The Annals of Statistics
2017, Vol. 45, No. 6, 2736–2763
https://doi.org/10.1214/17-AOS1546

OPTIMAL SEQUENTIAL DETECTION IN MULTI-STREAM DATA

BY HOCK PENG CHAN[1]

National University of Singapore

Consider a large number of detectors each generating a data stream. The task is to detect online, distribution changes in a small fraction of the data streams. Previous approaches to this problem include the use of mixture likelihood ratios and sum of CUSUMs. We provide here extensions and modifications of these approaches that are optimal in detecting normal mean shifts. We show how the (optimal) detection delay depends on the fraction of data streams undergoing distribution changes as the number of detectors goes to infinity. There are three detection domains. In the first domain for moderately large fractions, immediate detection is possible. In the second domain for smaller fractions, the detection delay grows logarithmically with the number of detectors, with an asymptotic constant extending those in sparse normal mixture detection. In the third domain for even smaller fractions, the detection delay lies in the framework of the classical detection delay formula of Lorden. We show that the optimal detection delay is achieved by the sum of detectability score transformations of either the partial scores or CUSUM scores of the data streams.

1. Introduction. Consider N data streams with X_{nt} the observation of the nth data stream at time t. We want to detect as quickly as we can a possible change-point $\nu \geq 1$, such that for some $\mathcal{N} \subset \{1, \dots, N\}$, the post-change observations X_{nt} for $n \in \mathcal{N}$ (and $t \geq \nu$) have distributions different from the pre-change observations. Applications for this multi-stream sequential change-point detection problem include hospital management, infectious-disease modeling and target detection.

Tartakovsky and Veervallli [18] consider distributed decision-making and optimal fusion, with minimax, uniform and Bayesian formulations for sequential detection in multi-stream data. Though optimal detection is achieved, the asymptotics involve N fixed as the average run lengths go to infinity.

Mei [12] considers distribution changes that do not affect all data streams, and recommends a sum of CUSUM approach. The advantages of his approach are that the distribution changes are not assumed to have occurred simultaneously, and the efficient computation of his stopping rule. However, as has been shown in an earlier

Received June 2015; revised July 2016.

[1]Suported by the National University of Singapore Grant R-155-000-158-112.
MSC2010 subject classifications. 62G10, 62L10.
Key words and phrases. Average run length, CUSUM, detectability score, detection delay, mixture likelihood ratio, sparse detection, stopping rule.

simulation study, the detection delay is relatively large when $\#\mathcal{N}$, the number of data streams undergoing change, is small.

Xie and Siegmund [19] are the first to look from the perspective of $\#\mathcal{N}$ small. They suggest a mixture likelihood ratio (MLR) approach and show via simulation studies the superiority of their MLR stopping rules in detecting over a wide range of $\#\mathcal{N}$, compared to other known approaches. They also provide analytical approximations to average run lengths and detection delays of their stopping rules that are accurate and useful. However, they do not give any small or moderate $\#\mathcal{N}$ optimality theory.

In parallel developments, motivated by applications in DNA copy-number samples, there have been advances made; see Siegmund, Yakir and Zhang [17], Jeng, Cai and Li [8] and Chan and Walther [4], on fixed-sample change-point detection in multiple sequences having a common location index. The work here also has connections with detection on spatial indices; see [1–3].

In this paper, we show that subject to an average run length constraint, a modified version of the MLR stopping rule achieves minimum detection delay, extending the classical single-stream optimal detection of Lorden [10], Pollak [14, 15] and Moustakides [13] to multiple data streams, in the detection of normal mean shifts. In Section 2, we provide the asymptotic lower bounds of the detection delays for different domains of $\mathcal{N}$. Under the first domain for large $\#\mathcal{N}$, the lower bound is trivially given by 1. Under the second domain for moderate $\#\mathcal{N}$, the lower bound grows logarithmically with N. Under the third domain for small $\#\mathcal{N}$, the detection delay grows polynomially with N. In Section 3, we show that a MLR stopping rule that tests against the limits of detectability achieves optimal detection on all three domains. A window-limited rule, suggested in Lai [9], is incorporated into the stopping rule for computational savings. In Section 4, a numerical study is performed to provide justification for using the MLR stopping rule for finite N. In Section 5, we extend the idea of testing against the limits of detectability on Mei's sum of CUSUM test. Rather than summing the CUSUM scores as in Mei [12], we suggest instead to sum the detectability score transformations of the CUSUM scores. Optimality of this procedure is shown but it occurs only when we select the assumed mean shift at a specific value between one to two times the true mean shift, surprisingly not at the true mean shift itself. In Sections 6–8, we provide the proofs of Theorems 1–3.

