging than the corresponding detection problem, \(\mu \geq \left(\sqrt{2 \log \frac{en}{|I_n^*|}} + b_n\right) / \sqrt{|I_n^*|}\) with \(b_n \to +\infty\) is also necessary for any procedure to be consistent in identifying the signal \(I_n^*\). Thus our procedure \(\mathcal{P}_n\) is optimal in block signal identification under our current setting. As a corollary, the block signal identification problem has the same statistical difficulty as the corresponding detection problem: whenever we can detect the signal, we can identify the location of the signal. In the computational aspect, our procedure runs in \(O(n \log n)\) time, even though there are \(O(n^2)\) number of possible intervals in the sequence.
Remark 2. If we define $d_l = \lceil \frac{cn}{c^l} \rceil$ for some $c > 0$ and $\zeta \geq 0.5$ in the definition of approximation set, then Theorem 1 still holds and the computational complexity is $O(nc^2 \log^{2\zeta} n)$.

We make the following comparison between our procedure $\mathcal{P}_n$ and the Likelihood Ratio Selector (LRS) procedure in Jeng, Cai, and Li (2010). (1) In LRS procedure, the authors assume the length of the signal $|I_n^r|$ satisfying $\log |I_n^r| = o(\log n)$. In contrast, we allow $|I_n^r| = n^{1-\beta}$ for $0 < \beta \leq 1$. (2) The identification boundary for LRS procedure is $\sqrt{2 \log n} / |I_n^r|$, which is optimal only for signals with a sufficiently small spatial extent. In contrast, the identification boundary for $\mathcal{P}_n$ is $\sqrt{2 \log \frac{en}{|I_n^r|}} / |I_n^r|$, which is optimal for a broad range of signals. (3) In the LRS procedure, one needs to pre-specify the parameter $L$ which is some number greater than the length of the signal $|I_n^r|$. However, since the length of the signal is unknown, it is not always easy to specify $L$ and the misspecification of $L$ may cause misidentification. Our procedure $\mathcal{P}_n$ has no need to pre-specify the parameter $L$ or any other unknown parameters. (4) The LRS procedure has a computational complexity of $O(nL)$. Depending on the choice of $L$, this complexity could be large. In contrast, our procedure $\mathcal{P}_n$ has a computational complexity of $O(n \log n)$, regardless of the length of the signal. We also note that besides LRS procedure in Jeng, Cai, and Li (2010), similar multiscale methods also appear in Arias-Castro, Donoho, and Huo (2005), Arias-Castro, Candès, and Durand (2011) and Neill et al. (2004). However, these methods do not lead to the same optimal results in our setting.

It is also interesting to compare block signal identification with block signal detection. The block signal detection problem is considered in Chan and Walther (2013); Walther (2010) and is defined in the same way except that the task is to test whether $H_0 : \mu = 0$ vs $H_1^{(n)} : \mu = \mu(n) > 0$. If $\mu = \mu(n) \leq (\sqrt{2} - \epsilon_n) \sqrt{\log \frac{en}{|I_n^r|}} / \sqrt{|I_n^r|}$ with $\epsilon_n \sqrt{\frac{\log \frac{en}{|I_n^r|}}{C_3}} \rightarrow \infty$, then no consistent test exists, see Chan and Walther (2013). In Chan and Walther (2013), the author also shows the test based on penalized scan,

\[
P_n = \max_{0 \leq j < k \leq n} \left( \frac{\sum_{i=j+1}^{k} X_i}{\sqrt{k-j}} - \sqrt{2 \log \frac{en}{k-j}} \right).
\]

is optimal for block signal detection for signal with arbitrary spatial extent and the detection boundary for the penalized scan is $\sqrt{2 \log \frac{en}{|I_n^r|}}$. Our identification procedure is built upon penalized scan and can achieve stronger results: we can not only determine whether the signal exists but also locate where is the signal.

3. Multi-dimensional rectangular signal identification

In this section, instead of one dimensional case, we consider the problem of identifying the rectangular signal in the multi-dimensional hyper-rectangle.

Consider the $D$-dimensional model

\[
Y_i = \mu I_{I_n^r}(i) + Z_i, \quad i = (i_1, ..., i^D) \in [1, ..., n]^D
\] (5)
where $Z_i$ are i.i.d standard normal random variable and $I_n^*$ are unknown rectangle with sides parallel to the axes and with arbitrary sizes and aspect ratios.

All of the results in this section can be easily extended to higher dimensions, but we will focus on the two-dimensional case $D=2$ to simplify our notation. We use the superscript $\{2\}$ to denote we are now considering the two-dimensional case.

We first introduce the approximation set for two-dimensional rectangle, which is a variation of the construction in Walther (2010). Using $(j_1, j_2, k_1, k_2)$ denotes the rectangle with vertices $(j_1, k_1), (j_2, k_1), (j_2, k_2)$ and $(j_1, k_2)$. For fixed $\ell$ and $0 \leq i \leq \ell$, define

$$T^{(2)}_{app}(\ell, i) = \{(j_1, j_2, k_1, k_2) : j_1, j_2 \in \{r^{(1)}d^{(1)}_{\ell, i}, r^{(1)} = 0, 1, \ldots\}$$

$$\text{and } k_1, k_2 \in \{r^{(2)}d^{(2)}_{\ell, i}, r^{(2)} = 0, 1, \ldots\},$$

$$0 \leq j_1 \leq \lfloor \ell 2^{\ell - i} \rfloor, 1 \leq j_2 - j_1 \leq \lfloor \ell \rfloor,$$

$$0 \leq k_1 \leq \lfloor \ell 2^i \rfloor, 1 \leq k_2 - k_1 \leq \lfloor 2\ell \rfloor$$

where $d^{(1)}_{\ell, i} = \lfloor \ell - \lfloor \ell \rfloor \rfloor n$, $d^{(2)}_{\ell, i} = \lfloor \ell - \lfloor \ell \rfloor \rfloor n$. Then let $T^{(2)}_{app}(\ell) = \bigcup_{i=0}^{\ell} T^{(2)}_{app}(\ell, i)$.

Define $\ell_{\max} = \lfloor \log_2 n^2 2^{-\ell} \rfloor$, $m_1 = n^2 2^{-\ell}$ and

$$T^{(2)}_{small} = \{I, |I| \leq m_{\ell_{\max}}\}$$

Then our approximation set $T^{(2)}_{app}$ is defined as:

$$T^{(2)}_{app} = \bigcup_{\ell=1}^{\ell_{\max}} T^{(2)}_{app}(\ell) \bigcup T^{(2)}_{small}$$

Define $Y(I) = \frac{\sum_{i=1}^{\|I\|} Y_i}{\sqrt{|I|}}$, now we quantify the two dimensional penalized scan statistic under the null hypothesis.

**Proposition 2.** Define

$$P^{(2)}_n = \max_{I \in T^{(2)}_{app}} \left( Y(I) - \sqrt{2 \log \frac{en^2}{|I|}} \right)$$

under the null hypothesis, $P^{(2)}_n = O_p(1)$, i.e., $P^{(2)}_n$ are uniformly bounded in probability.

We define $\gamma^{(2)}_n(\alpha) < \infty$ as the $(1 - \alpha)$ quantile of the null distribution of $P^{(2)}_n$, which is well defined by Proposition 2.

The identification procedure $\mathcal{P}^{(2)}_n$ in the two-dimensional case works as follows: If $\max_{I \in T^{(2)}_{app}} (Y(I) - \sqrt{2 \log \frac{en^2}{|I|}}) < \gamma^{(2)}_n(\alpha)$, we claim there is no signal, i.e., $\hat{I}_n = \emptyset$. Otherwise, our estimated rectangle

$$\hat{I}_n = \arg\max_{I \in T^{(2)}_{app}} \left( Y(I) - \sqrt{2 \log \frac{en^2}{|I|}} \right)$$

Similar to the one-dimensional case, one can establish the optimality for $\mathcal{P}^{(2)}_n$, which is given in the following theorem.
Theorem 3. Assume Model (5) and there exists a $\kappa > 0$ such that $|I_n^*| \ll n^{2-\kappa}$. If $\mu \geq \left( \sqrt{2\log \frac{en}{I_n}} + b_n \right) / \sqrt{|I_n^*|}$ with $b_n \rightarrow +\infty$, then our procedure $\mathcal{P}_n^{(2)}$ is consistent with $1 \gg \delta_n \gg \sqrt{\log \log n / \log n}$. In addition, this procedure can be computed in $O(n^2 \log^2 n)$ time.

For a very similar argument to Section 2 of Chan and Walther (2013), $\mu \geq \left( \sqrt{2\log \frac{en}{I_n}} + b_n \right) / \sqrt{|I_n^*|}$ is necessary for any test to be consistent in detecting the rectangular signal in the two-dimensional case and as a result, also necessary for any procedure to be consistent in identifying the signal in the two-dimensional case. Thus, as a corollary to Theorem 3, our procedure is optimal in the two-dimensional case and the identification problem and the corresponding detection problem has the same statistical difficulty. Although there are $O(n^4)$ number of possible rectangles, our algorithm runs in $O(n^2 \log^2 n)$. In general, in $d$ dimensional case, our algorithm runs in $O(n^d \log^d n)$.

