# MULTIPLE CHANGE-POINT DETECTION VIA A SCREENING AND RANKING ALGORITHM

Ning Hao[1], Yue Selena Niu[1] and Heping Zhang[2]

[1]*University of Arizona and* [2]*Yale University*

*Abstract:* Let $Y_1, \ldots, Y_n$ be a sequence whose underlying mean is a step function with an unknown number of the steps and unknown change points. The detection of the change points, namely the positions where the mean changes, is an important problem in such fields as engineering, economics, climatology and bioscience. This problem has attracted a lot of attention in statistics, and a variety of solutions have been proposed and implemented. However, there is scant literature on the theoretical properties of those algorithms. Here, we investigate a recently developed algorithm called the Screening and Ranking algorithm (SaRa). We characterize the theoretical properties of SaRa and show its superiority over other commonly used algorithms. In particular, we develop a false discovery rate approach to the multiple change-point problem and show a strong sure coverage property for the SaRa.

*Key words and phrases:* Change-point detection, copy number variation, false discovery rate, high dimensional data, screening and ranking algorithm.

## 1. Introduction

Studies of change-point detection date back to 1950s. In the past half century, the topic has attracted a great deal of attention in such fields as statistics, engineering, economics, climatology and bioscience. Specifically, given a sequence of ordered or time dependent random variables, denoted by $Y_1, \ldots, Y_n$, a change point is a position or time at which the structure of this sequence changes; the goal of change-point detection is to estimate the locations of change points and provide an assessment of accuracy. Thus, in climate data series, a change point is a time at which the climate changes dramatically, a threshold which is of interest to climatologists; in financial econometrics, change-point analysis can help identify the directions in the market or economy; in engineering, for a continuous production process, it is important to find out if there is a point where the quality of the products begins to deteriorate. A recent development is its application to genetics on DNA copy number variation detection. The DNA copy number of a region is the number of copies of the genomic DNA. Copy number variation (CNV) usually refers to deletion or duplication of a region of DNA sequences. It is shown by recent studies that CNVs account for an abundance of genetic variation and may influence phenotypic differences. It is then a fundamental problem

to identify the CNVs in genetics; see Zhang (2010) for a thorough introduction on the application of change-point model in CNV detection. Finding CNVs in the massive data produced by modern DNA array technologies is a recent development in change-point problems. The main challenges here is in finding multiple change points accurately and efficiently in an expansive sequence where the length is typically of hundreds of thousands in SNP genotyping data.

A normal mean, multiple change-point model has played an essential role in the statistical analysis of the CNV problem (Olshen et al. (2004), Huang et al. (2005), Zhang and Siegmund (2007), Tibshirani and Wang (2008), Jeng, Cai, and Li (2010)). Let

$$y_i = \theta_i + \varepsilon_i, \qquad \varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \qquad i = 1, \ldots, n, \qquad (1.1)$$

and assume that

$$\theta_1 = \theta_2 = \cdots = \theta_{\tau_1} \neq \theta_{\tau_1+1} = \cdots = \theta_{\tau_2} \neq \theta_{\tau_2+1} = \cdots = \theta_{\tau_J} \neq \theta_{\tau_J+1} = \cdots = \theta_n,$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)^T$ is the location vector of the change points. We call model (1.1) with piecewise constant mean $\boldsymbol{\theta}$ a normal mean change-point model.

In CNV detection, this model plays an essential role in Olshen et al. (2004), Huang et al. (2005), and Zhang and Siegmund (2007). Some authors have considered more restricted models: Tibshirani and Wang (2008) assume that $\boldsymbol{\theta}$ itself is sparse; Jeng, Cai, and Li (2010) assume, in addition, that the nonzero segments of $\boldsymbol{\theta}$ are short. While these additional assumptions are reasonable, the model considered here is more general. Specifically, we only assume that $n$ is large and any two change points are "not too close to each other"; this will be clarified later. We take $J \ll n$ so while the data are high dimensional in terms of the potential locations of the change points, the number of true change-points is limited. Thus, (1.1) is a high-dimensional sparse model with a sequential structure.

For multiple-change-point problem, the number of change points and their locations are to be estimated. Popular multiple change-point detection tools include exhaustive search (Yao (1988), Yao and Au (1989)), binary segmentation (Vostrikova (1981), Olshen et al. (2004)), $\ell_1$ penalization (Huang et al. (2005), Tibshirani and Wang (2008)), and hidden Markov model approach (Wang et al. (2007), Lai, Xing and Zhang (2008)), among others. See Lai and Xing (2011) for multiple change-point models for the exponential family and for a comprehensive review of related methods.

Recently, Niu and Zhang (2012) proposed the Screening and Ranking algorithm (SaRa) as an alternative approach for change-point detection. They observed that when determining whether there is a change at the $j$th position,

the information at positions far from $j$ is rarely useful. Therefore, it is more effi-cient to concentrate on a local neighborhood first, say the interval $(j - h, j + h)$, when making decisions about $j$ and nearby points. Avoiding complex optimiza-tion or iterative algorithm, the SaRa is simple to implement with complexity $O(n)$, which makes the SaRa suitable for analyzing high throughput data. Be-sides computational efficiency, Niu and Zhang (2012) showed that, under mild conditions, the SaRa satisfies a sure coverage property that implies that the SaRa can estimate $J$ and $\boldsymbol{\tau}$ consistently. The SaRa, to our best knowledge, is the only algorithm known to combine both computational simplicity and consistency. In this paper, we derive deeper and broader theoretical properties for the SaRa that are useful for our understanding of its performance.

First of all, we propose a novel false discovery rate (FDR) approach to change-point detection. The multiple change-point problem can be naturally stated as a multiple testing problem. Although many change-point detection tools have been proposed recently, few authors have examined the issue of FDR. Tibshirani and Wang (2008), and Efron and Zhang (2011) studied the FDR on a normal mean change-point model with sparse mean vector $\boldsymbol{\theta}$, and applied their theories to the CNV problem. Our theory is different. We do not assume the sparsity of $\boldsymbol{\theta}$ and focus on the FDR of change-point locations. Specifically, we concentrate on $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n-1})^T$ where $\beta_j = \theta_{j+1} - \theta_j$ and whose support corresponds to the set of true change-point locations. We demonstrate how to es-tablish a well-defined multiple testing framework for change-point problem (1.1) and assess the FDR for the SaRa estimator. In addition, we show how the FDR control procedure helps us select tuning parameters in the SaRa procedure.

Secondly, we characterize the convergence rate of the location estimator $\hat{\boldsymbol{\tau}}$ for the SaRa. The sure coverage property (Niu and Zhang (2012)) states that the SaRa estimator for the location vector satisfies, with probability tending to one, $||\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}||_\infty < h$, where $h$ is a tuning parameter and is of order $O(\log n)$ under some reasonable conditions. Here we give a sharper convergence rate and show that $||\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}||_\infty = O_P(1)$, where the convergence rate is the same as the best result known for the single change-point case (Csörgő and Horváth (1997)). Moreover, we show that our assumptions cannot be weakened further except for the constant 32 in condition (4.8), implying that the SaRa is a nearly optimal procedure.

The SaRa can be easily generalized to solve more general change-point prob-lems. For example, we derive sure coverage properties for some non-normal cases, although a comprehensive study of non-normal data requires further effort.

