

# Likelihood Scores for Sparse Signal and Change-Point Detection

Shouri Hu<sup>1</sup>, Jingyan Huang<sup>2</sup>, Hao Chen<sup>3</sup>, and Hock Peng Chan

**Abstract**—We consider here the identification of change-points on large-scale data streams. The objective is to find the most efficient way of combining information across data stream so that detection is possible under the smallest detectable change magnitude. The challenge comes from the sparsity of change-points when only a small fraction of data streams undergo change at any point in time. The most successful approach to the sparsity issue so far has been the application of hard thresholding such that only local scores from data streams exhibiting significant changes are considered and added. However the identification of an optimal threshold is a difficult one. In particular it is unlikely that the same threshold is optimal for different levels of sparsity. We propose here a sparse likelihood score for identifying a sparse signal. The score is a likelihood ratio for testing between the null hypothesis of no change against an alternative hypothesis in which the change-points or signals are barely detectable. By the Neyman-Pearson Lemma this score has maximum detection power at the given alternative. The outcome is that we have a scoring of data streams that is successful in detecting at the boundary of the detectable region of signals and change-points. The likelihood score can be seen as a soft thresholding approach to sparse signal and change-point detection in which local scores that indicate small changes are down-weighted much more than local scores indicating large changes. We are able to show sharp optimality of the sparsity likelihood score in the sense of achieving successful detection at the minimum detectable order of change magnitude as well as the best constant with respect this order of change.

**Index Terms**—Asymptotic optimality, change-point, sequence segmentation, signal detection.

## I. INTRODUCTION

CONSIDER a large number  $N$  of data streams containing change-points. We consider the situation in which all data up to a given time is available for analysis, so each data stream

is an observed sequence of length  $T$ . At each change-point one or more of the sequences undergo distribution change. The objective is to identify these change-points and the sequences undergoing distribution change. Of interest here is the identification of these change-points when there is sparsity, that is when the number of sequences undergoing change is small compared to  $N$ . More specifically we want to know the minimum magnitude of change for which the distribution change can be detected under sparsity. And secondly we want to have an algorithm that is able to detect, with high probability, change-points under the minimum detectable change. See Niu et al. [28] and Wang and Samworth [37] for applications to engineering, genomics and finance.

A typical strategy to deal with sparsity is to subject local scores to thresholding or penalization before summing them up across sequence. Algorithms employing this strategy include the Sparsified Binary Segmentation (SBS) [10], the double CUSUM (DC) [9], the Informative Sparse Projection (INSPECT) [37] and the scan algorithm of Enikeeva and Harachoui [15]. The strategy was also employed by Mei [26], Xie and Siegmund [40] and Wang and Mei [38] in sequential change-point detection on multiple sequences, and Zhang et al. [43] to detect distribution deviations from known baselines on multiple sequences. Thresholding and penalization suppress noise by removing small and moderate scores, mostly from the majority of sequences without change, thus enhancing the signals from the sparse sequences with changes. It is however unlikely that we are able to specify a threshold or penalization parameter that is optimal at all levels of sparsity.

The higher-criticism (HC) test statistic, proposed by Tukey [35] to check for significantly large number of small p-values, uses multiple thresholds for sparse mixture detection. The number of p-values below a threshold is transformed to a higher-criticism score and this score is maximized over all thresholds. The Berk and Jones [4] test statistic uses multiple thresholds as well but it applies a different scoring function. The HC test statistic was shown by Donoho and Jin [11] to be optimal in the detection of a sparse normal mixture. Cai and Wu [6] extended the optimality of the HC test statistic to sparse non-normal mixtures and Moscovich et al. [27] extended the optimality of the Berk-Jones test statistic. Cai et al. [5] applied the HC test statistic to detect intervals in multiple sequences where the means of a sparse fraction of the sequences deviate from a known baseline and showed that the HC test statistic is optimal. Chan and Walther [7] considered

Manuscript received 19 April 2022; revised 4 November 2022; accepted 31 January 2023. Date of publication 3 February 2023; date of current version 19 May 2023. This work was supported in part by the Ministry of Education, Singapore, under Grant A-0004820-00-00. The work of Hao Chen was supported in part by NSF under Award DMS-1848579. (Corresponding author: Hao Chen.)

Shouri Hu is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: hushouri@uestc.edu.cn).

Jingyan Huang and Hock Peng Chan are with the Department of Statistics and Data Science, National University of Singapore, Singapore 117546 (e-mail: jingyan\_huang@nus.edu.sg; stachp@nus.edu.sg).

Hao Chen is with the Department of Statistics, University of California at Davis, Davis, CA 95616 USA (e-mail: hxchen@ucdavis.edu).

Communicated by R. Talmon, Associate Editor for Signal Processing and Source Coding.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2023.3242297>.

Digital Object Identifier 10.1109/TIT.2023.3242297

0018-9448 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

sequence length much larger than number of sequences with detection boundaries that are more complex. They showed that the HC test statistic achieves detection at these boundaries and is optimal in more general settings. They also showed that the Berk-Jones test statistic achieves the same optimality.

Our approach here is to convert the p-values into likelihood scores for testing sparse sequences. The scoring applies on each p-value instead of on the number of p-values below a threshold. It can be considered to be a soft form of thresholding in which p-values that are close to zero are penalized less than p-values that are barely significant.

Since the likelihood scores are transformations of p-values, the proposed method can be applied to any type of distribution changes and it can handle data types that vary across sequences. Our theory however requires a specific distribution family for neat asymptotics and we consider here in particular either normal or Poisson data. We show optimality up to the correct asymptotic constants. For sparse normal change-points these constants are two-dimensional extensions of those in Ingster [20] and Donoho and Jin [11] for sparse normal mixture detection. These constants have been discussed in the context of sparse normal change-point detection assuming a known baseline in Chan and Walther [7] and Chan [8]. For sparse Poisson change-points the constants are new and different from sparse normal constants.

The optimality of multiple sequence identification of change-points up to the correct constant is new. Previous works on optimality for normal data are up to the correct order of magnitude though they go beyond the i.i.d. model, for example Pilliat et al. [31] considered sparse change-point detection in time-series with normal errors. Liu et al. [25] showed optimality up to the best order for normal errors, under the constraint of not more than one change-point.

As far as we are aware, there are currently no optimality theory in the literature on sparse change-point detection on Poisson data. For sparse Poisson mixtures of size  $N$ , Arias-Castro and Wang [3] showed that the HC test statistic is optimal when the Poisson means grow faster than  $\log N$ , and that a Bonferroni correction is optimal when the means grow slower than  $\log N$ . Donoho and Kipnis [12] characterized the asymptotic behavior of the HC test statistic on frequency tables with Poisson counts. Stoepker et al. [34] applied the HC test statistic to test against sparse alternatives in multiple data streams, with p-values obtained via permutation tests, and showed optimality for exponential families, covering both normal and Poisson data. They showed that optimality does not require the null distribution to be known, however their problem is different in that the observations are identically distributed and there are no change-points.

The algorithm we propose here has two steps in the identification of two change-points. The first detection screening step applies the Screening and Ranking (SaRa) idea of Niu and Zhang [29]. The second estimation step for more precise location of change-points uses the CUSUM-like procedure of Wild Binary Segmentation (WBS), cf. Fryzlewicz [17]. This two-step approach saves computation time because the fast screening step evaluates a large number of segments whereas the computationally intensive estimation step is only applied

when a change-point has been detected during screening. In contrast for WBS the estimation step is applied on a large number of randomly generated segments. Unlike in Niu and Zhang [29] we do not apply the BIC criterion of Zhang and Siegmund [42] to determine the number of change-points. Instead critical values are specified in advance and binary segmentation, cf. Olshen et al. [30], is applied to detect the change-points sequentially.

An alternative to binary segmentation is estimating the full set of change-points at one go by applying global optimization and making use of dynamic programming to manage the computational complexity. This was employed by the HMM algorithms of Yao [41] and Lai and Xing [24], the multi-scale SMUCE algorithm of Frick et al. [16] and the Bayesian Likelihood algorithm of Du et al. [13]. These methods are however designed for single sequence segmentation. Niu et al. [28] provides an excellent background of the historical developments.

The outline of this paper is as follows. In Section II we introduce the sparse likelihood (SL) scores and show that they are optimal in the detection of sparse normal mixtures. In Section III we extend SL scores to detect change-points in multiple sequences. In Section IV we show that SL scores are optimal for change-point detection when the observations are normal or Poisson. In Section V we discuss the assumptions, implications and contributions of the optimality results. In Section VI we perform simulation studies on the SL scores. In the appendices we prove the optimality of SL scores.

### A. Notations

We write  $a_n \sim b_n$  to denote  $\lim_{n \rightarrow \infty} (a_n/b_n) = 1$ . We write  $a_n = o(b_n)$  to denote  $\lim_{n \rightarrow \infty} (a_n/b_n) = 0$ . We write  $a_n \lesssim b_n$  to denote  $a_n \leq Cb_n$  for all  $n$  for some  $C > 0$  and  $a_n \gtrsim b_n$  to denote  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$ . We write  $X_n = O_p(a_n)$  to denote  $P(X_n \leq Ca_n) \rightarrow 1$  for some  $C > 0$ . Let  $\lfloor \cdot \rfloor$  ( $\lceil \cdot \rceil$ ) denote the greatest (least) integer function. Let  $\phi$  and  $\Phi$  denote the density and distribution function respectively of the standard normal. Let  $\mathbf{1}$  denote the indicator function. Let  $\emptyset$  denote the empty set and let  $\#A$  denote the number of elements in a set  $A$ . Let  $\|\cdot\|$  denote the  $L_2$ -norm of a vector and  $\|\cdot\|_0$  the number of non-zero entries of a vector.

## II. SPARSE MIXTURE DETECTION

We start with the simpler problem of detecting a sparse mixture, with the objective of motivating the sparse likelihood score.

Let  $\mathbf{p} = (p^1, \dots, p^N)$  be independent p-values of  $N$  null hypotheses and let  $p^{(1)} \leq \dots \leq p^{(N)}$  be the sorted p-values. Tukey proposed the higher-criticism test statistic

$$\text{HC}(\mathbf{p}) = \max_{n: Np^{(n)} \leq n} \frac{n - Np^{(n)}}{\sqrt{Np^{(n)}(1 - p^{(n)})}}, \quad (1)$$

with  $\text{HC}(\mathbf{p}) = 0$  if  $Np^{(n)} > n$  for all  $n$ , for the overall test that all null hypotheses are true.

Donoho and Jin [11] showed that the HC test statistic is optimal for detecting a sparse fraction of false null hypotheses.

Consider test scores  $Z^n \sim N(0, 1)$  when the  $n$ th null hypothesis is true and  $Z^n \sim N(\mu_N, 1)$  for some  $\mu_N > 0$  when the  $n$ th null hypothesis is false. Define

$$\rho_Z(\beta) = \begin{cases} \beta - \frac{1}{2} & \text{if } \frac{1}{2} < \beta < \frac{3}{4}, \\ (1 - \sqrt{1 - \beta})^2 & \text{if } \frac{3}{4} \leq \beta < 1. \end{cases} \quad (2)$$

Donoho and Jin [11] showed that on the sparse mixture  $(1 - \epsilon)N(0, 1) + \epsilon N(\mu_N, 1)$  no algorithm is able to achieve, as  $N \rightarrow \infty$ ,

$$P_0(\text{Type I error}) + P_{\mu_N}(\text{Type II error}) \rightarrow 0, \quad (3)$$

for testing  $H_0: \epsilon = 0$  versus  $H_1: \epsilon = N^{-\beta}$ , if  $\mu_N = \sqrt{2\nu \log N}$  for  $\nu < \rho_Z(\beta)$ . They also showed that the HC test statistic achieves (3) when  $\nu > \rho_Z(\beta)$  and is thus optimal. Type I error refers to the conclusion of  $H_1$  when  $H_0$  is true whereas Type II error refers to the conclusion of  $H_0$  when  $H_1$  is true. Ingster (1997, 1998) established the detection lower bound showing that (3) cannot be achieved when  $\nu < \rho_Z(\beta)$ .

Like the HC test statistic, the Berk and Jones [4] test statistic

$$\text{BJ}(\mathbf{p}) = \max_{n: Np^{(n)} \leq n} \left[ n \log \left( \frac{n}{Np^{(n)}} \right) + (N - n) \log \left( \frac{N - n}{N(1 - p^{(n)})} \right) \right] \quad (4)$$

achieves (3) when  $\nu > \rho_Z(\beta)$ .

We introduce the sparse likelihood scores in Section II-A and show that they achieve (3) in the detection of sparse mixtures, when  $\nu > \rho_Z(\beta)$ , in Section II-B.

### A. Sparse Likelihood

Let  $f_1(p) = \frac{1}{p(2 - \log p)^2} - \frac{1}{2}$  and  $f_2(p) = \frac{1}{\sqrt{p}} - 2$ . For both  $i = 1$  and  $2$ ,  $\int_0^1 f_i(p) dp = 0$  and  $f_i(p)$  increases as  $p$  decreases.

Define the sparse likelihood score

$$\ell_N(\mathbf{p}) = \sum_{n=1}^N \ell_N(p^n), \quad (5)$$

where  $\ell_N(p) = \log \left( 1 + \frac{\lambda_1 \log N}{N} f_1(p) + \frac{\lambda_2}{\sqrt{N \log N}} f_2(p) \right)$ ,

with  $\lambda_1 \geq 0$  and  $\lambda_2 > 0$ .

When  $\lambda_1 = 0$ , the sparse likelihood score is the log-likelihood ratio of the null hypothesis  $p^n \sim_{\text{i.i.d.}} \text{Uniform}(0, 1)$  versus the alternative hypothesis

$$p^n \sim_{\text{i.i.d.}} F(p) = p + \frac{2\lambda_2}{\sqrt{N \log N}} (\sqrt{p} - p).$$

Let the empirical distribution function  $\hat{F}(p) = \frac{\#\{n: p^n \leq p\}}{N}$ . Under the null hypothesis,

$$\hat{F}(p) - p = O_p(\sqrt{\frac{p}{N}}).$$

Since  $F(p) - p \sim 2\lambda_2 \sqrt{\frac{p}{N \log N}}$  as  $p \rightarrow 0$ , we are able to detect with small error probabilities if  $\frac{\lambda_2}{\sqrt{\log N}}$  is large. As  $\sqrt{\log N}$  increases slowly with  $N$ , we can view the density  $f(p) = 1 + \frac{\lambda_2}{\sqrt{N \log N}} f_2(p)$  as lying near the boundary where detection with asymptotically zero error probabilities is possible. That is, the sparse likelihood score is the most powerful test for some of the alternatives lying near this boundary.

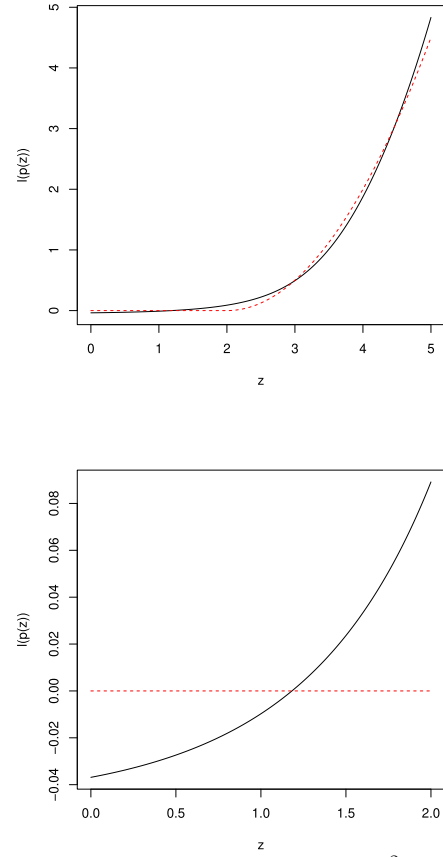


Fig. 1. Graphs of  $\ell_N(p(z))$  (black, —) and  $(z - 2)_+^2/2$  (red, - -), with  $p(z) = 2\Phi(-|z|)$ , for  $0 \leq z \leq 5$  (top) and  $0 \leq z \leq 2$  (bottom). The parameters of  $\ell_N$  are  $N = 500$ ,  $\lambda_1 = 1$  and  $\lambda_2 = 1.84 \left( \doteq \sqrt{\frac{\log T}{\log \log T}} \right)$  for  $T = 500$ . These parameters were applied in the simulation exercises in Section VI-A and correspond to those in (15) used to show optimality in sparse change-point detection.