4. Signal identification for multiple signals

In the previous two sections, we assume there exists only one signal in the sequence or hyper-rectangle. In this section, we generalize our theory to multiple signals case. We will only discuss the univariate case in this section for notation simplicity, but all our theory can be extended to multivariate case by using the corresponding approximation set. The proofs will also follow the same way as below except for changing all notations to multi-dimensional case.

To further simplify the notation and avoid confusion, we suppress the subscript $n$ in $I_n^*$ in this section. We denote the set of true block signals $I^* = \{I_1^*, \ldots, I_K^*\}$, where $K$ is the number of signals. Our model can be written as:

$$Y_i = \mu_{I^*}(i) + Z_i, \quad i = 1, \ldots, n$$

(6)

where $Z_i$ are i.i.d standard normal random variable, the unknown set of intervals $I^* = \{I_1^*, \ldots, I_K^*\} = \{(j_1, k_1), \ldots, (j_K, k_K), \ 0 \leq j_1 < k_1 < j_2 < \ldots < k_K \leq n\}$ and

$$\mu_{I^*}(i) = \begin{cases} \mu_{I_j^*} & i \in I_j^* \\ 0 & \text{otherwise} \end{cases}$$

In another word, $\mu_{I^*}(i)$ has constant value $\mu_{I_j^*}$ on each interval $I_j^*, j = 1, \ldots K$ and 0 otherwise.

Definition 2. We call a procedure $\mathcal{P}$ consistent in identifying Model (6) if its estimated set of signals (intervals) $\hat{I} = \{\hat{I}_1, \ldots, \hat{I}_K\}$ satisfies

$$P_{H_0}(\hat{I} \neq \emptyset) \leq \alpha$$

(7)

$$E_{H_1}(\bar{K}) \leq K + C(\alpha) + o(1)$$

(8)

and
for some $\delta_n = o(1)$, where $\emptyset$ denotes the empty set, $\alpha$ denotes the significant level, $C(\alpha) > 0$ is a function that depends on $\alpha$ and $\lim_{\alpha \to 0} C(\alpha) = 0$.

Equation (9) in fact implies that $P_{H_1}(\hat{K} \geq K) \to 1$ as $n \to \infty$. Together with Equation (8), we can conclude that if a procedure is consistent, it can identify all true intervals and may include some false intervals asymptotically. In expectation, the number of the false intervals identified goes to 0 as $\alpha \to 0$ and $n \to \infty$.

Note this slightly complicated definition reflects a fundamental difficulty in multiple signals case: even after we correctly identifying all signals, we get back to the null case and we cannot avoid the probability less than $\alpha$ that we include a false interval. One can take $\alpha$ to be small so this effect is minimal.

Theorem 1 can be generalized to Model (6) with a little modification as long as $K = O(\log^p n)$ for some $p > 0$ and the minimum distance between two signals $d_{\text{min}} \gg \max_{j=1}^K |I_j^*| \log n$.

Define $I_{\text{app}}$ and $\gamma_n(\alpha)$ as in Section 2. Our identification procedure for multiple signals case, denote by $\mathcal{P}_{n,\text{multi}}$, works as follows:

Initialize our result set $\hat{I}$ to be $\{\}$, where no signal has been identified.

Denote $\hat{I}_{\text{app}} = I_{\text{app}}$.

Let $i = 1$, repeat the following step until $\max_{I \in \hat{I}_{\text{app}}} (Y(I) - \sqrt{2 \log \frac{C}{|I|}}) < \gamma_n(\alpha)$:

\{ 

Denote $\hat{I}_i = \arg\max_{I \in \hat{I}_{\text{app}}} (Y(I) - \sqrt{2 \log \frac{C}{|I|}})$, and add interval $\hat{I}_i$ to our result set $\hat{I}$,

Let $\hat{I}_{\text{app}}^{i+1} = \hat{I}_{\text{app}}^i \setminus \{I \in \hat{I}_{\text{app}}^i : I \cap \hat{I}_i \neq \emptyset\}$.

Increase $i$ by 1.

\}

We have the following theorem regarding the consistency of the procedure $\mathcal{P}_{n,\text{multi}}$.

Theorem 4. Assume Model (6) and there exists a $\kappa > 0$ such that $|I_j^*| \ll n^{1-\kappa}$ for each $j = 1, ..., K$ and the minimum distance between two signals $d_{\text{min}} \gg \max_{j=1}^K |I_j^*| \log n$ and $K = O(\log^p n)$ for some $p > 0$, if for each $j = 1, ..., K$, $\mu_j \geq \left(\frac{2 \log \frac{C}{|I_j^*|} + b_{n,j}}{\sqrt{|I_j^*|}}\right)$ with $b_{n,j}/\sqrt{\log \log n} \to +\infty$, then our procedure $\mathcal{P}_{n,\text{multi}}$ is consistent with $1 \gg \delta_n \gg \sqrt{\log \log n}/\sqrt{\log n}$.

The main difference between Theorem 4 and Theorem 1 is that we require $b_{n,j}/\sqrt{\log \log n} \to +\infty$ rather than $b_{n,j} \to +\infty$. This stronger requirement will ensure that the probability of making mistakes for each iteration is small enough so that the identification procedure is consistent after all iterations. The reason we require the minimum spacing between the signals $d_{\text{min}} \gg \max_{j=1}^K |I_j^*| \log n$ is that if two signals are too close, it is difficult to identify the exact location of the signal under the influence of another signal. For example, with non-diminishing probability that two signals both with length
\[|I^*| \text{ and distance } |I^*| \text{ between them can be identified as one signal with length about 3}|I^*|\]. This is different from the detection problem where no such requirement is needed.

Similar to Section 2, \(\mu_{j} \geq \left(\sqrt{2\log \frac{\alpha}{|I^*_j|}} + b_{n,j}\right)/\sqrt{|I^*_j|}\) with \(b_{n,j} \rightarrow +\infty\) is necessary for any test to be consistent in detecting the signal \(I^*_j, 1 \leq j \leq k\). Since the identification problem is more challenging than the corresponding detection problem, this bound is also necessary for any procedure to be consistent in identifying all of the signal \(I^*_j, 1 \leq j \leq k\). Thus our procedure \(P_{n,\text{multi}}\) is optimal under our current setting.

**Remark 3.** If the true number of signals \(K\) is known, we can stop the procedure \(P_{n,\text{multi}}\) after \(K\) iterations. It can be shown that this modified procedure is still consistent and its estimated set does not contain any false discoveries.

### 5. Simulation study

We have shown that by adopting the correct penalty term, the procedure based on the penalized scan can be much more powerful, which is one of the major difference between our procedure and the LRS procedure in Jeng, Cai, and Li (2010). It would be difficult to compare these two procedures directly since LRS needs to specify additional parameter \(L\) and its optimality is established under some additional assumptions. It is also unclear how LRS works in high-dimensional case. Due to these reasons, in this section, we will illustrate the previous results by comparing the performance of the identification procedure \(P_n\) defined in Section 2 and the identification procedure without the penalty term, denote by \(P_{n,\text{unpen}}\). To be comparable, \(P_{n,\text{unpen}}\) is defined exactly the same as \(P_n\) except we set the penalty term to 0 instead of \(\sqrt{2\log \frac{\alpha}{|I^*_j|}}\). It can be shown that the identification boundary for \(P_{n,\text{unpen}}\) is \(2\sqrt{\log n}/\sqrt{|I^*_n|}\), which is the same as LRS. If \(|I^*_n| = n^{1-p}, p \in (0,1)\), this detection boundary is \(p^{-1/2}\) times larger than the optimal boundary.

Denote the (Hamming) similarity between interval \(I_1\) and \(I_2\) by \(1 - D(I_1, I_2) = \frac{|I_1 \cap I_2|}{\sqrt{|I_1||I_2|}}\) and \(\gamma_n(\alpha)\) as in Section 2. For a particular simulation, if \(\max_{I \in \mathcal{I}_{\text{app}}} \{Y(I) - \sqrt{2\log \frac{\alpha}{|I|}}\} > \gamma_n(\alpha)\), the similarity is calculated as \(\frac{|I_1 \cap I_2|}{\sqrt{|I_1||I_2|}}\), where \(\hat{I}_n\) is defined in Equation (4); otherwise the similarity is set to 0. We do the same for \(P_{n,\text{unpen}}\), except that \(\gamma_n(\alpha)\) is replaced by \(\tau_n(\alpha)\), which is defined as the \((1 - \alpha)\) quantile of \(\max_{I \in \mathcal{I}_{\text{app}}} Y(I)\). In all of our simulations for univariate setting, we choose \(n = 10,000\).