The rest of this paper is organized as follows. In Section 2, we briefly recall the SaRa procedure and introduce an FDR control approach to multiple change-point detection based on the SaRa. The numerical studies of the FDR approach

are demonstrated in Section 3. In Section 4, a strong sure coverage property of the SaRa is verified to illustrate the optimality of the SaRa, which is followed by some final remarks in Section 5. All proofs are in the Appendix.

## 2. False Discovery Rate for Change-point Detection

From the hypothesis testing perspective, the change-point problem can be viewed as a multiple testing problem by testing every data point as a potential change point. The false discovery rate (FDR) approach to multiple testing problems has been studied extensively since the seminal paper of Benjamini and Hochberg (1995). However, the multiple testing problem derived from change-point detection presents a problem that beyond the classical framework; it has a distinctive correlation, as well as the sequential, structure. We illustrate our approach to this problem.

### 2.1. Change-point detection as a multiple testing problem

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ be a sequence of independent random variables with probability distribution function $F_1, \ldots, F_n$, respectively. The multiple change-point problem can be stated as the hypothesis testing problem

$$H_0 : F_1 = F_2 = \cdots = F_n \text{ vs}$$
$$H_1 : F_1 = F_2 = \cdots = F_{\tau_1} \neq F_{\tau_1+1} = \cdots = F_{\tau_2} \neq F_{\tau_2+1} = \cdots F_{\tau_J} \neq F_{\tau_J+1} = \cdots = F_n, \tag{2.1}$$

where $J$ is an unknown number of change points and $0 < \tau_1 < \cdots < \tau_J < n$ are their unknown locations. The purpose is to make a decision between two hypotheses and, if the alternative is supported, to estimate the number of change points $J$ and the location vector $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_J)^T$. Although this single hypothesis test has been employed to formulate the multiple change-point problem in classical works, it does not address the accuracy of estimation of the change points here.

The testing problem (2.1) is naturally decomposed to the sequence of hypotheses

$$H_0(j) : j \text{ is a change point;} \quad \text{vs} \quad H_1(j) : j \text{ is not a change point,} \tag{2.2}$$

where $j = 1, \ldots, n-1$.

For the normal mean model where $F_j \sim \mathcal{N}(\theta_j, \sigma^2)$, $H_0(j)$ and $H_1(j)$ correspond to $\beta_j \equiv \theta_{j+1} - \theta_j = 0$ and $\beta_j \neq 0$, respectively. One can use $z_j \equiv y_{j+1} - y_j$ as the statistic for testing $H_0(j)$ against $H_1(j)$. However, its power may be limited by the fact that it does not fully utilize the sparse and sequential structures. Observing that the change points are apart from each other in many applications, we reformulate the hypotheses.

First, suppose that the minimal distance between two change points is at least $h$ and modify (2.2) to

$$
\begin{aligned}
&H_0(j) : F_{j+1-h} = \cdots = F_{j+h} \text{ vs} \\
&H_1(j) : F_{j+1-h} = \cdots = F_j \neq F_{j+1} = \cdots = F_{j+h}.
\end{aligned}
\tag{2.3}
$$

Second, it is impossible to recover the true change-point locations exactly in any reasonable asymptotic settings; see Section 4 for details on this aspect. For an FDR theory, we cannot treat the testing problems in (2.3) independently in terms of "true or false" as in the classical framework, so relaxed version of FDR is introduced.

**Definition 1.** Suppose that $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{J}})^T$ is the change point location estimator from some procedure such that $\min_{1 \leq i \leq \hat{J}} \{\hat{\tau}_i - \hat{\tau}_{i-1}\} \geq 2h$. Then $H_0(j)$ is rejected for all $j \in \hat{\boldsymbol{\tau}}$. We define $\hat{\tau}_i$ as a *true positive* if there exists a true change point $\tau_{i'}$ such that $|\tau_{i'} - \hat{\tau}_i| < h$, otherwise, $\hat{\tau}_i$ is a *false positive*. The *false discovery proportion* (FDP) is the number of false positives divided by $\hat{J}$. The FDR is defined as $\mathrm{E}(FDP)$.

For such methods as CBS and $\ell_1$ penalization, it is not straightforward to evaluate the quantity $\hat{\tau}_i - \tau_i$, or the FDR. In the next two subsections, we recall the SaRa and establish an FDR theory for the normal mean change-point model via the SaRa.

## 2.2. The screening and ranking algorithm

The SaRa was proposed by Niu and Zhang (2012) to detect change points in the normal mean model (1.1). For a position $i$, they considered the locally defined statistic,

$$
D_h(j) = \Big( \sum_{k=j+1}^{j+h} Y_k - \sum_{k=j-h+1}^{j} Y_k \Big) h^{-1},
\tag{2.4}
$$

where $h \ll n$. Intuitively, if $j$ is a local maximizer of $|D_h(\cdot)|$ and $|D_h(j)|$ is quite large, it is likely that there is a change point at or around $j$. Therefore, it is reasonable to consider all local maximizers of $|D_h(\cdot)|$ and search for change points among them. First, the SaRa calculates $D_h(\cdot)$ for all positions and finds all $h$-local maximizers of $|D_h(\cdot)|$. Here, $j$ is a $h'$-local maximizer of $|D_h(\cdot)|$ if $|D_h(j)| \geq |D_h(k)|$ for all $k \in (j - h', j + h')$. One can take $h'$ different from $h$, $h' = h$ was used as a default setting in Niu and Zhang (2012). Second, the SaRa estimator is obtained by a thresholding rule $|D_h(\cdot)| > \lambda$, that is applied to all local maxima. Thus

$$
\mathcal{J}_{h,\lambda} = \{\hat{\tau}_i : \hat{\tau}_i \text{ is a local maximizer of } |D_h(\cdot)|, \text{ and } |D_h(\hat{\tau}_i)| > \lambda\}
$$

is the SaRa location estimator. Let $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{J}})^T$, where $\hat{\tau}_1 < \hat{\tau}_2 < \cdots < \hat{\tau}_{\hat{J}}$, $\hat{J} = |\mathcal{J}_{h,\lambda}|$.

An important observation is that the local statistic $D_h(j)$ employed in the SaRa is the natural test statistic for the multiple testing problem (2.3) when the $F_j$'s are normal. If we define $\eta_j^{(h)} = \mathrm{E}(|D_h(j)|)$, then $H_0$ and $H_1$ correspond to $\eta_j^{(h)} = 0$ and $\eta_j^{(h)}$ is a local maximum, respectively. We illustrate how to combine the SaRa with our extended FDR criterion to establish an FDR theory below.

## 2.3. A false discovery rate approach via the SaRa

First, we restate the SaRa procedure from the viewpoint of hypothesis testing for the general setting (2.3). Let $D(j)$ be a test statistic that can be used to test (2.3), and $p(j)$ be the corresponding P-value when available. Without loss of generality, we assume that larger values of $D$ tend to support the alternative hypothesis. The SaRa is as follows. In Step 1, calculate the test statistic $D(j)$, or the P-value $p(j)$ for each $j$. Then pick out all local maximizers of $D(j)$ or, equivalently, local minimizers of $p(j)$. Here, for technical reasons, we temporarily use the $h'$-local extremizer for $h' = 2h$. For example, we call $j^*$ a local minimizer of $p(\cdot)$ if

$$p(j^*) \leq p(j) \quad \text{for all} \quad j \in (j^* - 2h, j^* + 2h). \tag{2.5}$$

We denote the set of the local extremizers by $\mathcal{LM}$. In Step 2, the SaRa estimator $\hat{\mathcal{J}}_{h,\lambda} = \{\hat{\tau}_1 < \cdots < \hat{\tau}_{\hat{J}}\} \subset \mathcal{LM}$ for the locations of change points is defined by a thresholding rule

$$D(j) > \lambda \text{ or } p(j) < p^*.$$

We can rank the elements in $\mathcal{LM}$ by their $D$ values or P-values and get a solution path.

