OPTIMAL DETECTION OF MULTI-SAMPLE ALIGNED SPARSE SIGNALS

BY HOCK PENG CHAN[1] AND GUENTHER WALTHER[2]

National University of Singapore and Stanford University

We describe, in the detection of multi-sample aligned sparse signals, the critical boundary separating detectable from nondetectable signals, and construct tests that achieve optimal detectability: penalized versions of the Berk–Jones and the higher-criticism test statistics evaluated over pooled scans, and an average likelihood ratio over the critical boundary. We show in our results an inter-play between the scale of the sequence length to signal length ratio, and the sparseness of the signals. In particular the difficulty of the detection problem is not noticeably affected unless this ratio grows exponentially with the number of sequences. We also recover the multiscale and sparse mixture testing problems as illustrative special cases.

1. Introduction. Consider a population of sequences having a common time (or location) index. Signals, when they occur, are present in a small fraction of the sequences and aligned in time. In the detection of copy number variants (CNV) in multiple DNA sequences, Efron and Zhang [11] used local f.d.r., Zhang et al. [29] and Siegmund, Yakir and Zhang [24] applied scans of weighted χ^2-statistics, Jeng, Cai and Li [17] applied higher-criticism test statistics. Tartakovsky and Veeravalli [25], Mei [20] and Xie and Siegmund [28] considered the analogous sequential detection of sparse aligned changes of distribution in parallel streams of data, with applications in communications, disease surveillance, engineering and hospital management. These advances have brought in an added multi-sample dimension to traditional scan statistics works (see, e.g., the papers in [12]) that consider a single stream of data.

In this paper, we tackle the problem of detectability of aligned sparse signals, extending sparse mixture detection (cf. [4, 6, 9, 13–15, 27]) to aligned signals, and extending multiscale detection (cf. [8, 10, 19, 23]) to multiple sequences. Hence not surprisingly, we incorporate ideas developed by the sparse mixture and multiscale detection communities to find the critical boundary separating detectable from nondetectable hypotheses. In Arias-Castro, Donoho and Huo [1, 2], there

are also links between sparse mixtures and multiscale detection methods in the
detection of a sparse component on an unknown low-dimensional curve within
a higher-dimensional space. Our work here is less geometrical in nature as the
aligned-signal assumption allows us to reduce the problem to one dimension by
summarizing across sample first.

We supply optimal adaptive max-type tests: penalized scans of the higher criti-
cism and Berk–Jones test statistics. We also supply an optimal Bayesian test: an
average likelihood ratio (ALR) that tests against alternatives lying on the critical
boundary. The rationale behind the ALR is to focus testing at the most sensitive
parameter values, where small perturbations can result in sharp differences of de-
tection powers.

We state the main results in Section 2. We describe the detectable region of
aligned sparse signals in the multi-sample setting, and show that the penalized
scans achieve asymptotic detection power 1 there. We learn from the detection
boundary the surprising result that the requirement to locate the signal in the time
domain does not affect the overall difficulty of the detection problem, unless the
sequence length to signal length ratio grows exponentially with the number of
sequences.

In Section 3, we show the optimality of the ALR and consider special cases of
our model that have been well studied in the literature using max-type tests: the
detection of a signal with unknown location and scale in a single sequence, and
the detection of a sparse mixture in many sequences of length 1. We show that
the general form of our ALR provides optimal detection in these important special
cases. We also illustrate the detectability and detection of multi-sample signals on
a CNV dataset.

In Section 4, the detection problem is extended to heteroscedastic signals. The
extension illustrates the adaptivity of the penalized scans. Even though the detec-
tion boundary has to be extended to take into account the heteroscedasticity, the
penalized scans as described in Section 2 are still optimal. On the other hand, the
ALR tests have to be re-designed to ensure optimality under heteroscedasticity.
The model set-up here is similar to that in Jeng, Cai and Li [17]. There optimal-
ity is possible without imposing penalties on the scan of the higher-criticism test
because the signal length was assumed to be very short.

2. Main results. Let $\{(X_{n1}, \ldots, X_{nT}) : 1 \leq n \leq N\}$ be a population of se-
quences. We consider the prototypical set-up

$$(2.1) \qquad X_{nt} = \mu_{nt} + Z_{nt} \qquad \text{where } Z_{nt} \text{ are i.i.d. N}(0, 1).$$

Under the null hypothesis H_0 of no signals, $\mu_{nt} = 0$ for all n and t. Under the
alternative hypothesis H_1 of aligned signals, there exists an unknown $q > 0$ of
disjoint intervals $(j_T^{(k)}, j_T^{(k)} + \ell_T^{(k)}]$ such that for the kth interval, $1 \leq k \leq q$, there

is a probability $\pi_N^{(k)} > 0$ that this interval has an elevated mean

(2.2)
$$\mu_{nt} = \begin{cases} \mu_N^{(k)} I_n^{(k)} / \sqrt{\ell_T^{(k)}}, & \text{if } j_T^{(k)} < t \le j_T^{(k)} + \ell_T^{(k)}, \\ 0, & \text{otherwise,} \end{cases}$$
$$I_n^{(k)} \sim \text{Bernoulli}\big(\pi_N^{(k)}\big),$$

with $\mu_N^{(k)} > 0$ and the $I_n^{(k)}$'s and Z_{nt}'s jointly independent. Let $\pi_N = \pi_N^{(1)}$, $\mu_N = \mu_N^{(1)}$ and so forth.

Model (2.2) extends sparse mixture detection by adding a time-dimension, and there is a similar extension in potential applications. For example, in the detection of bioweapons use, as introduced in [9], we can assume that there are N observational units in a geographical region, each accumulating information over time on bioweapons usage. The bioweapons are in use over a specific but unknown time period, and only a small fraction of the units are affected. Alternatively in covert communications detection, only a small fraction of N detectors, each tuned to a distinct signal spectrum, observes unusual activities during the period in which communications are taking place. In the detection of genes that are linked to cancer, readings of DNA copy numbers are taken from the chromosomes of N cancer patients, and only a small fraction of the patients exhibits copy number changes at the gene locations. In Section 4, we shall consider an extension of (2.1) and (2.2) to signals carrying a noise component.

In the detection of copy number changes, the common practice was to process samples one at a time; see Lai et al. [18]. In contrast, Efron and Zhang [11], Zhang et al. [29] and Jeng et al. [17] proposed procedures that pool across samples first. Our analysis here shows that the alignment information is important, and we should indeed pool across samples first. In Appendix C, we provide a comparison between pooling information across sample versus pooling information within sample first.

Consider $\pi_N = N^{-\beta}$ for some $\frac{1}{2} < \beta < 1$. Ingster [14, 15] and Donoho and Jin [9] showed that in the special case $T = 1$ (hence $q = 1$, $j_1 = 0$, $\ell_1 = 1$), as $N \to \infty$, the critical detectable value of μ_N is $b_N^*(\beta) := \sqrt{2\rho^*(\beta) \log N}$, where

(2.3)
$$\rho^*(\beta) = \begin{cases} \beta - \frac{1}{2}, & \text{if } \frac{1}{2} < \beta \le \frac{3}{4}, \\ (1 - \sqrt{1-\beta})^2, & \text{if } \frac{3}{4} < \beta < 1. \end{cases}$$

That is, if $\mu_N = \sqrt{2\rho \log N}$ with $\rho < \rho^*$, then no test can detect that $\mu_N \ne 0$ in the sense that the sum of Type I and Type II error probabilities tends to 1 for any test. Donoho and Jin [9] further showed that Tukey's higher criticism as well as the Berk–Jones statistic achieve the detection boundary b_N^*; that is, if $\rho > \rho^*$, then the sum of Types I and II error probabilities tends to 0. Jager and Wellner [16] showed that Tukey's test is a member of a family of goodness-of-fit ϕ-divergence tests that can each achieve the detection boundary.

When $T > 1$, we need to deal with the complication of multiple comparisons over j_T and ℓ_T, and the question arises of how much harder the detection problem becomes. The number of disjoint intervals in $(0, T]$ with common length ℓ_T is approximately T/ℓ_T. This ratio has to be factored into the computation of the detection boundary. The main message of Section 2.1 is that the difficulty of the detection problem is not noticeably affected unless this ratio of sequence length to signal length grows exponentially with the number N of sequences. Sections 2.2 and 2.3 provide optimal max-type tests that attain the detection boundary.

2.1. *Detectability of aligned signals.* Let $a_m \sim b_m$ if $\lim_{m\to\infty}(a_m/b_m) = 1$ and $a_m \stackrel{.}{\sim} b_m$ if $\lim_{m\to\infty}(a_m/b_m) = C$ for some constant $C > 0$. Let $\lfloor \cdot \rfloor$ be the greatest integer function and $\#B$ the number of elements in a set B. Let $E_0(E_1)$ denote expectation under $H_0(H_1)$. We are interested here in the signal length $\ell_T^{(k)}$ in (2.2) satisfying

$$(2.4) \qquad T/\ell_T^{(k)} \sim \exp(N^{\zeta^{(k)}} - 1) \qquad \text{for some } \zeta^{(k)} \geq 0.$$

The case of T varying sub-exponentially with N will be considered in Section 4.

We shall show that under (2.4) with $N \to \infty$, the asymptotic threshold detectable value of μ_N when $\pi_N = N^{-\beta}$ and $\beta \in (0, 1)$ is

$$(2.5)$$
$$b_N(\beta, \zeta)$$
$$= \begin{cases} \sqrt{\log(1 + N^{2\beta-1+\zeta})}, & \text{if } 0 \leq \zeta \leq 1 - 4\beta/3, \\ (\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})\sqrt{2\log N}, & \text{if } 1 - 4\beta/3 < \zeta \leq 1 - \beta, \\ \sqrt{N^{\beta+\zeta-1}}, & \text{if } \zeta > 1 - \beta. \end{cases}$$

The first case $0 \leq \zeta \leq 1 - 4\beta/3$ can be further sub-divided into: (a) $0 \leq \zeta \leq 1 - 2\beta$, under which

$$(2.6) \qquad b_N(\beta, \zeta) \stackrel{.}{\sim} N^{-(1-2\beta-\zeta)/2} \qquad \text{(decays polynomially with } N\text{)},$$

and (b) $1 - 2\beta < \zeta \leq 1 - 4\beta/3$, under which

$$(2.7) \qquad b_N(\beta, \zeta) \sim \sqrt{(2\beta - 1 + \zeta)\log N} \qquad \text{(grows at } \sqrt{\log N} \text{ rate).}$$

Formula (2.5) specifies the functional form of b_N as a function of β. Since β appears in the exponent in (2.6) and in the third case of (2.5), b_N is specified only up to multiplicative constants in these cases.

The boundary b_N is an extension of the Donoho–Ingster–Jin boundary b_N^*. In the case of a sparse mixture, $T = \ell_T = 1$, and (2.4) is satisfied with $\zeta = 0$. By the second case in (2.5) and by (2.7), $b_N(\beta, 0) \sim b_N^*(\beta)$ when $\frac{1}{2} < \beta < 1$. Furthermore, $b_N(\beta, 0)$ in (2.6) recovers the detection boundary in the dense case $0 < \beta \leq \frac{1}{2}$ established by Cai, Jeng and Jin [5].

Formula (2.5) likewise recovers the detection boundary for the special case of only one sequence. For the scaled mean μ_N in (2.2), this boundary is known to be

$\sqrt{2\log(eT/\ell_T)}$ and is attained by the penalized scan; see, for example, [8]. To see how this special case is subsumed in the general setting above, set $T \sim \exp(N-1)$ so that it suffices to consider $\zeta \in (0, 1)$ in (2.4) to parametrize the scale of the signal $\ell_T/T \in (0, 1)$. Then set $\beta = 0$ so that the signal is present in each of the N sequences. Since the signals are aligned and have the same means, by sufficiency one can equivalently consider the one sequence S_t of length T obtained by summing the X_{nt} over n. Dividing by $\sqrt{N}$ to restore unit variance and formally plugging $\beta = 0$ into (2.6) gives a detection threshold for $\sqrt{N}\mu_N$ of $\overset{.}{\sim} N^{\zeta/2} \sim \sqrt{\log(eT/\ell_T)}$. This yields the above detection threshold for the one sequence problem apart from the multiplicative constant $\sqrt{2}$, which can be recovered with a more refined analysis in (2.5).

The general formula (2.5) shows how the growth coefficient and the phase transitions of the $\sqrt{\log N}$ growth are altered by the effect of multiple comparisons in the location of signals. The formula also shows that in the case $\zeta > 0$, the signal detection thresholds can grow polynomially with N.

THEOREM 1. *Assume that (2.2) and (2.4) hold for $1 \le k \le q$, with $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)})$ and $\pi_N^{(k)} = N^{-\beta^{(k)} - \varepsilon^{(k)}}$ for some $0 < \beta^{(k)} < 1$ and $\varepsilon^{(k)} > 0$. Under these conditions, there is no test that can achieve, at all $j_T^{(k)}$, $1 \le k \le q$,*

$$(2.8) \qquad P(\text{Type I error}) + P(\text{Type II error}) \to 0.$$

The simple likelihood ratio of $(X_{n1}, \ldots, X_{nT})$, for H_0 against (2.2), is $L_{n\ell_T j_T}(\pi_N, \mu_N)$, where

$$(2.9) \qquad L_{n\ell j}(\pi^*, \mu) = 1 - \pi^* + \pi^* \exp(\mu Y_{n\ell j} - \mu^2/2),$$

with $Y_{n\ell j} = \ell^{-1/2} \sum_{t=j+1}^{j+\ell} X_{nt}$. The key to proving Theorem 1 (details in Section 5) is to show that under the conditions of Theorem 1,

$$(2.10) \qquad \prod_{n=1}^{N} L_{n\ell_T j_T}(\pi_N, \mu_N) = O_p(T/\ell_T).$$

That is, the likelihood ratio of the signal does not grow fast enough to overcome the noise due to the $\sim T/\ell_T$ independent comparisons of length ℓ_T. Theorem 1 follows because the likelihood ratio test is the most powerful test.