When  $p$  is of order smaller than  $N^{-1}$ ,  $\frac{\log N}{N} f_1(p)$  dominates  $\frac{1}{\sqrt{N \log N}} f_2(p)$  and the selection of  $\lambda_1 > 0$  is advantageous. This is relevant in the extension of sparse likelihood scores to detect change-points on long sequences where large number of likelihood comparisons is involved.

The sparse likelihood score can be viewed as a form of soft thresholding. To visualize this we compare in Figure 1 the plot of  $\ell_N(p(z))$  for  $p(z) = 2\Phi(-|z|)$ ,  $N = 500$ ,  $\lambda_1 = 1$  and  $\lambda_2 = 1.84$ , against that of  $(z - 2)_+^2/2$ . For  $0 \leq z \leq 5$ , the two functions are close to each other however within  $0 \leq z \leq 2$ ,  $\ell_N(p(z))$  is not constant but has a gentle upward curve. The sparsity likelihood score is negative for  $z \leq 1.18$  and  $\ell_N(p(Z))$  for  $Z$  standard normal has a mean of  $-0.004$ . This negative mean helps in controlling the sum of scores when  $N$  is large and  $p^n \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$ .

### B. Optimal Detection

We show here that the sparse likelihood score is optimal in the detection of change-points for a broad range of sparsity. Let  $E_0$  and  $P_0$  denote expectation and probability respectively with respect to  $p^n \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$ . Since

$$\begin{aligned} & E_0 \exp(\ell_N(\mathbf{p})) \\ &= \prod_{n=1}^N E_0 \left[ 1 + \frac{\lambda_1 \log N}{N} f_1(p^n) + \frac{\lambda_2}{\sqrt{N \log N}} f_2(p^n) \right] = 1, \end{aligned}$$

it follows from Markov's inequality that

$$P_0(\ell_N(\mathbf{p}) \geq c_N) \leq e^{-c_N}. \quad (6)$$

This exponential bound makes the sparsity likelihood score easy to work with when there are large number of likelihood comparisons, as critical values satisfying a required level of Type I error control can have a simple expression not depending on  $N$ . We show in Theorem 1 that by selecting

$$c_N \rightarrow \infty \text{ with } c_N = o(N^\delta) \text{ for all } \delta > 0, \quad (7)$$

the Type I and II error probabilities both go to zero at the detection boundary.

**Theorem 1:** Assume (7). Consider the test of  $H_0: Z^n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$  versus  $H_1: Z^n \stackrel{\text{i.i.d.}}{\sim} (1 - \epsilon)N(0, 1) + \epsilon N(\mu_N, 1)$ , for  $1 \leq n \leq N$ , with  $\epsilon = N^{-\beta}$  for some  $\frac{1}{2} < \beta < 1$ . Consider the likelihood score  $\ell_N(\mathbf{p})$  with parameters  $\lambda_1 \geq 0$  and  $\lambda_2 > 0$  not depending on  $N$ , and p-values  $p^n = \Phi(-Z^n)$ . If  $\mu_N = \sqrt{2\nu \log N}$  for  $\nu > \rho_Z(\beta)$ , then

$$P_0(\ell_N(\mathbf{p}) \geq c_N) + P_{\mu_N}(\ell_N(\mathbf{p}) < c_N) \rightarrow 0.$$

### III. CHANGE-POINT DETECTION

Let  $X_t^n$  denote the  $t$ th observation of the  $n$ th sequence for  $1 \leq t \leq T$  and  $1 \leq n \leq N$ . Consider first the model

$$X_t^n \sim_{\text{indep.}} N(\mu_t^n, 1). \quad (8)$$

We are interested in the detection and estimation of

$$\tau := \{t : \mu_t^n \neq \mu_{t+1}^n \text{ for some } n\}.$$

For  $s < t$ , let  $\bar{X}_{st}^n = (t - s)^{-1} \sum_{u=s+1}^t X_u^n$ . To check for a change of mean on the  $n$ th sequence at location  $t$ , select  $s < t < u$  and let p-value

$$p_{stu}^n = 2\Phi(-|Z_{stu}^n|), \text{ where } Z_{stu}^n = \frac{\bar{X}_{tu}^n - \bar{X}_{st}^n}{\sqrt{(u-t)^{-1} + (t-s)^{-1}}}.$$

In the sparse likelihood algorithm we combine these p-values using  $\ell_N(\mathbf{p}_{stu})$ , where  $\mathbf{p}_{stu} = (p_{stu}^1, \dots, p_{stu}^N)$ . When the data follow some other distributions, the corresponding likelihood ratio statistic and p-value can be computed accordingly.

Sparse likelihood scores detects well when only a small fraction of the sequences undergo change of mean. For  $T$  large computing the sparse likelihood score for all  $(s, t, u)$  is expensive. Instead we combine the approximating set idea of Arias-Castro et al. [1] and Walther [36] to first space out the  $(s, t, u)$  that are evaluated, and to apply the CUSUM-type scores used in WBS to estimate the change-point location accurately only when the first step indicates a change-point.

In addition to computational savings, through this two-step approach we are able to incorporate multi-scale penalization terms similar to those used in Dümbgen and Spokoiny [14] and the SMUCE algorithm of Frick et al. [16], to ensure optimality not only at all levels of sparse change-points, but also at all orders of change magnitudes.

Let  $1 \leq h_1 < h_2 < \dots$  and  $1 \leq d_1 < d_2 < \dots$  be integer-valued sequences with  $h_i \geq d_i$  for all  $i$ . Our grid approach uses segments of length  $2h_i$  spaced  $d_i$  apart, with segments near the two ends shortened due to edge effect. For

a dataset of length  $g$  there are  $K_i(g) = \lfloor \frac{g-1}{d_i} \rfloor$  segments of length  $2h_i$  and we consider all  $i$  from 1 to

$$i_g = \max\{i : h_i + d_i \leq g\}. \quad (9)$$

More specifically define

$$\begin{aligned} \mathcal{A}_i(g) &= \{(s(ik), t(ik), u(ik)) : 1 \leq k \leq K_i(g)\}, \\ s(ik) &= \max(0, kd_i - h_i), \\ t(ik) &= kd_i, \\ u(ik) &= \min(kd_i + h_i, g). \end{aligned}$$

The elements of  $\mathcal{A}_i(g)$  are the indices where sparse likelihood scores for segments of length  $2h_i$  are computed. Initially we have the full dataset  $\mathbf{X}_{1:T} = (X_t^n : 1 \leq t \leq T, 1 \leq n \leq N)$  and after one or more change-points have been estimated, it is split into sub-datasets  $\mathbf{X}_{b:e} = (X_t^n : b \leq t \leq e, 1 \leq n \leq N)$ , with length  $g = e - b + 1$ . We check for change-points in  $\mathbf{X}_{b:e}$  using segments specified by  $\mathcal{A}_i(g)$ .

Let the penalized sparse likelihood scores

$$\ell_N^{\text{pen}}(\mathbf{p}_{stu}) = \ell_N(\mathbf{p}_{stu}) - \log\left(\frac{T}{4} \left(\frac{1}{t-s} + \frac{1}{u-t}\right)\right). \quad (10)$$

The detection of change-points within  $\mathbf{X}_{b:e}$ , with segment lengths of at least  $2h_{i_0}$ , is as follows.

---

#### Algorithm 1 SL-Estimate

---

```

INPUT( $c, i_0, b, e$ )
 $\mathbf{X} \leftarrow \mathbf{X}_{b:e}$ 
 $g \leftarrow e - b + 1$ 
FOR  $i = i_0, \dots, i_g$ 
  IF  $\max_{1 \leq k \leq K_i(g)} \ell_N^{\text{pen}}(\mathbf{p}_{s(ik), t(ik), u(ik)}) \geq c$  THEN
     $j \leftarrow \arg\max_{k: 1 \leq k \leq K_i(g)} \ell_N^{\text{pen}}(\mathbf{p}_{s(ik), t(ik), u(ik)})$ 
     $\hat{\tau} \leftarrow \lceil \arg\max_{t: s(ij) < t < u(ij)} \ell_N^{\text{pen}}(\mathbf{p}_{s(ij), t, u(ij)}) \rceil + b - 1$ 
  OUTPUT  $(\hat{\tau}, i)$ 
STOP
END IF
END FOR
OUTPUT  $(0, 0)$ 

```

---

There are two steps in SL-estimate in the estimation of a change-point, when the largest penalized score exceeds the critical value  $c$ . The first is the identification of an interval  $(s(ij), u(ij))$ , associated with the largest penalized score, within which a change-point lies. The second is the estimation of the change-point within this interval. In the approximating set  $\mathcal{A}_i(g)$ , neighboring windows are located  $d_i$  apart, hence we are unable to estimate the change-points accurately in the first step. Accurate estimation is carried out, with more intensive computations within  $(s(ij), u(ij))$ , in the second step. Since the second step is performed only after an interval has been identified as containing a change-point, performing this two-step procedure saves computations in regions where scores are generally small and the likelihood of change-points is low.

After a change-point has been identified, we split the dataset into two and execute the same algorithm on each split dataset. To avoid repetitive computations, we start from segment length  $2h_{i_0}$  used in the evaluation of the change-point splitting the



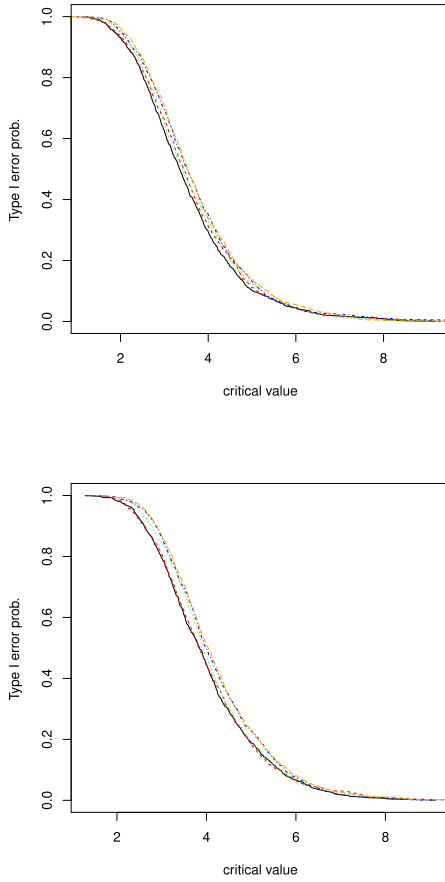


Fig. 2. Graphs of Type I error probability against critical value for the sparse likelihood detection algorithm, for independent unit variance normal observations. We consider parameters  $d_i$ ,  $h_i$ ,  $\lambda_1$  and  $\lambda_2$  as applied in the numerical studies in Section VI, with  $T = 2000$  (top),  $T = 20,000$  (bottom). and  $N = 50$  (black),  $N = 100$  (red),  $N = 200$  (green),  $N = 500$  (blue),  $N = 1000$  (orange).

dataset, instead of starting from the smallest segment length  $2h_1$ , on the split datasets. The use of a set of representative set of segment lengths for computational savings in change-point detection have been proposed in Willsky and Jones [39]. The recursive segmentation algorithm for the computation of the estimated change-point set  $\hat{\tau}$  is given below, with initialization at  $(c, 1, 1, T, \emptyset)$ .

---

**Algorithm 2** SL-Detect

---

```

INPUT( $c, i_0, b, e, \hat{\tau}$ )
  ( $\hat{\tau}, i$ )  $\leftarrow$  SL-estimate( $c, i_0, b, e$ )
  IF  $\hat{\tau} > 0$  THEN
     $\hat{\tau} \leftarrow \hat{\tau} \cup \{\hat{\tau}\}$ 
     $\hat{\tau} \leftarrow$  SL-detect( $c, i, b, \hat{\tau}, \hat{\tau}$ )
     $\hat{\tau} \leftarrow$  SL-detect( $c, i, \hat{\tau}, e, \hat{\tau}$ )
  END IF
OUTPUT  $\hat{\tau}$ 

```

---

In Figure 2 we show that the critical values of the sparse likelihood algorithm, for a specified Type I error probability, is stable over  $N$ . Contributing factors include  $\ell_N(p)$  having a mean that is close to zero and  $\ell_N(\mathbf{p})$  having exponential tail

probabilities not depending on  $N$ , see (6), when  $p$  and  $p^n$  are uniformly distributed.

#### IV. OPTIMAL DETECTION

Let  $\boldsymbol{\mu} = (\mu_t^n : 1 \leq t \leq T, 1 \leq n \leq N)$ ,  $\boldsymbol{\mu}_t = (\mu_t^n : 1 \leq n \leq N)$  and let  $J = (\#\tau)$  be the number of change-points. We show that the sparse likelihood algorithm is optimal for normal observations in Section IV-A and for Poisson observations in Section IV-B. Consider  $T$  growing exponentially with  $N$  in the sense that

$$\log T \sim N^\zeta \text{ for some } 0 < \zeta < 1. \quad (11)$$

The asymptotics in (11) are meant to highlight how the optimality constants are affected by the growth rate. We discuss the corresponding optimality theory for sub-exponential growth  $\log T = o(N^\zeta)$  for all  $\zeta > 0$ , in Section V-A.

In Theorems 2 and 4 we specify the detection boundary for asymptotically zero Type I and II error probabilities. Analogous detection boundaries for a single sequence is given in Arias-Castro et al. [1], [2].

In Theorems 3 and 5 we show that Type I and II error probabilities of the sparse likelihood algorithm go to zero at the detection boundary.

Recall from (9) that  $i_T = \max\{i : h_i + d_i \leq T\}$ . Consider the sparse likelihood algorithm with  $d_i$  and  $h_i$  satisfying

$$\frac{h_{i+1}}{h_i} \rightarrow 1 \text{ and } d_i = o(h_i) \text{ as } i \rightarrow \infty, \quad (12)$$

$$\log \left( \sum_{i=1}^{i_T} \frac{h_i}{d_i} \right) = o(\log T) \text{ as } T \rightarrow \infty, \quad (13)$$

and critical values  $c_T$  satisfying

$$c_T = o(\log T) \text{ and } c_T - \log \left( \sum_{i=1}^{i_T} \frac{h_i}{d_i} \right) \rightarrow \infty \text{ as } T \rightarrow \infty. \quad (14)$$

For the sparse likelihood algorithm select parameters  $\lambda_1 > 0$  and

$$\lambda_2 = \sqrt{\frac{\log T}{\log \log T}}. \quad (15)$$

We satisfy (12) when  $h_i \sim \exp(\frac{i}{\log i})$  and  $d_i \sim \frac{h_i}{i}$  as  $i \rightarrow \infty$ . Moreover (13) holds because

$$\log \left( \sum_{i=1}^{i_T} \frac{h_i}{d_i} \right) \sim 2 \log i_T \sim 2 \log \log T.$$

Condition (12) ensures that the set of  $(h_i, d_i)$  is sufficiently dense to detect change-points optimally. Condition (13) is required for (14) to hold. The first half of condition (14) ensures Type II error probability goes to 0. The second half ensures Type I error probability goes to 0.

##### A. Normal Model

Let

$$m_{j\Delta} = \#\{n : |\mu_{\tau_j+1}^n - \mu_{\tau_j}^n| \geq \Delta\}$$

be the number of sequences with change of mean of at least  $\Delta$  at the  $j$ th change-point. Let

$$\Omega_0 = \{\boldsymbol{\mu} : J = 0\},$$

$$\Omega_1(\Delta, V, h) = \{\mu : \text{there exists } j \text{ such that} \\ \min(\tau_j - \tau_{j-1}, \tau_{j+1} - \tau_j) \geq h \\ \text{and } m_{j\Delta} \geq V\},$$

with the convention  $\tau_0 = 0$  and  $\tau_{J+1} = T$ . We consider here the test of  $H_0: \mu \in \Omega_0$  versus  $H_1: \mu \in \Omega_1(\Delta, h, V)$ . Define

$$\rho_Z(\beta, \zeta) = \begin{cases} \beta - \frac{1-\zeta}{2} \\ \text{if } \frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}, \\ (\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2, \\ \text{if } \frac{3(1-\zeta)}{4} < \beta < 1-\zeta. \end{cases} \quad (16)$$

These constants are extensions of  $\rho_Z(\beta)$  in (2) to capture the effect of multiple testing in change-point detection.