From now on, we use thresholding rule $p(j) < p^*$ to conform with the literature on FDR theory, although calculating $D$ may be more convenient in practice. An advantage of the SaRa is that we can easily control the minimal distance $\min_i\{\hat{\tau}_i - \hat{\tau}_{i-1}\}$ of the SaRa estimator $\hat{\boldsymbol{\tau}}$ by choosing a proper neighborhood when defining the local extremizer. Under (2.5), for any $j_1, j_2 \in \mathcal{LM}$ with $j_1 \neq j_2$, we have $|j_1 - j_2| > 2h$. Because the P-values are derived from local statistics, $p(j_1)$ and $p(j_2)$ are independent. Therefore, we obtain a sequence of independent P-values $\{p(j)|j \in \mathcal{LM}\}$. When the null distribution, $F_0$, of the P-value at a local minimizer is known or can be estimated accurately, the standard FDR control procedure can be applied to the P-value sequence directly. Specifically, suppose $F_0$ is known and consider the set of modified P-values

$$\{F_0^{-1}(p(j))|j \in \mathcal{LM}\} = \{\tilde{p}_{(1)} < \tilde{p}_{(2)} < \cdots < \tilde{p}_{(m)}\},$$

where $m = |\mathcal{LM}|$. Under the null hypothesis $H_0(j)$, $F_0^{-1}(p(j)) \sim \text{Uniform}(0, 1)$. By the Benjamin–Hochberg procedure (Benjamini and Hochberg (1995)), for a target FDR $q^*$, let $k$ be the largest $i$ for which $\tilde{p}_{(i)} \leq \frac{i}{m} q^*$, and then reject all $H(j)$ corresponding to $\tilde{p}_{(1)}, \ldots, \tilde{p}_{(k)}$. In other words, to determine the SaRa estimator from the set $\mathcal{LM}$, a thresholding rule

$$p(j) \leq p^* = F_0(\tilde{p}_{(k)}) \tag{2.6}$$

is used in order to control the FDR at target rate $q^*$. Following the results in Benjamini and Hochberg (1995), and Benjamini and Yekutieli (2001) directly, we have our result.

**Theorem 1.** *If the SaRa procedure with the thresholding rule* (2.6) *is applied, then the FDR of Definition* 1 *is controlled at a level less than or equal to* $((n - J)/n)q^* \approx q^*$.

By definition, $F_0$ is the distribution of $p(j)$ under $H_0(j)$ provided $p(j)$ is a local minimizer. Given a change-point model, $F_0$ depends on only the bandwidth $h$ in (2.4) and (2.5). Although it is usually unknown, the distribution $F_0$ can be estimated in many situations. For the normal mean change-point model, we can generate a long sequence of *i.i.d.* $\mathcal{N}(0, \sigma^2)$ random variables and find the empirical distribution $\hat{F}_0$. In practice, $\sigma^2$ is unknown but can be estimated accurately because of the fact $J \ll n$. Without normality, we can use permutation methods. Let $\nu$ be the index set $\{1, \ldots, n\}$ and $\pi$ be a permutation of $\nu$. Take $\mathbf{y}_{\pi(\nu)} = (y_{\pi(1)}, \ldots, y_{\pi(n)})^T$. If $\boldsymbol{\theta}$ is sparse, the empirical distribution of the local minima of $p(j)$ can be calculated from a sequence of permutations $\mathbf{y}_{\pi_1(\nu)}, \ldots, \mathbf{y}_{\pi_N(\nu)}$. Otherwise, we can estimate $\boldsymbol{\theta}$ by local regression and apply permutation to the residuals.

## 3. Numerical Study

The contemporary genome-wide SNP genotyping array techniques, which can measure half a million SNPs along the whole genome, offer a more sensitive approach to CNV detection, compared to the aCGH techniques. For each subject, the SNP genotyping data usually consist of a sequence of the measurements of Log R ratios. The segments with concentrated high or low Log R ratios correspond to gains or losses of copy numbers. See Peiffer et al. (2006) for more details. The SaRa has been established in to analyze such data. The numerical properties of the SaRa has been investigated intensively by simulation and for data examples in Niu and Zhang (2012), so we focus on the FDR theory in this section.
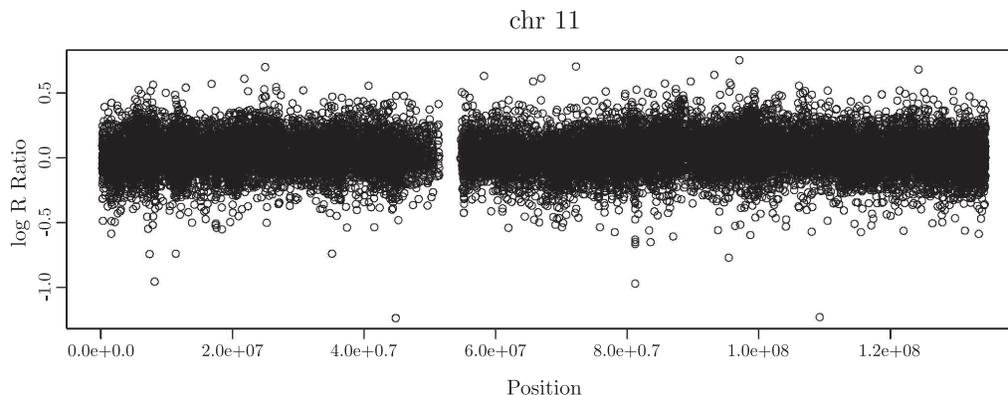
chr 11



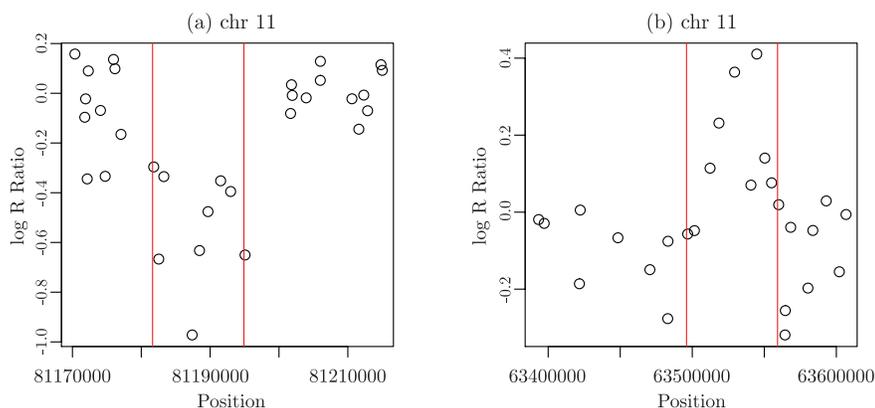Figure 1. Plot of Log R ratios along Chromosome 11 of the subject father.