For the first simulation, we give the similarity for different choices of \(\mu \sqrt{|I^*|}\) range from 1.5 to 5 with a step of 0.5 and signal length \(|I^*_n| = 100\) and \(|I^*_n| = 1000\), respectively. The result is given in Figure 1. The testing power of penalized scan is also given in the plot, which should be higher than the similarity of penalized scan identification. The difference between these two curves mainly comes from the approximation error. From Figure 1, we can see that when \(|I^*_n| = 1000\), which is relatively large, \(P_n\) performs
much better than $P_{\text{unpen}}$, while when $|I^*_n| = 100$, which is relatively small, the performance of $P_n$ is only slightly better than $P_{\text{unpen}}$.

For the second simulation, we give the similarity for different choices of the ratio $n/|I^*_n|$. For each choice of $n/|I^*_n|$, we choose $\mu \sqrt{|I^*|} = 1.2 * \sqrt{2 \log \frac{mn}{|I^*|}} + 0.1$. The simulation result is shown in Figure 2. We can see that the $P_{\text{unpen}}$ seems to be powerless when $n/|I^*_n|$ is small ($|I^*_n|$ is large). However, the gap between two procedures becomes smaller as the ratio $n/|I^*_n|$ increase. Again the testing power of penalized scan detection is given in the plot. The difference between the power of penalized scan detection and

Figure 1. Simulated similarities under different $\mu \sqrt{|I^*|}$. The left sub-graph gives the result for $|I^*_n| = 1000$ and the right sub-graph gives the result for $|I^*_n| = 100$. In both graphs, the black solid line denotes the penalized procedure $P_n$, the red dashed line denotes the unpenalized procedure $P_{\text{unpen}}$ and the green dash-dot line denotes the power of penalized scan detection. The $x$-axis is $\mu \sqrt{|I^*|}$ and $y$-axis is similarity/power.

Figure 2. Simulated similarities for penalized procedure (black solid line), unpenalized procedure (red dashed line) and power of penalized scan detection (green dash-dot line) under different choices of ratio $n/|I^*_n|$ and $\mu \sqrt{|I^*|} = 1.2 * \sqrt{2 \log \frac{mn}{|I^*|}} + 0.1$. The $x$-axis is the ratio in log-scale and $y$-axis is the similarity/power.
the similarity of penalized scan identification seems to be stable across different choices of $n/|I_n^*|$. For the third simulation, we compare the performance for Model (5) in the two-dimensional case between our procedure $\mathcal{P}_n^{(2)}$ defined in Section 3 and $\mathcal{P}_n^{(2),\text{unpen}}$, which is defined the same as $\mathcal{P}_n^{(2)}$ except the penalty term is set to 0. The simulation result is shown in Figure 3. In this simulation, we calculate the similarity for different choices of $l := \sqrt{|I^*|}/n$. We choose $n = 100$, so our space is a $100 \times 100$ rectangle. We have four different experiments with (height, width) pair of the signal set to (25, 30, 40, 48), (20, 60) and (15, 80). Note that the area of the rectangle is $|I_n^*| = 1200$ in all cases but the aspect ratios are different. The simulation result shows that $\mathcal{P}_n^{(2)}$ performs better than $\mathcal{P}_n^{(2),\text{unpen}}$ in all cases and the performance is robust to different aspect ratios.

For the fourth simulation, we repeat the first simulation with multiple signals. In a sequence of length 10,000, we have three blocks of signals with length 100, 300 and 1000, respectively. We assume the signals are at least $d_{\text{min}} = 2000$ away from each other.
to mimic the requirement that $d_{\text{min}} \gg \max_{j=1}^{K} |I_j^i| \log n$. In Figure 4, we give the similarity for each of the three blocks under different choices of $\mu_i \sqrt{|I_j^i|}$, $i = 1, 2, 3$ range from 1.5 to 5 with a step of 0.5. For penalized scan, we follow the identification procedure $P_{n, \text{multi}}$ defined in Section 4 and for direct scan, it would be the same without the penalization term. If more than three signals are identified in the procedure, we only take the first three. After identifying those signals, in order to calculate the similarity, we assign them to true signals in the following way: we first assign each identified signal to the true signal with the most similarity; if more than one identified signals are assigned to the same true signal, only the one with highest similarity one will be taken (and in that case at least one of the true signal will be completely missed and will be assigned zero similarity). From Figure 4, we can see that our procedure is doing a reasonably job in identifying all of the three signals. The longer the true signal, the better the penalized scan is over the direct scan.

For the fifth simulation, we repeat the first simulation under nonwhite background noises: (i) t-distribution with degree of freedom 20, (ii) t-distribution with degree of freedom 10, (iii) t-distribution with degree of freedom 5, (iv) Bernoulli distribution where the random variable can take value 1 or $-1$ with equal probability. To be
Figure 5. Simulated similarities for different $\mu \sqrt{p}$ under nonwhite noise. The left column gives the result for $|I_n^*| = 1000$ and the right column gives the result for $|I_n^*| = 100$. The first to the fourth row gives the results for $t$ distribution with $df = 20$, $df = 10$, $df = 5$ and Bernoulli distribution, respectively. In all plots, the black solid line denotes the penalized procedure $P_n$, the red dashed line denotes the unpenalized procedure $P_{n,unpen}$ and the green dash-dot line denotes the power of penalized scan detection. The x-axis is $\mu \sqrt{p}$ and y-axis is similarity/power.
comparable, we normalize these noises so they all have variance one. Results are given in Figure 5. In all cases, either both procedure are powerless or $P_n$ dominate $P_{n \text{unpen}}$. As the tail gets heavier, the similarity we can achieve gets much worse, and for t-distribution with degree of freedom 5, none of the procedure can achieve a meaningful detection/identification results.

All similarities/power in Figures 1–5 are with respect to a 5% significance level. The critical values were simulated with 10,000 Monte Carlo samples, and the similarities/powers were simulated with 2000 Monte Carlo samples. The location of the signal was sampled at random in each of these simulations to avoid confounding with the approximation scheme.

6. Extensions and discussion

6.1. Identification under the exponential family

In this section, we consider the block signal identification problem under the exponential family setting. Suppose we observe independent random variables $Y_i$, $i = 1, \ldots, n$ through the one-dimensional exponential family model $Y_i \sim F_{\nu_i}, i = 1, \ldots, n$, where

$$\nu_i = a + b I_{I_n}(i)$$

with baseline $a$ known, signal strength $b$ unknown, and unknown interval $I_n$ defined the same as in Model (1) and the task is to recover the support of $I_n$. This model recovers Gaussian noise Model (1) when $F$ is the standard Gaussian distribution and $a = 0$. Note that if $|I|$ is sufficiently large, then under $H_0$, $Y(I)$ would be approximately normally distributed, which suggest the optimality results in Section 2 would hold provided $|I_n|$ is large enough. Formally, if $|I_n| \geq \log^{3+\delta} n$ for some $\delta > 0$ and denote $I_{\text{app}}^\text{exp} = I_{\text{app}} \cap \{1 : |I| \geq \log^{3+\delta} n\}$, then our identification procedure in Section 2 is consistent by replacing $I_{\text{app}}$ with $I_{\text{app}}^\text{exp}$, see Frick, Munk, and Sieling (2014) and Arias-Castro, Candès, and Durand (2011) for some similar arguments. Alternatively, one can use $\sqrt{2 \log T(I)}$ instead of $Y(I)$ where $T(I)$ is the local likelihood ratio statistic for testing $H_0 : b = 0$ against $H_1 : b \neq 0$ on interval $I$, and the above statement still holds.

6.2. Identification with unspecified noise distribution

Distribution-free detection has been considered in Arias-Castro et al. (2018) and Tony Cai, Jeng, and Li (2012). In Tony Cai, Jeng, and Li (2012), the authors consider the identification problem with an unspecified noise distribution under the assumption $\log |I_n| = o(\log n)$. The idea is to apply the identification procedure LRS on the “local median transformed” data. For more details about local median transformation, we refer the reader to Brown, Cai, and Zhou (2008). It is worthwhile to note that our procedures $P_n$, $P_{n}^{(2)}$, $P_{n, \text{multi}}$ are also adapted to the local median transformation and would give a near-optimal solution over a broad range of noise distribution with a much milder assumption on the length of the signal.
6.3. Discussion

It is interesting to compare our results with other results in the change-point detection area. For most research in change-point detection, they typically seek to find an rate optimal solution rather than an exact optimal solution, due to the more complicated structure they consider, see for example Frick, Munk, and Sieling (2014) and Brunel (2014). Our signal identification problem has a simpler setting and we can then achieve the exact optimal constant. We have shown in the simulation that the constant actually matters and a suboptimal constant may lead to a significant loss of power. Last but not the least, in Theorem 4, we assume that the number of block signals $K = O(\log^p n)$ for some $p > 0$. If instead we assume $K = n^\delta$ for some $\delta > 0$, then our procedure may not be optimal anymore. In fact, in this case, block signal identification would be statistically more difficult than the block signal detection, as there are so many block signals. The intuition here is the lower bound of the block signal problem can be reduced to a heterogeneous mixture problem, which has been thoroughly studied by Donoho and Jin (2004, 2009) and Hall and Jin (2010) and a different lower bound between detection and identification has been established there given polynomially many signals. The detection version of the problem with polynomially many signals has been studied in Kou and Walther (2019). It would be interesting to develop an optimality theory for identification under this assumption, which is left as an open problem.