**Theorem 2:** Assume (11) and let  $0 < \epsilon < 1$ . Let  $\Delta = CT^{-\eta}$  for constants  $C > 0$  and  $0 \leq \eta < \frac{1}{2}$ . For normal observations, no algorithm is able to achieve, as  $N \rightarrow \infty$ ,

$$\sup_{\mu \in \Omega_0} P_\mu(\text{Type I error}) + \sup_{\mu \in \Omega_1(\Delta, V, h)} P_\mu(\text{Type II error}) \rightarrow 0, \quad (17)$$

under either of the following conditions.

- (a) When  $V = o(\frac{\log T}{\log N})$  and  $h\Delta^2 = 4(1-2\eta)(1-\epsilon)(\frac{\log T}{V})$ .
- (b) When  $V \sim N^{1-\beta}$  for some  $\frac{1-\zeta}{2} < \beta < 1-\zeta$  and  $h\Delta^2 = 4(1-\epsilon)\rho_Z(\beta, \zeta) \log N$ .

**Theorem 3:** Assume (11) and let  $\epsilon > 0$ . Let  $\Delta = CT^{-\eta}$  for constants  $C > 0$  and  $0 \leq \eta < \frac{1}{2}$ . For normal observations the sparse likelihood algorithm, with parameters satisfying (12)–(15) achieves (17) under either of the following conditions.

- (a) When  $V = o(\frac{\log T}{\log N})$  and  $h\Delta^2 = 4(1-2\eta)(1+\epsilon)(\frac{\log T}{V})$ .
- (b) When  $V \sim N^{1-\beta}$  for some  $\frac{1-\zeta}{2} < \beta < 1-\zeta$  and  $h\Delta^2 = 4(1+\epsilon)\rho_Z(\beta, \zeta) \log N$ .

### B. Poisson Model

Rivera and Walther [33] provided the asymptotics for optimal change-point detection on a single Poisson sequence. We show here the optimality of the sparse likelihood detection algorithm for detecting sparse change-points in multi-stream data, with

$$X_t^n \sim_{\text{indep.}} \text{Poisson}(\mu_t^n). \quad (18)$$

Let  $Y_{st}^n = \sum_{v=s+1}^t X_v^n$ . Consider  $s < t < u$ . Under the null hypothesis of no change-points in the interval  $(s, u)$ , conditioned on  $Y_{su}^n = y_{su}^n$ ,  $Y_{st}^n$  is binomial distributed with  $y_{su}^n$  trials and success probability  $\frac{t-s}{u-s}$ . Let  $p_{stu}^n$  be the two-sided p-value of this conditional binomial test, with randomization of p-values so that they are distributed as Uniform(0,1) under the null hypothesis. More specifically when  $Y_{st}^n = y_{st}^n$  and  $Y_{su}^n = y_{su}^n$  simulate

$$\psi_{stu}^n \sim \text{Uniform}(P(Y < y_{st}^n), P(Y \leq y_{st}^n)), \quad (19)$$

where  $P$  is probability with respect to  $Y \sim \text{Binomial}(y_{su}^n, \frac{t-s}{u-s})$ , and define  $p_{stu}^n = 2 \min(\psi_{stu}^n, 1 - \psi_{stu}^n)$ .

Let

$$m_{j\Delta} = \#\{n : |\log(\mu_{\tau_j+1}^n / \mu_{\tau_j}^n)| \geq \Delta\},$$

and for a given  $\mu_0 > 0$ , let

$$\Lambda = \{\mu : \mu_n^t \geq \mu_0 \text{ for all } n \text{ and } t\},$$

$$\begin{aligned} \Lambda_0 &= \{\mu \in \Lambda : J = 0\}, \\ \Lambda_1(\Delta, V, h) &= \{\mu \in \Lambda : \text{there exists } j \text{ such that} \\ &\quad \min(\tau_{j+1} - \tau_j, \tau_j - \tau_{j-1}) \geq h \\ &\quad \text{and } m_{j\Delta} \geq V\}. \end{aligned}$$

We consider here the test of  $H_0: \mu \in \Lambda_0$  vs  $H_1: \mu \in \Lambda_1(\Delta, V, h)$ .

For a given  $r > 1$ , let

$$I_r = r \log(\frac{2r}{r+1}) + \log(\frac{2}{r+1}). \quad (20)$$

Let  $g_r(\omega) = (\frac{1+r\omega}{2})^{\frac{1}{\omega}}$  and let

$$\rho_r(\beta, \zeta) = \max_{\frac{1-\zeta}{\beta} < \omega \leq 2} \left( \frac{\beta - \omega^{-1}(1-\zeta)}{2g_r(\omega) - 1 - r} \right) \text{ for } \frac{1-\zeta}{2} < \beta < 1-\zeta. \quad (21)$$

In Theorem 4 we show that (21) is the asymptotic constant in the detection boundary of Poisson random variables. In Theorem 5 we show that the sparse likelihood algorithm achieves detection at this boundary for a broad range of sparsity.

**Theorem 4:** Assume (11). Let  $r = e^\Delta$  for some  $\Delta > 0$  and  $0 < \epsilon < 1$ . For Poisson observations no algorithm is able to achieve, as  $N \rightarrow \infty$ ,

$$\sup_{\mu \in \Lambda_0} P_\mu(\text{Type I error}) + \sup_{\mu \in \Lambda_1(\Delta, V, h)} P_\mu(\text{Type II error}) \rightarrow 0 \quad (22)$$

under either of the following conditions.

- (a) When  $V = o(\frac{\log T}{\log N})$  and  $h\mu_0 = (1-\epsilon)I_r^{-1}(\frac{\log T}{V})$ .
- (b) When  $V \sim N^{1-\beta}$  for some  $\frac{1-\zeta}{2} < \beta < 1-\zeta$  and  $h\mu_0 = (1-\epsilon)\rho_r(\beta, \zeta) \log N$ .

**Theorem 5:** Assume (11). Let  $\epsilon > 0$ ,  $\Delta > 0$  and  $1 < r < e^\Delta$ . For Poisson observations the sparse likelihood algorithm, with parameters satisfying (12)–(15), achieves (22) under either of the following conditions.

- (a) When  $V = o(\frac{\log T}{\log N})$  and  $h\mu_0 = (1+\epsilon)I_r^{-1}(\frac{\log T}{V})$ .
- (b) When  $V \sim N^{1-\beta}$  for some  $\frac{1-\zeta}{2} < \beta < 1-\zeta$  and  $h\mu_0 = (1+\epsilon)\rho_r(\beta, \zeta) \log N$ .

## V. DISCUSSIONS

### A. On the Exponential Growth of $T$ With Respect to $N$

The exponential growth of  $T$  with respect to  $N$  in (11) is chosen to highlight how the asymptotic constants  $\rho_Z(\beta, \zeta)$  and  $\rho_r(\beta, \zeta)$  varies with this growth rate. If instead

$$\log T = o(N^\zeta) \text{ for all } \zeta > 0, \quad (23)$$

then the asymptotics in Theorems 2(b)–5(b) apply with  $\rho_Z(\beta, \zeta)$  replaced by  $\rho_Z(\beta, 0)$  and  $\rho_r(\beta, \zeta)$  replaced by  $\rho_r(\beta, 0)$ , provided  $\frac{T\Delta^2}{\log N} \rightarrow \infty$ . Note that  $\rho_Z(\beta, 0) = \rho_Z(\beta)$ , the constant for sparse normal mixture detection. For example Jeng et al. [22] showed in their Theorem 2 that the HC test statistic achieves (17) with asymptotic constant  $\rho_Z(\beta)$  in the detection boundary of sparse change-points in multi-stream normal data, when  $T$  satisfies (23).

Under the very sparse setting  $V = o(\frac{\log T}{\log N})$ , we only require, under (23), that  $\frac{T\Delta^2 V}{\log T} \rightarrow \infty$  for the asymptotics of Theorems 2(a)–5(a) to hold.

### B. Our Theoretical Contributions for the Normal Model

The minimax detection boundary of Pilliat et al. [31] for the normal model, in their equation (8), is expressed in terms of  $\|\mu_{\tau_j+1} - \mu_{\tau_j}\|$  and  $\|\mu_{\tau_j+1} - \mu_{\tau_j}\|_0$ . Rephrased using the notations of this paper, their results imply that there exists  $c_0$  large enough such that if

$$h\Delta^2 \geq c_0 \left[ \log \left( 1 + V^{-1} \sqrt{N \log \left( \frac{T}{h\delta} \right)} \right) + V^{-1} \log \left( \frac{T}{h\delta} \right) \right],$$

then with probability at least  $1 - 6\delta$ , their dyadic grid algorithm is able to detect all change-points, each with location error not more than  $\frac{h}{2}$ , and with no spurious change-points. Moreover no algorithm is able to improve upon their algorithm beyond a smaller  $c_0$ .

Under the sparse setting

$$\log T \sim N^\zeta, V \sim N^{1-\beta} \text{ for } 0 < \zeta < 1 \text{ and } \frac{1-\zeta}{2} < \beta < 1 - \zeta, \quad (24)$$

and  $\Delta = CT^{-\eta}$  for some  $C > 0$  and  $0 \leq \eta < \frac{1}{2}$ , their grid algorithm has Type I and II error probabilities tending to 0 when

$$h\Delta^2 \geq c'_0 \log N$$

for  $c'_0$  large enough. Under the very sparse setting  $V = o(\frac{\log T}{\log N})$ , the error probabilities tend to 0 when  $h\Delta^2 \geq c'_0(\frac{\log T}{V})$  for  $c'_0$  large enough. Our contribution is in showing how the best constant  $c'_0$  depends on the sparsity of  $V$  and the exponential growth rate of  $T$ .

Liu et al. [25] tackled the problem of deciding between the null hypothesis of no change-point and the alternative hypothesis of a single change-point at an unknown location  $\tau$ . As in [31], their minimax detection boundary is expressed in terms of  $\|\mu_{\tau+1} - \mu_\tau\|$  and  $\|\mu_{\tau+1} - \mu_\tau\|_0$ . Rephrased using the notations of this paper, under (24), their results imply Type I and II error probabilities both bounded by  $\epsilon$  when

$$\frac{\tau(T-\tau)\Delta^2}{T} \geq C_\epsilon \times \begin{cases} \log N & \text{if } \beta > \frac{1}{2}, \\ \log \log N & \text{if } \beta = \frac{1}{2}, \\ V^{-1} \sqrt{N \log N} & \text{if } \beta < \frac{1}{2}, \end{cases}$$

for  $C_\epsilon > 0$  large enough. Moreover no algorithm is able to achieve this error probability bound beyond a smaller  $C_\epsilon$ . Theorems 2 and 3 indicate that when the restriction of a single change-point is relaxed, the  $\log N$  growth of the boundary for  $\frac{1}{2} < \beta < 1 - \zeta$  is not affected provided there are sufficient spacings between change-points.

### C. Our Contributions for the Poisson Model

Theorems 2–5 highlight the similarities and differences in the asymptotics for the normal and Poisson model for fixed  $\Delta > 0$  ( $\eta = 0$  in Theorems 2 and 3). For the very sparse setting  $V = o(\frac{\log T}{\log N})$ , a  $\frac{\log T}{V}$  growth rate for the signal strength  $h\Delta^2$  (for the normal model) and  $h\mu_0$  (for the Poisson model) is required for detection with asymptotically zero error probabilities. For the sparse scenario given in (24), a  $\log N$  growth rate is required. The asymptotic constants differ however, with  $\rho_Z(\beta, \zeta)$  for the normal model and  $\rho_r(\beta, \zeta)$  for the Poisson model. While the constant  $\rho_Z(\beta, \zeta)$  has appeared

in earlier works [7] and [8], the constant  $\rho_r(\beta, \zeta)$  is new in the literature.

Unlike in Theorems 2 and 3, the asymptotics in Theorems 4 and 5 do not involve  $\Delta \rightarrow 0$ . If  $r(= e^\Delta) \rightarrow 1$  as  $T \rightarrow \infty$ , the Gaussian approximation of the Poisson distribution kicks in and the asymptotics for the Poisson model correspond to that of the normal model. In particular as

$$\rho_r(\beta, \zeta) \sim 4\Delta^{-2} \rho_Z(\beta, \zeta) \text{ and } I_r \sim \frac{1}{4} \Delta^2 \text{ as } r \rightarrow 1,$$

the proofs of Theorems 2–5 indicate that if  $\Delta = CT^{-\eta}$  for some  $0 < \eta < \frac{1}{2}$ , then the boundary of asymptotically zero Type I and II error probabilities is at

$$h\Delta^2 \mu_0 = \begin{cases} 4(1 - 2\eta)(\frac{\log T}{V}) & \text{if } V = o(\frac{\log T}{\log N}), \\ 4\rho_Z(\beta, \zeta) \log N & \text{if } V \sim N^{1-\beta}. \end{cases}$$

## VI. SIMULATION STUDIES

### A. Change-Point Detection

We follow here the simulation set-up in Sections V-A and V-C of Wang and Samworth [37]. Assume that the random variables are normal with variances that are unknown but equal within sequence. These variances are estimated using median absolute differences of adjacent observations and after normalization, the random variables are treated like unit variance normal.

In the first study there is exactly one change-point  $\tau_1$ . Consider  $\mu_t^n = 0$  for  $t \leq \tau_1$  and all  $n$ . For  $t > \tau_1$ , let

$$\mu_t^n = \begin{cases} 0.8 / \sqrt{n \sum_{m=1}^V m^{-1}} & \text{if } n \leq V, \\ 0 & \text{if } n > V. \end{cases}$$

The objective is to estimate  $\tau_1$  assuming we know there is exactly one change-point. We estimate  $\tau_1$  here by

$$\hat{\tau}_1 = \arg \max_{0 < t < T} \ell_N^{\text{pen}}(\mathbf{p}_{0tT}),$$

where  $\ell_N^{\text{pen}}$  is the penalized sparse score with  $\lambda_1 = 1$  and  $\lambda_2 = \sqrt{\frac{\log T}{\log \log T}}$ .

We simulate the probabilities that  $|\hat{\tau}_1 - \tau_1| \leq k$  for  $k = 3$  and  $10$ , and compare against the INSPECT algorithm and the scan algorithm of Enikeeva and Harchaoui [15]. These two algorithms have the best numerical performances in Wang and Samworth [37]. The comparisons in Table I show that the sparse likelihood algorithm performs well.

In the second study there are three change-points within  $N = 200$  sequences of length  $T = 2000$ , at  $\tau_1 = 500$ ,  $\tau_2 = 1000$  and  $\tau_3 = 1500$ . At each change-point exactly 40 sequences undergo mean changes. Six scenarios are considered, corresponding to

$$\mu_{\tau_j+1}^{k(j-1)+n} - \mu_{\tau_j}^{k(j-1)+n} = r / \sqrt{n \sum_{m=1}^{40} m^{-1}},$$

$1 \leq j \leq 3$  and  $1 \leq n \leq 40$ , for  $r = 0.4, 0.6$  and  $k = 0, 20, 40$ . For  $k = 0$ , the mean changes are within the same 40 sequences at all three change-points, whereas for  $k = 40$  the mean changes at all three change-points are on distinct sequences. For  $k = 20$ , there is partial overlap of the sequences having mean changes at adjacent change-points.