Figure 2. The CNVs found by the SaRa.

## 3.1. An application to SNP genotyping data

We illustrate the SaRa using SNP genotying data for a father-mother- off-spring trio produced by Illumina 550K platform, which is available `http://www.openbioinformatics.org/penncnv/`.

For each subject, the Log R ratios along Chromosomes 3, 11, and 20 are included in the data set. There are 37768, 27272, and 14296 SNPs on Chromosomes 3, 11, and 20, respectively. In Figure 1, we plot the sequence of Log R ratios along along Chromosome 11 for the father. Since there are 27272 points, it is very difficult to eyeball the changes. In Figure 2, we zoomed in on a short interval where a CNV was detected by the SaRa.

Before we applied the SaRa, we found the Log R ratios to be approximately normal. We chose the bandwidth parameter $h = 7$ and threshold $p^*$ by controlling the FDR to 5%, 10%, and 15%, respectively. Specifically, we calculated the local

Table 1. Number of change points and CNVs detected by the SaRa on Chromosome 11 of the father, mother, and offspring. (A CNV is suggested only when two adjacent change points are within 200 SNPs).

| Source | q = 0.05 | | q = 0.10 | | q = 0.15 | |
|---|---|---|---|---|---|---|
| | Change Points | CNV | Change Points | CNV | Change Points | CNV |
| father | 2 | 1 | 9 | 2 | 9 | 2 |
| mother | 4 | 1 | 5 | 1 | 5 | 1 |
| offspring | 3 | 1 | 3 | 1 | 4 | 1 |

statistic $D_h(\cdot)$ and $p(j) = 2\left(1 - \Phi(|D_7(j)|/(\hat{\sigma}\sqrt{2/h}))\right)$, where $\hat{\sigma}$ can be easily estimated because of the sparsity of the change points. Note that under the normality assumption the distribution of $F_0$ is independent of $\sigma^2$. Therefore, $F_0$ can be approximated empirically by the distribution of the local minimizers in a long $i.i.d.$ standard normal sequence. Then we calculated the corrected P-values and applied the Benjamini-Hochberg procedure to determine the threshold. The numbers of change points detected by the SaRa on Chromosome 11 for the father, mother and offspring are listed in Table 1. Because the CNVs are believed to be short, in dozens of SNPs for our data, (Zhang et al. (2009), Jeng, Cai, and Li (2010)), the intervals that are flanked by two relatively near change points are more likely to be CNVs. On the other hand, it is reasonable to treat those isolated change points as false positives. In Table 1, we report the number of the detected changes by the SaRa, requiring each suggested CNV be flanked by two change points within 200 SNPs. Therefore, the number of the suggested CNVs in Table 1 is fewer than the number of adjacent change points. In particular, on Chromosome 11 of the father, the SaRa suggested two short CNVs as plotted in Figure 2. The CNV in Figure 2(a) was also detected by `PennCNV`, although the one in Figure 2(b) was not detected before.

## 3.2. A simulation example on the FDR control

We tested our theory on FDR control through an example. Consider model (1.1) with $n = 30,000$ and $\sigma = 1$. We set $J = 50$ and drew 50 change points uniformly between 1 and 30,000 that were multiples of 5, to get $\boldsymbol{\tau} = (650, 855, \ldots, 29630)^T$. Here $L = \min(\tau_{j+1} - \tau_j) = 15$. We designed the mean vector $\boldsymbol{\theta}$ by letting $\theta_i = 0$ when $\tau_{2j} \leq i \leq \tau_{2j+1}$, and $\theta_i = \delta$ otherwise, where $\delta = 1.5$ or 3.

We set the SaRa with $h = 10$, 20, and 30, and chose threshold $p^*$ by controlling FDR via the Benjamini-Hochberg procedure BH(q) with $q = 0.05, 0.10, 0.15$. We counted $\hat{\tau}_k$ as a false positive if there was no $\tau_j$ such that

$$|\hat{\tau}_k - \tau_j| < 10. \tag{3.1}$$

Otherwise, $\hat{\tau}_k$ was counted as a true positive. We present our results in Table 2. We see that the average false discovery proportion was close to the target FDR,

Table 2. The average estimated number of change points $\hat{J}$, true positives (TP), and false discovery proportion (FDP). The results were based on 100 replications.

| | $q = 0.05$ | | | $q = 0.10$ | | | $q = 0.15$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{J}$ | TP | FDP | $\hat{J}$ | TP | FDP | $\hat{J}$ | TP | FDP |
| $\delta = 1.5$, $h = 10$ | 3.700 | 3.520 | 0.4% | 20.860 | 19.130 | 7.6% | 27.690 | 23.640 | 13.6% |
| $\delta = 1.5$, $h = 20$ | 45.730 | 43.600 | 4.5% | 50.710 | 45.600 | 9.9% | 54.620 | 46.560 | 14.5% |
| $\delta = 1.5$, $h = 30$ | 50.580 | 47.130 | 6.7% | 53.800 | 47.380 | 11.7% | 56.740 | 47.460 | 16.1% |
| $\delta = 3$, $h = 10$ | 51.500 | 49.920 | 3.0% | 53.680 | 49.970 | 6.7% | 57.040 | 49.980 | 12.1% |
| $\delta = 3$, $h = 20$ | 50.380 | 49.070 | 2.5% | 52.820 | 49.070 | 7.0% | 55.000 | 49.070 | 10.6% |
| $\delta = 3$, $h = 30$ | 50.770 | 48.650 | 4.1% | 53.000 | 48.650 | 8.0% | 55.490 | 48.650 | 12.1% |

suggesting that our methodology worked well. In particular, when the signal was strong ($\delta = 3$), the SaRa with bandwidth $h = 10$ achieved the best performance. The SaRa with bandwidths $h = 20$ and 30 performed well also. However, because there were two true change points at positions 11,070 and 11,085, it was difficult to detect both using a large bandwidth. When the signal was weak ($\delta = 1.5$), the SaRa with $h = 10$ was less powerful. Although the FDR can be controlled, only a small portion of true change-points can be detected. The SaRa with larger bandwidths are more powerful. When the bandwidth was too large ($h = 30$), the FDP was somewhat greater than the target FDR. There were two reasons. First, our assumption $h < L$ was not satisfied for large $h$; second, the rule in (3.1) for counting true positives was too strict.

## 4. Optimality of the Screening and Ranking Algorithm

We start this section with a brief summary of results on change-point detection when the number of change points is at most 2.

Among all change-point problems, the simplest case is the detection of the mean change for a sequence of independent Gaussian random variables with common variance. This has been investigated by Page (1955, 1957), Chernoff and Zacks (1964), Gardner (1969), and Sen and Srivastava (1975), among others. If $Y_i \sim \mathcal{N}(\theta_i, \sigma^2)$ for $i = 1, \ldots, n$, the change-point problem can be stated as the hypothesis test:

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_n; \quad \text{vs} \quad H_1 : \theta_1 = \cdots = \theta_j \neq \theta_{j+1} = \cdots = \theta_n.$$

For simplicity, we assume $\sigma^2$ is known and set to 1. We refer to Csörgö and Horváth (1997), Chen and Gupta (2000), and the references therein for the unknown variance and non-Gaussian cases. Among all methods, the likelihood ratio test is popular. Consider the change point at $j$, and let $\Lambda_j$ denote the

likelihood ratio statistic. Then,

$$-2 \log \Lambda_j = \frac{(\bar{Y}_{j+} - \bar{Y}_{j-})^2}{j^{-1} + (n-j)^{-1}}, \tag{4.1}$$

where $\bar{Y}_{j-} = \sum_{k=1}^{j} Y_k / j$ and $\bar{Y}_{j+} = \sum_{k=j+1}^{n} Y_k / (n-j)$.