2. Detection delay lower bound.

Let X_{nt}, $1 \le n \le N$, $t \ge 1$, be distributed as independent $\mathrm{N}(\mu_{nt}, 1)$. Assume that at some unknown time $\nu \ge 1$, there are mean shifts in a subset $\mathcal{N}$ of the data streams. More specifically, we assume that

$$\mu_{nt} = \mu \mathbf{I}_{\{t \ge \nu, n \in \mathcal{N}\}} \qquad \text{for some } \mu > 0, \tag{2.1}$$

with $\mathbf{I}_{\{1 \in \mathcal{N}\}}, \dots, \mathbf{I}_{\{N \in \mathcal{N}\}}$ i.i.d. Bernoulli(p) for some $0 < p < 1$. We shall let P_ν (E_ν) denote probability measure (expectation) with respect to distribution changes at time ν, with $\nu = \infty$ indicating no change.

A standard measure of the performance of a stopping rule T (see Pollak [14, 15]) is the (expected) detection delay:

$$D_N(T) := \sup_{1 \le \nu < \infty} E_\nu(T - \nu + 1 | T \ge \nu), \tag{2.2}$$

subject to the constraint that $\mathrm{ARL}(T)$ $(:= E_\infty T) \ge \gamma$ for some $\gamma \ge 1$.

In this section, we find (asymptotic) lower bounds of $D_N(T)$ under the conditions that as $N \to \infty$,

$$\log \gamma \sim N^\zeta \qquad \text{for some } 0 < \zeta < 1, \tag{2.3}$$

$$p \sim N^{-\beta} \qquad \text{for some } 0 < \beta < 1. \tag{2.4}$$

In Sections 3 and 5, we devise optimal detectability score stopping rules that achieve this lower bound. In Theorem 1 below, only $\beta > \frac{1-\zeta}{2}$ is considered. For $\beta < \frac{1-\zeta}{2}$, the detectability score stopping rules achieve asymptotic detection delay of 1, and are hence optimal.

For $\frac{1-\zeta}{2} < \beta < 1 - \zeta$, the detection delay lower bound grows logarithmically with N. The proportionality constant is

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

This is a two-dimensional extension of the Donoho–Ingster–Jin constants $\rho(\beta) := \rho(\beta, 0)$, which has appeared in connection with sparse normal mixture detection; see [5–7]. The extension results from the additional difficulty of detecting a normal mean shift when there are multiple comparisons, here for sequential change-point detection, and in [4] for fixed-sample change-point detection.

THEOREM 1. *Let T be a stopping rule such that* $\mathrm{ARL}(T) \ge \gamma$, *with γ satisfying* (2.3). *If* (2.4) *holds with $\frac{1-\zeta}{2} < \beta < 1 - \zeta$, then*

$$\liminf_{N \to \infty} \frac{D_N(T)}{\log N} \ge 2\mu^{-2} \rho(\beta, \zeta). \tag{2.5}$$

For the case $\beta > 1 - \zeta$, if (2.4) holds and $\mathrm{ARL}(T) \ge \gamma$ with γ satisfying (2.3), then

$$P_1\{\#\mathcal{N} = 0\} = (1-p)^N = \exp\{-[1 + o(1)]Np\} \gg \gamma^{-1} \log N,$$

and $D_N(T)$ is therefore large compared to $\log N$. To provide a more meaningful analysis that avoids the conclusion of a large detection delay from the consideration of $\#\mathcal{N} = 0$ alone, we consider the following minimax formulation.