TABLE I

THE FRACTION OF SIMULATION RUNS (OUT OF 1000) FOR WHICH  $\hat{\tau}_1$  IS WITHIN DISTANCE  $k$  FROM  $\tau_1$  FOR  $k = 3$  AND 10. THE SAME DATASETS ARE USED TO COMPARE SPARSE LIKELIHOOD (SL), INSPECT AND THE SCAN TEST, WITH  $\tau_1 = 200$  FOR  $T = 500$  AND  $\tau_1 = 800$  FOR  $T = 2000$

	$k$	$V$	3		10		3		10		3		10	
			SL		INSPECT		scan							
$T = 500$	3		0.511	0.801	0.478	0.785	0.520	0.804						
	$N = 500$	5	0.466	0.740	0.427	0.718	0.463	0.722						
		10	0.393	0.645	0.370	0.637	0.362	0.599						
		22	0.319	0.553	0.282	0.547	0.256	0.465						
		50	0.244	0.462	0.211	0.453	0.197	0.378						
		500	0.177	0.339	0.148	0.335	0.112	0.240						
$T = 500$	3		0.481	0.748	0.410	0.667	0.480	0.730						
	$N = 2000$	5	0.423	0.673	0.344	0.584	0.394	0.633						
		10	0.320	0.546	0.246	0.480	0.261	0.456						
		20	0.237	0.431	0.198	0.403	0.188	0.332						
		45	0.186	0.344	0.136	0.311	0.130	0.242						
		200	0.114	0.227	0.095	0.235	0.074	0.153						
$T = 2000$	3		0.603	0.859	0.587	0.855	0.589	0.854						
	$N = 500$	5	0.604	0.865	0.595	0.855	0.558	0.832						
		10	0.565	0.827	0.569	0.833	0.487	0.764						
		22	0.522	0.789	0.522	0.795	0.438	0.714						
		50	0.472	0.748	0.468	0.745	0.384	0.652						
		500	0.378	0.643	0.336	0.609	0.273	0.524						
$T = 2000$	3		0.607	0.866	0.608	0.861	0.599	0.858						
	$N = 2000$	5	0.594	0.864	0.586	0.857	0.557	0.829						
		10	0.553	0.847	0.558	0.847	0.476	0.780						
		20	0.494	0.807	0.498	0.789	0.435	0.726						
		45	0.447	0.747	0.451	0.746	0.377	0.657						
		200	0.362	0.649	0.342	0.604	0.297	0.554						
		2000	0.274	0.532	0.241	0.471	0.225	0.457						

TABLE II

NUMBER OF CHANGE-POINTS ESTIMATED BY THE SPARSE LIKELIHOOD ALGORITHM AND THE AVERAGE ARI OVER 100 SIMULATED DATASETS

$r$	$k$	# change-points				ARI
		2	3	4	5	
0.6	0	11	80	8	1	0.91
0.4	0	61	35	4	0	0.74
0.6	20	12	80	8	0	0.91
0.4	20	66	31	2	1	0.74
0.6	40	10	78	12	0	0.91
0.4	40	68	26	6	0	0.75

The number of estimated change-points over 100 simulated datasets on each sequence is recorded, as well as the adjusted Rand index (ARI), see Rand [32] and Hubert and Arabie [19], to measure the quality of the change-point estimation.

In the application of the sparse likelihood algorithm, we select  $h_1 = 1$  and  $h_{i+1} = \lceil 1.1 h_i \rceil$  for  $i \geq 1$ , and  $d_i = \lfloor h_i/i \rfloor$ , for a total of  $i_T = 61$  window lengths. We select critical value  $c_T = 5$  and parameters  $\lambda_1 = 1$ ,  $\lambda_2 = \sqrt{\frac{\log T}{\log \log T}} \doteq 1.94$ .

Wang and Samworth [37] showed that INSPECT achieves average ARI of 0.90 when  $r = 0.6$  and either 0.73 (for  $k = 20$ ) or 0.74 (for  $k = 0$  and 40) when  $r = 0.4$ , comparable to sparse likelihood, see Table II.

In addition to INSPECT, Wang and Samworth [37] considered DC, SBS and scan, as well as the CUSUM aggregation algorithms of Jirak [23] and Horváth and Hušková [18], with average ARI in the range 0.77–0.87 when  $r = 0.6$  and 0.68–0.72 when  $r = 0.4$ .

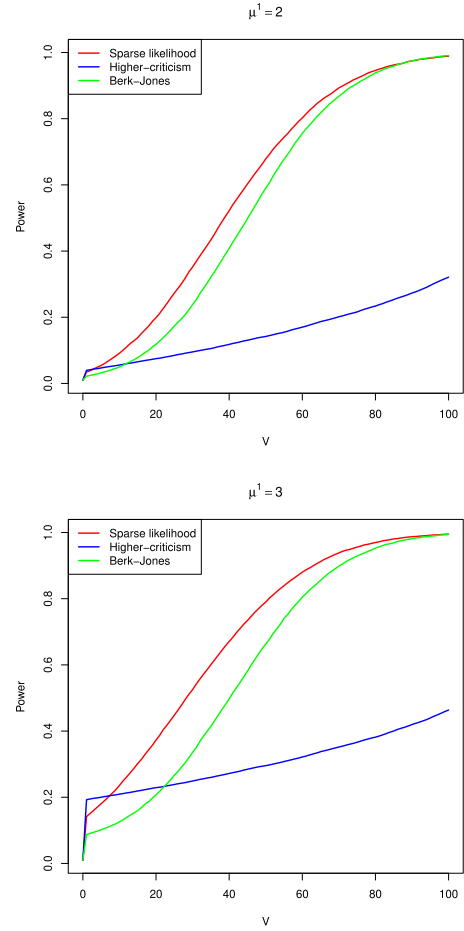


Fig. 3. Power of the likelihood score, HC and Berk-Jones test statistics for Gaussian mixtures, with means  $\mu^1 = 2$  (or 3),  $\mu^n = 1$  for  $2 \leq n \leq V$  and  $\mu^n = 0$  for  $V < n \leq 100$ .

### B. Normal Mixtures

We compare here the performance of the likelihood score, HC and Berk-Jones test statistics when testing a normal mixture  $Z^n \stackrel{\text{indep.}}{\sim} N(\mu^n, 1)$ , with  $\mu^n = 0$  for all  $n$  under the null hypothesis and

$$\mu^n = \begin{cases} 2 & \text{if } n = 1, \\ 1 & \text{if } 2 \leq n \leq V, \\ 0 & \text{if } V < n \leq N, \end{cases} \quad (25)$$

under the alternative hypothesis. Two-sided p-values  $p^n = 2\Phi(-|Z^n|)$  are applied. This exercise is repeated with  $\mu^1 = 3$  in place of  $\mu^1 = 2$ .

We see from (1) and (4) that the HC and Berk-Jones test statistics consider only the most significant deviation of  $p^{(n)}$  from  $\frac{n}{N}$ . They do not take into account significant deviations occurring over a broad range of  $n$ , which can happen when signals are of unequal strengths, as in (25). The likelihood score, which sums transformed scores of all p-values, may have an advantage here.

The simulations are performed with  $N = 100$  and  $V$  ranging from 1 to 100, with critical values chosen to satisfy Type I error probability  $\alpha = 0.01$ . For the likelihood score we consider (5) with parameters  $\lambda_1 = 0$  and  $\lambda_2 = 1$ . Figure 3 shows that whereas the HC test statistic does better



for smaller  $V$ , both the Berk-Jones test statistic and likelihood score have more power for larger  $V$ . The likelihood score is moderately better compared to the Berk-Jones test statistic over a broad range of  $V$ .

#### APPENDIX A PROOF OF THEOREM 1

Since  $c_N \rightarrow \infty$ , by Markov's inequality  $P_0(\ell_N(\mathbf{p}) \geq c_N) \leq e^{-c_N} \rightarrow 0$ . The proof of  $P_{\mu_N}(\ell_N(\mathbf{p}) < c_N) \rightarrow 0$  applies Lemmas 1 and 2 below. Lemma 1 says that the sum of sparse likelihood scores under  $q^n \sim \text{Uniform}(0,1)$  is bounded below by a value close to zero, with large probability. Lemma 2 provides a lower bound to the increase in score when the p-value is divided by at least 2. Their proofs are at the end of Appendix A.

**Lemma 1:** Let  $\mathbf{q} = (q^1, \dots, q^N)$ , with  $q^n \sim_{\text{i.i.d.}} \text{Uniform}(0,1)$ . For fixed  $\lambda_1 \geq 0$  and  $\delta > 0$ ,

$$\sup_{\delta \leq \lambda_2 \leq \sqrt{N}} P(\ell_N(\mathbf{q}) \leq -\lambda_2^2 \sqrt{\log N}) \rightarrow 0.$$

**Lemma 2:** For  $\lambda_1 > 0$  fixed,  $\delta \leq \lambda_2 \leq \sqrt{N}$  for some  $\delta > 0$  and  $\xi_N = o(N^{-\eta})$  for some  $\eta > 0$  such that  $\xi_N \geq \frac{\lambda_2^2}{2N}$ ,

$$\inf_{\substack{(p,q): p \leq \xi_N, \\ q \geq \lambda_2^2/N, p \leq q/2}} [\ell_N(p) - \ell_N(q)] \geq \frac{\lambda_2}{4\sqrt{N}\xi_N \log N}$$

for large  $N$ .

**Proof of Theorem 1:** Let  $\frac{1}{2} < \beta < 1$ ,  $\lambda_1 \geq 0$  and  $\lambda_2 > 0$  be fixed. Let  $\nu$  be such that

$$\begin{aligned} (1 - \sqrt{1 - \beta})^2 &< \nu < 1 & \text{if } \frac{3}{4} \leq \beta < 1, \\ \beta - \frac{1}{2} < \nu < 4(\beta - \frac{1}{2}) & & \text{if } \frac{1}{2} < \beta < \frac{3}{4}, \end{aligned}$$

and let

$$\begin{aligned} \mu_N &= \sqrt{2\nu \log N}, \\ Q^n &\sim \text{Bernoulli}(N^{-\beta}), \\ Z^n | Q^n &\sim N(\mu_N Q^n, 1), \\ p^n &= \Phi(-Z^n), \\ q^n &= \Phi(-Z^n + \mu_N Q^n). \end{aligned}$$

The additional assumptions of  $\nu < 1$  for  $\frac{3}{4} \leq \beta < 1$  and  $\nu < 4(\beta - \frac{1}{2})$  for  $\frac{1}{2} < \beta < \frac{3}{4}$  is not restrictive because  $\ell_N(\mathbf{p})$  increases stochastically with  $\mu_N$ .

**Case 1:**  $\frac{3}{4} \leq \beta < 1$ . Let

$$\Gamma = \{n : Q^n = 1, Z^n \geq \sqrt{2 \log N}, q^n \geq \frac{\lambda_2^2}{N}\}.$$

For  $N$  large,  $p^n \leq \frac{q^n}{2}$  for  $n \in \Gamma$ . Moreover  $\ell_N(p^n) \geq \ell_N(q^n)$  for all  $n$ . Hence by Lemma 1 and Lemma 2 with  $\xi_N = N^{-1}$ , with probability tending to 1,

$$\begin{aligned} \ell_N(\mathbf{p}) &\geq \ell_N(\mathbf{q}) + \sum_{n \in \Gamma} [\ell_N(p^n) - \ell_N(q^n)] \\ &\geq -\lambda_2^2 \sqrt{\log N} + (\#\Gamma) \frac{\lambda_2}{4\sqrt{\log N}}. \end{aligned}$$

Since  $\#\Gamma$  is binomial with mean

$$\begin{aligned} E_{\mu_N}(\#\Gamma) &= N^{1-\beta} [\Phi(-\sqrt{2 \log N} + \sqrt{2\nu \log N}) - \frac{\lambda_2^2}{N}] \\ &\gtrsim \frac{N^{1-\beta-(1-\sqrt{\nu})^2}}{\sqrt{\log N}}, \end{aligned} \quad (26)$$

with  $1 - \beta - (1 - \sqrt{\nu})^2 > 0$  for  $(1 - \sqrt{1 - \beta})^2 < \nu < 1$ , and since  $c_N$  is subpolynomial in  $N$ , we conclude  $P_{\mu_N}(\ell_N(\mathbf{p}) \geq c_N) \rightarrow 1$ .

**Case 2:**  $\frac{1}{2} < \beta < \frac{3}{4}$ . Let

$$\Gamma = \{n : Q^n = 1, Z^n \geq 2\sqrt{(2\beta - 1) \log N}, q^n \geq \frac{\lambda_2^2}{N}\}.$$

For  $N$  large,  $p^n \leq \frac{q^n}{2}$  for  $n \in \Gamma$ . Hence by Lemma 1 and Lemma 2 with  $\xi_N = N^{2-4\beta}$ , with probability tending to 1,

$$\begin{aligned} \ell_N(\mathbf{p}) &\geq \ell_N(\mathbf{q}) + \sum_{n \in \Gamma} [\ell_N(p^n) - \ell_N(q^n)] \\ &\geq -\lambda_2^2 \sqrt{\log N} + (\#\Gamma) \frac{\lambda_2}{4N^{\frac{3}{2}-2\beta} \sqrt{\log N}}. \end{aligned}$$

Since  $\#\Gamma$  is binomial with mean

$$\begin{aligned} E_{\mu_N}(\#\Gamma) &= N^{1-\beta} [\Phi(-2\sqrt{(2\beta - 1) \log N} + \sqrt{2\nu \log N}) - \frac{\lambda_2^2}{N}] \\ &\gtrsim \frac{N^{1-\beta-(\sqrt{4\beta-2}-\sqrt{\nu})^2}}{\sqrt{\log N}}, \end{aligned} \quad (27)$$

and

$$1 - \beta - (\sqrt{4\beta - 2} - \sqrt{\nu})^2 > \frac{3}{2} - \beta \text{ for } \beta - \frac{1}{2} < \nu < 4(\beta - \frac{1}{2}),$$

we conclude  $P_{\mu_N}(\ell_N(\mathbf{p}) \geq c_N) \rightarrow 1$ .  $\square$

**Proof of Lemma 1:** Let

$$x_N(p) = \frac{\lambda_1 \log N}{N} f_1(p) + \frac{\lambda_2}{\sqrt{N \log N}} f_2(p),$$

where  $f_1(p) = \frac{1}{p(2-\log p)^2} - \frac{1}{2}$ ,  $f_2(p) = \frac{1}{\sqrt{p}} - 2$ ,  $\lambda_1 \geq 0$  and  $\delta \leq \lambda_2 \leq N^{\frac{1}{2}}$  for some  $\delta > 0$ . Let  $r_N = \frac{1}{N \log N}$ . Since  $x_N(r_N) \geq 0$  and  $x_N(1) \geq -\frac{1}{2}$  for  $N$  large and  $\log(1+x) \geq x - x^2$  for  $x \geq -\frac{1}{2}$ ,

$$\ell_N(\mathbf{q}) = \sum_{n=1}^N \log(1 + x_N(q^n)) \geq \sum_{n=1}^N h_N(q^n) - \sum_{n=1}^N h_N^2(q^n), \quad (28)$$

where  $h_N(q) = x_N(q) \mathbf{1}_{\{q \geq r_N\}}$ .