When $j$ is unknown, a commonly used test statistic is $T_1 = \max_{1 \le j \le n-1} (-2 \log \Lambda_j)$, and

$$\hat{j} = \operatorname*{argmax}_{1 \le j \le n-1} (-2 \log \Lambda_j) \tag{4.2}$$

serves as the location estimator.

The distribution of the statistic $T_1$ is quite complicated. However, as in Csörgö and Horváth (1997), under the null hypothesis the limiting distribution of $\sqrt{T_1(n)}$ satisfies

$$\lim_{n \to \infty} P\{a_n \sqrt{T_1(n)} - b_n \le t\} = \exp(-2e^{-t}) \text{ for all } t, \tag{4.3}$$

where $a_n = \sqrt{2 \log \log n}$ and $b_n = 2 \log \log n + (1/2) \log \log \log n - (1/2) \log \pi$.

If the experiment is designed in a way such that, at the true change point position $j(n)$ with the mean change of $\delta(n) = \theta_j - \theta_{j+1}$, either of the following two conditions holds

$$0 < \frac{j(n)}{n} \to t < 1, \quad \delta(n) \to 0, \text{ with } \lim_{n \to \infty} \frac{n\delta^2}{\log \log n} = \infty; \tag{4.4}$$

$$\frac{j(n)}{n} \to 0, \qquad \delta(n) \to 0, \text{ with } \lim_{n \to \infty} \frac{j(n)\delta^2}{\log \log n} = \infty, \tag{4.5}$$

it is shown in Csörgö and Horváth (1997) that

$$\delta^2 |\hat{j} - j| = O_P(1). \tag{4.6}$$

In the less challenging case that $\delta(n) \to c > 0$, it can be shown that $|\hat{j} - j| = O_P(1)$, implying $|\hat{j}/n - j/n| = O_P(1/n)$. If all data points are collected from a fixed interval (e.g. $[0, 1]$ or a chromosome), this implies a convergence rate of $1/n$. From (4.3), it is easy to see that (4.4) and (4.5) cannot be relaxed further. Otherwise, the signal is too weak to be detectable.

The second simplest case for the change-point problem has two change points with an epidemic alternative Yao (1993); Arias-Castro, Donoho, and Huo (2005). Specifically, test

$$H_0 : \theta_1 = \cdots = \theta_n = \theta; \quad \text{vs} \tag{4.7}$$

$$H_1 : \theta_1 = \cdots = \theta_l = \theta_{r+1} = \cdots = \theta_n = \theta, \quad \theta_{l+1} = \cdots = \theta_r = \theta + \delta.$$

Arias-Castro, Donoho, and Huo (2005) showed that no method can detect the segment reliably if $\delta^2(r-l) < 2\log n$, which offers a benchmark for necessary conditions to solve the general multiple change-point problems. In the next two subsections, we derive the theoretical properties of the SaRa.

### 4.1. Normal case

Consider the normal mean change-point model (1.1) and define

$$L = \min_{1 \le j \le J+1}(\tau_j - \tau_{j-1}), \quad \delta = \min_{1 \le j \le J} \frac{|\theta_{\tau_j+1} - \theta_{\tau_j}|}{\sigma},$$

where $J$, $\boldsymbol{\tau}$, $\boldsymbol{\theta}$ and $\sigma$ depend on $n$. Here $L$ is the minimal distance among all change points, $\delta$ measures the ratio of the minimum jump size to the standard deviation at change points. The key quantity that reflects the strength of the signal is $S^2 = \delta^2 L$. When the signal is too weak, it is not distinguishable from the noise and cannot be recovered by any methods. Here we consider the setting

$$S^2 = \delta^2 L > 32\log n. \tag{4.8}$$

In the conventional setting of multiple change-point analysis, it is usually assumed that $J$ is constant, $\boldsymbol{\tau}/n$ converge to a constant vector $\mathbf{t}$ as $n \to \infty$, which implies $S^2 = cn \gg \log n$. However, in some applications, $L$ is quite small compared to $n$. Hence it is useful to consider the more flexible condition in (4.8).

**Theorem 2.** *Under* (4.8), *there exist* $h = h(n)$ *and* $\lambda = \lambda(n)$ *such that* $\hat{\mathcal{J}} = \hat{\mathcal{J}}_{h,\lambda} = \{\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{j}}\}$ *satisfies*

$$\lim_{n \to \infty} \mathbf{P}\left(\left\{\hat{J} = J\right\}\right) = 1;$$

*Moreover, conditional on* $\hat{J} = J$,

$$\delta^2(\hat{\tau}_i - \tau_i) = O_P(1). \tag{4.9}$$

*In particular, taking* $h = L/2$ *and* $\lambda = \delta\sigma/2$, *we have*

$$\mathbf{P}\left(\left\{\hat{J} = J\right\} \bigcap \bigcap_i \left\{|\hat{\tau}_i - \tau_i| < h\right\}\right) > 1 - \frac{4\sqrt{2}}{\sqrt{\pi}} S^{-1} \exp\{\log n - \frac{S^2}{32}\}.$$

**Remark 1.** In Niu and Zhang (2012), it was shown that

$$\lim_{n \to \infty} \mathbf{P}\left(\left\{\hat{J} = J\right\} \bigcap \bigcap_i \left\{|\hat{\tau}_i - \tau_i| < h\right\}\right) = 1. \tag{4.10}$$

This is called the sure coverage property since it basically says that the true change points locations $\tau_i$'s are covered by neighborhoods $(\hat{\tau}_i - h, \hat{\tau}_i + h)$ with probability tending to one.

**Remark 2.** The main improvement from Theorem 1 is the convergence rate (4.9), the same as the convergence rate of the likelihood ratio estimator in the single change-point case.

**Remark 3.** The choices of $h$ and $\lambda$ in Theorem 1 may not be optimal in general, especially when $S^2 \gg \log n$. Basically, it is enough to use an $h$ that is slightly greater than $16 \log n / \delta^2$.

**Corollary 1.** *If there is only one change point at position $\tau \leq n/2$ with $\delta^2 \tau > 32 \log n$, then there exist $h$, such that $\hat{\tau} = \operatorname{argmax}_j |D_h(j)|$ satisfies $\delta^2(\hat{\tau} - \tau) = O_P(1)$.*

Note that in this corollary, the convergence rate of the location estimator $\hat{\tau}$ is the same as for the likelihood ratio estimator. The assumption $\delta^2 \tau > 32 \log n$ is slightly stronger than (4.5). Since the likelihood ratio test statistic $-2 \log \Lambda_j$ involves all the data points while the local statistic $D_h(j)$ uses only $2h$ data points, where $h$ is usually $O(\log n / \delta^2)$.