Acknowledgments

The author is thankful to Professor G. Walther for supervision and helpful suggestions.

References

Arias-Castro, E., E. J. Candès, and A. Durand. 2011. Detection of an anomalous cluster in a network. The Annals of Statistics 39 (1):278–304.

Arias-Castro, E., R. M. Castro, E. Tánzos, and M. Wang. 2018. Distribution-free detection of structured anomalies: Permutation and rank-based scans. Journal of the American Statistical Association 113 (522):789–801. doi: 10.1080/01621459.2017.1286240.

Arias-Castro, E., D. L. Donoho, and X. Huo. 2005. Near-optimal detection of geometric objects by fast multiscale methods. IEEE Transactions on Information Theory 51 (7):2402–25.

Brown, L. D., T. T. Cai, and H. H. Zhou. 2008. Robust nonparametric estimation via wavelet median regression. The Annals of Statistics 36 (5):2055–84.

Brunel, V.-E. 2014. Convex set detection. arXiv preprint arXiv:1404.6224.

Chan, H. P. 2009. Detection of spatial clustering with average likelihood ratio test statistics. The Annals of Statistics 37 (6B):3985–4010. doi: 10.1214/09-AOS701.

Chan, H. P., and G. Walther. 2013. Detection with the scan and the average likelihood ratio. Statistica Sinica 23:409–28. doi: 10.5705/ss.2011.169.

Delaigle, A., and P. Hall. 2009. Higher criticism in the context of unknown distribution, non-independence and classification. In Perspectives in mathematical sciences I: Probability and statistics, edited by N. S. Narasimha Sastry, T. S. S. R. K. Rao, M. Delampady and B. Rajeev, 109–38. World Scientific Publishing Company.

Donoho, D., and J. Jin. 2004. Higher criticism for detecting sparse heterogeneous mixtures. Annals of Statistics 32 (3):962–94.

Donoho, D., and J. Jin. 2009. Feature selection by higher criticism thresholding achieves the optimal phase diagram. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 367 (1906):4449–70.
Donoho, D., and J. Jin. 2015. Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science* 30 (1):1–25. doi: 10.1214/14-STS506.

Dumbgen, L., and G. Walther. 2008. Multiscale inference about a density. *The Annals of Statistics* 36 (4):1758–85.

Dumbgen, L., and V. G. Spokoiny. 2001. Multiscale testing of qualitative hypotheses. *Annals of Statistics* 29 (1):124–52.

Frick, K., A. Munk, and H. Sieling. 2014. Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76 (3):495–580. doi: 10.1111/rssb.12047.

Gangnon, R. E., and M. K. Clayton. 2001. A weighted average likelihood ratio test for spatial clustering of disease. *Statistics in Medicine* 20 (19):2977–87. doi: 10.1002/sim.917.

Glaz, J., and N. Balakrishnan. 2012. *Scan statistics and applications*. Basel: Birkhäuser.

Hall, P., and J. Jin. 2010. Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* 38 (3):1686–732. doi: 10.1214/09-AOS764.

Jeng, X. J., T. T. Cai, and H. Li. 2010. Optimal sparse segment identification with application in copy number variation analysis. *Journal of the American Statistical Association* 105 (491):1156–66. doi: 10.1198/jasa.2010.tm10083.

Kou, J., and G. Walther. 2019. Large-scale inference with block structure. Manuscript.

Neill, D. B. 2009. An empirical comparison of spatial scan statistics for outbreak detection. *International Journal of Health Geographics* 8 (1):20. doi:10.1186/1476-072X-8-20.

Neill, D. B., A. W. Moore, F. Pereira, and T. M. Mitchell. 2004. Detecting significant multidimensional spatial clusters. In *Advances in Neural Information Processing Systems*, edited by L. K. Saul, Y. Weiss, and L. Bottou, 969–76.

Rivera, C., and G. Walther. 2013. Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. *Scandinavian Journal of Statistics* 40 (4):752–69. doi: 10.1111/sjos.12027.

Rufibach, K., and G. Walther. 2010. The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics* 19 (1):175–90. doi: 10.1080/10618600902779023.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science (New York, N.Y.)* 315 (5813):848–53. doi: 10.1126/science.1136678.

Tony Cai, T., X. J. Jeng, and H. Li. 2012. Robust detection and identification of sparse segments in ultrahigh dimensional data analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74 (5):773–97. doi: 10.1111/j.1467-9868.2012.01028.x.

Walther, G. 2010. Optimal and fast detection of spatial clusters with scan statistics. *The Annals of Statistics* 38 (2):1010–33. doi: 10.1214/09-AOS732.

**Appendix**

**A.1. Proofs for Section 2**

Before presenting the proof of Theorem 1, we first introduce some useful lemmas.

The following lemma is proved in Chan and Walther (2013), which is the consequence of a result in Dumbgen and Spokoiny (2001).

**Lemma 5.** Define $Z_n(I) = \sum_{i \in I} z_i$. Let $J \in \mathcal{I}_n$, where $\mathcal{I}_n = \{(j,k), 0 \leq j < k \leq n\}$ and $J$ does not depend on $Z_n$, then

$$\max_{I \in \mathcal{I}_n, |J| \leq n} \left( Z_n(I) - \sqrt{2 \log \frac{|J|}{|I|}} \right) \overset{d}{=} L < \infty \quad \text{a.s.}$$
here the random variable $L$ defined above is universally applicable for all $I$ and $n$ and is finite almost surely.

Throughout the proof of this article, we ignore the rounding issues in the definition of $d_i$ whenever this does not affect our results. To simplify the notation, in the following proof we use $I^*$ rather than $I_n^*$ to denote the true signal whenever this does not cause confusion.

The following lemma shows that we can approximate every interval well using our approximation set $I_{app}$ defined in Section 2.

**Lemma 6.** For each $I^*$ with $|I^*| \ll n$, there exists an $\tilde{I} \in I_{app}$ such that $D(\tilde{I}, I^*) \leq \frac{1}{3\log \frac{n}{|I^*|}}$ and $\sqrt{2\log \frac{en}{|\tilde{I}|}} - \sqrt{2\log \frac{en}{|I^*|}} = o(1)$.

**Proof.** There are two cases: $|I^*| > m_{\ell_{\max}}$ and $|I^*| \leq m_{\ell_{\max}}$.

Consider first when $|I^*| > m_{\ell_{\max}}$. Let $\ell'$ be the integer satisfying $m_{\ell'} < |I^*| \leq 2m_{\ell'}$. By our construction, unless $|I^*| < m_{\ell'} + 2\ell'$, there exists an interval $\tilde{I} \in I_{app}$, such that $\tilde{I} \subset I^*$ and

$$|\tilde{I}| > \left(1 - \frac{2\ell'}{m_{\ell'}}\right)|I^*| = \left(1 - \frac{1}{3\ell'}\right)|I^*| \geq \left(1 - \frac{1}{3\log \frac{n}{|I^*|}}\right)|I^*|$$

Thus

$$D(\tilde{I}, I^*) = 1 - \sqrt{\frac{|\tilde{I}|}{|I^*|}} \leq 1 - \sqrt{1 - \frac{1}{3\log \frac{n}{|I^*|}}} \leq \frac{1}{3\log \frac{n}{|I^*|}}$$

and

$$0 \leq \sqrt{2\log \frac{en}{|\tilde{I}|}} - \sqrt{2\log \frac{en}{|I^*|}} \leq \sqrt{2\log \frac{en}{|I^*|}} - \sqrt{2\log \frac{en}{|I^*|}} - \log \left(1 - \frac{1}{3\log \frac{n}{|I^*|}}\right) \leq \frac{\log \left(1 - \frac{1}{3\log \frac{n}{|I^*|}}\right)}{2\log \frac{en}{|I^*|}}$$

As we assume $|I^*| \ll n$, then numerator goes to zero and the denominator goes to infinity, so $\sqrt{2\log \frac{en}{|\tilde{I}|}} - \sqrt{2\log \frac{en}{|I^*|}} = o(1)$.

If $|I^*| < m_{\ell'} + 2\ell'$, then there exists an interval $\tilde{I} \in I_{app}$, such that $\tilde{I} \supset I^*$,

$$|\tilde{I}| < \left(1 + \frac{2\ell'}{m_{\ell'}}\right)|I^*| = \left(1 + \frac{1}{3\ell'}\right)|I^*| \leq \left(1 + \frac{1}{3\log \frac{n}{|I^*|}}\right)|I^*|$$

The remaining proof for bounding $D(\tilde{I}, I^*)$ and $\sqrt{2\log \frac{en}{|\tilde{I}|}} - \sqrt{2\log \frac{en}{|I^*|}}$ are similar to the case $\tilde{I} \subset I^*$.

Now consider the case $|I^*| \leq m_{\ell_{\max}}$. In this situation, $I^* \in I_{app}$, and thus we can simply let $\tilde{I} = I^*$ and the lemma holds trivially. $\square$

**Proof of Theorem 1:**
**Proof.** Denote \( X(I) = Y(I) - \sqrt{2 \log \frac{en}{|I|}} \) for interval \( I \).

By the definition of \( \gamma_n(x) \),

\[
P_{H_0}(I \neq \emptyset) = P_{H_0}(\max_{I \in \mathcal{I}_{app}} X(I) > \gamma_n(x)) \leq x
\]

which proves Equation (2).