2.2. *Optimal detection with the penalized higher-criticism test.* As an illustration, first consider sparse mixture detection. That is, let $T = 1$ and test

$$(2.11) \qquad X_n \overset{\text{i.i.d.}}{\sim} (1 - \pi_N)N(0, 1) + \pi_N N(\mu_N, 1), \qquad 1 \le n \le N,$$

for H_0: $\pi_N = 0$ against H_1: $\pi_N > 0$ and $\mu_N > 0$. Let $p_{(1)} \le \cdots \le p_{(N)}$ be the ordered p-values of the X_n's.

Donoho and Jin [9] proposed to separate H_0 from H_1 by applying Tukey's higher-criticism test statistic

$$(2.12) \qquad \mathrm{HC}_N := \max_{1 \le n \le (N/2)\,:\, p_{(n)} \ge N^{-1}} \frac{n/N - p_{(n)}}{\sqrt{p_{(n)}(1 - p_{(n)})/N}}.$$

They showed that the higher-criticism test is optimal for sparse mixture detection. Under H_0, $\mathrm{HC}_N \sim \sqrt{2 \log \log N}$; see [9], Theorem 1. Under H_1, the argument of HC_N at some $p_{(n)}$ is asymptotically larger than $\sqrt{2 \log \log N}$, when $\pi_N = N^{-\beta}$ for some $\frac{1}{2} < \beta < 1$, and μ_N lies above the detection boundary $b_N^*(\beta) = \sqrt{2\rho^*(\beta) \log N}$. For μ_N lying below the detection boundary, it is not possible to separate H_0 from H_1. Cai et al. [5] showed that optimality extends to $\beta \in (0, \frac{1}{2})$.

We motivate the extension of the higher-criticism test to $T > 1$ by first considering a fixed, known signal on the interval $(j, j + \ell]$. By sufficiency, testing for an aligned signal there is the same as testing H_0 against H_1 for the sample $Y_{1\ell j}, \ldots, Y_{N\ell j}$. Let $p_{(1)\ell j} \le \cdots \le p_{(N)\ell j}$ be the ordered p-values of the sample, and let $s_{\ell T} = \log(eT/\ell)$. We define the higher-criticism test statistic on this interval to be

$$(2.13) \qquad \mathrm{HC}_{N\ell j} := \max_{1 \le n \le (N/2)\,:\, p_{(n)} \ge s_{\ell T}/N} \frac{n/N - p_{(n)\ell j}}{\sqrt{p_{(n)\ell j}(1 - p_{(n)\ell j})/N}}.$$

For $\ell = T$, the constraint in (2.13) becomes $p_{(n)} \ge N^{-1}$, which agrees with the constraint in (2.12). As explained in [9], Section 3, the standardization of $p_{(n)}$ given in (2.13) has increasingly heavy tails as n becomes smaller, so if $\mathrm{HC}_{N\ell j}$ is defined without constraints on $p_{(n)}$, then it has large values frequently due to the smallest $p_{(n)}$. For $\ell < T$, the multiple comparisons when maximizing $\mathrm{HC}_{N\ell j}$ over j necessitates a more restrictive constraint of $p_{(n)} \ge s_{\ell T}/N$.

The term $s_{\ell T}$ appears also in the scan of the higher-criticism test statistic

$$(2.14) \qquad \mathrm{PHC}_{NT} := \max_{(j, j+\ell) \in B_T} (\mathrm{HC}_{N\ell j} - \sqrt{s_{\ell T} \log s_{\ell T}}),$$

as a penalty that increases with T/ℓ to counter-balance the generally higher scores under H_0 for larger T/ℓ when maximizing $\mathrm{HC}_{N\ell j}$ over j.

We will now specify the scanning set B_T in (2.14). In applications T is often large, so maximizing $\mathrm{HC}_{N\ell j}$ over all j and ℓ is computationally expensive; the cost is NT^2. We construct below an approximating set B_T, similar to that in Walther [26] and Rivera and Walther [22], which has a computation cost of $NT \log T$.

Construction of B_T: Let $d_{r,T} = \lfloor T/(r^{1/2} e^r) \rfloor + 1$, and let

$$(2.15) \qquad B_{r,T} = \{(j, j + \ell) \in (d_{r,T} \mathbf{Z})^2 : 0 \le j \le T - \ell, T/e^r < \ell \le T/e^{r-1}\}.$$

We define $B_T = \bigcup_{r=1}^{r_T} B_{r,T}$, where $r_T = \lfloor \log T \rfloor$. The specification of $d_{r,T}$ is so that for any (j_T, ℓ_T), we can find $(j_T^*, \ell_T^*) \in B_{r,T}$ for some r such that $j_T \leq j_T^* < j_T^* + \ell_T^* \leq j_T + \ell_T$ and

$$(2.16) \qquad 1 - \ell_T^*/\ell_T = O(r^{-1/2}).$$

This property plays a part in ensuring that the loss of information due to restriction to B_T is negligible.

THEOREM 2. *Assume* (2.2) *and that for some* $1 \leq k \leq q$, (2.4) *holds and* $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)})$, $\pi_N^{(k)} = N^{-\beta^{(k)}+\varepsilon^{(k)}}$ *for some* $0 < \beta^{(k)} < 1 - \zeta^{(k)}$ *and* $0 < \varepsilon^{(k)} \leq \beta^{(k)}$. *Under these conditions,* $P(\text{Type I error}) + P(\text{Type II error}) \to 0$ *can be achieved by testing with* PHC_{NT}.

For signal identification, when applying the penalized higher-criticism test statistics at a threshold c:

(1) Rank the pairs $(j, j + \ell) \in B_T$ in order of descending values of $\text{HC}_{N\ell j} - \sqrt{s_{\ell T} \log s_{\ell T}}$, and remove those pairs with values less than c.

(2) Starting with the highest-ranked pair and moving downward, remove a pair from the list if its interval overlaps with that of a higher-ranked pair still on the list by more than a fraction $f \geq 0$ of its length.

Jeng et al. [17] focused on the detection of signal segments that are well separated. Hence their signal identification procedure is restricted to $f = 0$. Zhang et al. [29] focused on both the detection of signal segments that are well separated, as well as the detection of overlapping or nested signal segments. Hence their procedure allows for $f > 0$. If there are a finite number of well-separated signal segments, then intuitively, all the segments are identified with probability converging to 1, in the sense that a segment with local maximum score, in a suitably defined neighborhood of each signal segment, is identified.

2.3. *Optimal detection with the penalized Berk–Jones test.* Let $K(x, t) = x \log(\frac{x}{t}) + (1 - x) \log(\frac{1-x}{1-t})$ if $x \geq t$ and $K(x, t) = 0$ otherwise. This is the Berk–Jones [4] test statistic that was first proposed as a more powerful alternative to the Kolmogorov–Smirnov test statistic for testing a distribution function; see also Owen [21]. Jager and Wellner [16] showed that there is a class of test statistics that includes the Berk–Jones and higher-criticism test statistics as special cases that can be used to detect sparse mixtures (2.11) optimally. Specifically for $T = 1$, the testing of $\pi_N > 0$ and $\mu_N > 0$ in (2.11) can be detected optimally by

$$(2.17) \qquad \text{BJ}_N := N \max_{1 \leq n \leq (N/2) \,:\, p_{(n)} < n/N} K(n/N, p_{(n)}).$$

Therefore, analogously to (2.13),

$$(2.18) \qquad \text{BJ}_{N\ell j} := N \max_{1 \leq n \leq (N/2) \,:\, p_{(n)\ell j} < n/N} K(n/N, p_{(n)\ell j})$$

can optimally detect aligned signals on the interval $(j, j + \ell]$. In Theorem 3 below, we shall show that analogously to (2.14), the penalized Berk–Jones test statistic

$$(2.19) \qquad \text{PBJ}_{NT} := \max_{(j,j+\ell) \in B_T} (\text{BJ}_{N\ell j} - s_{\ell T} \log s_{\ell T})$$

is optimal for aligned signals detection when the signal locations are unknown.

THEOREM 3. *Assume* (2.2) *and that for some* $1 \le k \le q$, (2.4) *holds and* $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)})$, $\pi_N^{(k)} = N^{-\beta^{(k)} + \varepsilon^{(k)}}$ *for some* $0 < \beta^{(k)} < 1$ *and* $0 < \varepsilon^{(k)} \le \beta^{(k)}$. *Under these conditions,*

$$P(\text{Type I error}) + P(\text{Type II error}) \to 0$$

can be achieved by testing with PBJ_{NT}.

As in Section 2.2, a sequential approach can be used to identify signals when the penalized Berk–Jones exceeds a specified threshold.

3. Optimal detection with ALR tests. We shall introduce in Section 3.1 an ALR that is optimal for detecting multi-sample aligned signals. We then consider the special cases of detecting a sparse mixture ($T = 1$ with $N \to \infty$) in Section 3.2 and multiscale detection in a single sequence ($N = 1$ with $T \to \infty$) in Section 3.3.

3.1. *Detecting multi-sample aligned signals.* The ALR builds upon the likelihood ratios $L_{n\ell_T j_T}(\pi_N, \mu_N)$ as defined in (2.9), first by substituting μ_N by its asymptotic threshold detectable value, followed by integrating $\pi_N = N^{-\beta}$ over β and finally by summing over an approximating set for ℓ_T and j_T. In view of (2.4), let $\zeta_{\ell,NT} = \log_N[\log(T/\ell) + 1]$, and let

$$(3.1) \qquad L_{n\ell j}(\beta) = L_{n\ell j}(N^{-\beta}, b_N(\beta, \zeta_{\ell,NT}));$$

see (2.5) for the definition of b_N.

In the case of the ALR, we consider

$$(3.2) \qquad A_{NT} := \frac{6}{\pi^2} \sum_{r=1}^{r_T \vee 1} \frac{1}{r^3 e^{r+1}} \sum_{(j,j+\ell) \in B_{r,T}} \int_0^1 \left[\prod_{n=1}^{N} L_{n\ell j}(\beta) \right] d\beta,$$

where r_T and $B_{r,T}$ are given in Section 2.2. By (2.15), $\#B_{r,T} \le r e^{r+1}$.

The weights in (3.2) are chosen for the following reason: Since $L_{n\ell j}(\beta)$ is a likelihood ratio for H_0, it has expectation 1 under H_0. Hence it follows from (3.2) that

$$(3.3) \qquad E_0(A_{NT}) = \frac{6}{\pi^2} \sum_{r=1}^{r_T} \frac{1}{r^3 e^{r+1}} (\#B_{r,T}) \le \frac{6}{\pi^2} \sum_{r=1}^{\infty} \frac{1}{r^2} = 1.$$

From (3.3), it follows that under H_0, $A_{NT} = O_p(1)$ uniformly over N and T. If the aligned signals under H_1 are such that $A_{NT} \overset{p}{\to} \infty$ as $N \to \infty$, then $P(\text{Type I error}) + P(\text{Type II error}) \to 0$ is achieved by simply selecting rejection thresholds going to infinity slowly enough.

The ALR (3.2) is, by its construction, optimal when $\mu_N = b_N(\beta, \zeta_{\ell_T, NT})$. It is not designed to be optimal at other μ_N. However, there is really no point in being optimal at smaller μ_N, where the maximum power that can be attained is small. At larger μ_N, the ALR test has power close to 1, so there is not much more to be gained in being optimal there. By focusing only on the boundary detectable values, we remove the noise due to the consideration of unproductive likelihood ratios associated with too large and too small μ_N.

THEOREM 4. *Assume* (2.2) *and that for some* $1 \leq k \leq q$ (2.4) *holds and* $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)})$, $\pi_N^{(k)} = N^{-\beta^{(k)} + \varepsilon^{(k)}}$ *for some* $0 < \beta^{(k)} < 1$ *and* $0 < \varepsilon^{(k)} \leq \beta^{(k)}$. *Under these conditions,* $A_{NT} \overset{p}{\to} \infty$. *Hence* $P(\text{Type I error}) + P(\text{Type II error}) \to 0$ *can be achieved by testing with* A_{NT}.

Among the three optimal tests that we propose here, the ALR is the most intuitive, and the proof of its optimality is also the most straightforward. However, its computation involves the evaluation of a nonstandard integral, and its form is closely linked to normal errors. On the other hand, the penalized scans involve no integrations in their computations, and the p-values in their expressions are not tied to normal errors.