**Remark 4.** The assumption (4.8) cannot be relaxed further except for the constant 32. For example, in the epidemic case (4.7) even if we know that there exists, at most, one segment whose mean is shifted from zero, Arias-Castro, Donoho, and Huo (2005) suggested that (4.8) should be required except for the constant. In our setting, the sparsity of $\boldsymbol{\theta}$ and number of change points are not restricted. Therefore, we do not expect that the 32 can be significantly improved. We conclude that the SaRa is a nearly optimal procedure.

## 4.2. Beyond normality

Suppose
$$y_i = \theta_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{4.11}$$
where the mean vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ is a piecewise constant with change points at $0 < \tau_1 < \cdots < \tau_J < n$, and the noise is not necessarily normal but satisfies the following.

(C0) The noises are *i.i.d.* with $E(\varepsilon_i) = 0$ and $\operatorname{Var}(\varepsilon_i) = \sigma^2 < \infty$. Moreover, the density function of $\varepsilon_i$ is symmetric. With

$$L = \min_{1 \leq j \leq J+1} (\tau_j - \tau_{j-1}), \quad \delta = \min_{1 \leq j \leq J} \frac{|\theta_{\tau_j+1} - \theta_{\tau_j}|}{\sigma},$$

and, for easy presentation, $\sigma^2 = 1$, we consider three cases:

(C1) $\varepsilon_i \sim \mathcal{N}(0,1)$, $S^2 = \delta^2 L > 32 \log n$.

(C2) $E(\exp|\varepsilon_i|/a) \leq b$, $\delta^2 L / (8ab^2 + 2a\delta) \gg \log n$.

(C3) $E(|\varepsilon_i|^t) = m_t < \infty$ for some $t > 2$, $\delta^t L^{t-1} \gg n$ and $\delta^2 L \gg \log n$.

The case (C1) has been characterized in Theorem 1. We give results for the other two cases.

**Theorem 3.** *If* (C0) *and one of* (C2) *and* (C3) *hold, there exist* $h = h(n)$ *and* $\lambda = \lambda(n)$ *such that* $\hat{\mathcal{J}} = \hat{\mathcal{J}}_{h,\lambda} = \{\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{\jmath}}\}$ *satisfies*

$$\lim_{n \to \infty} \mathbf{P}\Big(\big\{\hat{J} = J\big\} \bigcap \bigcap_i \big\{|\hat{\tau}_i - \tau_i| < h\big\}\Big) = 1;$$

*In particular, taking* $h = L/2$ *and* $\lambda = \delta/2$, *we have, under* (C0) *and* (C2),

$$\mathbf{P}\Big(\big\{\hat{J} = J\big\} \bigcap \bigcap_i \big\{|\hat{\tau}_i - \tau_i| < h\big\}\Big) > 1 - 2\exp\Big\{\log n - \frac{\delta^2 L}{64a^2b + 8a\delta}\Big\},$$

*and, under* (C0) *and* (C3),

$$\mathbf{P}\Big(\big\{\hat{J} = J\big\} \bigcap \bigcap_i \big\{|\hat{\tau}_i - \tau_i| < h\big\}\Big) > 1 - C_1\frac{n}{\delta^t L^{t-1}} + \exp\{\log n - C_2\delta^2 L\},$$

*where* $C_1 = (1/2)m_t(4 + 8/t)^t$ *and* $C_2 = (1/8)(t+2)^{-2}e^{-t}$.

## 5. Discussion

Change-point detection is a classic problem with emerging applications across a spectrum of fields from finance to engineering and bioscience. In this work, we focused on the detection of multiple change points, arising from the need to identify CNVs from genomic data. We invoked the recently developed algorithm SaRa of Niu and Zhang (2012), evaluating it as the only algorithm that is shown to reach optimal computational complexity and to possess the sure coverage property. We introduced a concept of false discovery proportion to address the specific setting in the detection of multiple change points, and established FDR theory for the dependent and sequence data pertinent to CNV. We proved a stronger sure coverage property and showed that the convergence rate is optimal, and the same as the existing one for the detection of a single change point. When we applied our procedure to a well-known data set, we confirmed a known CNV and detected a new one, highlighting the potential of the method to detect CNVs that are difficult to detect by existing methods.

## Acknowledgement

## Appendix. Proofs

In this appendix, we prove Theorems 1 and 2, and Corollary 1. Although Corollary 1 is implied directly by Theorem 1, providing its proof might help readers understand the technique.

**A direct proof of Corollary 1.** First note that the condition $\tau \leq n/2$ does not put any restrictions on the position of the change point, by symmetry. The only condition we need is that the position $\tau$ is not too close to the boundary, which is guaranteed by $\delta^2 \tau > 32 \log n$.

Without loss of generality, we assume $\sigma^2 = 1$ and $\theta_{\tau+1} - \theta_\tau > 0$. Fix an integer $h$ such that $16 \log n/\delta < h < \tau$. By definition,

$$D_h(\tau) = \frac{1}{h} \left( \sum_{k=\tau+1}^{\tau+h} y_k - \sum_{k=\tau-h+1}^{\tau} y_k \right) = \delta + \frac{1}{h} \left( \sum_{k=\tau+1}^{\tau+h} \varepsilon_k - \sum_{k=\tau-h+1}^{\tau} \varepsilon_k \right) \sim \mathcal{N}(\delta, \frac{2}{h}).$$

We consider the local statistic $D_h$ at points right of $\tau$ and study the behavior of estimator $\hat{\tau} = \underset{j \geq \tau}{\operatorname{argmax}} \, D_h(j)$. Let $W_m = \varepsilon_{\tau+h+m} - 2\varepsilon_{\tau+m} + \varepsilon_{\tau-h+m}$, $m = 1, \ldots, h$. Note the following.

(1) For $m \geq 2h$, $D_h(\tau + m) \sim \mathcal{N}(0, 2/h)$ is independent of $D_h(\tau)$.
(2) For $h < m < 2h$, $D_h(\tau + m) \sim \mathcal{N}(0, 2/h)$ and $\operatorname{corr}(D_h(\tau + m), D_h(\tau)) = -(2h - m)/2h$.
(3) For $m \leq h$, $D_h(\tau + m) \sim \mathcal{N}((h - m)/m, 2/h)$ and $D_h(\tau + m) - D_h(\tau) = -(m/h)\delta + (1/h) \sum_{\ell=1}^{m} W_\ell$.

To show $\lim_{M \to \infty} \lim_{n \to \infty} \mathbf{P}(\delta^2 |\hat{\tau} - \tau| < M) = 1$, it suffices to verify that

$$\lim_{M \to \infty} \lim_{n \to \infty} \mathbf{P} \left( \max_{m > M/\delta^2} \{D_h(\tau + m) - D_h(\tau)\} < 0 \right) = 1. \tag{A.1}$$

Note that in (A.1), $h > 16 \log n/\delta^2 \gg M/\delta^2$ as $n \to \infty$. Therefore, we can bound the probabilities

$$P_1 = \mathbf{P} \left( \max_{m > h} \{D_h(\tau + m) - D_h(\tau)\} < 0 \right),$$

$$P_2 = \mathbf{P} \left( \max_{h \geq m > M/\delta^2} \{D_h(\tau + m) - D_h(\tau)\} < 0 \right).$$

By (1) and (2), we have

$$D_h(\tau + m) - D_h(\tau) \sim \mathcal{N}(-\delta, \frac{4}{h}), \qquad m \geq 2h,$$

$$D_h(\tau + m) - D_h(\tau) \sim \mathcal{N}(-\delta, \frac{6 - m/h}{h}), \quad h < m < 2h.$$