By Chebyshev's inequality and the bounds in (29)–(31) below,

$$\begin{aligned} P(\ell_N(\mathbf{q}) \leq -\lambda_2^2 \sqrt{\log N}) &\leq P\left(\sum_{n=1}^N h_N(q^n) \leq -\frac{\lambda_2^2 \sqrt{\log N}}{2}\right) \\ &\leq P\left(\sum_{n=1}^N h_N^2(q^n) \geq \frac{\lambda_2^2 \sqrt{\log N}}{2}\right) \\ &\leq \frac{N \text{Var}(h_N(q^n))}{(N E h_N(q^n) + \frac{\lambda_2^2 \sqrt{\log N}}{2})^2} + \frac{N \text{Var}(h_N^2(q^n))}{(\frac{\lambda_2^2 \sqrt{\log N}}{2} - N E h_N^2(q^n))^2} \rightarrow 0. \end{aligned}$$

Since  $E x_N(q^n) = 0$ ,

$$\begin{aligned} E h_N(q^n) &= -E[x_N(q^n) \mathbf{1}_{\{q^n < r_N\}}] \\ &= -\frac{\lambda_1 \log N}{N} \left(\frac{1}{2-\log r_N} - \frac{r_N}{2}\right) \\ &\quad - \frac{\lambda_2}{\sqrt{N \log N}} (2\sqrt{r_N} - 2r_N) \\ &\geq -\frac{\lambda_1}{N} - \frac{2\lambda_2}{N \log N}. \end{aligned} \quad (29)$$

$$\text{Let } s_N = \frac{(\log N)^2}{N}.$$

$$\begin{aligned} \text{Var}(h_N(q^n)) &\leq E h_N^2(q^n) \end{aligned} \quad (30)$$

$$\begin{aligned}
&\leq \frac{2\lambda_1^2(\log N)^2}{N^2} \int_{r_N}^1 \frac{dq}{q^2(2-\log q)^4} + \frac{2\lambda_2^2}{N \log N} \int_{r_N}^1 \frac{dq}{q} \\
&\leq \frac{2\lambda_1^2(\log N)^2}{N^2} \left( \int_{s_N}^1 \frac{dq}{q^2} + \frac{1}{(2-\log s_N)^4} \int_{r_N}^{s_N} \frac{dq}{q^2} \right) \\
&\quad + \frac{2\lambda_2^2 \log(\frac{1}{r_N})}{N \log N} \\
&\lesssim \frac{\lambda_1^2 + \lambda_2^2}{N} \\
&\quad \text{Var}(h_N^2(q^n)) \\
&\leq E h_N^4(q^n) \\
&\leq \frac{8\lambda_1^4(\log N)^4}{N^4} \left( \int_{s_N}^1 \frac{dq}{q^4} + \frac{1}{(2-\log s_N)^8} \int_{r_N}^{s_N} \frac{dq}{q^4} \right) \\
&\quad + \frac{8\lambda_2^4}{(N \log N)^4} \int_{r_N}^1 \frac{dq}{q^2} \\
&\lesssim \frac{\lambda_1^4 + \lambda_2^4}{N}.
\end{aligned}$$

*Proof of Lemma 2:* For  $\frac{\lambda_2^2}{2N} \leq r \leq 2\xi_N$ ,  $|\log r| \asymp \log N$  and therefore

$$\frac{\frac{\lambda_1 \log N}{N} f_1(r)}{\frac{\lambda_2}{\sqrt{N \log N}} f_2(r)} \asymp \frac{1}{\lambda_2 \sqrt{N r \log N}} \rightarrow 0.$$

Moreover,

$$\frac{\lambda_2}{N \log N} f_2(r) \sim \frac{\lambda_2}{\sqrt{N r \log N}} \rightarrow 0.$$

Hence by  $\log(1+x) \sim x$  as  $x \rightarrow 0$ ,

$$\ell_N(r) \sim \frac{\lambda_2}{\sqrt{N r \log N}}. \quad (32)$$

*Case 1:*  $\frac{\lambda_2^2}{2N} \leq p \leq \xi_N$ . By (32) and  $q \geq 2p$ ,

$$\begin{aligned}
\ell_N(p) - \ell_N(q) &\geq \ell_N(p) - \ell_N(2p) \\
&\sim \left(1 - \frac{1}{\sqrt{2}}\right) \frac{\lambda_2}{\sqrt{N p \log N}} \\
&> \frac{\lambda_2}{4\sqrt{N \xi_N \log N}}.
\end{aligned}$$

*Case 2:*  $p < \frac{\lambda_2^2}{2N}$ . By (32),  $q \geq \frac{\lambda_2^2}{N}$  and  $\xi_N \geq \frac{\lambda_2^2}{2N}$ ,

$$\begin{aligned}
\ell_N(p) - \ell_N(q) &\geq \ell_N\left(\frac{\lambda_2^2}{2N}\right) - \ell_N\left(\frac{\lambda_2^2}{N}\right) \\
&\sim \left(1 - \frac{1}{\sqrt{2}}\right) \frac{\lambda_2}{\sqrt{N\left(\frac{\lambda_2^2}{2N}\right) \log N}} \\
&> \frac{\lambda_2}{4\sqrt{N \xi_N \log N}}.
\end{aligned}$$

## APPENDIX B PROOF OF THEOREM 2

*Proof of Theorem 2(a):* Consider first  $\eta = 0$ , that is  $\Delta > 0$  not varying with  $T$ . Let  $h = \lfloor \frac{4(1-\epsilon) \log T}{\Delta^2 V} \rfloor$  for some  $0 < \epsilon < 1$ . Let  $P_0$  denote probability with respect to  $\mu_t^n = 0$  for all  $n$  and  $t$ . Let  $t_k = (2k-1)h$  and let  $P_k$ ,  $1 \leq k \leq K := \lfloor \frac{T}{2h} \rfloor$ , denote probability under which, for  $n \leq V$ ,

$$\begin{aligned}
\mu_{t_k-h+1}^n &= \dots = \mu_{t_k}^n = -\frac{\Delta}{2}, \\
\mu_{t_k+1}^n &= \dots = \mu_{t_k+h}^n = \frac{\Delta}{2}, \\
\mu_t^n &= 0 \text{ for } t \leq t_k - h \text{ and } t > t_k + h,
\end{aligned} \quad (33)$$

and  $\mu_1^n = \dots = \mu_T^n = 0$  for  $n > V$ . Let  $E_k$  denote expectation with respect to  $P_k$ .

Let  $P_* = \frac{1}{K} \sum_{k=1}^K P_k$  and let  $L = \frac{1}{K} \sum_{k=1}^K L_k$ , where  $L_k = \frac{dP_k}{dP_0}(\mathbf{X})$  with  $\mathbf{X} = (X_t^n : 1 \leq n \leq N, 1 \leq t \leq T)$ . Hence

$$\log L_1 = \frac{h\Delta}{2} \sum_{n=1}^V (\bar{X}_{h,2h}^n - \bar{X}_{0h}^n) - \frac{hV\Delta^2}{4}. \quad (34)$$

Let  $A_i = \{L \leq 3\} \cap \{\text{conclude } H_i\}$ . Since  $P(A_1) = E_0(L1_{A_1}) \leq 3 P_0(A_1)$ ,

$$\begin{aligned}
&\sup_{\mu \in \Omega_0} P_\mu(\text{Type I error}) \\
&\quad + \sup_{\mu \in \Omega_1(\Delta, V, h)} P_\mu(\text{Type II error}) \\
&\geq P_0(\text{conclude } H_1) + P_*(\text{conclude } H_0) \\
&\geq P_0(A_1) + P_*(A_0) \geq \frac{1}{3} P_*(L \leq 3) = \frac{1}{3} P_1(L \leq 3),
\end{aligned} \quad (35)$$

with the last equality due to  $L$  having the same distribution under all  $P_k$  and  $P_*$ .

Since  $E_1 L_k = 1$  for  $k \geq 2$ , it follows that  $P_1(\frac{1}{K} \sum_{k=2}^K L_k \leq 2) \geq \frac{1}{2}$ . Hence by (35), to show that  $\sup_{\mu \in \Omega_0} P_\mu(\text{Type I error}) + \sup_{\mu \in \Omega_1(\Delta, V, h)} P_\mu(\text{Type II error}) \rightarrow 0$  is not possible, it suffices to show that

$$P_1(L_1 \leq K) \rightarrow 1 \text{ as } T \rightarrow \infty. \quad (36)$$

By (34),  $\log L_1 \sim N(\frac{hV\Delta^2}{4}, \frac{hV\Delta^2}{2})$ , and indeed

$$P_1(L_1 \leq K) = \Phi\left(\frac{\log K - \frac{1}{4}hV\Delta^2}{\sqrt{\frac{1}{2}hV\Delta^2}}\right) \rightarrow 1. \quad (37)$$

For  $\Delta = CT^{-\eta}$  with  $C > 0$  and  $0 < \eta < \frac{1}{2}$ , proceed as above with  $h = \lfloor \frac{4(1-2\eta)(1-\epsilon)}{V\Delta^2} \log T \rfloor$ . Here

$$\log K = \log(\lfloor \frac{T}{2h} \rfloor) \sim (1-2\eta) \log T,$$

and (37) still holds.  $\square$

*Proof of Theorem 2(b):* Consider first  $\eta = 0$ , that is  $\Delta > 0$  not varying with  $T$ . Proceed as in the proof of Theorem 2(a), but with  $h = \lfloor \frac{4(1-\epsilon)\rho_Z(\beta, \zeta) \log N}{\Delta^2} \rfloor$ , and  $P_k$  probability under which, independently for  $1 \leq n \leq N$ ,  $Q^n = 1$  with probability  $2N^{-\beta}$  and  $Q^n = 0$  otherwise. When  $Q^n = 1$ , (33) holds. When  $Q^n = 0$ ,  $\mu_1^n = \dots = \mu_T^n = 0$ .

By the law of large numbers,  $P_1(\mu \in \Omega_1(h, \Delta, V)) = P_1(\sum_{n=1}^N Q^n \geq V) \rightarrow 1$ . Hence by (35) it suffices to show (36) with

$$L_1 = \prod_{n=1}^N [1 + 2N^{-\beta} (e^{Z^n \Delta \sqrt{\frac{h}{2}} - \frac{h\Delta^2}{4}} - 1)], \quad (38)$$

$$Z^n = \sqrt{\frac{h}{2}} (\bar{X}_{h,2h}^n - \bar{X}_{0h}^n) \sim N\left(Q^n \Delta \sqrt{\frac{h}{2}}, 1\right). \quad (39)$$

*Case 1:*  $\frac{1-\zeta}{2} < \beta < \frac{3(1-\zeta)}{4}$ . Recall that  $\rho_Z(\beta, \zeta) = \beta - \frac{1-\zeta}{2}$ . By (38) and (39),

$$\begin{aligned}
E_1 L_1 &= (1 + 4N^{-2\beta} [\exp(\frac{h\Delta^2}{2}) - 1])^N \\
&\leq \exp(4N^{1-2\beta+2(1-\epsilon)\rho_Z(\beta, \zeta)}) \\
&= \exp(4N^{\zeta-2\epsilon\rho_Z(\beta, \zeta)}).
\end{aligned}$$

Since  $\log K = \log(\lfloor \frac{T}{2h} \rfloor) \sim N^\zeta$ , it follows that  $P_1(L_1 \leq K) \geq 1 - K^{-1} E_1 L_1 \rightarrow 1$  and (36) holds.

*Case 2:*  $\frac{3(1-\zeta)}{4} \leq \beta < 1 - \zeta$ . Recall that  $\rho_Z(\beta, \zeta) = (\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2$ . Express  $\log L_1 = \sum_{i=0}^3 R_i$ , where

$$\begin{aligned} R_i &= \sum_{n \in \Gamma_i} \log \left( 1 + 2N^{-\beta} \left[ \exp \left( Z^n \Delta \sqrt{\frac{h}{2}} - \frac{\Delta^2 h}{4} \right) - 1 \right] \right), \\ \Gamma_0 &= \{n : Q^n = 0\}, \\ \Gamma_1 &= \{n : Q^n = 1, Z^n \leq \sqrt{2(1-\zeta) \log N}\}, \\ \Gamma_2 &= \{n : Q^n = 1, \sqrt{2(1-\zeta) \log N} < Z^n \leq 2\sqrt{2 \log N}\}, \\ \Gamma_3 &= \{n : Q^n = 1, Z^n > 2\sqrt{2 \log N}\}. \end{aligned}$$

We show (36) by showing that

$$P_1(R_i \geq \frac{1}{4} \log K) \rightarrow 0 \text{ for } 0 \leq i \leq 3. \quad (40)$$

$$i = 3: \text{ Since } \Delta \sqrt{\frac{h}{2}} \leq \sqrt{2 \log N},$$

$$P_1(R_3 > 0) \leq 2N^{1-\beta} \Phi(-\sqrt{2 \log N}) \rightarrow 0.$$

$i = 2$ : Since

$$\begin{aligned} &\Delta \sqrt{\frac{h}{2}} \\ &\leq \sqrt{2(1-\zeta) \log N} - \sqrt{2(1-\zeta-\beta) \log N} - \sqrt{2\delta \log N} \end{aligned}$$

for some  $\delta > 0$ , it follows that

$$\Phi\left(\Delta \sqrt{\frac{h}{2}} - \sqrt{2(1-\zeta) \log N}\right) = o(N^{\zeta+\beta-1-\delta}).$$

Hence

$$\begin{aligned} &E_1 R_2 \\ &\leq E_1(\#\Gamma_2) \log(1 + 2N^{4-\beta}) \\ &\lesssim (N^{1-\beta} \log N) \Phi\left(\Delta \sqrt{\frac{h}{2}} - \sqrt{2(1-\zeta) \log N}\right) \\ &= o(N^{\zeta-\delta} \log N), \end{aligned} \quad (41)$$

and (40) follows from  $\log K \sim N^\zeta$ .

$i = 1$ : Since  $\log(1+x) \leq x$ ,

$$\begin{aligned} E_1 R_1 &\leq 4N^{1-2\beta} e^{-h\Delta^2/4} \\ &\times \int_{-\infty}^{\sqrt{2(1-\zeta) \log N}} \frac{1}{\sqrt{2\pi}} e^{-(z - \Delta \sqrt{\frac{h}{2}})^2/2 + z\Delta \sqrt{\frac{h}{2}}} dz \\ &= 4N^{1-2\beta} \Phi\left(\sqrt{2(1-\zeta) \log N} - 2\Delta \sqrt{\frac{h}{2}}\right) e^{h\Delta^2/2} \\ &\leq 4N^{1-2\beta - (\sqrt{1-\zeta} - 2\sqrt{(1-\epsilon)\rho_Z(\beta, \zeta)})^2 + 2(1-\epsilon)\rho_Z(\beta, \zeta)} \\ &= 4N^{\zeta-\delta} \text{ for some } \delta > 0. \end{aligned} \quad (42)$$

The last step above is shown below. Since

$$R_1 \geq (\#\Gamma_1) \log(1 - 2N^{-\beta}) \gtrsim -2N^{1-2\beta} = o(N^\zeta),$$

and  $\log K \sim N^\zeta$ , (40) follows from (42) and Markov's inequality.

$i = 0$ : Since  $E_1 e^{R_0} = 1$ ,

$$P_1(R_0 \geq \frac{1}{4} \log K) \leq K^{-\frac{1}{4}} \rightarrow 0.$$

For  $\Delta = CT^{-\eta}$  with  $C > 0$  and  $0 < \eta < \frac{1}{2}$ ,

$$\log K = \log(\lfloor \frac{T}{2h} \rfloor) \sim (1 - 2\eta)N^\zeta,$$

and the same arguments above can be applied to show (36).  $\square$

*Proof of (42):* It suffices to show that

$$1 - 2\beta - (\sqrt{1-\zeta} - 2\sqrt{(1-\epsilon)\rho_Z(\beta, \zeta)}) + 2(1-\epsilon)\rho_Z(\beta, \zeta) < \zeta. \quad (43)$$

Let  $m(\rho) = -(\sqrt{1-\zeta} - 2\sqrt{\rho})^2 + 2\rho$ . Inequality (43) follows from

$$\begin{aligned} &m(\rho_Z(\beta, \zeta)) \\ &= -(\sqrt{1-\zeta} - 2\sqrt{\rho_Z(\beta, \zeta)})^2 + 2\rho_Z(\beta, \zeta) \\ &= -(2\sqrt{1-\zeta-\beta} - \sqrt{1-\zeta})^2 \\ &\quad + 2(\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2 \\ &= 1 - \zeta - 2(1-\zeta-\beta) = \zeta - 1 + 2\beta, \end{aligned}$$

and

$$\begin{aligned} \frac{d}{d\rho} m(\rho) &= 2\rho^{-\frac{1}{2}}(\sqrt{1-\zeta} - 2\sqrt{\rho}) + 2 \\ &= 2\rho^{-\frac{1}{2}}\sqrt{1-\zeta} - 2 > 0 \text{ for } \rho < 1 - \zeta. \end{aligned}$$

$\square$

## APPENDIX C PROOF OF THEOREM 3

For  $(s, t, u) \in \mathcal{A}_i(T)$ , the penalty of the SL scores is

$$\log\left(\frac{T}{4}\left(\frac{1}{t-s} + \frac{1}{u-t}\right)\right) \geq \log\left(\frac{T}{2h_i}\right).$$

Moreover  $\#\mathcal{A}_i(T) \leq \frac{T}{d_i}$ . Hence by (6) and  $c_T - \log(\sum_{i=1}^{i_T} \frac{h_i}{d_i}) \rightarrow \infty$ , for  $\mu \in \Omega_0$ ,

$$\begin{aligned} &P_\mu(\text{Type I error}) \\ &\leq \sum_{i=1}^{i_T} \sum_{(s,t,u) \in \mathcal{A}_i(T)} P_\mu(\ell_N(\mathbf{p}_{stu}) \geq c_T + \log(\frac{T}{2h_i})) \\ &\leq \sum_{i=1}^{i_T} \frac{T}{d_i} \exp(-c_T - \log(\frac{T}{2h_i})) \\ &= 2e^{-c_T} \sum_{i=1}^{i_T} \frac{h_i}{d_i} \rightarrow 0. \end{aligned} \quad (44)$$

Consider  $\mu \in \Omega_1(\Delta, h, V)$  and let  $\tau_j$  be the change-point satisfying the conditions in the definition of  $\Omega_1(\Delta, h, V)$ . Let  $Q^n = 1$  if  $|\mu_{\tau_j+1}^n - \mu_{\tau_j}^n| \geq \Delta$  and  $Q^n = 0$  otherwise. We assume without loss of generality that  $0 < \epsilon < 1$ .