Let $E_{\nu,\mathcal{N}}$ denote expectation with respect to $X_{nt} \sim \mathrm{N}(\mu_{nt}, 1)$, with $\mu_{nt} = \mu \mathbf{I}_{\{t \ge \nu, n \in \mathcal{N}\}}$ for some $\mu > 0$, that is, $E_{\nu,\mathcal{N}}$ denotes expectation with respect to a fixed $\mathcal{N}$ [whereas E_ν as defined earlier denotes expectation with respect to $\#\mathcal{N} \sim \text{Binomial}(N, p)$]. For a given stopping rule T, define

$$D_N^{(m)}(T) = \sup_{1 \le \nu < \infty} \Big[\max_{\mathcal{N}:\#\mathcal{N}=m} E_{\nu,\mathcal{N}}(T - \nu + 1 | T \ge \nu) \Big].$$

We show in Theorem 4 (Appendix A) that if $m \sim N^{1-\beta}$ for $\beta > 1 - \zeta$, then

$$\liminf_{N\to\infty} \frac{\log D_N^{(m)}(T)}{\log N} \ge \beta + \zeta - 1. \tag{2.6}$$

By (2.5) and (2.6), we see that the phase transition between logarithmic and polynomial growth of the detection delay boundary is at $N^{1-\beta} = N^\zeta$, that is, at $(m =)\#\mathcal{N} \doteq \log\gamma$. By (2.5) for larger $\#\mathcal{N}$ the detection delay lower bound grows at a $\log N$ rate. By (2.6) for smaller $\#\mathcal{N}$, the lower bound is roughly $(\log\gamma)/m$. The detection delay lower bound (2.5) in the logarithmic domain is closely linked to the Donoho–Ingster–Jin detection boundary for sparse normal mixture detection, whereas the lower bound (2.6) in the polynomial domain lies in the framework of the classical lower bound established by Lorden [10] for N fixed as $\gamma \to \infty$.

We shall first establish the connection between (2.5) and the Donoho–Ingster–Jin detection boundary $\sqrt{2\rho(\beta)\log N}$. Let $t \ge \nu \ge 1$ and $k = t - \nu + 1$. If $p \sim N^{-\beta}$, $\frac{1}{2} < \beta < 1$, then as

$$k^{-1/2}\sum_{i=\nu}^{t} X_{ni}$$
$$\sim \begin{cases} \mathrm{N}(0,1), & \text{under } P_\infty, \\ (1-p)\mathrm{N}(0,1) + p\mathrm{N}(\mu\sqrt{k}, 1), & \text{under } P_\nu, \end{cases} \qquad 1 \le n \le N,$$

sparse normal mixture detection theory dictates that k should satisfy

$$\mu\sqrt{k} \ge [1 + o(1)]\sqrt{2\rho(\beta)\log N} \qquad \big(\text{i.e., } k \ge [2\mu^{-2}\rho(\beta) + o(1)]\log N\big),$$

in order for it to be possible that the sum of Type I and II error probabilities goes to zero, when testing P_ν against P_∞ with observations up to time t. By (2.2), this leads to

$$D_N(T) \ge [2\mu^{-2}\rho(\beta) + o(1)]\log N, \tag{2.7}$$

for any stopping rule T satisfying $\text{ARL}(T) \ge \gamma$ with $\gamma/\log N \to \infty$. What Theorem 1 says is that under (2.3) with ζ small enough ($<1-\beta$), $\log N$ detection is still possible with a larger asymptotic constant.

The link between (2.6) and the classical lower bound formula of Lorden is best established via the inequality in Mei [11], Proposition 2.1, that for N fixed,

$$D_N^{(m)}(T) \ge 2\mu^{-2}(\log\gamma)/m + O(1) \qquad \text{as } \gamma \to \infty. \tag{2.8}$$

What (2.6) says is that if $\log\gamma$ is large compared to m, then the right-hand side of (2.8) provides the correct order for the attainable detection delay. Whereas if $\log\gamma$ is small compared to m, then the right-hand side of (2.8) does not provide the correct order for the attainable detection delay as we have already noted in the previous paragraph situations under which a $\log N$ detection delay is required. Therefore, the $O(1)$ in (2.8) is more appropriately $O(\log N)$ if the dependence on N in $O(1)$ is made explicit. What Theorem 1 and (2.6) also say is that the transition is sharp. Once we get out of the classical $(\log\gamma)/m$ domain, we fall into the $\log N$ domain, and there are no intermediate asymptotics.