Now turn to prove Equation (3). Define

\[
S_0 = \{ I \in \mathcal{I}_{app} : I \cap I^* = \emptyset \}
\]

and

\[
S_1 = \{ I \in \mathcal{I}_{app} : I \cap I^* \neq \emptyset \text{ and } D(I, I^*) > \delta_n \}
\]

By Lemma 6, there exists an \( \tilde{I} \in \mathcal{I}_{app} \) such that \( D(\tilde{I}, I^*) \leq \frac{1}{\sqrt{\log |I|}} \frac{1}{2 \log |I|} - \sqrt{2 \log \frac{en}{|I|}} = o(1) \). Now we have,

\[
P_{H_1}(D(\tilde{I}, I^*) > \delta_n) \leq P_{H_1}(\max_{I \in \mathcal{I}_{app}, D(I, I^*) > \delta_n} \max_{I \in S_0} X(I), \gamma_n(x) \geq X(\tilde{I}))
\]

\[
\leq P_{H_1}(\max_{I \in S_0} X(I) \geq X(\tilde{I})) + P_{H_1}(\max_{I \in S_1} X(I) \geq X(\tilde{I})) + P_{H_1}(\gamma_n(x) \geq X(\tilde{I}))
\]

By Lemma 5, \( \max_{I \in S_0} X(I) \leq d L < \infty \) a.s. and \( \limsup_{n \to \infty} \gamma_n(x) < \infty \) a.s. Also notice that under \( H_1 \), by Lemma 6, we have

\[
X(\tilde{I}) = Y(\tilde{I}) - \sqrt{2 \log \frac{en}{|I|}}
\]

\[
\geq -|Z(\tilde{I})| + (1 - D(\tilde{I}, I^*)) \left( \sqrt{2 \log \frac{en}{|I|}} + b_n \right) - \sqrt{2 \log \frac{en}{|I|}}
\]

\[
\geq -|Z(\tilde{I})| + (1 - D(\tilde{I}, I^*)) b_n - \left( \sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I^*|}} \right) - D(\tilde{I}, I^*) \sqrt{2 \log \frac{en}{|I^*|}}
\]

\[
\geq -|Z(\tilde{I})| + (1 - D(\tilde{I}, I^*)) b_n - \frac{2 \log \frac{en}{|I^*|}}{3 \log_2 \frac{n}{|I^*|}} - o(1)
\]

\[
\xrightarrow{p} \infty
\]

so

\[
P_{H_1}(\max_{I \in S_0} X(I) \geq X(\tilde{I})) \to 0 \quad (A1)
\]

and

\[
P_{H_1}(\gamma_n(x) \geq X(\tilde{I})) \to 0 \quad (A2)
\]

To finish the proof, it remains to show that \( P_{H_1}(\max_{I \in S_1} X(I) \geq X(\tilde{I})) \to 0 \). Denote

\[
S_{near} = \{ I \in S_1 : \frac{|I|}{\log n} \leq |I^*| \leq |I| \log n \}
\]
and notice that
\[
P_{H_1}(\max_{I \in S_1} X(I) \geq X(\bar{I})) \leq \sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(\bar{I})) + P_{H_1}(\max_{I \in S_1 \setminus \text{near}} X(I) \geq X(\bar{I}))
\]
\[=: (A) + (B)
\]

We need the following two lemmas to bound part (A) and (B), respectively. The proof of these lemmas is given after this theorem.

**Lemma 7.**
\[
\sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(\bar{I})) \to 0
\]

**Lemma 8.**
\[
P_{H_1}(\max_{I \in S_1 \setminus \text{near}} X(I) \geq X(\bar{I})) \to 0
\]

Combining Lemma 7, Lemma 8, Equations (A1) and (A2) will lead to 
\[
P_{H_1}(D(\bar{I}, I^*) > \delta_n) \to 0
\]
and we finish our proof.

**Proof of Lemma 7:**

For each interval \(I \in S_{\text{near}}\), we have
\[
P_{H_1}(X(I) \geq X(\bar{I})) = P_{H_1}(Y(I) - Y(\bar{I}) \geq \sqrt{2 \log \frac{en}{|I|}} - \sqrt{2 \log \frac{en}{|I'|}})
\]

Simple calculation shows that \(Y(I) - Y(\bar{I})\) has a normal distribution with mean
\[-\mu \sqrt{|I'|}(D(I, I') - D(\bar{I}, I'))\]
and variance
\[2D(I, \bar{I}) \leq 2\]

Thus,
\[
\sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(\bar{I})) \leq \sum_{I \in S_{\text{near}}} \Phi\left(\left(\frac{\sqrt{|I'|}}{2} \frac{D(I, I') - D(\bar{I}, I')} - \sqrt{2 \log \frac{en}{|I|}} + \sqrt{2 \log \frac{en}{|I'|}}\right)/\sqrt{2}\right)
\]
where \(\Phi\) denotes the upper cumulative distribution function of the normal distribution.

Note that \(D(I, I') > \delta_n, D(\bar{I}, I') \leq \frac{1}{3 \sqrt{\log \frac{en}{|I'|}}}, \) and \(\mu \sqrt{|I'|} \geq \sqrt{2 \log \frac{en}{|I'|}} + b_n\), so
\[
\mu \sqrt{|I'|}(D(I, I') - D(\bar{I}, I')) \geq \left(\delta_n - \frac{1}{3 \sqrt{\log \frac{en}{|I'|}}} \sqrt{2 \log \frac{en}{|I'|}} + b_n\right) \approx \sqrt{\log \log n}\]  
(A3)

where the last inequality follows from the assumption that \(\delta_n \gg \sqrt{\log \log n}/\sqrt{\log n}\) and \(b_n \to +\infty\).
When \( I \in S_{\text{near}} \), we have \(|I| \leq |I^*| \leq |I| \log n \), then
\[
\left| \sqrt{2 \log \frac{en}{|I|}} - \sqrt{2 \log \frac{en}{|I^*|}} \right| \leq \sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I|}} + \sqrt{2 \log \frac{en}{|I|}} - \sqrt{2 \log \frac{en}{|I^*|}}
\]
\[
\leq \sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I|}} + \sqrt{2 \log \frac{en}{|I^*|}}
\]
\[
= o(1).
\]

Since we assume there exists a \( \kappa > 0 \), such that \(|I^*| \ll n^{1-\kappa} \), then \( \frac{2 \log \log n}{2 \log \frac{en}{|I^*|}} \to 0 \). By Lemma 6, we have \( \left| \sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I|}} \right| = o(1) \). Thus,
\[
\left| \sqrt{2 \log \frac{en}{|I|}} - \sqrt{2 \log \frac{en}{|I^*|}} \right| = o(1)
\]
(A4)

Combine (A3, A4) and using the inequality \( \Phi(x) \leq \exp \left( -\frac{x^2}{2} \right) \) for \( x > 1 \), we have for each \( I \in S_{\text{near}} \)
\[
\Phi \left( \left( \mu \sqrt{|I^*|} (D(I, I^*) - D(\bar{I}, \bar{I}^*)) - \sqrt{2 \log \frac{en}{|I^*|}} + \sqrt{2 \log \frac{en}{|I|}} \right) / \sqrt{2} \right) \leq \log^{-\eta} n
\]
for all \( \eta > 0 \).

Now we bound the cardinality of set \( S_{\text{near}} \). Since increasing the level \( l \) by one decreases the length of the shortest intervals for each level by a factor of 2, there are at most \( O(\log \log n) \) levels that contains intervals in \( S_{\text{near}} \). For those levels, the smallest grid \( d_l \) is of order \( O\left( \frac{|I^*|}{\log n \sqrt{\log n}} \right) = O\left( \frac{|I^*|}{\log n} \right) \) which happens at the level corresponding to the shortest intervals in \( S_{\text{near}} \) and by noticing \( \bar{l} \) is \( O(\log n) \). Thus, in each of such level, there are at most \( O\left( |I^*| / |I^*| \right) = O(\log n) \) intervals in \( S_{\text{near}} \). So the cardinality of set \( S_{\text{near}} \) is \( O(\log n \log \log n) = O(\log^2 n) \). Thus
\[
\sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(I^*)) \to 0
\]