To aid in the checking of the proof of Theorem 3, we provide here the key ideas. Let  $j$  be such that

$$\min(\tau_j - \tau_{j-1}, \tau_{j+1} - \tau_j) \geq h \text{ and } m_j \Delta \geq V.$$

Consider  $\Delta > 0$  fixed and  $V \sim N^{1-\beta}$  for some  $\frac{1-\zeta}{2} < \beta < 1 - \zeta$ . Since  $h \rightarrow \infty$ , it follows from (12) that for  $N$  large we are able to find  $(s, t, u) = (s(ik), t(ik), u(ik))$  close to  $(\tau_j - h, \tau_j, \tau_j + h)$  such that

$$E_\mu Z_{stu}^n \geq [1 + o(1)] \frac{h\Delta^2}{2} \text{ for } n \text{ satisfying } |\mu_{\tau_j+1}^n - \mu_{\tau_j}^n| \geq \Delta. \quad (45)$$

Recall that  $p_{stu}^n = 2\Phi(-|Z_{stu}^n|)$  and let  $q_{stu}^n = \Phi(-|Z_{stu}^n| + E_\mu Z_{stu}^n) + \Phi(-|Z_{stu}^n| - E_\mu Z_{stu}^n)$ . Let

$$\begin{aligned} \Gamma &= \{n : |Z_{stu}^n| \geq \sqrt{2\omega \log N}, q_{stu}^n \geq N^{\zeta-1}, \\ &\quad |\mu_{\tau_j+1}^n - \mu_{\tau_j}^n| \geq \Delta\}, \end{aligned} \quad (46)$$

with  $\omega = 1 - \zeta$  when  $\frac{3(1-\zeta)}{4} < \beta < 1 - \zeta$  and  $\omega = 4(\beta - \frac{1-\zeta}{2})$  when  $\frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}$ . It follows from Lemmas 1 and 2 that with probability tending to 1,

$$\begin{aligned}\ell_N(\mathbf{p}_{stu}) &\geq \ell_N(\mathbf{q}_{stu}) + (\#\Gamma) \frac{\lambda_2}{4\sqrt{N\xi_N \log N}} \\ &\geq -\lambda_2^2 \sqrt{\log N} + (\#\Gamma) \frac{\lambda_2}{4\sqrt{N\xi_N \log N}}\end{aligned}$$

for  $\xi_N = N^{-\omega}$ .

Since the penalty  $\log(\frac{T}{4}(\frac{1}{t-s} + \frac{1}{u-t})) \leq \log T \sim N^\zeta$ ,  $c_T = o(\log T)$  and  $\lambda_2 \sim \frac{N^{\frac{\zeta}{2}}}{\sqrt{\zeta \log N}}$ , to show  $P_\mu(\ell_{stu}^{\text{pen}}(\mathbf{p}) \geq c_T) \rightarrow 1$ , it suffices to show that there exists  $\delta > 0$  such that

$$E_\mu(\#\Gamma) \gtrsim \begin{cases} N^{\zeta+\delta} & \text{if } \frac{3(1-\zeta)}{4} < \beta < 1 - \zeta, \\ N^{\frac{3}{2}-2\beta-\frac{\zeta}{2}+\delta} & \text{if } \frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}. \end{cases} \quad (47)$$

□

*Proof of Theorem 3(a):* Consider first  $\eta > 0$ , that is  $\Delta > 0$  not varying with  $T$ , and  $V = o(\frac{\log T}{\log N})$ . Since  $h = 4(1+\epsilon)(\frac{\log T}{\Delta^2 V}) \rightarrow \infty$ ,  $\frac{h_{i+1}}{h_i} \rightarrow 1$  and  $d_i = o(h_i)$ , for large  $T$  there exists

$$h_i \geq 4(1+\epsilon)^{\frac{1}{2}} \left( \frac{\log T}{\Delta^2 V} \right)$$

such that for all  $\mu \in \Omega_1(h, \Delta, V)$ , there exists  $k$  satisfying

$$\tau_{j-1} < s(ik) < u(ik) < \tau_{j+1} \text{ and } |t(ik) - \tau_j| \leq \frac{d_i}{2}. \quad (48)$$

Hence when  $Q^n = 1$ ,

$$|E_\mu Z_{stu}^n| \geq \Delta(1 - \frac{d_i}{2h_i}) \sqrt{\frac{h_i}{2}} \geq \sqrt{2(1+\epsilon)^{\frac{1}{3}} V^{-1} \log T}, \quad (49)$$

where  $(s, t, u) = (s(ik), t(ik), u(ik))$ .

Let  $\Gamma = \{n : Q^n = 1, |Z_{stu}^n| \geq \sqrt{2(1+\epsilon)^{\frac{1}{4}} (\frac{\log T}{V})}\}$ . Let  $p_{stu}^n = 2\Phi(-|Z_{stu}^n|)$  and  $q_{stu}^n = \Phi(-|Z_{stu}^n| + E_\mu Z_{stu}^n) + \Phi(-|Z_{stu}^n| - E_\mu Z_{stu}^n)$ . Since  $q^n \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0,1)$  and

$$P_\mu(\ell_N(q_{stu}^n) \geq 2 \log N) \leq N^{-2} E_\mu[\exp(\ell_N(q_{stu}^n))] = N^{-2}, \quad (50)$$

by Lemma 1, with probability tending to 1,

$$\begin{aligned}\ell_N(\mathbf{p}_{stu}) &\geq \ell_N(\mathbf{q}_{stu}) \\ &\quad + (\#\Gamma) \left[ \ell_N \left( 2\Phi \left( -\sqrt{2(1+\epsilon)^{\frac{1}{4}} \frac{\log T}{V}} \right) \right) - 2 \log N \right] \\ &\geq -\lambda_2^2 \sqrt{\log N} + V[(1+\epsilon)^{\frac{1}{5}} \frac{\log T}{V} - 2 \log N] \\ &\geq (1+\epsilon)^{\frac{1}{6}} \log T.\end{aligned} \quad (51)$$

Since the penalty  $\log(\frac{T}{4}(\frac{1}{t-s} + \frac{1}{u-t})) \leq \log T$  and  $c_T = o(\log T)$ , it follows that  $P_\mu(\ell_N^{\text{pen}}(\mathbf{p}_{stu}) \geq c_T) \rightarrow 1$ .

Consider next  $\Delta = CT^{-\eta}$  for  $0 < \eta < \frac{1}{2}$  and  $V = o(\frac{\log T}{\log N})$ . Let  $h_i \geq (1-2\eta)(1+\epsilon)^{\frac{1}{2}} (\frac{\log T}{\Delta^2 V})$  be such that for all  $\mu \in \Omega_1(h, \Delta, V)$ , (48) holds for some  $k$ . Let

$$\Gamma = \{n : Q^n = 1, |Z_{stu}^n| \geq \sqrt{2(1-2\eta)(1+\epsilon)^{\frac{1}{4}} \frac{\log T}{V}}\}$$

and define  $p_{stu}^n$  and  $q_{stu}^n$  as above.

By the arguments in (51), with probability tending to 1,

$$\ell_N(\mathbf{p}_{stu}) \geq (1-2\eta)(1+\epsilon)^{\frac{1}{6}} \log T.$$

Since  $\Delta = CT^{-\eta}$ , it follows that  $h_i \gtrsim T^{2\eta} \log N$  and the penalty  $\log(\frac{T}{4}(\frac{1}{u-t} + \frac{1}{t-s})) \leq (1-2\eta) \log T$  for  $T$  large. Hence by  $c_T = o(\log T)$  we conclude  $P_\mu(\ell_N^{\text{pen}}(\mathbf{p}_{stu}) \geq c_T) \rightarrow 1$ . □

*Proof of Theorem 3(b): Case 1:*  $V \sim N^{1-\beta}$  for  $\frac{3(1-\zeta)}{4} \leq \beta < 1 - \zeta$ . Since  $h\Delta^2 = 4(1+\epsilon)(\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2 \log N$  and  $d_i = o(h_i)$ , for large  $N$  there exists  $i$  satisfying  $h_i \geq (1+\epsilon)^{-\frac{1}{2}} h$  such that whenever  $Q^n = 1$ ,

$$\begin{aligned}|E_\mu Z^n| &\geq \Delta(1 - \frac{d_i}{2h_i}) \sqrt{\frac{h_i}{2}} \geq \sqrt{2\nu \log N}, \quad (52) \\ \nu &= (1+\epsilon)^{\frac{1}{3}} (\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2,\end{aligned}$$

with  $(s, t, u) = (s(ik), t(ik), u(ik))$  for  $k$  satisfying (48).

For  $\Gamma$  defined in (46),

$$\begin{aligned}E_\mu(\#\Gamma) &\geq V[\Phi(-\sqrt{2(1-\zeta) \log N} + \sqrt{2\nu \log N}) \\ &\quad - N^{\zeta-1}] \\ &\gtrsim N^{1-\beta-(\sqrt{1-\zeta}-\sqrt{\nu})^2} (\log N)^{-\frac{1}{2}},\end{aligned}$$

and (47) follows from

$$\sqrt{1-\zeta} > \sqrt{\nu} > \sqrt{1-\zeta} - \sqrt{1-\zeta-\beta}.$$

*Case 2:*  $V \sim N^{1-\beta}$  for  $\frac{1-\zeta}{2} < \beta < \frac{3(1-\zeta)}{4}$ . Since  $h\Delta^2 = 4(1+\epsilon)(\beta - \frac{1-\zeta}{2}) \log N$ , for large  $N$  there exists  $h_i \geq (1+\epsilon)^{-\frac{1}{2}} h$  such that whenever  $Q^n = 1$ ,

$$\begin{aligned}|E_\mu Z_{stu}^n| &\geq \Delta(1 - \frac{d_i}{2h_i}) \sqrt{\frac{h_i}{2}} \geq \sqrt{2\nu \log N}, \quad (53) \\ \nu &= (1+\epsilon)^{\frac{1}{3}} (\beta - \frac{1-\zeta}{2}),\end{aligned}$$

with  $(s, t, u) = (s(ik), t(ik), u(ik))$  for  $k$  satisfying (48).

For  $\Gamma$  defined in (46),

$$\begin{aligned}E_\mu(\#\Gamma) &\geq V[\Phi(-2\sqrt{(2\beta-1+\zeta) \log N} + \sqrt{2\nu \log N}) \\ &\quad - N^{\zeta-1}] \\ &\gtrsim N^{1-\beta-(2\sqrt{\beta-\frac{1-\zeta}{2}}-\sqrt{\nu})^2} (\log N)^{-\frac{1}{2}},\end{aligned}$$

and (47) follows from

$$2\sqrt{\beta - \frac{1-\zeta}{2}} > \sqrt{\nu} > \sqrt{\beta - \frac{1-\zeta}{2}}.$$

□

## APPENDIX D PROOF OF THEOREM 4

*Proof of Theorem 4(a):* Let  $h = \lfloor \frac{(1-\epsilon) \log T}{\mu_0 V I_r} \rfloor$  for some  $0 < \epsilon < 1$ . Let  $P_0$  denote probability with respect to  $\mu_t^n = (\frac{1+r}{2})\mu_0$  for all  $n$  and  $t$ . Let  $t_k = (2k-1)h$ . Let  $P_k$ ,  $1 \leq k \leq K := \lfloor \frac{T}{2h} \rfloor$ , denote probability under which for  $n \leq V$ ,

$$\mu_t^n = \begin{cases} \mu_0 & \text{for } t_k - h < t \leq t_k, \\ r\mu_0 & \text{for } t_k < t \leq t_k + h, \\ (\frac{1+r}{2})\mu_0 & \text{for } t \leq t_k - h \text{ and } t > t_k + h, \end{cases} \quad (54)$$

and  $\mu_1^n = \dots = \mu_T^n = (\frac{1+r}{2})\mu_0$  for  $n > V$ . Let  $E_k$  and  $\text{Var}_k$  denote expectation and variance respectively with respect to  $P_k$ . Let

$$U^n = S_{0h}^n \log(\frac{2}{1+r}) + S_{h,2h}^n \log(\frac{2r}{r+1}), \quad (55)$$

$$L_1 = \frac{dP_1}{dP_0}(\mathbf{X}) = \prod_{n=1}^V \exp(U^n). \quad (56)$$



By (35)–(36), it suffices to show that

$$P_1(L_1 \leq K) \rightarrow 1 \text{ as } T \rightarrow \infty. \quad (57)$$

Since  $E_1(\log L_1) = h\mu_0 V I_r$  and  $\text{Var}_1(\log L_1) = h\mu_0 V C_r$ , where  $C_r = r[\log(\frac{2r}{r+1})]^2 + [\log(\frac{2}{r+1})]^2$ , by Chebyshev's inequality,

$$P_1(L_1 \leq K) \geq 1 - \frac{hV\mu_0 C_r}{(\log K - hV\mu_0 I_r)^2} \rightarrow 1,$$

and (57) holds.  $\square$

We preface the proof of Theorem 4(b) with Lemma 3, which provides an alternative representation of  $\rho_r(\beta, \zeta)$ . Let

$$\begin{aligned} D(\omega) &= \frac{1}{1+r\omega} \log\left(\frac{2}{1+r\omega}\right) + \frac{r\omega}{1+r\omega} \log\left(\frac{2r\omega}{1+r\omega}\right), \\ g(\omega) &= \left(\frac{1+r\omega}{2}\right)^{\frac{1}{\omega}}. \end{aligned} \quad (58)$$

Let  $\xi(\omega) = \frac{\beta - \omega^{-1}(1-\zeta)}{2g(\omega) - 1 - r}$ . Recall from (21) that

$$\rho_r(\beta, \zeta) = \max_{\frac{1-\zeta}{\beta} < \omega \leq 2} \xi(\omega) \text{ for } \frac{1-\zeta}{2} < \beta < 1 - \zeta. \quad (59)$$

**Lemma 3:** For  $\frac{1}{2} < \frac{\beta}{1-\zeta} \leq \frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}]$ ,  $\xi$  achieves its maximum at  $\omega = 2$  and

$$\rho_r(\beta, \zeta) = \frac{\beta - \frac{1}{2}(1-\zeta)}{2g(2) - 1 - r}. \quad (60)$$

For  $\frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}] < \frac{\beta}{1-\zeta} < 1$ ,  $\xi$  achieves its maximum at some  $\omega < 2$  and

$$\rho_r(\beta, \zeta) = \frac{1-\zeta}{2g(\omega)D(\omega)}. \quad (61)$$

*Proof:* Since

$$\begin{aligned} \frac{d}{d\omega} \log \xi(\omega) &= \frac{\omega^{-2}(1-\zeta)}{\beta - \omega^{-1}(1-\zeta)} - \frac{2\frac{d}{d\omega}g(\omega)}{2g(\omega) - 1 - r}, \\ \frac{d}{d\omega} g(\omega) &= \frac{d}{d\omega} \exp\left[\frac{1}{\omega} \log\left(\frac{1+r\omega}{2}\right)\right] \\ &= \left[\frac{r\omega \log r}{\omega(1+r\omega)} - \frac{1}{\omega^2} \log\left(\frac{1+r\omega}{2}\right)\right]g(\omega) \\ &= \frac{D(\omega)g(\omega)}{\omega^2}, \end{aligned}$$

it follows that  $\frac{d}{d\omega} \log \xi(\omega) = 0$  when

$$\omega^{-2}(1-\zeta)[2g(\omega) - 1 - r] = 2[\beta - \omega^{-1}(1-\zeta)] \frac{D(\omega)g(\omega)}{\omega^2}, \quad (62)$$

that is when

$$\frac{\beta}{1-\zeta} = \omega^{-1} + \frac{2g(\omega) - 1 - r}{2g(\omega)D(\omega)}. \quad (63)$$

For  $\frac{1}{2} < \frac{\beta}{1-\zeta} \leq \frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}]$ , the solution of  $\omega$  to (63) is at least 2 and the maximum in (59) is attained at  $\omega = 2$ . We conclude (60). For  $\frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}] < \frac{\beta}{1-\zeta} < 1$ , the solution of  $\omega$  to (63) lies in the interval  $(\frac{1-\zeta}{\beta}, 2)$ . We conclude (61) from (59) and a rearrangement of (62).  $\square$

*Proof of Theorem 4(b):* For  $\frac{1-\zeta}{2} < \beta < 1 - \zeta$ , let  $\omega$  be the maximizer in

$$\rho_r(\beta, \zeta) = \max_{\frac{1-\zeta}{\beta} < \omega \leq 2} \left( \frac{\beta - \omega^{-1}(1-\zeta)}{2g(\omega) - 1 - r} \right). \quad (64)$$

Let  $h = \lfloor \frac{(1-\epsilon)\rho_r(\beta, \zeta) \log N}{\mu_0} \rfloor$  for some  $\epsilon > 0$ . Let  $P_0$  denote probability with respect to  $\mu_t^n = g(\omega)\mu_0$  for all  $n$  and  $t$ . Let  $t_k = (2k-1)h$ . Let  $P_k$ ,  $1 \leq k \leq K := \lfloor \frac{T}{2h} \rfloor$ , denote probability under which, independently for  $1 \leq n \leq N$ ,

$Q^n = 1$  with probability  $2N^{-\beta}$ , and  $Q^n = 0$  otherwise. When  $Q^n = 1$ ,

$$\mu_t^n = \begin{cases} \mu_0 & \text{for } t_k - h < t \leq t_k, \\ r\mu_0 & \text{for } t_k < t \leq t_k + h, \\ g(\omega)\mu_0 & \text{for } t \leq t_k - h \text{ and } t > t_k + h. \end{cases} \quad (65)$$

When  $Q^n = 0$ ,  $\mu_1^n = \dots = \mu_T^n = g(\omega)\mu_0$ . Let  $E_1$  denote expectation with respect to  $P_1$ . Let  $P_Q = P_1(\cdot | Q^1 = 1)$  and let  $E_Q$  denote expectation with respect to  $P_Q$ .