3.2. *Detecting sparse mixtures.* This setting has been studied in [9] and discussed briefly in Section 2. It corresponds to the special case $T = 1$ in the above theory, and our test statistic A_{NT} simplifies as follows: $T = 1$ implies $r_T = \zeta = \zeta_{\ell, NT} = 0$, and B_T contains only $j = 0, \ell = 1$. Hence

$$A_{N1} = \frac{6}{\pi^2 e^2} \int_0^1 \prod_{n=1}^N \left(1 - N^{-\beta} + N^{-\beta} \exp\{b_N(\beta, 0) X_n - b_N^2(\beta, 0)/2\}\right) d\beta,$$

where

$$b_N(\beta, 0) = \begin{cases} \sqrt{\log(1 + N^{2\beta - 1})}, & \text{if } 0 < \beta \leq \frac{3}{4}, \\ (1 - \sqrt{1 - \beta})\sqrt{2 \log N}, & \text{if } \frac{3}{4} < \beta < 1, \end{cases}$$

is essentially the Cai–Jeng–Jin detection boundary $b_N^*(\beta) := N^{-1/2+\beta}$ for $\beta \in (0, \frac{1}{2})$, and the Donoho–Ingster–Jin detection boundary $b_N^*(\beta) = \sqrt{2\rho^*(\beta) \log N}$ for $\beta \in (\frac{1}{2}, 1)$.

COROLLARY 1. *Assume* (2.11) *with* $\mu_N = b_N(\beta, 0)$ *and* $\pi_N = N^{-\beta + \varepsilon}$ *for some* $0 < \beta < 1$ *and* $0 < \varepsilon \leq \beta$. *Under these conditions,* $A_{N1} \overset{p}{\to} \infty$ *and* (2.8) *can be achieved by testing with* A_{N1}.

3.3. *Signal detection in a single sequence.* Let $N = 1$ and $\pi_1 = 1$. The resulting $\beta = 0$ is not covered by our general theory, but it is a boundary case, and therefore it is of interest to see whether our general statistic A_{NT} still allows optimal detection in this important special case. For this testing problem in a single sequence with $T \to \infty$, it is known that the critical detectable value of μ_T is $b_T(\ell_T)$, where $b_T(\ell) = \sqrt{2\log(eT/\ell)}$, and that the popular scan statistic is suboptimal except for signals on the smallest scales; see Chan and Walther [8]. It is also shown there that optimal detection can be achieved by modifying the scan with the penalty method introduced by Dümbgen and Spokoiny [10], or by employing the condensed average likelihood ratio.

Note that when analyzing a single sequence we know a priori that $\beta = 0$, and therefore it makes sense to set β to 0 in the definition of A_{NT} rather than integrating β over $(0, 1)$. The resulting statistic is

$$(3.4) \qquad A_T := \frac{6}{\pi^2} \sum_{r=1}^{r_T \vee 1} \frac{1}{r^3 e^{r+1}} \sum_{(j,j+\ell) \in B_{r,T}} \exp[b_T(\ell)Y_{\ell j} - b_T^2(\ell)/2],$$

where $Y_{\ell j} = \ell^{-1/2} \sum_{t=j+1}^{j+\ell} X_{1t}$.

The test statistic A_T is able to achieve the detection boundary $b_T(\ell)$ simply because it optimizes detection power there:

THEOREM 5. *If there exist ℓ_T and j_T such that $E_1(Y_{\ell_T j_T}) = b_T(\ell_T) + c_T$, with $c_T \to \infty$ as $T \to \infty$, then $A_T \xrightarrow{p} \infty$ and* (2.8) *can be achieved by testing with A_T.*

3.4. *An example.* Efron and Zhang [11] applied local f.d.r. to detect CNV in multi-sample DNA sequences. Measurements from $T = 42,075$ probes were taken on each chromosome 1 of $N = 207$ glioblastoma subjects from the Cancer Genome Atlas Project [7]. At each probe on each sequence, the moving averages of the readings over windows of length $\ell = 51$ were normalized. These normalized averages correspond to the $Y_{n\ell j}$ scores defined just before (2.9). The computed local f.d.r. of the scores at each j determined the conclusion of an aligned signal there. The scientific purpose is to detect rare inherited CNV that may occur in a small fraction, perhaps 5%, of the population.

Consider, for example, $\ell_T = 51$ and $\pi_N = 0.05$. The solution of $T/\ell_T = \exp(N^\zeta - 1)$ [see (2.4)] is $\zeta = 0.383$. The solution of $N^{-\beta} = 0.05$ is $\beta = 0.568$. Since $1 - 4\beta/3 < \zeta \le 1 - \beta$, we are under the second case in (2.5). Based on (2.5), the signal-to-noise ratio (for a single observation in a variant segment) required for successful detection in a mixture with 5% variant is then

$$b_N(\beta, \zeta)/\sqrt{\ell_T} = 0.258.$$

It is known from earlier studies that the sequence between probes 8800 and 8900 contains two genes that enhance cell death. Copy number losses of these
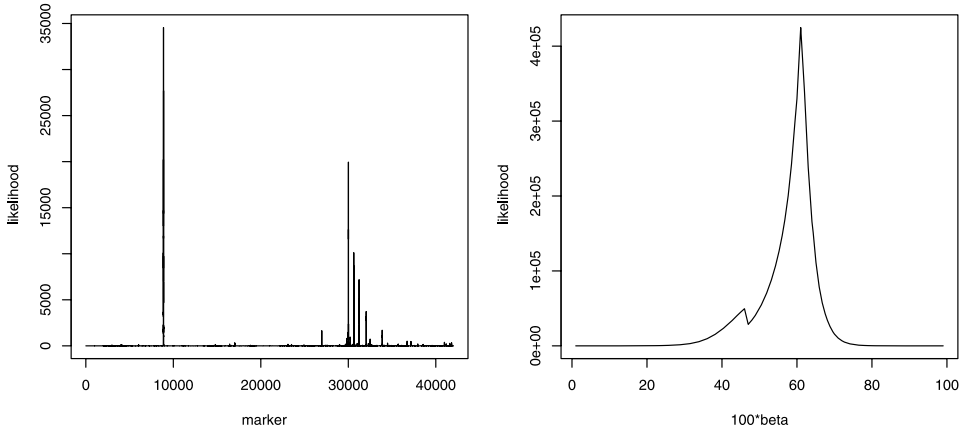
FIG. 1. (*Left*) *Plot of likelihood against marker position. The tallest peak is at marker* 8852. (*Right*) *Plot of likelihood against variant fraction* $\pi_N = N^{-\beta}$ *at marker* 8852.

genes promote unregulated cell growth, leading to tumor. The display in Figure 1 (left) shows that the likelihood at marker j (for a signal on the probe interval $j < t \leq j + \ell$),

$$L_{\ell j} := \int_0^1 \left[\prod_{n=1}^{N} L_{n\ell j}(\beta) \right] d\beta,$$

is maximized at $j = 8852$. The display in Figure 1 (right) shows that the likelihood at marker $j = 8852$ for variant fraction $\pi_N = N^{-\beta}$,

$$L_{\ell j}(\beta) := \prod_{n=1}^{N} L_{n\ell_T j}(\beta),$$

is maximized at $\beta = 0.61$. This translates to an estimated 4% of the population tested having copy number losses in the probe interval $8852 < t \leq 8903$.

4. Extensions. Cai, Jeng and Jin [5] and Cai and Wu [6] showed that the HC test is optimal for heteroscedastic and more general mixtures, respectively. Arias-Castro and Wang [3] analyzed the detection capabilities of distribution-free tests for null hypotheses that are not fully specified. Jeng, Cai and Li [17] showed that the HC test statistic is optimal for detecting heteroscedastic aligned sparse signals when assuming that the signal length is very small and that T does not grow rapidly with N. However, when the aligned signals may range over multiple scales, the penalty terms introduced in Section 2.2 are critical in ensuring optimality of the HC test.

Below, we shall show how the detection boundary of Cai et al. [5] looks for general T/ℓ_T asymptotics, and show that the adaptive optimality of the HC and BJ

tests extends to heteroscedastic signals when their penalties, as given in Section 2, are applied. This brings home the point that the penalties are not tied down to a particular model. Following Jeng et al. [17], we assume in place of (2.1) that

$$(4.1) \qquad X_{nt} = U_{nt} + Z_{nt} \qquad \text{where } Z_{nt} \text{ are i.i.d. N}(0, 1).$$

Under the null hypothesis H_0 of no signals, $U_{nt} \equiv 0$ for all n and t. Under the alternative hypothesis H_1 of aligned signals, there exists an unknown $q > 0$ of disjoint intervals $(j_T^{(k)}, j_T^{(k)} + \ell_T^{(k)}]$, $1 \le k \le q$, such that for the kth interval,

$$
\begin{aligned}
U_{nt} &= \text{N}\Big(\mu_N^{(k)}/\sqrt{\ell_T^{(k)}}, \tau^{(k)}\Big) \qquad \text{if } I_n^{(k)} = 1 \text{ and } t \in (j_T^{(k)}, j_T^{(k)} + \ell_T^{(k)}], \\
I_n^{(k)} &\sim \text{Bernoulli}\big(\pi_N^{(k)}\big),
\end{aligned}
$$
(4.2)

and $U_{nt} = 0$ otherwise, with $\pi_N^{(k)} > 0$, $\mu_N^{(k)} > 0$ and $\tau^{(k)} \ge 0$. We shall denote $\mu_N^{(1)}$ by μ_N, $\ell_T^{(1)}$ by ℓ_T and so forth. Let $b_N(\beta, \zeta, \tau)$ be such that $b_N(\beta, \zeta, 0) = b_N(\beta, \zeta)$, and for $\tau > 0$ and $0 \le \zeta < 1 - \beta$, let

$$
b_N(\beta, \zeta, \tau)/\sqrt{\log N}
$$

$$
= \begin{cases}
0, & \text{if } \zeta \le 1 - 2\beta \text{ or } \tau \ge \dfrac{\beta}{1 - \zeta - \beta}, \\
\sqrt{(1 - \tau)(2\beta + \zeta - 1)}, & \\
\qquad \text{if } 1 - 2\beta < \zeta \le 1 - \dfrac{4\beta}{3 - \tau}, \\
\sqrt{2(1 - \zeta)} - \sqrt{2(1 + \tau)(1 - \zeta - \beta)}, & \\
\qquad \text{if } 1 - \min\Big(2\beta, \dfrac{4\beta}{3 - \tau}\Big) < \zeta \text{ and } \tau < \dfrac{\beta}{1 - \zeta - \beta}.
\end{cases}
$$

THEOREM 6. *Assume* (4.1) *and* (4.2). *If for all* $1 \le k \le q$, (2.4) *holds and* $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)}, \tau^{(k)})$, $\pi_N^{(k)} = N^{-\beta^{(k)} - \varepsilon^{(k)}}$ *for some* $0 < \beta^{(k)} < 1 - \zeta^{(k)}$ *and* $\varepsilon^{(k)} > 0$, *then there is no test that can achieve, at all* $j_T^{(k)}$, $1 \le k \le q$,

$$(4.3) \qquad P(\text{Type I error}) + P(\text{Type II error}) \to 0.$$

Conversely, if for some $1 \le k \le q$, (2.4) *holds and* $\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta^{(k)}, \tau^{(k)})$, $\pi_N^{(k)} = N^{-\beta^{(k)} + \varepsilon^{(k)}}$ *for some* $0 < \beta^{(k)} < 1 - \zeta^{(k)}$ *and* $0 < \varepsilon^{(k)} \le \beta^{(k)}$, *then* (4.3) *can be achieved by the penalized HC and BJ tests.*

It can be checked that setting $\zeta^{(k)} = 0$ will recover for us the boundary for aligned signals in Jeng et al. [17]. Incidentally, they assumed that

$$(4.4) \qquad \log T = o(N^C) \qquad \text{for all } C > 0,$$

which effectively brings us to the case $\zeta^{(k)} = 0$. Corollary 2 below extends the optimality of the HC test in Jeng et al. [17] to multiscale signal lengths, by

introducing the penalty terms as described in Section 2. In place of (2.4), let
$\zeta_N^{(k)} = \log\log(eT/\ell_T^{(k)})/\log N$.

COROLLARY 2. *Assume* (4.1) *and* (4.2). *Theorem* 6 *holds under* (4.4) *with*
$\mu_N^{(k)} = b_N(\beta^{(k)}, \zeta_N^{(k)}, \tau^{(k)})$ *and* $0 < \beta^{(k)} < 1$.

5. Proofs of Theorems 1, 4 and 5. We say that $U_m \overset{p}{\sim} V_m$ if $U_m = O_p(V_m)$ and $V_m = O_p(U_m)$, and that $a_m \gg b_m$ if $a_m/b_m \to \infty$. We start with the proof of Theorem 1 in Section 5.1, that detection is asymptotically impossible below the detection boundary b_N, followed by the proofs of Theorem 4 (in Section 5.2) and Theorem 5 (in Section 5.3), that the average likelihood ratio test is optimal. These proofs are consolidated in this section as they are unified by a likelihood ratio approach. Since the detection problem is easier when $q > 1$ compared to $q = 1$, we may assume without loss of generality that $q = 1$ under H_1 in all the proofs.

5.1. *Proof of Theorem* 1. For $\zeta = 0$ the claim of the theorem reduces to theorems proved by Ingster [15] in the sparse case and by Cai, Jeng and Jin [5] in the dense case.