**Proof of Lemma 8:**

**Proof.** When \( I \in S_1, S_{\text{near}} \), we have \( 1 - D(I, I^*) \leq \min \left( \sqrt{\frac{|I|}{|I^*|}}, \sqrt{\frac{|I^*|}{|I|}} \right) \leq 1 / \sqrt{\log n} \), so \( D(\bar{I}, \bar{I}^*) - D(I, I^*) \leq \left( \frac{1}{\sqrt{\log n}} + \frac{1}{\sqrt{\log n}} - 1 \right) < 0 \) for \( n \) large enough. Thus,
\[
(D(\bar{I}, \bar{I}^*) - D(I, I^*)) \mu \sqrt{|I^*|} \leq \left( \frac{1}{3 \sqrt{\log 2 \frac{n}{|I|}}} + \frac{1}{\sqrt{\log n}} - 1 \right) \sqrt{2 \log \frac{en}{|I^*|} + b_n}
\]
for \( n \) large enough.
Thus,
\[
P_{H_1}(\max_{I \in S_{\text{new}}} X(I) \geq X(\tilde{I})) \\
= P_{H_1}(\max_{I \in S_{\text{new}}} (Z(I) + (1 - D(I, I^*)) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en}{|I|}}) \\
\geq Z(\tilde{I}) + (1 - D(\tilde{I}, I^*)) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en}{|I|}} \\
= P_{H_1}(\max_{I \in S_{\text{new}}} (Z(I) + (D(\tilde{I}, I^*) - D(I, I^*)) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en}{|I|}}) \\
\geq Z(\tilde{I}) - \sqrt{2 \log \frac{en}{|I|}} \\
\leq P_{H_1}(\max_{I \in S_{\text{new}}} (Z(I) - \sqrt{2 \log \frac{en}{|I|} }) + \left( \frac{1}{3 \sqrt{\log 2 \frac{n}{|I^*|}}} + \frac{1}{\sqrt{\log n}} - 1 \right) \left( \sqrt{2 \log \frac{en}{|I^*|}} + b_n \right) \\
\geq Z(\tilde{I}) - \sqrt{2 \log \frac{en}{|I|}} \\
\leq P_{H_1}(\max_{I \in S_{\text{new}}} (Z(I) - \sqrt{2 \log \frac{en}{|I|} }) \\
\geq Z(\tilde{I}) - \left( \sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I|}} \right) \\
- \left( \frac{1}{3 \sqrt{\log 2 \frac{n}{|I^*|}}} + \frac{1}{\sqrt{\log n}} - 1 \right) \left( \sqrt{2 \log \frac{en}{|I^*|}} + \left( 1 - \frac{1}{3 \sqrt{\log 2 \frac{n}{|I^*|}}} - \frac{1}{\sqrt{\log n}} \right) b_n \right) \\
\right)
\]

Since \(\max_{I \in S_{\text{new}}} (Z(I) - \sqrt{2 \log \frac{en}{|I|}}) = O_p(1)\) by Lemma 5, \(Z(\tilde{I}) = O_p(1)\), \(\frac{1}{3 \sqrt{\log 2 \frac{n}{|I|}}} + \frac{1}{\sqrt{\log n}} \left( \sqrt{2 \log \frac{en}{|I|}} \right) = O(1)\), \(\sqrt{2 \log \frac{en}{|I^*|}} - \sqrt{2 \log \frac{en}{|I|}} = o(1)\) by Lemma 6 and \((1 - \frac{1}{3 \sqrt{\log 2 \frac{n}{|I^*|}}} - \frac{1}{\sqrt{\log n}}) b_n \rightarrow \infty\), then \(P_{H_1}(\max_{I \in S_{\text{new}}} X(I) \geq X(\tilde{I})) \rightarrow 0\) and we finish our proof. \(\Box\)

### A.2. Proofs For Section 3

**Proof of Proposition 2:**

**Proof.** Denote the number of rectangles in \(I^{(2)}_{\text{app}}(\ell)\) by \(\# I^{(2)}_{\text{app}}(\ell)\). A simple counting shows that when \(\ell \leq [\log_2 \frac{n^2}{\log n}], I^{(2)}_{\text{app}}(\ell) \leq 2 \ell^3 2^\ell\).

For \(I^{(2)}_{\text{small}}\) part, let’s abuse the notation a bit: for \([\log_2 n^2] \geq \ell > [\log_2 \frac{n^2}{\log n}]\), let \(I^{(2)}_{\text{app}}(\ell) = \{ I \in I^{(2)}_{\text{small}}, n^2 2^{-\ell} < |I| \leq 2n^2 2^{-\ell} \}\). Then simple counting again gives \(I^{(2)}_{\text{app}}(\ell) \leq n^2 \sum_{|I| = n^2 2^{-\ell}} |I| \leq 2n^4 2^{-\ell} \log n \leq 2 \log^3 n 2^\ell \leq 2^{3} 2^\ell\), where the second to the last inequality comes from \(\ell > [\log_2 \frac{n^3}{\log n}]\), and last inequality comes from \(\ell > [\log_2 \frac{n^2}{\log n}] \geq \log n\).
For \( \kappa > 2 \), we obtain:

\[
P(\max_{I \in \mathcal{I}^{(2)}_{\text{app}}} (Y(I) - \sqrt{2 \log |I|} > \kappa))
\]

\[
\leq \sum_{\ell = 1}^{\lfloor \log 2 n \rfloor} \mathcal{I}^{(2)}_{\text{app}}(\ell) \max_{I \in \mathcal{I}^{(2)}_{\text{app}}(\ell)} \exp \left( -\frac{1}{2} \left( \sqrt{2 \log \frac{en^2}{|I|}} + \kappa \right)^2 \right)
\]

\[
\leq \sum_{\ell = 1}^{\lfloor \log 2 n \rfloor} 2\ell^3 2^{-\ell} \exp \left( -\kappa \sqrt{\ell} - \kappa^2 / 2 \right)
\]

\[
= \sum_{\ell = 1}^{\lfloor \log 2 n \rfloor} 2\ell^3 \exp \left( -\kappa \sqrt{\ell} - \kappa^2 / 2 \right)
\]

\[
\leq C \exp \left( -\kappa^2 / 2 \right)
\]

for some constant \( C > 0 \) not depending on \( n \). Thus, we have

\[
\max_{I \in \mathcal{I}^{(2)}_{\text{app}}} \left( Y(I) - \sqrt{2 \log \frac{en^2}{|I|}} \right) = O_p(1)
\]

Analogously to Lemma 6, the following lemma shows that in the two-dimensional case, we can also approximate every rectangle well enough using \( \mathcal{I}^{(2)}_{\text{app}} \).

**Lemma 9.** For each \( I^* \) with \( |I^*| \ll n^2 \), there exists an \( \tilde{I} \in \mathcal{I}^{(2)}_{\text{app}} \) such that

\[
D(\tilde{I}, I^*) \leq \frac{6}{\sqrt{\log \frac{en^2}{|I|}}} \quad \text{and} \quad \sqrt{2 \log \frac{en^2}{|I|}} - \sqrt{2 \log \frac{en^2}{|\tilde{I}|}} = o(1).
\]

The proof of this lemma is very similar to Lemma 6, and thus is omitted. See also Walther (2010).

**Proof of Theorem 3:**

**Proof.** For rectangle \( I \), denote \( X(I) = Y(I) - \sqrt{2 \log \frac{en^2}{|I|}} \).

By the definition of \( \gamma_n^{(2)}(x) \),

\[
P_{H_0}(\tilde{I} \neq \phi) = P_{H_0}(\max_{I \in \mathcal{I}^{(2)}_{\text{app}}} X(I) \geq \gamma_n^{(2)}(x)) \leq x
\]

which proves (2).

Now we turn to prove (3). Again, let \( I^* \) denote the true rectangle with length \( a^* \) and width \( b^* \). Define

\[
S_0 = \{ I \in \mathcal{I}^{(2)}_{\text{app}} : I \cap I^* = \emptyset \}
\]

and

\[
S_1 = \{ I \in \mathcal{I}^{(2)}_{\text{app}} : I \cap I^* \neq \emptyset \text{ and } D(I, I^*) > \delta_n \}
\]

By Lemma 9, there exists an rectangle \( \tilde{I} \in \mathcal{I}^{(2)}_{\text{app}} \), such that

\[
D(\tilde{I}, I^*) \leq \frac{6}{\sqrt{\log \frac{en^2}{|I^*|}}} \quad \text{and} \quad \sqrt{2 \log \frac{en^2}{|I^*|}} - \sqrt{2 \log \frac{en^2}{|\tilde{I}|}} = o(1). \]

Then,
\[ P_{H_1}(D(\bar{I}, I^*) > \delta_n) \leq P_{H_1}\left(\max_{I \in I_{\text{app}}} \max_{D(I, I^*) > \delta_n} X(I, \gamma_n^{(2)}(x)) \geq X(\bar{I})\right) \]
\[ \leq P_{H_1}\left(\max_{I \in S_0} X(I) \geq X(\bar{I})\right) + P_{H_1}\left(\max_{I \in S_1} X(I) \geq X(\bar{I})\right) + P_{H_1}(\gamma_n^{(2)}(x) \geq X(\bar{I})) \]

By Proposition 2, \(\max_{I \in S_0} X(I) = O_p(1)\) and \(\gamma_n^{(2)}(x) < \infty\) a.s. Notice that under \(H_1\), by Lemma 9 and the same reasoning as in the proof of Theorem 1, we have \(X(\bar{I}) \overset{p}{\to} \infty\). Thus \(P_{H_1}(\max_{I \in S_0} X(I) \geq X(\bar{I})) \to 0\) and \(P_{H_1}(\gamma_n^{(2)}(x) \geq X(\bar{I})) \to 0\).