Therefore,

$$
\begin{aligned}
1 - P_1 &= \mathbf{P}\left(\max_{m>h}\{D_h(\tau+m) - D_h(\tau)\} > 0\right) \\
&< \sum_{m=h+1}^{2h-1} \mathbf{P}\left(D_h(\tau+m) - D_h(\tau) > 0\right) + \sum_{m\geq 2h} \mathbf{P}\left(D_h(\tau+m) - D_h(\tau) > 0\right) \\
&< (h-1)\Phi(-\delta\frac{\sqrt{h}}{\sqrt{5}}) + (n-m-2h+1)\Phi(-\delta\frac{\sqrt{h}}{2}) \\
&< n\Phi(-\delta\frac{\sqrt{h}}{\sqrt{5}}) \\
&< \sqrt{\frac{5}{2\pi}}\frac{1}{\delta\sqrt{h}}\exp\left\{\log n - \frac{\delta^2 h}{10}\right\}.
\end{aligned}
\tag{A.2}
$$

In the last step, we have used the bound on the Gaussian tail probability

$$
\Phi(-t) = 1 - \Phi(t) < \frac{1}{\sqrt{2\pi}}t^{-1}e^{-t^2/2}.
$$

Obviously, $P_1 \to 1$ as $n \to \infty$ when $\delta^2 h > 10\log n$.

To bound the second part, the Bonferroni inequality is not sharp enough. First, by 3),

$$
P_2 = \mathbf{P}\left(\max_{h\geq m>M/\delta^2}\left\{-\frac{m}{h}\delta + \frac{1}{h}\sum_{\ell=1}^{m}W_\ell\right\} < 0\right),
$$

where $W_1,\ldots,W_h$ are $i.i.d.$ $\mathcal{N}(0,6)$. If $B(t)$ is a standard Brownian motion, observe that

$$
\begin{aligned}
P_2 &= \mathbf{P}\left(\max_{h\geq m\geq M/\delta^2}\left\{\frac{1}{m}\sum_{\ell=1}^{m}\frac{W_\ell}{\sqrt{6}} - \frac{\delta}{\sqrt{6}}\right\} < 0\right) \\
&\geq \mathbf{P}\left(\max_{t\geq M/\delta^2}\left\{\frac{B(t)}{t} - \frac{\delta}{\sqrt{6}}\right\} < 0\right) \\
&= \mathbf{P}\left(\max_{t\geq M/\delta^2}\left\{\frac{tB(1/t)}{t} - \frac{\delta}{\sqrt{6}}\right\} < 0\right) \\
&= \mathbf{P}\left(\max_{t\leq\delta^2/M}\left\{B(t) - \frac{\delta}{\sqrt{6}}\right\} < 0\right) \\
&= \mathbf{P}\left(\max_{t\leq\delta^2/M}\left\{\delta B(t/\delta^2) - \frac{\delta}{\sqrt{6}}\right\} < 0\right) \\
&= \mathbf{P}\left(\max_{t\leq 1/M}\left\{B(t) - \frac{1}{\sqrt{6}}\right\} < 0\right) \\
&= \mathbf{P}\left(\max_{t\leq 1/M}\left\{B(t)\right\} < \frac{1}{\sqrt{6}}\right).
\end{aligned}
$$

Therefore, $\lim_{M\to\infty} \lim_{n\to\infty} P_2 = 1$. By symmetry, the behavior of $\hat{\tau} = \operatorname*{argmax}_{j \leq \tau} D_h(j)$ is exactly the same. The proof is finished.

**Proofs of Theorems 1 and 2.** The only difference between Theorems 1 and 2 is that they assume different conditions on the noise distribution. We give a unified proof of the nonasymptotic result for both. We first give notation and two lemmas.

For $L$ an even number fix $h = L/2$ and $\lambda = \delta/2$. Recall that we set $\sigma^2 = 1$. We say a point $j$ is a flat point if there is no change point in the neighborhood $[j - h + 1, j + h]$ so $H_0(j)$ is true at (2.3). Let $\mathcal{F}_h = \{j : H_0(j) \text{ is true }\}$ be the set of all flat points and $\mathcal{J} = \{\tau_i : i = 1, \ldots, n\}$ be the set of all change points. Consider the event $\mathcal{A}_\tau = \{|D(\tau, h)| > \lambda\}$ for a change point $\tau \in \mathcal{J}$ and the event $\mathcal{B}_j = \{|D(j, h)| < \lambda\}$ for a flat point $j \in \mathcal{F}_h$. Let

$$\mathcal{E}_n = \Big( \bigcap_{\tau \in \mathcal{J}} \mathcal{A}_\tau \Big) \bigcap \Big( \bigcap_{j \in \mathcal{F}_h} \mathcal{B}_j \Big).$$

**Lemma A.1.** On $\mathcal{E}_n$, $\hat{J} = J$ and $|\tau_i - \hat{\tau}_i| < h$ for all $i$.

This lemma, shown in Niu and Zhang (2012), says that $\mathcal{E}_n$ implies $\tau_i \in (\hat{\tau}_i - h, \hat{\tau}_i + h)$. The next lemma shows the nonasymptotic properties of Theorem 1 and 2.

**Lemma A.2.** If (C0) and one of (C1), (C2), and (C3) hold, then $\mathbf{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$. In particular, under (C1),

$$\mathbf{P}(\mathcal{E}_n) > 1 - \frac{4\sqrt{2}}{\sqrt{\pi}} S^{-1} \exp\Big\{ \log n - \frac{S^2}{32} \Big\}; \tag{A.3}$$

under (C2),

$$\mathbf{P}(\mathcal{E}_n) > 1 - 2 \exp\Big\{ \log n - \frac{\delta^2 L}{64a^2b + 8a\delta} \Big\}; \tag{A.4}$$

under (C3),

$$\mathbf{P}(\mathcal{E}_n) > 1 - C_1 \frac{n}{\delta^t L^{t-1}} + \exp\{\log n - C_2 \delta^2 L\}, \tag{A.5}$$

where $C_1 = (1/2) m_t (4 + 8/t)^t$ and $C_2 = (1/8)(t + 2)^{-2} e^{-t}$.

Proof of Lemma 2. Under (C1), we have

$$D_h(j) \sim \mathcal{N}(0, \frac{2}{h}) \qquad \text{for} \quad j \in \mathcal{F}_h,$$

$$D_h(\tau) \sim \mathcal{N}(\delta_\tau, \frac{2}{h}) \qquad \text{for} \quad \tau \in \mathcal{J},$$

where $\delta_\tau = \theta_{\tau+1} - \theta_\tau$.