Let $\zeta > 0$, and set $i_T = \lfloor T/\ell_T \rfloor - 1$, so $i_T \sim \exp(N^\zeta - 1)$, set $\mu_N = b_N(\beta, \zeta)$ and $\pi_N = N^{-\beta-\varepsilon}$. Let $Y_{n1} = \ell_T^{-1/2}(X_{n,j_T+1} + \cdots + X_{n,j_T+\ell_T})$, and let each Y_{ni}, $2 \le i \le i_T$, be of the form $\ell_T^{-1/2}(X_{n,j+1} + \cdots + X_{n,j+\ell_T})$, with all $(j, j+\ell_T]$ disjoint from each other, and from $(j_T, j_T + \ell_T]$. Let

$$(5.1) \qquad L_{ni} = 1 + \pi_N[\exp(\mu_N Y_{ni} - \mu_N^2/2) - 1], \qquad L_i = \prod_{n=1}^{N} L_{ni}.$$

Since $(j, j + \ell_T]$ are disjoint, $L_1, \ldots, L_{i_T}$ are independent. We take note that $L_2, \ldots, L_{i_T}$ have identical distributions which are unchanged when we switch from H_0 to H_1. In contrast, the distribution of L_1 changes when we switch from H_0 to H_1. Consider

$$L = \frac{1}{i_T}L_1 + \frac{1}{i_T}\sum_{i=2}^{i_T} L_i,$$

which is the likelihood ratio when j_T is equally likely to take one of i_T possible values spaced at least ℓ_T apart, as explained above. This we assume without loss of generality.

If we are able to find λ_N such that both

$$(5.2) \quad L_1 = O_p(\lambda_N) \qquad \text{under } H_1 \quad \text{and}$$

$$(5.3) \quad P\left(a_N + M\lambda_N > \sum_{i=2}^{i_T} L_i > a_N\right) \nrightarrow 1 \qquad \text{for all } a_N \in \mathbf{R} \text{ and } M > 0$$

are satisfied, then L is unable to achieve (2.8). If so, then no test is able to achieve (2.8) because the likelihood ratio test is the most powerful test.

Case 1: $0 < \zeta \leq 1 - 4\beta/3$, $\mu_N = \sqrt{\log(1 + N^{2\beta-1+\zeta})}$, $\pi_N = N^{-\beta-\varepsilon}$ with $0 < \varepsilon < \zeta/2$. Under H_1, $Y_{11}, \ldots, Y_{N1}$ are i.i.d. $(1 - \pi_N)\mathrm{N}(0, 1) + \pi_N\mathrm{N}(\mu_N, 1)$. Let

$$(5.4) \qquad \lambda_N = E_1(L_1) = \left[1 + \pi_N^2(e^{\mu_N^2} - 1)\right]^N = \exp\{[1 + o(1)]N^{\zeta-2\varepsilon}\}.$$

Hence (5.2) holds.

Let $i \geq 2$. Since $Y_{ni} \sim \mathrm{N}(0, 1)$,

$$(5.5) \qquad E_0(L_i) = 1, \qquad E_0(L_i^2) = \left[1 + \pi_N^2(e^{\mu_N^2} - 1)\right]^N = \lambda_N.$$

We check in Appendix A that Lyapunov's condition holds. Hence

$$(5.6) \qquad \frac{1}{\sqrt{(\lambda_N - 1)(i_T - 1)}} \sum_{i=2}^{i_T} (L_i - 1) \Rightarrow \mathrm{N}(0, 1).$$

Since $\sqrt{(\lambda_N - 1)(i_T - 1)} = \exp\{[1 + o(1)](N^{\zeta-2\varepsilon} + N^\zeta)/2\} \gg \lambda_N$, (5.3) follows from (5.6).

Case 2: $1 - 4\beta/3 < \zeta \leq 1 - \beta$, $\mu_N = (x - y)\sqrt{2\log N}$ where $x = \sqrt{1 - \zeta}$, $y = \sqrt{1 - \zeta - \beta}$, $\pi_N = N^{-\beta-\varepsilon}$. Let

$$\mathcal{N}^0 = \{n : I_n = 0 \text{ or } Y_{n1} < x\sqrt{2\log N}\},$$

$$\mathcal{N}^1 = \{n : I_n = 1 \text{ and } Y_{n1} \geq x\sqrt{2\log N}\},$$

and define, for $h = 0, 1$,

$$(5.7) \quad L_1^h = \prod_{n \in \mathcal{N}^h} L_{n1} \qquad \text{where } L_{n1} = 1 + \pi_N[\exp(\mu_N Y_{n1} - \mu_N^2/2) - 1].$$

Check that

$$m_0 := E_1(L_{n1}|I_n = 0) = 1,$$

$$m_1 := E_1[1 + (L_{n1} - 1)\mathbf{I}_{\{Y_{n1} \leq x\sqrt{2\log N}\}}|I_n = 1]$$

$$= 1 + \pi_N[e^{\mu_N^2}\Phi(x\sqrt{2\log N} - 2\mu_N) - \Phi(x\sqrt{2\log N} - \mu_N)].$$

Since $y^2 - x^2 = -\beta$ and $x < 2(x - y)$ when $1 - 4\beta/3 < \zeta \leq 1 - \beta$, it follows that

$$(5.8) \quad \begin{aligned} E_1(L_1^0) &= \left[(1 - \pi_N)m_0 + \pi_N m_1\right]^N \\ &\leq \left[1 + \pi_N^2 e^{\mu_N^2}\Phi(x\sqrt{2\log N} - 2\mu_N)\right]^N \\ &= \left[1 + O(N^{2(y^2-x^2)-2\varepsilon+2(x-y)^2-(2y-x)^2})/\sqrt{\log N}\right]^N \\ &= \exp[O(N^{1-x^2-2\varepsilon})/\sqrt{\log N}] = \exp[O(N^{\zeta-2\varepsilon})/\sqrt{\log N}]. \end{aligned}$$

Next we apply $\max_{n \in \mathcal{N}^1} Y_{n1} = O_p(\sqrt{\log N})$ to show that

$$\begin{aligned}
\log L_1^1 &= O_p((\#\mathcal{N}^1) \log N) = O_p(N \pi_N \Phi(-y\sqrt{2 \log N}) \log N) \\
&= O_p(N^{\zeta - \varepsilon} \sqrt{\log N}).
\end{aligned} \tag{5.9}$$

By (5.8), (5.9) and $L_1 = L_1^0 L_1^1$, (5.2) holds for $\lambda_N = \exp(N^{\zeta - \varepsilon} \log N)$.

Let $i \geq 2$. Let $\widetilde{L}_{ni} = 1 + \pi_N[\exp(\mu_N Y_{ni} - \mu_N^2/2) - 1]\mathbf{I}_{\{Y_{ni} \leq x\sqrt{2 \log N}\}}$ and $\widetilde{L}_i = \prod_{n=1}^N \widetilde{L}_{ni}$. We check that

$$\begin{aligned}
E_0(\widetilde{L}_{ni}) &= 1 + \pi_N\big[\Phi(x\sqrt{2 \log N} - \mu_N) - \Phi(x\sqrt{2 \log N})\big] \\
&= 1 - [C + o(1)]N^{-\beta - y^2 - \varepsilon}/\sqrt{\log N},
\end{aligned} \tag{5.10}$$

where $C = (2y\sqrt{\pi})^{-1}$. From this and $1 - \beta - y^2 = \zeta$, we conclude that

$$\kappa_N := E_0(\widetilde{L}_i) = \exp\{-[C + o(1)]N^{\zeta - \varepsilon}/\sqrt{\log N}\}. \tag{5.11}$$

Since $E_0(\widetilde{L}_{ni}^2) \geq [E_0(\widetilde{L}_{ni})]^2$ for $n \geq 2$, it follows that

$$\begin{aligned}
v_N := \mathrm{Var}_0(\widetilde{L}_i) &= \prod_{n=1}^N E_0(\widetilde{L}_{ni}^2) - \kappa_N^2 \\
&\geq \left(\frac{E_0(\widetilde{L}_{1i}^2)}{[E_0(\widetilde{L}_{1i})]^2} - 1\right)\kappa_N^2 \sim \mathrm{Var}_0(\widetilde{L}_{1i})\kappa_N^2,
\end{aligned} \tag{5.12}$$

and by (5.10),

$$\begin{aligned}
\mathrm{Var}_0(\widetilde{L}_{1i}) &= \pi_N^2\big[e^{\mu_N^2}\Phi(x\sqrt{2 \log N} - 2\mu_N) \\
&\quad - 2\Phi(x\sqrt{2 \log N} - \mu_N) + \Phi(x\sqrt{2 \log N})\big] \\
&\quad - \big[E_0(\widetilde{L}_{1i} - 1)\big]^2 \\
&\sim C_1 N^{\zeta - 1 - 2\varepsilon}/\sqrt{\log N},
\end{aligned} \tag{5.13}$$

where $C_1 = [(2x - 4y)\sqrt{\pi}]^{-1}$. We check that Lyapunov's condition holds and conclude that

$$\frac{1}{\sqrt{v_N(i_T - 1)}} \sum_{i=2}^{i_T} (\widetilde{L}_i - \kappa_N) \Rightarrow \mathrm{N}(0, 1). \tag{5.14}$$

By (5.11)–(5.13), $\sqrt{v_N(i_T - 1)} \geq \exp\{[1 + o(1)]N^{\zeta}/2\} \gg \lambda_N$, and so (5.3) holds, but with L_i replaced by $\widetilde{L}_i$. The variability of L_i is larger than that of $\widetilde{L}_i$, and hence (5.3) for L_i holds as well.

Case 3: $\zeta > 1 - \beta$, $\mu_N = \sqrt{N^{\zeta-1+\beta}}$, $\pi_N = N^{-\beta-\varepsilon}$ with $0 < \varepsilon < \zeta + \beta - 1$. Let $\mathcal{N}^h = \{n : I_n = h\}$, $h = 0, 1$, and define L_1^h as in (5.7). Since $\max_{n \in \mathcal{N}^0} Y_{n1} = O_p(\sqrt{\log N}) = o_p(\mu_N)$, it follows that

$$(5.15) \qquad\qquad P_1(L_1^0 \leq 1) \to 1.$$

We next apply the inequality

$$\log(1 - \pi_N + \pi_N e^{\mu_N Y_{n1} - \mu_N^2/2}) \leq 1 + \max(\mu_N Y_{n1} - \mu_N^2/2 - \beta \log N, 0)$$

to show that

$$(5.16) \quad \log L_1^1 \leq [1 + o_p(1)] N \pi_N E(\mu_N Y_{11} - \mu_N^2/2 | I_1 = 1) \overset{P}{\sim} N^{\zeta-\varepsilon}.$$

It follows from (5.15), (5.16) and $L_1 = L_1^0 L_1^1$ that (5.2) holds for $\lambda_N = \exp(N^{\zeta-\varepsilon} \log N)$.

Let $i \geq 2$ and $\Gamma_i = \{Y_{ni} \geq \mu_N$ for all $1 \leq n \leq N^{1-\beta}/2 + 1\}$. Then

$$\log P(\Gamma_i) \sim -N^{1-\beta} \mu_N^2/4 \sim -N^{\zeta}/2.$$

Hence $i_T \sim \exp(N^{\zeta} - 1) \gg [P(\Gamma_i)]^{-1}$ and

$$(5.17) \qquad\qquad P(\#\{i : \Gamma_i \text{ occurs}\} = k_N) \not\to 1 \qquad \text{for any } k_N.$$

If Γ_i occurs, then

$$
\begin{aligned}
L_i &\geq (1 - \pi_N)^N \left(1 - \pi_N + \pi_N e^{\mu_N^2/2}\right)^{N^{1-\beta}/2} \\
(5.18) \qquad &= \exp\{[1 + o(1)][-N^{1-\beta-\varepsilon} + (N^{1-\beta}/2)N^{\zeta-1+\beta}/2]\} \\
&= \exp\{[1 + o(1)]N^{\zeta}/4\} \gg \lambda_N.
\end{aligned}
$$

Since typically $L_i \ll \lambda_N$, we can conclude (5.3) from (5.17) and (5.18).

5.2. *Proof of Theorem* 4.

Let (2.2) and (2.4) hold with $\mu_N = b_N(\beta, \zeta)$ and $\pi_N = N^{-\beta+\varepsilon}$. Let $(j_T^*, j_T^* + \ell_T^*) \in B_{r,T}$ satisfy (2.16), and note that by (2.4) and (2.15), $r \sim N^{\zeta}$. Hence for $\zeta > 0$ and N large,

$$
\begin{aligned}
(5.19) \qquad E_1(Y_{n\ell_T^* j_T^*} | I_n = 1) &= \mu_N \sqrt{\ell_T^*/\ell_T} \geq [1 - O(N^{-\zeta/2})]b_N(\beta, \zeta) \\
&\geq b_N(\eta, \zeta_{\ell_T^*, NT}), \qquad \beta_1 \leq \eta \leq \beta_2,
\end{aligned}
$$

for any (β_1, β_2) lying in the interior of $(\beta - \varepsilon, \beta)$, and this inequality can also be checked for $\zeta = 0$. Let

$$(5.20) \qquad\qquad L(\eta) = \prod_{n=1}^{N} L_{n\ell_T^* j_T^*}(\eta).$$

Since $r^3 e^{r+1} = O(N^{3\zeta} \exp(N^\zeta))$, in view of (3.2) and (5.19), Theorem 4 follows from

$$(5.21) \qquad N^{-3\zeta} \exp(-N^\zeta) \int_{\beta_1}^{\beta_2} L(\eta)\, d\eta \xrightarrow{p} \infty.$$

To show (5.21), it suffices to check that

$$(5.22) \qquad N^{-3\zeta} \exp(-N^\zeta) L(\eta) \xrightarrow{p} \infty,$$

when $E_1(Y_{n\ell_T^* j_T^*} | I_n = 1) \geq b_N(\eta, \zeta_{\ell_T^*, NT})$, and $\pi_N = N^{-\beta+\varepsilon}$ with $\beta - \varepsilon < \eta$. To see why (5.22) leads to (5.21), define $C_{N\zeta} = N^{3\zeta} \exp(N^\zeta)$, let $M > 0$ and let

$$\chi_N = \text{Leb. meas.}\{\eta \in [\beta_1, \beta_2] : L(\eta) < M C_{N\zeta}\}.$$

By (5.22), $E_1 \chi_N \to 0$ and hence $P_1\{\chi_N < (\beta_2 - \beta_1)/2\} \to 1$. If $\chi_N < (\beta_2 - \beta_1)/2$, then $C_{N\zeta}^{-1} \int_{\beta_1}^{\beta_2} L(\eta)\, d\eta \geq M(\beta_2 - \beta_1)/2$. Since $M > 0$ can be chosen arbitrarily large, (5.21) holds.