Now consider the term \(P_{H_1}(\max_{I \in S_1} X(I) \geq X(\bar{I}))\). Denote
\[ S_{\text{near}} = \{I \in S_1 : \frac{a^*}{\log n} \leq a \leq a^* \log n \text{ and } \frac{b^*}{\log n} \leq b \leq b^* \log n\} \]

then
\[ P_{H_1}(\max_{I \in S_1} X(I) \geq X(\bar{I})) \leq \sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(\bar{I})) + P_{H_1}(\max_{I \in S_1 S_{\text{near}}} X(I) \geq X(\bar{I})) \]
\[ =: (A) + (B) \]

Consider part (A) first. Similar to the proof of Theorem 1, one can show that for all \(I \in S_{\text{near}}\)
\[ P_{H_1}(X(I) \geq X(\bar{I})) \leq \log^{-\eta} n \]
for all \(\eta > 0\). Simple counting shows the cardinality of the set \(S_{\text{near}}\) is \(O(\log^4 n)\). Thus
\[ \sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(I^*)) \to 0 \]

Now consider part (B), in this case, \(1 - D(I, I^*) \leq 1/\sqrt{\log n}\). Similar to the one-dimension case, now we have for \(n\) large enough,
Similar to the proof of Proposition 2, we know that when depending on $n$ and such that

$$
(D(I', I^*) - D(I, I^*)) \mu \sqrt{|I^*|} \leq \left( \frac{1}{\sqrt{\log n}} + \frac{6}{\sqrt{\log^2 |I^*|}} - 1 \right) \left( \sqrt{2 \log \frac{en^2}{|I^*|}} + b_n \right)
$$

$$
P_{H_I}( \max_{j \in S_{near}} X(I) \geq X(I'))
$$

$$
= P_{H_I}( \max_{j \in S_{near}} (Z(I) + (1 - D(I, I'))) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en^2}{|I|}})
$$

$$
\geq Z(I) + (1 - D(I, I')) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en^2}{|I|}})
$$

$$
= P_{H_I}( \max_{j \in S_{near}} (Z(I) + (D(I, I') - D(I, I'))) \mu \sqrt{|I^*|} - \sqrt{2 \log \frac{en^2}{|I|}})
$$

$$
\geq Z(I) - \sqrt{2 \log \frac{en^2}{|I|}}
$$

$$
\leq P_{H_I}( \max_{j \in S_{near}} (Z(I) + \left( \frac{1}{\sqrt{\log n}} + \frac{6}{\sqrt{\log^2 |I^*|}} - 1 \right) \left( \sqrt{2 \log \frac{en^2}{|I^*|}} + b_n \right) - \sqrt{2 \log \frac{en^2}{|I|}})
$$

$$
\geq Z(I) - \sqrt{2 \log \frac{en^2}{|I|}}
$$

$$
\leq P_{H_I}( \max_{j \in S_{near}} (Z(I) - \sqrt{2 \log \frac{en^2}{|I|}}) \geq Z(I) - \left( \sqrt{2 \log \frac{en^2}{|I^*|}} - \sqrt{2 \log \frac{en^2}{|I|}} \right)
$$

$$
- \left( \frac{6}{\sqrt{\log^2 |I^*|}} + \frac{1}{\sqrt{\log n}} \right) \left( \sqrt{2 \log \frac{en^2}{|I^*|}} + \left( 1 - \frac{6}{\sqrt{\log^2 |I^*|}} - \frac{1}{\sqrt{\log n}} \right) b_n \right)
$$

Since $\max_{j \in S_{near}} (Z(I) - \sqrt{2 \log \frac{en^2}{|I|}}) = O_p(1)$ by Proposition 2, $Z(I) = O_p(1)$,

$$
\left( \frac{6}{\sqrt{\log^2 |I^*|}} + \frac{1}{\sqrt{\log n}} \right) \left( \sqrt{2 \log \frac{en^2}{|I^*|}} \right) = O(1),
$$

$$
\left( \sqrt{2 \log \frac{en^2}{|I^*|}} - \sqrt{2 \log \frac{en^2}{|I|}} \right) = o(1) \text{ by Lemma 9,}
$$

and

$$
\left( 1 - \frac{6}{\sqrt{\log^2 |I^*|}} - \frac{1}{\sqrt{\log n}} \right) b_n \rightarrow \infty, \text{ we have } P_{H_I}( \max_{j \in S_{near}} X(I) \geq X(I')) \rightarrow 0. \quad \square
$$

### A.3. Proof For Section 4

We need the following lemma in the proof of Theorem 4.

**Lemma 10.** Denote $L = \max_{i \in S_{near}} (Y(I) - \sqrt{2 \log \frac{en^2}{|I|}})$. Let $Z$ be a standard normal random variable, not necessarily independent with $L$, then for all $\kappa > 4$, there exists a constant $C > 0$ not depending on $n$ and $\kappa$ such that

$$
P(L + Z > \kappa) \leq C \exp(-\kappa^2/8)
$$

**Proof.** Similar to the proof of Proposition 2, we know that when $\kappa > 2$, $P(L > \kappa) \leq C' \exp(-\kappa^2/2)$ for some $C' > 0$ not depending on $n$ and $\kappa$. Thus,
Proof of Theorem 4:

**Proof.** As before, denote \( X(I) = Y(I) - \sqrt{2 \log \frac{n}{I}} \) for interval \( I \).

When there exists no signal, by the definition of \( M_n \),

\[
P_{\mathcal{H}_0}(\hat{I} \neq \emptyset) = P_{\mathcal{H}_0}(\max_{I \in \mathcal{I}_{app}} X(I) > \gamma_n) \leq \alpha
\]

which proves (7).

Now we turn to prove (9). It is enough to show that with probability approaching 1, we will not stop before the \( K \)th iteration and for each of the first \( K \) iterations, we can correctly identify one of the true signals with precision \( \delta_n \).

Recall the true set of signals \( \mathcal{I}^* = \{ I_{1n}, I_{2n}, ..., I_{Kn} \} \). By **Lemma 6**, for each \( j = 1, ..., K \), there exists an interval \( I_j \), such that \( D(I_j, I_j^*) \leq \frac{1}{3} \left( \frac{1}{\log |I_j^*|} \right) \) and \( \sqrt{2 \log \frac{e}{|I_j|}} - \sqrt{2 \log \frac{e}{|I_j^*|}} = o(1) \).

Consider the event

\[
E_1 = \left\{ \max_{I \in \mathcal{I}_{app}^1} \max_{j = 1}^{K} X(I), \gamma_n(z) < \min_{j = 1}^{K} X(I_j) \right\} := \{ \text{LHS} < \text{RHS} \}
\]

If \( E_1 \) holds, then the interval \( \hat{I}_1 \) identified by the first iteration satisfies \( D(\hat{I}_1, I_j^*) < \delta_n \) for some \( j_1 \) from 1 to \( K \). \( D(\hat{I}_1, I_j^*) < \delta_n \) also implies that these two intervals intersect and their length cannot differ by more than a factor of \( \log n \), so by the assumption that each true signal are at least \( d_{\text{min}} \gg \max_{j=1}^{K} |I_j^*| \log n \) away, we have \( \hat{I}_1 \cap I_j^* = \emptyset \) for all \( j \neq j_1 \). After the first iteration, consider the event

\[
E_2 = \left\{ \max_{I \in \mathcal{I}_{app}^2} \max_{j = 1, j \neq j_1}^{K} X(I), \gamma_n(z) < \min_{j = 1, j \neq j_1}^{K} X(I_j) \right\} := \{ \text{LHS} < \text{RHS} \}
\]

the LHS of \( E_2 \) is non-increasing while the RHS of \( E_2 \) is non-decreasing compared to those of \( E_1 \). Thus, if \( E_1 \) holds, \( E_2 \) must hold, and the interval \( \hat{I}_2 \) identified by the second iteration satisfies \( D(\hat{I}_2, I_{j_2}^*) \leq \delta_n \) for some \( j_2 \) from 1 to \( K \), \( j_2 \neq j_1 \), and \( \hat{I}_2 \cap I_j^* = \emptyset \) for all \( j \neq j_2 \). If this procedure can be repeated for \( K \) times, then we can identify all \( K \) signals with precision \( \delta_n \). Thus,

\[
P_{\mathcal{H}_1} \left( \min_{j = 1}^{K} \max_{I \in \mathcal{I}_{app}^j} D(\hat{I}_j, I_j^*) > \delta_n \right)
\]

\[
\leq P_{\mathcal{H}_1}(E_1^C)
\]

\[
= P_{\mathcal{H}_1} \left( \max_{I \in \mathcal{I}_{app}^1} \max_{j = 1}^{K} X(I), \gamma_n(z) \geq \min_{j = 1}^{K} X(I_j) \right)
\]