By (35)–(36), it suffices to show (57) for

$$\begin{aligned} L_1 &= \frac{dP_1}{dP_0}(\mathbf{X}) = \prod_{n=1}^N (1 + 2N^{-\beta}[\exp(U^n) - 1]), \\ U^n &= S_{0h}^n \log\left(\frac{1}{g(\omega)}\right) + S_{h,2h}^n \log\left(\frac{r}{g(\omega)}\right) \\ &\quad - h\mu_0[1 + r - 2g(\omega)]. \end{aligned} \quad (66)$$

For notational simplicity, let  $S_{0h} = S_{0h}^1$  and  $S_{h,2h} = S_{h,2h}^1$ .

For  $X \sim \text{Poisson}(\lambda)$  and constant  $C > 0$ ,

$$E(C^X) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{(C\lambda)^x}{x!} = e^{\lambda(C-1)}. \quad (67)$$

This identity is applied in (68), (71) and (72).

*Case 1:*  $\frac{1}{2} < \frac{\beta}{1-\zeta} \leq \frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}]$ ,  $\omega = 2$ . By Lemma 2, (65)–(67) and  $[g(2)]^2 = \frac{1+r^2}{2}$ ,

$$\begin{aligned} E_Q \exp(U^1) &= E_Q \left[ \left( \frac{1}{g(2)} \right)^{S_{0h}} \left( \frac{r}{g(2)} \right)^{S_{h,2h}} e^{-h\mu_0[1+r-2g(2)]} \right] \\ &= \exp(h\mu_0 \left[ \frac{1}{g(2)} - 1 + \frac{r^2}{g(2)} - r \right] - h\mu_0[1+r-2g(2)]) \\ &= \exp(2h\mu_0[2g(2) - 1 - r]) \\ &= \exp\left(\frac{h\mu_0(2\beta-1+\zeta)}{\rho_r(\beta, \zeta)}\right) \leq N^{(1-\epsilon)(2\beta-1+\zeta)}. \end{aligned} \quad (68)$$

Hence

$$\begin{aligned} E_1 L_1 &= (1 + 4N^{-2\beta} [E_Q \exp(U^1) - 1])^N \\ &\leq \exp(4N^{\zeta-\delta}) = o(K), \end{aligned}$$

where  $\delta = \epsilon(2\beta - 1 + \zeta)$ , and (57) holds.

*Case 2:*  $\frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}] < \frac{\beta}{1-\zeta} < 1$ . Express

$$\log L_1 = R_0 + R_1, \text{ where} \quad (69)$$

$$R_i = \sum_{n \in \Gamma_i} \log(1 + 2N^{-\beta}[\exp(U^n) - 1]),$$

$$\Gamma_0 = \{n : Q^n = 0\} \cup \{n : Q^n = 1, \exp(U^n) \leq N^\beta\},$$

$$\Gamma_1 = \{n : Q^n = 1, \exp(U^n) > N^\beta\}.$$

We conclude (57) from

$$P_1(R_i \leq \frac{1}{2} \log K) \rightarrow 1 \text{ for } i = 0 \text{ and } 1. \quad (70)$$

*i = 0:* Let  $a = \omega - 1$  with  $\omega$  the maximizer in (64). Since  $g(\omega) = \frac{1+r^{a+1}}{2g^a(\omega)}$ , by (64), (66) and (67),

$$\begin{aligned} &E_Q[\exp(U^1) \mathbf{1}_{\{1 \in \Gamma_0\}}] \\ &\leq N^{\beta(1-a)} E_Q \exp(aU^1) \\ &= N^{\beta(1-a)} \exp(h\mu_0 \left[ \frac{1}{g^a(\omega)} - 1 + \frac{r^{a+1}}{g^a(\omega)} - r \right]) \end{aligned} \quad (71)$$

$$\begin{aligned}
& -ah\mu_0[1+r-2g(\omega)]) \\
& = N^{\beta(1-a)} \exp(\omega h\mu_0[2g(\omega)-1-r]) \\
& = N^{\beta(1-a)} \exp\left(\frac{h\mu_0(\beta\omega-1+\zeta)}{\rho_r(\beta,\zeta)}\right) \leq N^{2\beta-1+\zeta-\delta},
\end{aligned}$$

where  $\delta = \epsilon(\beta\omega - 1 + \zeta)$ . Since  $E_0 \exp(U^n) = 1$ , it follows from (71) that

$$\begin{aligned}
E_1 \exp(R_0) & \leq (1 + 4N^{-2\beta} E_Q[\exp(U^1) \mathbf{1}_{\{1 \in \Gamma_0\}}])^N \\
& \leq \exp(4N^{\zeta-\delta}),
\end{aligned}$$

and (70) holds.

$i = 1$ : Express  $U^1 = v_1 S_{0h} + v_2 S_{h,2h} - z$ , where  $v_1 = \log(\frac{1}{g(\omega)})$ ,  $v_2 = \log(\frac{r}{g(\omega)})$  and  $z = h\mu_0[1+r-2g(\omega)]$ . Since  $g(\omega) = \frac{1+r^{a+1}}{2g^a(\omega)}$ , by Markov's inequality and (67),

$$\begin{aligned}
& E_1(\#\Gamma_1) \\
& = 2N^{1-\beta} P_Q(e^{aU^1} > N^{a\beta}) \\
& \leq 2N^{1-\beta-a\beta} e^{-az} E_Q(e^{v_1 a S_{0h}} e^{v_2 a S_{h,2h}}) \\
& = 2N^{1-\omega\beta} \exp(-az + h\mu_0[e^{v_1 a} - 1 + r e^{v_2 a} - r]) \\
& = 2N^{1-\omega\beta} \exp(\omega h\mu_0[2g(\omega)-1-r]) \\
& = 2N^{1-\omega\beta} \exp\left(\frac{h\mu_0(\beta\omega-1+\zeta)}{\rho_r(\beta,\zeta)}\right) \leq N^{\zeta-\delta},
\end{aligned} \tag{72}$$

where  $\delta = \epsilon(\beta\omega - 1 + \zeta)$ . Since

$$R_1 \leq (\#\Gamma_1) \max_{n \in \Gamma_1} U_1^n \text{ and } P_1(\max_n U^n \geq N^{\frac{\delta}{2}}) \rightarrow 0,$$

we conclude (70) from (72) and Markov's inequality.  $\square$

#### APPENDIX E PROOF OF THEOREM 5

It follows from (44) that  $\sup_{\mu \in \Lambda_0} P_\mu(\text{Type I error}) \rightarrow 0$ .

Consider  $\mu \in \Lambda_1(h, \Delta, V)$  and let  $\tau_j$  be a change-point such that

$$\min(\tau_{j+1} - \tau_j, \tau_j - \tau_{j-1}) \geq h \text{ and } m_{j\Delta} \geq V,$$

where  $m_{j\Delta} = \#\{n : |\log(\mu_{\tau_j+1}^n / \mu_{\tau_j}^n)| \geq \Delta\}$ .

Let  $Q^n = 1$  if  $|\log(\mu_{\tau_j+1}^n / \mu_{\tau_j}^n)| \geq \Delta$  and  $Q^n = 0$  otherwise.

*Proof of Theorem 5(a):* Consider  $V = o(\frac{\log T}{\log N})$  and recall from (20) that  $I_r = r \log(\frac{2r}{r+1}) + \log(\frac{2}{r+1})$ . Let  $r_1$  and  $\mu_1$  be such that  $e^\Delta > r_1 > r$  and  $\mu_0/(1+\epsilon)^{\frac{1}{2}} < \mu_1 < \mu_0$ . Since  $hV I_r \mu_0 = (1+\epsilon) \log T$ ,  $\frac{h_{i+1}}{h_i} \rightarrow 1$  and  $d_i = o(h_i)$ , for  $T$  large there exists

$$h_i \geq (1+\epsilon)^{\frac{1}{2}} I_r^{-1} \left(\frac{\log T}{\mu_1 V}\right), \tag{73}$$

such that for all  $\mu \in \Lambda_1(h, \Delta, V)$ , there exists  $k$  such that

$$\tau_{j-1} < s(ik) < u(ik) < \tau_{j+1}, \quad |t(ik) - \tau_j| \leq \frac{d_i}{2}.$$

Moreover when  $Q^n = 1$ ,

$$|\log(E_\mu Y_{tu}^n / E_\mu Y_{st}^n)| \geq \log r_1, \tag{74}$$

where  $(s, t, u) = (s(ik), t(ik), u(ik))$ . Let

$$\Gamma = \{n : Q^n = 1, Y_{su}^n \geq (1+r)h_i \mu_1, |\log(Y_{tu}^n / Y_{st}^n)| \geq \log r\}.$$

By (73), for  $n \in \Gamma$ ,

$$p_{stu}^n \leq 2 \exp(-\mu_1 h_i I_r) \leq 2 \exp(-(1+\epsilon)^{\frac{1}{2}} \frac{\log T}{V}). \tag{75}$$

Since  $\frac{\log T}{V \log N} \rightarrow \infty$ , for  $N$  large,

$$\ell(p_{stu}^n) \geq (1+\epsilon)^{\frac{1}{3}} \left(\frac{\log T}{V}\right).$$

Hence as  $P_\mu(\ell_N(q_{stu}^n) \geq 2 \log N) \leq N^{-2}$ , see (50), by Lemma 1, with probability tending to 1,

$$\begin{aligned}
\ell_N(\mathbf{p}_{stu}) & \geq \ell_N(\mathbf{q}_{stu}) + (\#\Gamma)[(1+\epsilon)^{\frac{1}{3}} \left(\frac{\log T}{V}\right) - 2 \log N] \\
& \geq -\lambda_2^2 \sqrt{\log N} + (1+\epsilon)^{\frac{1}{4}} \log T.
\end{aligned}$$

Since the penalty  $\log(\frac{T}{4}(\frac{1}{t-s} + \frac{1}{u-t})) \leq \log T$ ,  $\lambda_2^2 \sqrt{\log N} = o(\log T)$  and  $c_T = o(\log T)$ , we can conclude that  $P_\mu(\ell_N^{\text{pen}}(\mathbf{p}_{stu}) \geq c_T) \rightarrow 1$ .  $\square$

*Proof of Theorem 5(b):* Consider  $V \sim N^{1-\beta}$  for  $\frac{1-\zeta}{2} < \beta < 1 - \zeta$ . For  $N$  large, there exists

$$\log N \gtrsim h_i \geq (1+\epsilon)^{\frac{1}{2}} \rho_r(\beta, \zeta) \left(\frac{\log N}{\mu_0}\right) \tag{76}$$

such that for all  $\mu \in \Lambda_1(h, \Delta, V)$ , there exists  $k$  such that

$$\tau_{j-1} < s(ik) < u(ik) < \tau_{j+1}, \quad |t(ik) - \tau_j| \leq \frac{d_i}{2},$$

and conditioned on  $Q^n = 1$ , either

$$E_\mu Y_{tu}^n \geq r E_\mu Y_{st}^n \text{ or } E_\mu Y_{st}^n \geq r E_\mu Y_{tu}^n, \tag{77}$$

where  $(s, t, u) = (s(ik), t(ik), u(ik))$ .

By Stirling's approximation  $x! \sim \sqrt{2\pi x} (\frac{x}{e})^x$ , for  $X \sim \text{Poisson}(\eta)$ , as  $x \rightarrow \infty$ ,

$$P(X = x) = e^{-\eta} \frac{\eta^x}{x!} \sim \frac{1}{\sqrt{2\pi x}} \exp[-\eta + x - x \log(\frac{x}{\eta})]. \tag{78}$$

By apply this in (80) and (85).

*Case 1:*  $\frac{1}{2} < \frac{\beta}{1-\zeta} \leq \frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}]$  and  $\rho_r(\beta, \zeta) = \frac{\beta - \frac{1}{2}(1-\zeta)}{2g(2)-1-r}$ . Let

$$\begin{aligned}
\Gamma & = \{n : Q^n = 1, Y_{su}^n \geq \sqrt{2(1+r^2)} h_i \mu_0 - 1, \\
& \quad |\log(Y_{tu}^n / Y_{st}^n)| \geq 2 \log r, q_{stu}^n \geq N^{\zeta-1}\}.
\end{aligned} \tag{79}$$

Consider  $Y_1 \sim \text{Poisson}(h_i \mu_0)$  and  $Y_2 \sim \text{Poisson}(r h_i \mu_0)$ . By (78) and  $h_i \lesssim \log N$ ,

$$\begin{aligned}
& P(Y_1 = \lfloor (\frac{2}{1+r^2})^{\frac{1}{2}} h_i \mu_0 \rfloor) \\
& \gtrsim \frac{1}{\sqrt{\log N}} \exp(h_i \mu_0 [-1 + (\frac{2}{1+r^2})^{\frac{1}{2}} \\
& \quad - (\frac{2}{1+r^2})^{\frac{1}{2}} \log((\frac{2}{1+r^2})^{\frac{1}{2}})]), \\
& P(Y_2 = \lceil (\frac{2}{1+r^2})^{\frac{1}{2}} r^2 h_i \mu_0 \rceil) \\
& \gtrsim \frac{1}{\sqrt{\log N}} \exp(h_i \mu_0 [-r + r^2 (\frac{2}{1+r^2})^{\frac{1}{2}} \\
& \quad - r^2 (\frac{2}{1+r^2})^{\frac{1}{2}} \log(r (\frac{2}{1+r^2})^{\frac{1}{2}})]).
\end{aligned} \tag{80}$$

Recall that  $g(2) = (\frac{1+r^2}{2})^{\frac{1}{2}}$  and  $D(2) = \frac{1}{1+r^2} \log(\frac{2}{1+r^2}) + \frac{r^2}{1+r^2} \log(\frac{2r^2}{1+r^2})$  [see (58)]. By (80),

$$\begin{aligned}
& E_\mu(\#\Gamma) \\
& \geq V[P(Y_1 = \lfloor (\frac{2}{1+r^2})^{\frac{1}{2}} h_i \mu_0 \rfloor) P(Y_2 = \lceil (\frac{2}{1+r^2})^{\frac{1}{2}} r^2 h_i \mu_0 \rceil) \\
& \quad - N^{\zeta-1}] \\
& \gtrsim \frac{N^{1-\beta}}{\log N} \exp(h_i \mu_0 [2g(2) - 1 - r - g(2)D(2)]).
\end{aligned} \tag{81}$$

By (79), for  $n \in \Gamma$ ,

$$p_{stu}^n \leq 2 \exp(-Y_{su}^n D(2)) \leq \xi_N \tag{82}$$

where  $\xi_N = C_2 \exp(-2h_i \mu_0 g(2)D(2))$  for  $C_2 = 2e^{D(2)}$ .