To cross-reference the results in the proof of Theorem 1 more easily, we relabel j_T^* and ℓ_T^* in (5.20) by j_T and ℓ_T, respectively, and rephrase (5.22) as

$$(5.23) \qquad N^{-3\zeta} \exp(-N^\zeta) L(\beta) \xrightarrow{p} \infty,$$

when $\mu_N = b_N(\beta, \zeta)$ and $\pi_N = N^{-\beta+\varepsilon}$.

Case 1: $0 \leq \zeta \leq 1 - 4\beta/3$, $\mu_N = \sqrt{\log(1 + N^{2\beta-1+\zeta})}$, $\pi_N = N^{-\beta+\varepsilon}$ with $0 < \varepsilon < (1 - \zeta)/2$. Under H_1, $Y_{n1} = \mu_N \mathbf{I}_{\{I_n=1\}} + Z_{n1}$, where $Z_{11}, \ldots, Z_{N1}$ are i.i.d. N(0, 1). Let $L_{n1} = 1 + N^{-\beta}[\exp(\mu_N Z_{n1} - \mu_N^2/2) - 1]$, and

$$(5.24) \qquad L_1^0 = \prod_{n=1}^N L_{n1},$$

$$(5.25) \qquad L_1^1 = \prod_{n: I_n = 1} \left(\frac{1 + N^{-\beta}[\exp(\mu_N Y_{n1} - \mu_N^2/2) - 1]}{1 + N^{-\beta}[\exp(\mu_N Z_{n1} - \mu_N^2/2) - 1]} \right).$$

Since for $v \geq 0$,

$$(5.26) \qquad f(v) := \frac{1 + N^{-\beta}(v e^{\mu_N^2} - 1)}{1 + N^{-\beta}(v - 1)} \qquad \text{is increasing and } f(v) \geq 1,$$

it follows that

$$\log L_1^1 \geq \#\{n : I_n = 1, Y_{n1} \geq 2\mu_N\} \log\left(\frac{1 + N^{-\beta}(e^{3\mu_N^2/2} - 1)}{1 + N^{-\beta}(e^{\mu_N^2/2} - 1)} \right)$$

$$(5.27)$$

$$\overset{p}{\sim} N\pi_N \Phi(-\mu_N) \times \begin{cases} N^{\beta-1+\zeta}, & \text{if } 0 \leq \zeta < 1 - 2\beta, \\ N^{-\beta}\sqrt{2}, & \text{if } \zeta = 1 - 2\beta, \\ N^{-\beta} e^{3\mu_N^2/2}, & \text{if } 1 - 2\beta < \zeta < 1 - 4\beta/3, \\ \log 2, & \text{if } \zeta = 1 - 4\beta/3. \end{cases}$$

In the above, we apply the relation $\log[1 + N^{-\beta}(e^{\kappa\mu_N^2/2} - 1)] \sim N^{-\beta}(e^{\kappa\mu_N^2/2} - 1)$ for $\kappa = 1, 3$, with the exception

$$\log[1 + N^{-\beta}(e^{3\mu_N^2/2} - 1)] \sim \log 2 \qquad \text{when } \zeta = 1 - 4\beta/3.$$

Since $\Phi(-\mu_N) \sim \frac{1}{2}$ when $0 \le \zeta < 1 - 2\beta$, $\Phi(-\mu_N) = \Phi(-\sqrt{\log 2})$ when $\zeta = 1 - 2\beta$ and $\Phi(-\mu_N) \sim e^{-\mu_N^2/2}/(\mu_N\sqrt{2\pi})$ when $1 - 2\beta < \zeta \le 1 - 4\beta/3$, it follows from checking each of the cases in (5.27) that

$$(5.28) \qquad \frac{\log L_1^1}{N^{\zeta+\varepsilon}(\log N)^{-1}} \xrightarrow{p} \infty.$$

We shall next obtain lower bounds of $\log L_1^0$. We apply Taylor's expansion $\log(1 + u) = u - [\frac{1}{2} + o(1)]u^2$ to show that

$$(5.29) \qquad E_1(\log L_{n1}) \sim -N^{-2\beta}(e^{\mu_N^2} - 1)/2 = -N^{\zeta-1}/2,$$

and $\log(1 + u) \sim u$ to show that

$$(5.30) \qquad E_1(\log L_{n1})^2 \sim N^{-2\beta}(e^{\mu_N^2} - 1) = N^{\zeta-1}.$$

It follows from (5.29) and (5.30) that $\log L_1^0 = \sum_{n=1}^N \log L_{n1} \xrightarrow{p} -N^\zeta/2$. Since $L(\beta) = L_1^0 L_1^1$, we can conclude (5.23) from (5.28).

Case 2: $1 - 4\beta/3 < \zeta \le 1 - \beta$, $\mu_N = (x - y)\sqrt{2\log N}$ where $x = \sqrt{1 - \zeta}$, $y = \sqrt{1 - \zeta - \beta}$, $\pi_N = N^{-\beta+\varepsilon}$. Define

$$\widetilde{L}_{n1} = 1 + N^{-\beta}[\exp(\mu_N Z_{n1} - \mu_N^2/2) - 1]\mathbf{I}_{\{Z_{n1} \le x\sqrt{2\log N}\}},$$

$$\widetilde{L}_1^0 = \prod_{n=1}^N \widetilde{L}_{n1},$$

$$\widetilde{L}_1^1 = \prod_{n:I_n=1} \left(\frac{1 + N^{-\beta}[\exp(\mu_N Y_{n1} - \mu_N^2/2) - 1]\mathbf{I}_{\{Z_{n1} \le x\sqrt{2\log N}\}}}{1 + N^{-\beta}[\exp(\mu_N Z_{n1} - \mu_N^2/2) - 1]\mathbf{I}_{\{Z_{n1} \le x\sqrt{2\log N}\}}}\right).$$

By (5.26),

$$\log \widetilde{L}_1^1 \ge \#\{n : I_n = 1, x\sqrt{2\log N} \le Y_{n1} \le 2x\sqrt{2\log N}\}$$

$$(5.31) \qquad \times \log\left(\frac{1 + N^{-\beta}[\exp(\mu_N x\sqrt{2\log N} - \mu_N^2/2) - 1]}{1 + N^{-\beta}[\exp(\mu_N x\sqrt{2\log N} - 3\mu_N^2/2) - 1]}\right)$$

$$\sim N\pi_N \Phi(-y\sqrt{2\log N})\log 2 \sim C(\log 2)N^{\zeta+\varepsilon}/\sqrt{\log N},$$

where $C = (2y\sqrt{\pi})^{-1}$. Recall that $C_1 = [(2x - 4y)\sqrt{\pi}]^{-1}$.

Apply Taylor's expansion $\log(1 + u) = u - [\frac{1}{2} + o(1)]u^2$ to show that

$$
\begin{aligned}
E_1(\log \widetilde{L}_{n1}) \\
&= N^{-\beta}[\Phi(y\sqrt{2\log N}) - \Phi(x\sqrt{2\log N})] \\
&\quad - [1/2 + o(1)]N^{-2\beta}[e^{\mu_N^2}\Phi((2y - x)\sqrt{2\log N}) \\
&\qquad\qquad\qquad\qquad - 2\Phi(y\sqrt{2\log N}) + \Phi(x\sqrt{2\log N})] \\
&= -[1 + o(1)]CN^{-\beta-y^2}/\sqrt{\log N} - [1/2 + o(1)]C_1N^{-x^2}/\sqrt{\log N} \\
&= -[1 + o(1)](C + C_1/2)N^{\zeta-1}/\sqrt{\log N},
\end{aligned}
$$

(5.32)

and $\log(1 + u) \sim u$ to show that

$$
(5.33) \qquad\qquad E_1(\log \widetilde{L}_{n1})^2 \sim C_1 N^{\zeta-1}/\sqrt{\log N}.
$$

It follows from (5.32) and (5.33) that

$$
\log \widetilde{L}_1^0 = \sum_{n=1}^{N} \log \widetilde{L}_{n1} \overset{p}{\sim} -(C + C_1/2)N^{\zeta}/\sqrt{\log N}
$$

if $\zeta > 0$ and $|\log \widetilde{L}_1^0| = O_p(1)$ if $\zeta = 0$. Since $L(\beta) \geq \widetilde{L}_1^0 \widetilde{L}_1^1$, we can conclude (5.23) from (5.31).

Case 3: $\zeta > 1 - \beta$, $\mu_N = \sqrt{N^{\zeta-1+\beta}}$, $\pi_N = N^{-\beta+\varepsilon}$. The inequality

$$
L(\beta) \geq (1 - N^{-\beta})^N \prod_{n:I_n=1} \{1 + N^{-\beta}[\exp(\mu_N Y_{n1} - \mu_N^2/2) - 1]\}
$$

leads to

$$
\begin{aligned}
\log L(\beta) &\geq -2N^{1-\beta} + [1 + o_p(1)]N\pi_N[\mu_N E_1(Y_{n1}) - \mu_N^2/2 - \beta\log N] \\
&= -2N^{1-\beta} + [1 + o_p(1)]N^{\zeta+\varepsilon}/2 \overset{p}{\sim} N^{\zeta+\varepsilon}/2,
\end{aligned}
$$

and from this, we can conclude (5.23).

5.3. *Proof of Theorem* 5. Let

$$
E_1(Y_{\ell_T j_T}) = b_T(\ell_T) + c_T \qquad \text{with } c_T \to \infty,
$$

and let $(j_T^*, j_T^* + \ell_T^*) \in B_{r,T}$ be such that (2.16) holds. Hence

$$
(5.34) \quad E_1(Y_{\ell_T^* j_T^*}) = [b_T(\ell_T) + c_T]\sqrt{\ell_T^*/\ell_T} = [1 - O(r^{-1/2})][b_T(\ell_T) + c_T].
$$

Since $b_T(\ell_T^*) = \sqrt{2\log(eT/\ell_T^*)} = \sqrt{b_T^2(\ell_T) + O(1)} = b_T(\ell_T) + O(1)$ and $r^{-1/2}b_T(\ell_T) = O(b_T(\ell_T)/\sqrt{\log(T/\ell_T)}) = O(1)$, it follows from (5.34) that

$$
(5.35) \qquad\qquad E_1(Y_{\ell_T^* j_T^*}) = b_T(\ell_T^*) + c_T' \qquad \text{with } c_T' \to \infty.
$$

We check that under (5.35),

$$\frac{\exp[b_T(\ell_T^*)Y_{\ell_T^* j_T^*} - b_T^2(\ell_T^*)/2]}{[\log(T/\ell_T^*)]^3 (T/\ell_T^*)} \xrightarrow{P} \infty.$$

Since $r^3 e^{r+1} = O([\log(T/\ell_T)]^3 (T/\ell_T))$, it follows from (3.4) that $A_T \xrightarrow{P} \infty$.

6. Proofs of Theorems 2 and 3. We shall prove Theorem 3 in Section 6.1, that the penalized Berk–Jones test is optimal, and Theorem 2 in Section 6.2, that the penalized higher criticism test is optimal as well.

6.1. *Proof of Theorem* 3. In Lemma 1 below, we show that the Type I error probability of the penalized Berk–Jones test statistic goes to zero for the threshold $h_N := 2 \log N$. We do this in more generality than is required for proving Theorem 3. For each ℓ and j, we assume only that $p_{1\ell j}, \ldots, p_{N\ell j}$ are i.i.d. Uniform$(0, 1)$ random variables under H_0. Hence we allow for X_{nt} to be non-Gaussian, and for $X_{n1}, \ldots, X_{nT}$ to be dependent random variables under H_0.