For a fixed flat point $j$,

$$\mathbf{P}(\mathcal{B}_j^c) = \mathbf{P}\Big(|D_h(j)| > \frac{\delta}{2}\Big) = 2\Big(1-\Phi\Big(\frac{\delta\sqrt{h}}{2\sqrt{2}}\Big)\Big) = 2\Big(1-\Phi\Big(\frac{\delta\sqrt{L}}{4}\Big)\Big) = 2\Big(1-\Phi\Big(\frac{S}{4}\Big)\Big),$$

where $\Phi$ is the standard normal distribution function. Similarly, for a change point $\tau$, since $|\delta_\tau| > \delta$ by definition, we have

$$\mathbf{P}(\mathcal{A}_\tau^c) = \mathbf{P}\Big(|D_h(\tau)| < \frac{\delta}{2}\Big) < 1 - \Phi\Big(\frac{S}{4}\Big).$$

By the Bonferroni Inequality and the bound on the Gaussian tail probability, we have

$$\mathbf{P}(\mathcal{E}_n^c) < \sum_{\tau\in\mathcal{J}}\mathbf{P}(\mathcal{A}_\tau^c) + \sum_{j\in\mathcal{F}_h}\mathbf{P}(\mathcal{B}_j^c) < 2n(1-\Phi(\frac{S}{4})) < \frac{4\sqrt{2}}{\sqrt{\pi}}S^{-1}\exp\{\log n - \frac{S^2}{32}\}.$$

Therefore, (A.3) holds.

Under (C0) and (C2), the probabilities $\mathbf{P}(\mathcal{A}_\tau^c)$ and $\mathbf{P}(\mathcal{B}_j^c)$ can be bounded using Lemma 2.2.11 in van der Vaart and Wellner (1996).

Note that $\mathrm{E}(\exp|\varepsilon_i|/a) \le b$ implies $\mathrm{E}|\varepsilon_i|^m < bm!a^m = m!a^{m-2}(2a^2b/2)$.

For a flat point $j$,

$$h \cdot D_h(j) = \sum_{i=j-h+1}^{j} Y_i - \sum_{i=j+1}^{j+h} Y_i = \sum_{i=j-h+1}^{j} \varepsilon_i - \sum_{i=j+1}^{j+h} \varepsilon_i.$$

Because $\varepsilon_i$ is symmetric, $h \cdot D_h(j)$ is a sum of $2h$ i.i.d. random variables. By the Bernstein's Inequality,

$$
\begin{aligned}
\mathbf{P}(\mathcal{B}_j^c) &= \mathbf{P}(|D_h(j)| > \frac{\delta}{2}) \\
&= \mathbf{P}(|h \cdot D_h(j)| > \frac{\delta h}{2}) \\
&\le 2\exp\Big\{-\frac{1}{2}\frac{\delta^2 h^2/4}{2a^2 b \cdot 2h + a\delta h/2}\Big\} \\
&= 2\exp\Big\{-\frac{\delta^2 L}{64a^2 b + 8a\delta}\Big\}
\end{aligned}
$$

$\mathbf{P}(\mathcal{A}_\tau^c)$ can be bounded by the same probability. By the Bonferroni Inequality, we have

$$\mathbf{P}(\mathcal{E}_n^c) < 2n\exp\Big\{-\frac{\delta^2 L}{64a^2 b + 8a\delta}\Big\} = 2\exp\Big\{\log n - \frac{\delta^2 L}{64a^2 b + 8a\delta}\Big\}.$$

Under (C0) and (C3), the inequality (A.5) can be obtained via results on large deviation of sums of independent random variables. For example, applying Corollary 1.8. in Nagaev (1979),

$$
\begin{aligned}
\mathbf{P}(\mathcal{B}_j^c) = \mathbf{P}\Big(|h \cdot D_h(j)| > \frac{\delta h}{2}\Big) \\
\leq C_1 \delta^{-t} L^{1-t} + \exp\{-C_2 \delta^2 L\},
\end{aligned}
$$

where $C_1 = (1/2)m_t(4 + 8/t)^t$ and $C_2 = (1/8)(t+2)^{-2}e^{-t}$. The same is true for $\mathbf{P}(\mathcal{A}_\tau^c)$, $\tau \in \mathcal{J}$. By the Bonferroni Inequality, we have

$$
\mathbf{P}(\mathcal{E}_n^c) < n(C_1 \delta^{-t} L^{1-t} + \exp\{-C_2 \delta^2 L\}) = C_1 \frac{n}{\delta^t L^{t-1}} + \exp\{\log n - C_2 \delta^2 L\}.
$$

Theorem 2 and part of Theorem 1 are straightforward corollaries of Lemmas 1 and 2. The remaining part of Theorem 1, $\delta^2(\hat{\tau}_i - \tau_i) = O_P(1)$, can be shown in the same way as the proof of Corollary 1. In fact, from the discussion above, we know $|\hat{\tau}_i - \tau_i| < h$ on event $\mathcal{E}_n$, which holds with probability close to 1. To sharpen our result to $\delta^2(\hat{\tau}_i - \tau_i) = O_P(1)$, it suffices to repeat the procedure used to bound $P_2$ in Corollary 1.

# References

Arias-Castro, E., Donoho, D. L. and Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51**.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289-300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165-1188.

Chen, J. and Gupta, A. (2000). *Parametric Statistical Change Point Analysis*. Birkhäuser.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35**, 999-1018.

Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.

Efron, B. and Zhang, N. R. (2011). False discovery rates and copy number variation. *Biometrika* **98**, 251-271.

Gardner, L. A. J. (1969). On detecting changes in the mean of normal variates. *Ann. Math. Statist.* **40**, 116-126.

Huang, T., Wu, B., Lizardi, P. and Zhao, H. (2005). Detection of DNA copy number alterations using penalized least squares regression. *Bioinformatics* **21**, 3811-3817.

Jeng, X. J., Cai, T. T., and Li, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.* **105**, 1056-1066.

Lai, T. L. and Xing, H. (2011). A simple bayesian approach to multiple change-points. *Statist. Sinica* **21**, 539-569.

Lai, T. L., Xing, H., and Zhang, N. (2008). Stochastic segmentation models for array-based comparative genomic hybridization data analysis. *Biostatistics* **9**, 290-307.

Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745-789.

Niu, S. Y. and Zhang, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **6**, 1306-1326.

Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557-572.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42**, 523-527.

Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44**, 248-252.

Peiffer, D. A., Le, J. M., Steemers, F. J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., Shaw, C. A., Belmont, J., Cheung, S. W. W., Shen, R. M., Barker, D. L. and Gunderson, K. L. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* **16**, 1136-1148.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3**, 98-108.

Tibshirani, R. and Wang, P. (2008). Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics* **9**, 18-29.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer.

Vostrikova, L. Y. (1981). Detecting "disorder" in multidimensional random processes. *Soviet Mathematics Doklady*, 55-59.

Wang, K. et al. (2007). PennCNV: An integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665-1674.

Yao, Q. (1993). Tests for change-points with epidemic alternatives. *Biometrika* **80**, 179-191.

Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters* **6**, 181-189.

Yao, Y.-C. and Au, S. T. (1989). Least-squares estimation of a step function. *Sankhya A* **51**, 370-381.

Zhang, F., Gu, W., Hurles, M. E. and Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Ann. Rev. Genomics and Human Genetics* **10**, 451-481.

Zhang, N. R. (2010). DNA copy number profiling in normal and tumor genomes. *Frontiers in Computational and Systems Biology*, 259-281.

Zhang, N. R. and Siegmund, D. O. (2007). A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63**, 22-32.

Department of Mathematics, The University of Arizona, Tucson AZ, 85721, USA.

E-mail: nhao@math.arizona.edu

Department of Mathematics, The University of Arizona, Tucson AZ, 85721, USA.

E-mail: yueniu@math.arizona.edu

Department of Epidemiology and Public Health, Yale University, New Haven CT, 06520, USA.

E-mail: heping.zhang@yale.edu