Let  $q_{stu}^n = F_n(p_{stu}^n)$  where  $F_n$  is the distribution function of  $p_{stu}^n$ . It follows from Lemmas 1 and 2 that with probability tending to 1,

$$\begin{aligned} \ell_N(\mathbf{p}_{stu}) &= \ell_N(\mathbf{q}_{stu}) + (\#\Gamma) \frac{\lambda_2}{4\sqrt{N}\xi_N \log N} \\ &\geq -\lambda_2^2 \sqrt{\log N} + \frac{\lambda_2 N^{\frac{1}{2}-\beta}}{(\log N)^{\frac{3}{2}}} \exp(h_i \mu_0 [2g(2) - 1 - r]). \end{aligned} \quad (83)$$

Since  $\lambda_2 \asymp \frac{N^{\frac{\zeta}{2}}}{\sqrt{\log N}}$  and by (76)

$$h_i \mu_0 \geq (1 + \epsilon)^{\frac{1}{2}} \rho_r(\beta, \zeta) \log N = (1 + \epsilon)^{\frac{1}{3}} \left( \frac{\beta - \frac{1}{2}(1-\zeta)}{2g(2) - 1 - r} \right) \log N,$$

it follows from (83) that  $\ell_N(\mathbf{p}_{stu}) \gtrsim \frac{N^{\zeta+\delta}}{(\log N)^2}$  for  $\delta = [(1 + \epsilon)^{\frac{1}{2}} - 1][\beta - \frac{1}{2}(1 - \zeta)]$ . Since the penalty  $\log(\frac{T}{4}(\frac{1}{t-s} + \frac{1}{u-t})) \leq \log T \sim N^\zeta$  and  $c_T = o(N^\zeta)$ , we conclude  $P_\mu(\ell_N^{\text{pen}}(\mathbf{p}_{stu}) \geq c_T) \rightarrow 1$ .

Case 2:  $\frac{1}{2}[1 + \frac{2g(2)-1-r}{g(2)D(2)}] < \frac{\beta}{1-\zeta} < 1$  and  $\rho_r(\beta, \zeta) = \frac{1-\zeta}{2g(\omega)D(\omega)} = \frac{\beta - \frac{1}{2}(1-\zeta)}{2g(\omega) - 1 - r}$  with  $\omega$  achieving the maximum in (59). Let

$$\Gamma = \{n : Q^n = 1, Y_{su}^n \geq 2g(\omega)h_i \mu_0 - 1, |\log(Y_{tu}^n/Y_{st}^n)| \geq \omega \log r, q_{stu}^n \geq N^{\zeta-1}\}. \quad (84)$$

Consider  $Y_1 \sim \text{Poisson}(h_i \mu_0)$  and  $Y_2 \sim \text{Poisson}(r h_i \mu_0)$ .

By (78) and  $h_i \lesssim \log N$ ,

$$\begin{aligned} P(Y_1 = \lfloor \frac{2g(\omega)}{r^\omega+1} h_i \mu_0 \rfloor) &\gtrsim \frac{1}{\sqrt{\log N}} \exp(h_i \mu_0 [-1 + \frac{2g(\omega)}{r^\omega+1} \\ &\quad - \frac{2g(\omega)}{r^\omega+1} \log(\frac{2g(\omega)}{r^\omega+1})]), \\ P(Y_2 = \lceil \frac{2r^\omega g(\omega)}{r^\omega+1} h_i \mu_0 \rceil) &\gtrsim \frac{1}{\sqrt{\log N}} \exp(h_i \mu_0 [-r + \frac{2r^\omega g(\omega)}{r^\omega+1} \\ &\quad - \frac{2r^\omega g(\omega)}{r^\omega+1} \log(\frac{2r^\omega g(\omega)}{r^\omega+1})]). \end{aligned} \quad (85)$$

Recall that  $g(\omega) = (\frac{1+r^\omega}{2})^{\frac{1}{\omega}}$  and  $D(\omega) = \frac{1}{1+r^\omega} \log(\frac{2}{1+r^\omega}) + \frac{r^\omega}{1+r^\omega} \log(\frac{2r^\omega}{1+r^\omega})$  [see (58)]. By (85),

$$\begin{aligned} E_\mu(\#\Gamma) &\geq V[P(Y_1 = \lfloor \frac{2g(\omega)}{r^\omega+1} h_i \mu_0 \rfloor)P(Y_2 \geq \lceil \frac{2r^\omega g(\omega)}{r^\omega+1} h_i \mu_0 \rceil) \\ &\quad - N^{\zeta-1}] \\ &\gtrsim \frac{N^{1-\beta}}{\log N} \exp(h_i \mu_0 [2g(\omega) - 1 - r - 2(\frac{\omega-1}{\omega})g(\omega)D(\omega)]). \end{aligned} \quad (86)$$

By (84), for  $n \in \Gamma$ ,

$$p^n \leq 2 \exp(-Y_{s(ik),u(ik)}^n D(\omega)) \leq \xi_N \quad (87)$$

where  $\xi_N = C_\omega \exp(-2h_i \mu_0 g(\omega) D(\omega))$  for  $C_\omega = 2e^{D(\omega)}$ .

Let  $q^n = F_n(p^n)$  where  $F_n$  is the distribution function of  $p^n$ . It follows from Lemmas 1 and 2 that with probability tending to 1,

$$\begin{aligned} \ell_N(\mathbf{p}_{stu}) &\geq \ell_N(\mathbf{q}_{stu}) + (\#\Gamma) \frac{\lambda_2}{4\sqrt{N}\xi_N \log N} \\ &\gtrsim -\lambda_2^2 \sqrt{\log N} + \frac{\lambda_2 N^{\frac{1}{2}-\beta}}{(\log N)^{\frac{3}{2}}} \\ &\quad \times \exp(h_i \mu_0 [2g(\omega) - 1 - r - (\frac{\omega-2}{\omega})g(\omega)D(\omega)]). \end{aligned} \quad (88)$$

Since  $\lambda_2 \asymp \frac{N^{\frac{\zeta}{2}}}{\sqrt{\log N}}$  and by (76),

$$\begin{aligned} h_i \mu_0 &\geq (1 + \epsilon)^{\frac{1}{2}} \rho_r(\beta, \zeta) \log N \\ &= (1 + \epsilon)^{\frac{1}{2}} \left( \frac{1-\zeta}{2g(\omega)D(\omega)} \right) \log N \\ &= (1 + \epsilon)^{\frac{1}{2}} \left( \frac{\beta - \frac{1}{2}(1-\zeta)}{2g(\omega) - 1 - r} \right) \log N, \end{aligned}$$

it follows from (88) that  $\ell_N(\mathbf{p}_{stu}) \gtrsim \frac{N^{\zeta+\delta}}{(\log N)^2}$  for  $\delta = [(1 + \epsilon)^{\frac{1}{2}} - 1][\beta - \frac{1}{2}(1 - \zeta)]$ . Since the penalty  $\log(\frac{T}{4}(\frac{1}{t-s} + \frac{1}{u-t})) \leq \log T \sim N^\zeta$  and  $c_T = o(\log T)$ , we conclude  $P_\mu(\ell_N^{\text{pen}}(\mathbf{p}_{stu}) \geq c_T) \rightarrow 1$ .  $\square$

## ACKNOWLEDGMENT

The authors would like to thank the associate editor and two referees for their constructive comments and suggestions.

## REFERENCES

- [1] E. Arias-Castro, D. L. Donoho, and X. Huo, "Near-optimal detection of geometric objects by fast multiscale methods," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2402–2425, Jul. 2005.
- [2] E. Arias-Castro, D. L. Donoho, and X. Huo, "Adaptive multiscale detection of filamentary structures in a background of uniform random points," *Ann. Statist.*, vol. 34, no. 1, pp. 326–349, Feb. 2006.
- [3] E. Arias-Castro and M. Wang, "The sparse Poisson means model," *Electron. J. Statist.*, vol. 9, no. 2, pp. 2170–2201, Jan. 2015.
- [4] R. H. Berk and D. H. Jones, "Goodness-of-fit test statistics that dominate the Kolmogorov statistics," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 47, no. 1, pp. 47–59, 1979.
- [5] T. Tony Cai, X. Jessie Jeng, and J. Jin, "Optimal detection of heterogeneous and heteroscedastic mixtures," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 73, no. 5, pp. 629–662, Nov. 2011.
- [6] T. T. Cai and Y. Wu, "Optimal detection of sparse mixtures against a given null distribution," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2217–2232, Apr. 2014.
- [7] H. P. Chan and G. Walther, "Optimal detection of multi-sample aligned sparse signals," *Ann. Statist.*, vol. 43, no. 5, pp. 1865–1895, Oct. 2015.
- [8] H. P. Chan, "Optimal sequential detection in multi-stream data," *Ann. Statist.*, vol. 45, no. 6, pp. 2736–2763, Dec. 2017.
- [9] H. Cho, "Change-point detection in panel data via double CUSUM statistic," *Electron. J. Statist.*, vol. 10, no. 2, pp. 2000–2038, Jan. 2016.
- [10] H. Cho and P. Fryzlewicz, "Multiple-change-point detection for high dimensional time series via sparsified binary segmentation," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 77, no. 2, pp. 475–507, Mar. 2015.
- [11] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Ann. Statist.*, vol. 32, no. 3, pp. 962–994, Jun. 2004.
- [12] D. L. Donoho and A. Kipnis, "Higher criticism to compare two large frequency tables, with sensitivity to possible rare and weak differences," *Ann. Statist.*, vol. 50, no. 3, pp. 1447–1472, Jun. 2022.
- [13] C. Du, C.-L.-M. Kao, and S. C. Kou, "Stepwise signal extraction via marginal likelihood," *J. Amer. Stat. Assoc.*, vol. 111, no. 513, pp. 314–330, Jan. 2016.
- [14] L. Dümbgen and V. G. Spokoiny, "Multiscale testing of qualitative hypotheses," *Ann. Statist.*, vol. 29, no. 1, pp. 124–152, Feb. 2001.
- [15] F. Enikeeva and Z. Harchaoui, "High-dimensional change-point detection under sparse alternatives," *Ann. Statist.*, vol. 47, no. 4, pp. 2051–2079, Aug. 2019.
- [16] K. Frick, A. Munk, and H. Sieling, "Multiscale change point inference," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 76, no. 3, pp. 495–580, Jun. 2014.
- [17] P. Fryzlewicz, "Wild binary segmentation for multiple change-point detection," *Ann. Statist.*, vol. 42, no. 6, pp. 2243–2281, Dec. 2014.
- [18] L. Horváth and M. Hušková, "Change-point detection in panel data," *J. Time Anal.*, vol. 33, no. 4, pp. 631–648, Jul. 2012.
- [19] L. Hubert and P. Arabie, "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, Dec. 1985.
- [20] Y. I. Ingster, "On some problems of hypothesis testing leading to infinitely divisible distributions," *Math. Methods Statist.*, vol. 6, pp. 47–69, Jan. 1997.
- [21] Y. I. Ingster, "Minimax detection of a signal for  $\ell^n$  balls," *Math. Methods Statist.*, vol. 7, no. 4, pp. 401–428, 1998.

- [22] X. J. Jeng, T. T. Cai, and H. Li, "Simultaneous discovery of rare and common segment variants," *Biometrika*, vol. 100, no. 1, pp. 157–172, Mar. 2013.
- [23] M. Jirak, "Uniform change point tests in high dimension," *Ann. Statist.*, vol. 43, pp. 2451–2483, Dec. 2015.
- [24] T. L. Lai and H. Xing, "A simple Bayesian approach to multiple change-points," *Statistica Sinica*, vol. 21, no. 2, pp. 539–569, Apr. 2011.
- [25] H. Liu, C. Gao, and R. J. Samworth, "Minimax rates in sparse, high-dimensional change point detection," *Ann. Statist.*, vol. 49, no. 2, pp. 1081–1112, Apr. 2021.
- [26] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, Jun. 2010.
- [27] A. Moscovich, B. Nadler, and C. Spiegelman, "On the exact Berk–Jones statistics and their  $p$ -value calculation," *Electron. J. Statist.*, vol. 10, no. 2, pp. 2329–2354, Jan. 2016.
- [28] Y. S. Niu, N. Hao, and H. Zhang, "Multiple change-point detection: A selective overview," *Stat. Sci.*, vol. 31, no. 4, pp. 611–623, Nov. 2016.
- [29] Y. S. Niu and H. Zhang, "The screening and ranking algorithm to detect DNA copy number variations," *Ann. Appl. Statist.*, vol. 6, no. 3, pp. 1306–1326, Sep. 2012.
- [30] A. B. Olshen, E. S. Venkatraman, R. Lucito, and M. Wigler, "Circular binary segmentation for the analysis of array-based DNA copy number data," *Biostatistics*, vol. 5, no. 4, pp. 557–572, Oct. 2004.
- [31] E. Pilliat, A. Carpentier, and N. Verzelen, "Optimal multiple change-point detection for high-dimensional data," 2020, *arXiv:2011.07818*.
- [32] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.
- [33] C. Rivera and G. Walther, "Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics," *Scandin. J. Statist.*, vol. 40, no. 4, pp. 752–769, Dec. 2013.
- [34] I. V. Stoepker, R. M. Castro, E. Arias-Castro, and E. Van Den Heuvel, "Anomaly detection for a large number of streams: A permutation-based higher criticism approach," *J. Amer. Stat. Assoc.*, pp. 1–25, Nov. 2022. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/01621459.2022.2126361>, doi: [10.1080/01621459.2022.2126361](https://doi.org/10.1080/01621459.2022.2126361).
- [35] J. W. Tukey, "T13 N: The higher criticism," Course Notes. (1976). Princeton, NJ, USA: Princeton Univ.
- [36] G. Walther, "Optimal and fast detection of spatial clusters with scan statistics," *Ann. Statist.*, vol. 38, no. 2, pp. 1010–1033, Apr. 2010.
- [37] T. Wang and R. J. Samworth, "High dimensional change point estimation via sparse projection," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 80, no. 1, pp. 57–83, Jan. 2018.
- [38] Y. Wang and Y. Mei, "Large-scale multi-stream quickest change detection via shrinkage post-change estimation," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6926–6938, Dec. 2015.
- [39] A. Willsky and H. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Trans. Autom. Control*, vol. AC-21, no. 1, pp. 108–112, Feb. 1976.
- [40] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *Ann. Statist.*, vol. 41, no. 2, pp. 670–692, Apr. 2013.
- [41] Y.-C. Yao, "Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches," *Ann. Statist.*, vol. 12, no. 4, pp. 1434–1447, Dec. 1984.
- [42] N. R. Zhang and D. Siegmund, "A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization," *Biometrics*, vol. 63, pp. 22–52, Mar. 2007.
- [43] N. R. Zhang, D. O. Siegmund, H. Ji, and J. Z. Li, "Detecting simultaneous change-points in multiple sequences," *Biometrika*, vol. 97, no. 3, pp. 631–645, Sep. 2010.

**Shouri Hu** received the B.Sc. and Ph.D. degrees in statistics from the National University of Singapore in 2016 and 2020, respectively. Since 2022, he has been a Faculty Member with the School of Mathematical Sciences, University of Electronic Science and Technology of China, where he is currently a Lecturer. His research interests include multi-armed bandit, change-point detection, and Bayesian optimization.

**Jingyan Huang** received the B.Sc. degree in statistics and operation research from Hong Kong Baptist University in 2018. She is currently pursuing the Ph.D. degree in statistics and applied probability with the National University of Singapore. Her research interests include change-point detection, sequential testing, and change-point detection in time-series, with the application in single-cell DNA sequencing data and S&P 500 data.

**Hao Chen** received the Ph.D. degree in statistics from Stanford University, Stanford, CA, USA, in 2013. Since then, she has been a Faculty Member with the Department of Statistics, University of California at Davis, Davis, CA, USA, where she is currently an Associate Professor (with tenure). Her research interests include developing practical and robust methods that can deal with various data types, including high-dimensional data, image data, and network data, for hypothesis testing, signal detection, classification, and clustering.

**Hock Peng Chan** received the Ph.D. degree in statistics from Stanford University in 1998. In 1998, he joined the Department of Statistics and Applied Probability, National University of Singapore, as an Assistant Professor. He is currently a Professor with the Department of Statistics and Data Science, National University of Singapore. His research interests include sequential analysis, change-point detection, and Monte Carlo methods.