LEMMA 1. *Assume that for each ℓ and j, $p_{(1)\ell j} \leq \cdots \leq p_{(N)\ell j}$ in* (2.18) *and* (2.19) *are the ordered values of i.i.d.* Uniform$(0, 1)$ *random variables. Then*

$$(6.1) \qquad P_0\{\mathrm{PBJ}_{NT} \geq h_N\} \to 0 \qquad \text{as } N \to \infty.$$

PROOF. Let $a_{N\ell} = h_N + s_{\ell T} \log s_{\ell T}$. For each $1 \leq \ell \leq T$ and $1 \leq n \leq N$, let ρ_ℓ be such that $K(\frac{n}{N}, \rho_\ell) = a_{N\ell}/N$. Let $\bar{S}_N(t) \,[= \bar{S}_{N\ell j}(t)] = N^{-1} \sum_{n=1}^{N} \mathbf{I}_{\{Y_{n\ell j} \geq z(t)\}}$, where $z(t)$ denotes the upper t-quantile of the standard normal. By the Chernoff–Hoeffding inequality,

$$P_0\{K(n/N, p_{(n)}) \geq a_{N\ell}/N\}$$
$$(= P_0\{\bar{S}_N(\rho_\ell) \geq n/N\}) \leq e^{-a_{N\ell}} = N^{-2} e^{-s_{\ell T} \log s_{\ell T}},$$

and hence by Bonferroni's inequality,

$$(6.2) \qquad P_0\{\mathrm{BJ}_{N\ell j} - s_{\ell T} \log s_{\ell T} \geq h_N\} \leq N^{-1} e^{-s_{\ell T} \log s_{\ell T}}.$$

By (2.15), $\#B_{r,T} \leq re^{r+1}$. Since $\ell \leq T/e^{r-1}$ for $(j, j + \ell) \in B_{r,T}$, so $s_{\ell T} = \log(eT/\ell) \geq r$, and by (2.19) and (6.2),

$$(6.3) \qquad P_0\{\mathrm{PBJ}_{NT} \geq h_N\} \leq N^{-1} \sum_{r=1}^{\infty} re^{r+1-r\log r},$$

and (6.1) holds. $\square$

PROOF OF THEOREM 3. Let $(j^*, j^* + \ell^*) \,[= (j_T^*, j_T^* + \ell_T^*)] \in B_{r,T}$ be such that $j_T \leq j^* < j^* + \ell^* \leq j_T + \ell_T$ and $1 - \ell^*/\ell_T = O(r^{-1/2})$; see (2.16). Since

$a_{N\ell^*} = O(N^\zeta \log N)$, in view of Lemma 1, it remains for us to show that if $\mu_N = b_N(\beta, \zeta)$ and $\pi_N = N^{-\beta+\varepsilon}$ for $\varepsilon > 0$, then we can find t_N such that in each case below,

$$(6.4) \qquad \frac{K(\bar{S}_N, t_N)}{N^{\zeta-1}\log N} \to \infty \qquad \text{where } \bar{S}_N = \bar{S}_{N\ell^* j^*}(t_N).$$

Case 1(a): $0 \le \zeta \le 1 - 2\beta$, $b_N(\beta, \zeta) = \sqrt{\log(1 + N^{2\beta-1+\zeta})}$. Let $t_N = \Phi(-2b_N(\beta, \zeta))$. Except when $\zeta = 1 - 2\beta$, we have $b_N(\beta, \zeta) \to 0$ and

$$(6.5) \qquad E_1\bar{S}_N - t_N \sim (2\pi)^{-1/2}\pi_N b_N(\beta, \zeta) \sim (2\pi)^{-1/2} N^{(\zeta-1)/2+\varepsilon}.$$

By Taylor's expansion, $K(t, x) \sim 2(t - x)^2$ when $t \to \frac{1}{2}$ and $x \to \frac{1}{2}$. Moreover, the standard error of $\bar{S}_N$ $[\sim (4N)^{-1/2}]$ is small relative to (6.5). Hence (6.4) holds because

$$(6.6) \qquad K(\bar{S}_N, t_N) \overset{p}{\sim} 2(\bar{S}_N - t_N)^2 \overset{p}{\sim} \pi^{-1} N^{\zeta-1+2\varepsilon}.$$

When $\zeta = 1 - 2\beta$, $b_N(\beta, \zeta) = \sqrt{\log 2}$ and

$$E_1\bar{S}_N - t_N \sim \widetilde{C}\pi_N \qquad \text{where } \widetilde{C} = \Phi(-2\sqrt{\log 2}) - \Phi(-\sqrt{\log 2}).$$

This leads to (6.6) with π^{-1} replaced by $2\widetilde{C}^2$, and then to (6.4).

Case 1(b): $1 - 2\beta < \zeta < 1 - 4\beta/3$, $b_N(\beta, \zeta) = \sqrt{\log(1 + N^{2\beta-1+\zeta})}$ $(\sim C\sqrt{\log N}$, where $C = \sqrt{2\beta - 1 + \zeta})$. Let $t_N = \Phi(-2b_N(\beta, \zeta))$. Then

$$(6.7) \qquad t_N \sim (C\sqrt{8\pi \log N})^{-1} N^{-4\beta-2(\zeta-1)},$$

$$(6.8) \qquad \begin{aligned} E_1\bar{S}_N - t_N &\sim \pi_N \Phi(-b_N(\beta, \zeta)) \\ &\sim (C\sqrt{2\pi \log N})^{-1} N^{-2\beta-(\zeta-1)/2+\varepsilon}, \end{aligned}$$

$$(6.9) \qquad \text{Var}_1\,\bar{S}_N \sim N^{-1}\big[t_N + \pi_N \Phi(-b_N(\beta, \zeta))\big].$$

We claim that a consequence of (6.7)–(6.9) is that

$$(6.10) \qquad \sqrt{t_N N^{\zeta-1}\log N} + \sqrt{\text{Var}_1\,\bar{S}_N} = o(E_1\bar{S}_N - t_N).$$

By (6.10), $\sqrt{t_N N^{\zeta-1}\log N} = o_p(|\bar{S}_N - t_N|)$ and hence

$$(6.11) \qquad \frac{(\bar{S}_N - t_N)^2/(2t_N)}{N^{\zeta-1}\log N} \overset{p}{\to} \infty.$$

By (6.7), the solution in y of $y^2/(2t_N) = a_{N\ell^*}/N$ satisfies

$$y \sim \left(\frac{\zeta^*}{C}\sqrt{\frac{\log N}{2\pi}}\right)^{1/2} N^{-2\beta-(\zeta-1)/2} \qquad [= o(t_N) \text{ because } \zeta < 1 - 4\beta/3],$$

where $\zeta^* = \zeta$ if $\zeta \neq 0$ and $\zeta^* = 2$ if $\zeta = 0$. Hence by (6.11) and $K(x, t) \sim \frac{(x-t)^2}{2t}$, as $t \to 0$ and $\frac{x}{t} \to 1$, (6.4) holds.

It remains for us to show (6.10). By (6.7), the exponent of N in $\sqrt{t_N N^{\zeta-1}}$ is $-2\beta - (\zeta - 1)/2$, which is smaller than the exponent in N of $E_1 \bar{S}_N - t_N$; see (6.8). Therefore,

$$(6.12) \qquad \sqrt{t_N N^{\zeta-1} \log N} = o(E_1 \bar{S}_N - t_N).$$

The leading exponent of N in $\mathrm{Var}_1 \bar{S}_N$ is

$$\max(-4\beta - 2\zeta + 1, -2\beta - (\zeta + 1)/2 + \varepsilon) \qquad [< -4\beta - (\zeta - 1) + 2\varepsilon],$$

and therefore by (6.8), $\mathrm{Var}_1 S_N = o((E_1 \bar{S}_N - t_N)^2)$. This, together with (6.12), implies (6.10).

Case 2: $1 - \frac{4}{3}\beta < \zeta \leq 1 - \beta$, $b_N(\beta, \zeta) = (x - y)\sqrt{2 \log N}$, where $x = \sqrt{1 - \zeta}$, $y = \sqrt{1 - \beta - \zeta}$. Let $t_N = \Phi(-x\sqrt{2 \log N})$ $[\sim (2x\sqrt{\pi \log N})^{-1} N^{\zeta-1}]$. Then

$$E_1 \bar{S}_N \sim (1 - \pi_N)t_N + \pi_N \Phi(-y\sqrt{2 \log N}) \sim (2y\sqrt{\pi \log N})^{-1} N^{\zeta-1+\varepsilon},$$

which is large relative to t_N, and

$$\mathrm{Var}_1 \bar{S}_N \sim N^{-1}\big[\Phi(-x\sqrt{2 \log N}) + \pi_N \Phi(-y\sqrt{2 \log N})\big]$$

$$\sim (2y\sqrt{\pi \log N})^{-1} N^{\zeta-2+\varepsilon} = o((E_1 \bar{S}_N)^2).$$

Therefore $\bar{S}_N \overset{p}{\sim} (2y\sqrt{\log N})^{-1} N^{\zeta-1+\varepsilon}$, and by $K(x, t) \sim x \log \frac{x}{t}$, as $x \to 0$ and $\frac{x}{t} \to \infty$,

$$K(\bar{S}_N, t_N) \overset{p}{\sim} \bar{S}_N \log(\bar{S}_N/t_N) \overset{p}{\sim} C'(\log N)^{1/2} N^{\zeta-1+\varepsilon},$$

for some $C' > 0$, and (6.4) therefore holds.

Case 3: $\zeta > 1 - \beta$, $b_N(\beta, \zeta) = \sqrt{N^{\beta+\zeta-1}}$. Let $t_N = \Phi(-b_N(\beta, \zeta)/2)$. Then

$$t_N \overset{p}{\sim} (\pi N^{\beta+\zeta-1}/2)^{-1/2} \exp(-N^{\beta+\zeta-1}/8) \quad \text{and} \quad \bar{S}_N \overset{p}{\sim} \pi_N = N^{-\beta+\varepsilon}.$$

Therefore $\bar{S}_N/t_N \overset{p}{\to} \infty$ and by $K(x, t) \sim x \log \frac{x}{t}$, as $x \to 0$ and $\frac{x}{t} \to \infty$,

$$K(\bar{S}_N, t_N) \overset{p}{\sim} \bar{S}_N \log(\bar{S}_N/t_N) \overset{p}{\sim} N^{\zeta-1+\varepsilon}/8,$$

and (6.4) therefore holds. □

6.2. *Proof of Theorem* 2. In Lemma 2 below, we show that the Type I error probability of the penalized higher criticism test statistic goes to zero for the threshold $h_N = 2 \log N$. Again as in Lemma 1, we do this more generally than is required for proving Theorem 2.

LEMMA 2. *Assume that for each ℓ and j, $p_{(1)\ell j} \leq \cdots \leq p_{(N)\ell j}$ in (2.13) and (2.14) are the ordered values of i.i.d. Uniform(0, 1) random variables. Then*

$$P_0\{\text{PHC}_{NT} \geq h_N\} \to 0 \qquad \text{as } N \to \infty.$$

PROOF. We first modify (6.2)–(6.3), in the proof of Lemma 1, step-by-step to show that for $c_{N\ell} := 2\log N + s_{\ell T} + 3\log s_{\ell T}$,

$$P_0\{\text{BJ}_{N\ell j} \geq c_{N\ell} \text{ for some } (j, j + \ell) \in B_T\} \to 0.$$

We then combine this with (B.2) in Appendix B to show that

(6.13)
$$P_0\{\text{HC}_{N\ell j} \geq (6c_{N\ell})^{1/2}(4 + s_{\ell T}^{-1} \log N)^{1/4} \text{ for some } (j, j + \ell) \in B_T\}$$
$$\to 0.$$

Therefore it suffices to show that

(6.14) $\quad (6c_{N\ell})^{1/2}(4 + s_{\ell T}^{-1} \log N)^{1/4} \leq h_N + (s_{\ell T} \log s_{\ell T})^{1/2} \qquad$ for N large,

uniformly over $1 \leq \ell \leq T$. The left-hand side of (6.14) is $o(h_N)$ uniformly over $s_{\ell T} \leq \log N$, and $o((s_{\ell T} \log s_{\ell T})^{1/2})$ uniformly over $s_{\ell T} > \log N$, so (6.14) indeed holds. $\square$

PROOF OF THEOREM 2. We apply the proofs for cases 1 and 2 in Theorem 3 to show that there exist $(j^*, j^* + \ell^*) \in B_T$ and $t_N^* \in [\frac{s_{\ell^* T}}{N}, \frac{1}{2}]$ [$t_N^* = t_N$ for case 1 and $t_N^* = \frac{s_{\ell^* T}}{N}$ ($\sim N^{\zeta-1}$) for case 2] such that

$$P_1\{K(\bar{S}_{N\ell^* j^*}(t_N^*), t_N^*) \geq N^{\zeta-1+\varepsilon/2}\} \to 1.$$

Hence by $\frac{x-t}{\sqrt{t(1-t)}} \geq \sqrt{2K(x, t)}$ for $x \geq t$,

$$P_1\{\text{PHC}_{NT} \geq \sqrt{2N^{\zeta+\varepsilon/2}} - \sqrt{s_{\ell^* T} \log s_{\ell^* T}}\} \to 1.$$

Since $h_N + \sqrt{s_{\ell^* T} \log s_{\ell^* T}} = o(\sqrt{N^{\zeta+\varepsilon/2}})$, the Type II error probability indeed goes to zero. $\square$

7. Proofs of Theorem 6 and Corollary 2.

We prove Theorem 6 here and in Sections 7.1 and 7.2. In Section 7.3, we prove Corollary 2. Let $\mu_N = b_N(\beta, \zeta, \tau)$, $Y_{n1} = \ell_T^{-1/2} \sum_{\ell=1}^{\ell_T} X_{n, j_T+\ell}$, and for $2 \leq i \leq i_T$ ($= \lfloor T/\ell_T \rfloor - 1$), let $Y_{ni} = \ell_T^{-1/2} \sum_{\ell=1}^{\ell_T} X_{n, j+\ell}$, with all $(j, j + \ell_T]$ disjoint from each other, and from $(j_T, j_T + \ell_T]$. Let $L_i = \prod_{n=1}^N L_{ni}$, where

(7.1) $\qquad L_{ni} = 1 + \pi_N\left\{\frac{1}{\sqrt{1+\tau}} \exp\left[-\frac{(Y_{ni} - \mu_N)^2}{2(1+\tau)} + \frac{Y_{ni}^2}{2}\right] - 1\right\}$

is the likelihood ratio of $Y_{ni} \sim (1 - \pi_N)\mathrm{N}(0, 1) + \pi_N\mathrm{N}(\mu_N, 1 + \tau)$ and $Y_{ni} \sim \mathrm{N}(0, 1)$. Below, we go over the relevant cases to show that there exists λ_N satisfying (5.2) and (5.3) when $\pi_N = N^{-\beta-\varepsilon}$. This implies that there is no test able to achieve (4.3). We shall only consider $\tau > 0$ as the case $\tau = 0$ has been covered in Theorem 1.