Define

\[
S_0 = \{ I \in \mathcal{I}_{app} : I \cap I_j^* = \emptyset \text{ for all } j = 1, ..., K \}
\]

and

\[
S_1 = \{ I \in \mathcal{I}_{app} : I \cap I_j^* \neq \emptyset \text{ for some } j = 1, ..., K \text{ and } \min_{j = 1}^{K} D(I, I_j^*) > \delta_n \}
\]
Then

\[ P_{H_1}(\max_{i \in S_0} \max_{j=1}^K \min_{i' \in \mathcal{X}_i} \log D(I, I_i^*) > \delta_n) \]

\leq P_{H_1}(\max_{i \in S_0} X(I) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) + P_{H_1}(\max_{i \in S_0} X(I) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) + P_{H_1}(\gamma_n(x) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*))

By Lemma 5, \( \max_{i \in S_0} X(I) \leq d \log n \) a.s. and \( \gamma_n(x) < \infty \) a.s. Notice that under \( H_1 \) and our assumption \( b_n \gg \sqrt{\log \log n} \), each \( \log X(I_i^*) \) is Gaussian distributed with mean greater than

\[
\left( 2 \log \frac{en}{I_i^*} + b_n \right) 
\left( 1 - \frac{1}{3 \sqrt{\log \frac{b_n}{I_i^*}}} \right) - \sqrt{2 \log \frac{en}{I_i^*}} \gg \sqrt{\log \log n}
\]

and variance 1. Then similar to the proof of Theorem 1, for each \( j = 1, \ldots, K \) and \( \eta > 0 \), we have

\[ P_{H_1}(\max_{i \in S_0} X(I) \geq X(I_i^*)) \leq \log^{-\eta} n \]

and

\[ P_{H_1}(\gamma_n(x) \geq X(I_i^*)) \leq \log^{-\eta} n \]

As a result, under the assumption that \( K = \log^p n \) for some \( p > 0 \),

\[ P_{H_1}(\max_{i \in S_0} X(I) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) \leq KP_{H_1}(\max_{i \in S_0} X(I) \geq X(I_i^*)) \rightarrow 0 \]

and

\[ P_{H_1}(\gamma_n(x) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) \leq KP_{H_1}(\gamma_n(x) \geq X(I_i^*)) \rightarrow 0 \]

To prove Equation (9), it remains to show that \( P_{H_1}(\max_{i \in S_0} X(I) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) \rightarrow 0 \). Denote

\[ S_{\text{near}} = \left\{ I \in S_1 : \min_{j=1}^K D(I, I_i^*) < 1 - 1/\sqrt{\log n} \right\} \]

then

\[ P_{H_1}(\max_{i \in S_1} X(I) \geq \min_{i' \in \mathcal{X}_i} \log X(I_i^*)) \leq \sum_{j=1}^K \sum_{I \in S_{\text{near}}} P_{H_1}(X(I) \geq X(I_i^*)) + \sum_{j=1}^K P_{H_1}(\max_{i \in S_1_{\text{near}}} X(I) \geq X(I_i^*)) \]

\[ =: (A) + (B) \]

For part (A), as in the proof of Theorem 1, for each \( j = 1, \ldots, K \), \( P_{H_2}(X(I) \geq X(I_i^*)) \leq \log^{-\eta} n \) for all \( \eta > 0 \). Also, in the proof of Theorem 1, we have showed that for each interval \( I_i^* \), there are at most \( O(\log^2 n) \) intervals \( I \) in \( S_1 \) satisfying \( |I|/\log n \leq |I_i^*| \leq |I|/\log n \) and \( I \cap I_i^* \neq \emptyset \). Notice that \( D(I, I_i^*) < 1 - 1/\sqrt{\log n} \) implies \( |I|/\log n \leq |I_i^*| \leq |I|/\log n \) and \( I \cap I_i^* \neq \emptyset \), as otherwise we would have,

\[ D(I, I_i^*) = 1 - \frac{|I \cap I_i^*|}{|I||I_i^*|} \geq 1 - \frac{\min(|I||I_i^*|)}{\sqrt{|I||I_i^*|}} = 1 - \frac{\sqrt{|I||I_i^*|}}{\sqrt{|I||I_i^*|}} \geq 1 - \frac{1}{\sqrt{\log n}} \]

So for each true signal \( I_i^* \), there are at most \( O(\log^2 n) \) intervals \( I \) in \( S_1 \) satisfying \( D(I, I_i^*) < 1 - 1/\sqrt{\log n} \). Thus

\[ S_{\text{near}} \leq \sum_{j=1}^K \{ I \in S_1 : D(I, I_i^*) < 1 - 1/\sqrt{\log n} \} = O(K \log^2 n) = O(\log^{p+2} n) \]

where the last equality comes from \( K = O(\log^p n) \) for some \( p > 0 \). As a result,
\[ (A) := \sum_{j=1}^{K} \sum_{I \in \mathcal{S}_{\text{near}}} P_{H_1}(X(I) \geq X(I_j^*)) \to 0 \]

For part (B), in this situation \( 1 - D(I, I_j^*) \leq 1/\sqrt{\log n} \) for all \( j = 1, ..., K \). Similar to Lemma 8, for \( n \) large enough, we have

\[
(D(\tilde{I}_j, I_j^*) - D(I, I_j^*)) \mu_{I_j^*} \sqrt{|I_j^*|} \leq \left( \frac{1}{\sqrt{\log n}} + \frac{1}{3 \sqrt{\log \frac{n}{|I_j^*|}}} - 1 \right) \left( \sqrt{2 \log \frac{en}{|I_j|}} + b_{n,j} \right)
\]

Thus,

\[
P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} X(I) \geq X(\tilde{I}_j) \right)
\]
\[
\geq Z(\tilde{I}_j) + (1 - D(\tilde{I}_j, I_j^*)) \mu_{I_j^*} \sqrt{|I_j^*|} - \sqrt{2 \log \frac{en}{|I_j^*|}}
\]
\[
\geq Z(\tilde{I}_j) - \sqrt{2 \log \frac{en}{|I_j^*|}}
\]
\[
\leq P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} Z(I) \right) + \left( \frac{1}{\sqrt{\log n}} + \frac{1}{3 \sqrt{\log \frac{n}{|I_j|}}} - 1 \right) \left( \sqrt{2 \log \frac{en}{|I_j|}} + b_{n,j} \right) - \sqrt{2 \log \frac{en}{|I_j|}}
\]
\[
\leq Z(\tilde{I}_j) - \sqrt{2 \log \frac{en}{|I_j|}}
\]
\[
\leq P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} Z(I) \right) - \sqrt{2 \log \frac{en}{|I_j|}}
\]
\[
\geq Z(\tilde{I}_j) - \left( \sqrt{2 \log \frac{en}{|I_j|}} - \sqrt{2 \log \frac{en}{|I_j|}} \right)
\]
\[
- \left( \frac{1}{\sqrt{\log n}} + \frac{1}{3 \sqrt{\log \frac{n}{|I_j|}}} \right) \left( \sqrt{2 \log \frac{en}{|I_j|}} \right) + \left( 1 - \frac{1}{3 \sqrt{\log \frac{n}{|I_j|}}} - \frac{1}{\sqrt{\log n}} \right) b_{n,j}
\]

By Lemma 10, \( P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} Z(I) - \sqrt{2 \log \frac{en}{|I_j|}} - Z(\tilde{I}_j) > \kappa \right) \leq C \exp(-\kappa^2/8) \). Notice that

\[
\left( \frac{1}{\sqrt{\log n}} + \frac{1}{3 \sqrt{\log \frac{n}{|I_j|}}} \right) \left( \sqrt{2 \log \frac{en}{|I_j|}} \right) = O(1) \quad \text{and} \quad \left( 1 - \frac{1}{3 \sqrt{\log \frac{n}{|I_j|}}} - \frac{1}{\sqrt{\log n}} \right) b_{n,j} \gg \sqrt{\log \log n},
\]

thus

\[
P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} X(I) \geq X(\tilde{I}_j) \right) \leq \log^{-\eta} n \quad \text{for all} \quad \eta > 0.
\]

So

\[
(B) := \sum_{j=1}^{K} P_{H_1} \left( \max_{I \in \mathcal{S}_{\text{near}}} X(I) \geq X(\tilde{I}_j) \right) \to 0
\]

and we finish our proof for (9).
Note that for each iteration, we may only remove an interval very close to $I_j^*$ for some $j = 1, \ldots, K$. So after $K$ iteration, there may still exist intervals $I \in \mathcal{I}_{app}^K$ such that $I \cap I_j^* \neq \emptyset$ for some $j$, denote these intervals by $\mathcal{I}_{\text{left}}$. By similar argument as above, we can show that $P(\max_{I \in \mathcal{I}_{\text{left}}} X(I) > \gamma_n(x)) \to 0$. Now consider all intervals in $\mathcal{I}_{app}^K \mathcal{I}_{\text{left}}$. By the definition of $\gamma_n(x)$ and noticing the fact that $\gamma_n(x)$ is non-decreasing in $n$, we can conclude that the number of the false intervals our procedure identifies is controlled by a geometric distribution with parameter $x$. Thus $E\hat{K} \leq K + \frac{x}{1-x} + o(1)$ and (8) follows by letting $C(x) = \frac{x}{1-x}$. $\square$