Case 1: $1 - 2\beta < \zeta \leq 1 - \frac{4\beta}{3-\tau}$ ($\Rightarrow \tau < 1$), $\mu_N = \sqrt{(1 - \tau)(2\beta + \zeta - 1)\log N}$. Let $C = (1 - \tau^2)^{-1/2}$. By (7.1),

$$E_1(L_{n1}) = 1 + \pi_N^2(Ce^{\mu_N^2/(1-\tau)} - 1) = 1 + [C + o(1)]N^{\zeta-1-2\varepsilon},$$

and therefore (5.2) holds with $\lambda_N = E_1(L_1)$ ($= \exp\{[C + o(1)]N^{\zeta-2\varepsilon}\}$). For $i \geq 2$, $Y_{ni} \sim \mathrm{N}(0, 1)$, $E_0(L_i) = 1$ and $E_0(L_i^2) = E_1(L_1) = \lambda_N$. We check Lyapunov's conditions to conclude (5.6) and (5.3).

Case 2: $1 - \min(2\beta, \frac{4\beta}{3-\tau}) < \zeta \leq 1 - \beta$, $\tau < \frac{\beta}{1-\zeta-\beta}$, $\mu_N = (x - y)\sqrt{2\log N}$ where $x = \sqrt{1 - \zeta}$ and $y = \sqrt{(1 + \tau)(1 - \zeta - \beta)}$. Let

$$\mathcal{N}^0 = \{n : I_n = 0 \text{ or } |Y_{n1}| < x\sqrt{2\log N}\},$$

$$\mathcal{N}^1 = \{n : I_n = 1 \text{ and } |Y_{n1}| \geq x\sqrt{2\log N}\},$$

$$L_1^h = \prod_{n\in\mathcal{N}^h} L_{n1}, \qquad h = 0, 1.$$

Check that $m_0 := E_1(L_{n1}|I_n = 0) = 1$ and

$$
\begin{aligned}
(7.2) \quad m_1 &:= E_1\big[1 + (L_{n1} - 1)\mathbf{I}_{\{|Y_{n1}|\leq x\sqrt{2\log N}\}}|I_n = 1\big] \\
&= 1 + [C_1 + o(1)]\pi_N N^{\zeta-1+2\beta}/\sqrt{\log N}
\end{aligned}
$$

for some $C_1 > 0$, hence

$$
\begin{aligned}
E_1(L_1^0) &= \big[(1 - \pi_N)m_0 + \pi_N m_1\big]^N \\
(7.3) \quad &= \{1 + [C_1 + o(1)]\pi_N^2 N^{\zeta-1+2\beta}/\sqrt{\log N}\}^N \\
&= \exp\{[C_1 + o(1)]N^{\zeta-2\varepsilon}/\sqrt{\log N}\}.
\end{aligned}
$$

Next, we apply $\max_{n\in\mathcal{N}^1}|Y_{n1}| = O_p(\sqrt{\log N})$ to show that

$$
\begin{aligned}
(7.4) \quad \log L_1^1 &= O_p\big((\#\mathcal{N}^1)\sqrt{\log N}\big) \\
&= O_p\big(N\pi_N\Phi(-\sqrt{2(1 - \zeta - \beta)\log N})\sqrt{\log N}\big) = O_p(N^{\zeta-\varepsilon}).
\end{aligned}
$$

By (7.3), (7.4) and $L_1 = L_1^0 L_1^1$, (5.2) holds for $\lambda_N = \exp(N^{\zeta-\varepsilon}\log N)$. For $i \geq 2$, let

$$\widetilde{L}_{ni} = L_{ni}\mathbf{I}_{\{|Y_{ni}|\leq x\sqrt{2\log N}\}}, \qquad \widetilde{L}_i = \prod_{n=1}^N \widetilde{L}_{ni}.$$

Then by a change-of-measure argument,

$$E_0(\widetilde{L}_{ni}) = 1 + \pi_N\big[P_1\{|Y_{n1}| \leq x\sqrt{2\log N}\} - P_0\{|Y_{n1}| \leq x\sqrt{2\log N}\}\big]$$

$$= 1 - [C_2 + o(1)]N^{-\zeta-1-\varepsilon}/\sqrt{\log N},$$

where $C_2 = \sqrt{1+\tau}/(2y\sqrt{\pi})$, therefore

$$\kappa_N := E_0(\widetilde{L}_i) = \exp\{-[C_2 + o(1)]N^{\zeta-\varepsilon}/\sqrt{\log N}\}.$$

Moreover, $E_0(\widetilde{L}_{ni}^2) \geq [E_0(\widetilde{L}_{ni})]^2$ for $i \geq 2$, and therefore, by (5.12),

$$\mathrm{Var}_0(\widetilde{L}_i) \geq [1 + o_p(1)]\kappa_N^2 \, \mathrm{Var}_0(\widetilde{L}_{1i}).$$

By (7.2) and a change-of-measure argument,

$$\mathrm{Var}_0(\widetilde{L}_{1i}) \sim E_0(\widetilde{L}_{1i}^2) = E_1(\widetilde{L}_{11}) \sim C_1 N^{\zeta-1-2\varepsilon}/\sqrt{\log N}.$$

We check Lyapunov's condition to conclude (5.14) and (5.3).

7.1. *Optimal detection using the penalized BJ test.* By Lemma 1, setting $h_N = 2\log N$ leads to $P(\text{Type I error}) \to 0$. To show $P(\text{Type II error}) \to 0$, it suffices to find $(j^*, j^* + \ell^*) \in B_{r,T}$ such that $j_T \leq j^* < j^* + \ell^* \leq j_T + \ell_T$, $1 - \ell^*/\ell_T = O(r^{-1/2})$ and

$$(7.5) \qquad P_1\{K(\bar{S}_N, t_N) \geq N^{\zeta-1+\delta}\} \to 1$$

for some $0 < t_N < 1$ and $\delta > 0$, where $\bar{S}_N = N^{-1}\sum_{n=1}^N \mathbf{I}_{\{Y_{n\ell^* j^*} \geq z(t_N)\}}$ and $\pi_N = N^{-\beta+\varepsilon}$, $0 < \varepsilon < \beta$.

Case 1(a): $0 \leq \zeta \leq 1 - 2\beta$, $\mu_N = 0$. Let $t_N = \Phi(-N^{-\varepsilon/2})$. Then

$$E_1\bar{S}_N - t_N = \pi_N\big[\Phi\big(-N^{-\varepsilon/2}/\sqrt{1+\tau}\big) - \Phi\big(-N^{-\varepsilon/2}\big)\big] \sim C_3 N^{-\beta+\varepsilon/2},$$

where $C_3 = \frac{1}{\sqrt{2\pi}}(1 - \frac{1}{\sqrt{1+\tau}})$, and since

$$\mathrm{Var}_1(\bar{S}_N) \to (4N)^{-1} = o\big((E_1\bar{S}_N - t_N)^2\big),$$

therefore $\bar{S}_N - t_N \overset{p}{\to} C_3 N^{-\beta+\varepsilon/2}$. Since $K(t,x) \sim 2(t-x)^2$ when $t \to \frac{1}{2}$ and $x \to \frac{1}{2}$, (7.5) holds for $\delta = \varepsilon/2$.

Case 1(b): $1 - 2\beta < \zeta < 1 - \frac{4\beta}{3-\tau}$ ($\Rightarrow \tau < 1$), $\mu_N = C_4\sqrt{\log N}$, where $C_4 = \sqrt{(1-\tau)(2\beta+\zeta-1)}$. Let $t_N = \Phi(-2\mu_N/(1-\tau))$. Then

$$(7.6) \qquad t_N \sim C_5 N^{-2C_4^2/(1-\tau)^2}/\sqrt{\log N},$$

where $C_5 = (1-\tau)/(C_4\sqrt{8\pi})$, and

$$(7.7) \qquad \begin{aligned} E_1\bar{S}_N - t_N &= \pi_N\left[\Phi\left(-\frac{\mu_N\sqrt{1+\tau}}{1-\tau}\right) - t_N\right] \\ &\sim C_6 N^{-C_4^2(1+\tau)/[2(1-\tau)]^2-\beta+\varepsilon}/\sqrt{\log N}, \end{aligned}$$

where $C_6 = (1 - \tau)/(C_4 \sqrt{2\pi(1 + \tau)})$.

We claim that

$$(7.8) \qquad \operatorname{Var}_1 \bar{S}_N \left(\sim N^{-1} \left[t_N + \pi_N \Phi \left(-\frac{\mu_N \sqrt{1 + \tau}}{1 - \tau} \right) \right] \right) = o((E_1 \bar{S}_N)^2).$$

By (7.6)–(7.8) and $t_N = o(E_1 \bar{S}_N)$,

$$(\bar{S}_N - t_N)^2/(2t_N) \overset{p}{\sim} C_7 N^{C_4^2/(1-\tau)-2\beta+2\varepsilon}/\sqrt{\log N}$$

$$(7.9)$$

$$= C_7 N^{\zeta - 1 + 2\varepsilon}/\sqrt{\log N}$$

for some $C_7 > 0$. Check that the inequality $-\frac{2C_4^2}{(1-\tau)^2} > \zeta - 1$ reduces to $\zeta < 1 - \frac{4\beta}{3-\tau}$. Therefore by (7.6), $t_N \sim C_5 N^{\zeta-1+2\delta}/\sqrt{\log N}$ for some $\delta > 0$, and the root of $y^2/(2t_N) = N^{\zeta-1+\delta}$ satisfies $y = o(t_N)$. Since $K(x, y) \sim \frac{(x-t)^2}{2t}$ as $t \to 0$ and $\frac{x}{t} \to 1$, (7.9) implies (7.5).

It remains to show (7.8) by comparing the leading exponent in N of the terms. That is, it remains to show that

$$-1 + \max\left(-\frac{2C_4^2}{(1-\tau)^2}, -\beta + \varepsilon - \frac{C^2(1+\tau)}{2(1-\tau)^2} \right) < -\frac{C_4^2(1+\tau)}{(1-\tau)^2} - 2\beta + 2\varepsilon,$$

summarized as $-1 + \max(A, B) < D$. The inequality $-1 + A < D$ reduces to $\zeta > -2\varepsilon$, which holds trivially, whereas $-1 + B < D$ reduces to

$$\zeta < \frac{3 - \tau}{1 + \tau} \left(1 - \frac{4\beta}{3 - \tau} \right) + \frac{2(1 - \tau)}{1 + \tau} \varepsilon,$$

which holds because $\frac{3-\tau}{1+\tau} > 1$ when $\tau < 1$, and it is assumed that $\zeta < 1 - \frac{4\beta}{3-\tau}$. Hence (7.8) holds.

Case 2: $1 - \min(2\beta, \frac{4\beta}{3-\tau}) \le \zeta < 1 - \beta$, $\tau < \frac{\beta}{1-\zeta-\beta}$, $\mu_N = (x - y)\sqrt{2 \log N}$ where $x = \sqrt{1 - \zeta}$ and $y = \sqrt{(1 + \tau)(1 - \zeta - \beta)}$. Let $t_N = \Phi(-x\sqrt{2 \log N})$ $[\sim N^{\zeta-1}/(2x\sqrt{\pi \log N})]$. Then

$$E_1 \bar{S}_N = (1 - \pi_N)t_N + \pi_N \Phi(-\sqrt{2(1 - \zeta - \beta) \log N})$$

$$\sim C_8 N^{\zeta-1+\varepsilon}/\sqrt{\log N} \qquad (\gg t_N),$$

where $C_8 = (2\sqrt{\pi(1 - \zeta - \beta)})^{-1}$. Moreover,

$$\operatorname{Var}_1 \bar{S}_N \sim E_1 \bar{S}_N/N \sim C_8 N^{\zeta-2+\varepsilon}/\sqrt{\log N} = o((E_1 \bar{S}_N)^2).$$

Therefore $\bar{S}_N \overset{p}{\sim} C_8 N^{\zeta-1+\varepsilon}/\sqrt{\log N}$, and since $K(x, t) \sim x \log(\frac{x}{t})$ as $x \to 0$ and $\frac{x}{t} \to \infty$,

$$(7.10) \qquad K(\bar{S}_N, t_N) \overset{p}{\sim} \bar{S}_N \log(\bar{S}_N/t_N) \overset{p}{\sim} C_9 N^{\zeta-1+\varepsilon}\sqrt{\log N}$$

for some $C_9 > 0$, (7.5) holds for $\delta = \varepsilon$. By similar arguments, (7.5) holds for $\delta < \varepsilon$ when $t_N \sim N^{\zeta-1}$.

Case 3: $1 - \min(2\beta, \frac{4\beta}{3-\tau}) \leq \zeta < 1 - \beta$, $\tau \geq \frac{\beta}{1-\zeta-\beta}$, $\mu_N = 0$. Let $t_N = \Phi(-\sqrt{2(1-\zeta)\log N})\ [\sim N^{\zeta-1}/(2x\sqrt{\pi \log N})]$. Then

$$(7.11) \qquad E_1 \bar{S}_N \sim \pi_N \Phi\left(-\sqrt{\frac{2(1-\zeta)\log N}{1+\tau}}\right) \sim C_{10} N^{-\beta+\varepsilon-(1-\zeta)/(1+\tau)}$$

for some $C_{10} > 0$. Since $\tau \geq \frac{\beta}{1-\zeta-\beta}$, therefore the exponent of N in (7.11) is at least $\zeta - 1 + \varepsilon$, and so $E_1 \bar{S}_N \gg t_N$. We apply the first relation in (7.10) to conclude (7.5), for both $t_N = \Phi(-\sqrt{2(1-\zeta)\log N})$ and $t_N \sim N^{\zeta-1}$.

7.2. *Optimal detection using the penalized HC test.* By Lemma 2, setting $h_N = 2\log N$ leads to $P(\text{Type I error}) \to 0$. Let

$$t_N = \begin{cases} \Phi(-N^{-\varepsilon/2}), & \text{if } \zeta \leq 1 - 2\beta, \\ \Phi\left(-2\sqrt{\frac{2\beta-1+\zeta}{1-\tau}\log N}\right), & \text{if } 1 - 2\beta < \zeta \leq 1 - \frac{4\beta}{3-\tau}, \\ \frac{s_{\ell^*T}}{N}\ (\sim N^{\zeta-1}), & \text{if } 1 - \min\left(2\beta, \frac{4\beta}{3-\tau}\right) < \zeta \leq 1 - \beta. \end{cases}$$

It was shown in (7.5) that in each case above, $P_1\{K(\bar{S}_N, t_N) \geq N^{\zeta-1+\delta}\} \to 1$ for some $\delta > 0$. Since $\frac{x-t}{\sqrt{t(1-t)}} \geq \sqrt{2K(x,t)}$ for $x \geq t$ and $h_N + \sqrt{s_{\ell^*T}\log s_{\ell^*T}} = o(N^{(\zeta+\delta)/2})$, $P(\text{Type II error}) \to 0$.

7.3. *Proof of Corollary* 2. Consider first the penalized HC test. By Lemma 2, setting $h_N = 2\log N$ leads to $P(\text{Type I error}) \to 0$. In the case $\pi_N = N^{-\beta+\varepsilon}$, the arguments above and in Theorem 2 show that

$$P_1\left\{\frac{\bar{S}_N - t_N}{\sqrt{t_N(1-t_N)/N}} \geq \sqrt{2N^\delta \log(T/\ell^*)}\right\} \to 1 \qquad \text{for some } \delta > 0.$$

By (4.4), $h_N + \sqrt{s_{\ell^*T}\log s_{\ell^*T}} = o(\sqrt{N^\delta})$, and therefore $P(\text{Type II error}) \to 0$, and (4.3) holds. By similar arguments, (4.3) holds for the penalized BJ test.

APPENDIX A: VERIFICATION OF LYAPUNOV'S CONDITION

We check in particular Lyapunov's condition to conclude (5.6). Let $\delta > 0$ to be specified. It follows from Taylor's expansion that

$$(A.1) \qquad (1+u)^{2+\delta} \leq 1 + (2+\delta)u + C^* u^2, \qquad |u| \leq 1/2,$$

for some $C^* > 0$ chosen large enough. If $u > \frac{1}{2}$, then

$$(A.2) \qquad (1+u)^{2+\delta} \leq \left[\sup_{v>1/2}\left(\frac{1+v}{v}\right)^{2+\delta}\right]u^{2+\delta} = (3u)^{2+\delta}.$$

By combining (A.1) and (A.2), we conclude that

(A.3) $(1 + u)^{2+\delta} \leq 1 + (2 + \delta)u + C^* u^2 + C_2 |u|^{2+\delta}$ for all $u \geq -1/2$,

where $C_2 = 3^{2+\delta}$. We apply (A.3) with $u = \pi_N[\exp(\mu_N Y_{n2} - \mu_N^2/2) - 1]$ on (5.1) to show that

$$
\begin{aligned}
\text{(A.4)} \quad & E_0(L_2^{2+\delta}) \\
& \leq [1 + C^* p_N^2 \exp(\mu_N^2) + C_2 \pi_N^{2+\delta} \exp(\mu_N^2(1 + \delta)(2 + \delta)/2)]^N.
\end{aligned}
$$

Since $\mu_N = \sqrt{\log(1 + N^{2\beta - 1 + \zeta})}$ and $\pi_N = N^{-\beta - \varepsilon}$, by (A.4),

(A.5) $\quad E_0(L_2^{2+\delta}) \leq \exp\{[1 + o(1)](C^* N^{\zeta - 2\varepsilon} + C_2 N^{\zeta - 2\varepsilon + \kappa})\},$

where $\kappa = -\delta(\beta + \varepsilon) + \frac{3\delta + \delta^2}{2}(2\beta - 1 + \zeta)$. Let $\delta > 0$ be small enough such that $\kappa < \varepsilon$. Since $i_T - 1 \sim \exp(N^\zeta - 1)$ and $L_i - 1 \geq -1$, we get from (A.5) that

$$(i_T - 1)E_0(L_i - 1)^{2+\delta} \leq \exp\{[1 + o(1)](N^\zeta + C_2 N^{\zeta - \varepsilon})\}.$$

On the other hand, $\mathrm{Var}_0(L_i) = \lambda_N - 1$ and

$$\big[(i_T - 1)(\lambda_N - 1)\big]^{1+\delta/2} = \exp\{[1 + \delta/2 + o(1)](N^\zeta + N^{\zeta - 2\varepsilon})\},$$

so Lyapunov's condition is satisfied.

APPENDIX B: QUADRATIC BOUNDS FOR THE FUNCTION K

For given ℓ, T, N, $\frac{s_{\ell T}}{N} \leq t \leq \frac{1}{2}$ and $0 < \gamma \leq \frac{c_{N\ell}}{N}$ (recall that $c_{N\ell} = \log N + s_{\ell T} + 3 \log s_{\ell T}$), let $x_t > t$ be such that

(B.1) $$Q(x_t, t) \left[:= \frac{(x_t - t)^2}{2t(1 - t)} \right] = \gamma.$$

We claim that

(B.2) $$K(x_t, t) \leq Q(x_t, t) \leq \left(3\sqrt{4 + s_{\ell T}^{-1} \log N} \right) K(x_t, t).$$

The left inequality of (B.2) is known. To obtain the right inequality, first note that by (B.1),

$$x_t = t(1 + y) \qquad \text{where } y = \sqrt{2\gamma(1 - t)/t}.$$

Since $\frac{d}{dx}[(1 - x) \log(\frac{1-x}{1-t})] = -1 - \log(\frac{1-x}{1-t}) \geq -1$ for $x \geq t$, therefore

(B.3) $$K(x_t, t) \geq x_t \log(x_t/t) - (x_t - t) = t[f(y)],$$

where $f(y) = (1 + y) \log(1 + y) - y$. Check that $\frac{d}{dy} f(y) = \log(1 + y)$, and that $\frac{d^2}{dy^2} f(y) = \frac{1}{1+y}$.

For $\gamma \le t$, apply (B.1), (B.3) and $f(y) \ge \frac{y^2}{4}$ on $0 < y \le 1$ to show that

$$\frac{Q(x_t, t)}{K(x_t, t)} \le \frac{4\gamma}{ty^2} = \frac{2}{1-t} \le 4.$$

The right inequality of (B.2) holds.

For $\gamma > t$, apply (B.1), (B.3) and $f(y) \ge y/3$ on $y > 1$ to show that

$$\frac{Q(x_t, t)}{K(x_t, t)} \le \frac{3\gamma}{ty} = 3\sqrt{\frac{\gamma}{2t(1-t)}} \le 3\sqrt{\frac{\gamma}{t}} \le 3\sqrt{\frac{c_N\ell}{s_\ell T}}.$$

The right inequality of (B.2) again holds.

APPENDIX C: DETECTABILITY OF NONALIGNED SIGNALS

Let $X_{nt} = \mu_{nt} + Z_{nt}$, where Z_{nt} are i.i.d. N(0, 1). Assume that there exists $1 \le \ell_T \le T$ such that for the nth sequence, $1 \le n \le N$, there is an unknown interval $(j_{nT}, j_{nT} + \ell_T]$ with probability $\pi_N > 0$ of having an elevated mean

(C.1)
$$\mu_{nt} = \begin{cases} \mu_N I_n/\sqrt{\ell_T}, & \text{if } j_{nT} < t \le j_{nT} + \ell_T, \\ 0, & \text{otherwise,} \end{cases}$$
$$I_n \sim \text{Bernoulli}(\pi_N),$$

with $\mu_N > 0$ and the I_n's and Z_{nt}'s jointly independent. This model is distinct from (2.2) in that we do not now assume that the signals are aligned.

We claim that if $\pi_N = N^{-\beta-\varepsilon}$ for some $0 < \beta < 1$ and $\varepsilon > 0$,

(C.2)
$$T/\ell_T \sim N^\zeta \qquad \text{for some } \zeta > \max(0, 1-2\beta),$$

and $\mu_N = \sqrt{(2\log N)(\zeta+1)\rho^*(\frac{\zeta+\beta}{\zeta+1})}$, then there is no test that can achieve at all such j_{nT},

(C.3)
$$P(\text{Type I error}) + P(\text{Type II error}) \to 0.$$

Note that though $\mu_N \overset{.}{\sim} \sqrt{\log N}$, as in the boundary for cases 1(b) and 2 in (2.5), the growth of T/ℓ_T that is allowed in (C.2) is considerably smaller than that of (2.4).

As in the proof of Theorem 1, set $i_T = \lfloor T/\ell_T \rfloor - 1$, so that $i_T \sim N^\zeta$. Let $Y_{n1} = \ell_T^{-1/2}(X_{n,j_T+1} + \cdots + X_{n,j_T+\ell_T})$ and each Y_{ni}, $2 \le i \le i_T$ be of the form $\ell_T^{-1/2}(X_{n,j+1} + \cdots + X_{n,j+\ell_T})$, with all $(j, j + \ell_T]$ disjoint from each other, and from $(j_{nT}, j_{nT} + \ell_T]$. Assume without loss of generality each j_{nT} is equally likely to take one of the i_T possible values spaced at least ℓ_T apart, as given above. Then when $\varepsilon = 0$, the detection of (nonaligned) signals satisfying (C.2) is at least as difficult as detecting a mixture of $\sim N^{\zeta+1}$ normal random variables $\{Y_{ni} : 1 \le i \le i_T\}$, with a sparse fraction $\sim N^{-(\zeta+\beta)}$ $[= (N^{\zeta+1})^{-(\zeta+\beta)/(\zeta+1)}]$ of them having mean μ_N. Therefore by the results in [14, 15], the critical detectable

μ_N is $\sqrt{(2 \log N^{\zeta+1}) \rho^*(\frac{\zeta+\beta}{\zeta+1})}$. Hence when $\varepsilon > 0$, (C.3) cannot be achieved. Note that the assumption $\zeta > 1 - 2\beta$ in (C.2) implies that $\frac{\zeta+\beta}{\zeta+1} > \frac{1}{2}$, so $\rho^*(\frac{\zeta+\beta}{\zeta+1})$ is well-defined.

REFERENCES

[1] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. MR2246369

[2] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2006). Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.* **34** 326–349. MR2275244

[3] ARIAS-CASTRO, E. and WANG, M. (2013). Distribution-free tests for sparse heteroscedastic mixtures. Preprint.

[4] BERK, R. H. and JONES, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531

[5] CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452

[6] CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inform. Theory* **60** 2217–2232. MR3181520

[7] THE CANCER GENOME ATLAS (2008). Comprehensive genomic characterization defines human gliobastoma genes losses and core pathways. *Nature* **455** 1061–1068.

[8] CHAN, H. P. and WALTHER, G. (2013). Detection with the scan and the average likelihood ratio. *Statist. Sinica* **23** 409–428. MR3076173

[9] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

[10] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[11] EFRON, B. and ZHANG, N. R. (2011). False discovery rates and copy number variation. *Biometrika* **98** 251–271. MR2806427

[12] GLAZ, J., POZDNYAKOV, V. and WALLENSTEIN, S., eds. (2009). *Scan Statistics*: *Methods and Applications*. Birkhäuser, Boston, MA. MR2761094

[13] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357

[14] INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

[15] INGSTER, Y. I. (1998). Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** 401–428. MR1680087

[16] JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962

[17] JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. MR3034330

[18] LAI, W. R., JOHNSON, M. D., KUCHERLAPATI, R. and PARK, P. J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* **21** 3763–3770.

[19] LEPSKI, O. V. and TSYBAKOV, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117** 17–48. MR1759508

[20] MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. MR2650748

[21] OWEN, A. B. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90** 516–521. MR1340504

[22] RIVERA, C. and WALTHER, G. (2013). Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. *Scand. J. Stat.* **40** 752–769. MR3145116

[23] ROHDE, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. MR2418660

[24] SIEGMUND, D., YAKIR, B. and ZHANG, N. R. (2011). Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Stat.* **5** 645–668. MR2840169

[25] TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2008). Asymptotically optimal quickest change detection in distributed sensor systems. *Sequential Anal.* **27** 441–475. MR2460208

[26] WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. MR2604703

[27] WALTHER, G. (2013). The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes. Inst. Math. Stat.* (*IMS*) *Collect.* **9** 317–326. IMS, Beachwood, OH. MR3202643

[28] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. MR3099117

[29] ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488

DEPARTMENT OF STATISTICS
  AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE 117546
E-MAIL: stachp@nus.edu.sg

STATISTICS DEPARTMENT
STANFORD UNIVERSITY
390 SERRA MALL
STANFORD, CALIFORNIA 94305
USA
E-MAIL: gwalther@stat.stanford.edu