3. Optimal detection using detectability score. The detectability score stopping rule to be introduced in (3.2) below is motivated by the MLR stopping rules of Xie and Siegmund [19]. In their formulation Xie and Siegmund first consider the ideal situation in which p and μ are known. The most powerful test at time t, for testing the hypothesis that change-point $\nu = s$ for some $s \leq t$, is the log likelihood ratio:

$$\ell_{\bullet st} := \sum_{n=1}^{N} \ell_{nst} \qquad \text{where } \ell_{nst} = \log\bigl(1 - p + pe^{\mu S_{nst} - k\mu^2/2}\bigr),$$

with $k = t - s + 1$ and $S_{nst} = \sum_{i=s}^{t} X_{ni}$.

Since the change-point ν is unknown, they suggest to maximize $\ell_{\bullet st}$ over s. The unknown μ (or more precisely μ_n) in ℓ_{nst} is substituted by S^+_{nst}/k, and a small p_0 is substituted for the unknown p. In summary, their stopping rule can be expressed as

$$T_{\mathrm{XS}}(p_0) = \inf\Bigl\{t : \max_{k=t-s+1\in\mathcal{K}} \widehat{\ell}_{\bullet st}(p_0) \geq b\Bigr\}, \tag{3.1}$$

where $\widehat{\ell}_{\bullet st}(p_0) = \sum_{n=1}^{N} \widehat{\ell}_{nst}(p_0)$ and

$$\widehat{\ell}_{nst}(p_0) = \log\bigl(1 - p_0 + p_0 e^{(Z^+_{nst})^2/2}\bigr), \qquad Z_{nst} = S_{nst}/\sqrt{k}.$$

The set $\mathcal{K}$ in (3.1) refers to a pre-determined set of window sizes. By applying nonlinear renewal theory, Xie and Siegmund derive accurate analytical approximations of $\mathrm{ARL}(T)$ and $D_N(T)$ for $T = T_{\mathrm{XS}}(p_0)$ and related stopping rules.

Our stopping rule is also a mixture likelihood ratio but based instead on the limits of detectability. Let

$$T_S(p_0) = \inf\Biggl\{t : \max_{k=t-s+1\in\mathcal{K}} \sum_{n=1}^{N} g\bigl(Z^+_{nst}\bigr) \geq b\Biggr\}, \tag{3.2}$$

where $g(z) = \log[1 + p_0(\lambda e^{z^2/4} - 1)]$ and $\lambda = 2(\sqrt{2} - 1)$. Following Lai [9], we consider window sizes:

$$\mathcal{K} = \{1, \ldots, k_1\} \cup \bigl\{\lfloor r^j k_1 \rfloor : j \geq 1\bigr\}, \qquad k_1 \geq 1, r > 1. \tag{3.3}$$

THEOREM 2. *Consider stopping rule* $T_S(p_0)$, $0 < p_0 \leq 1$, *with window sizes* (3.3). *If* $\mathrm{ARL}(T_S(p_0)) = \gamma$, *then the stopping rule threshold* $b \leq \log(4\gamma^2 + 2\gamma)$. *In addition, if* (2.3), (2.4) *hold and* $k_1/\log N \to \infty$, $p_0 = c[(\log\gamma)/N]^{1/2}$ *for some* $c > 0$, *then the following hold as* $N \to \infty$:

(a) *If* $\beta < \frac{1-\zeta}{2}$, *then* $D_N(T_S(p_0)) \to 1$.

(b) *If* $\frac{1-\zeta}{2} < \beta < 1 - \zeta$, *then*

$$\frac{D_N(T_S(p_0))}{\log N} \to 2\mu^{-2}\rho(\beta, \zeta). \tag{3.4}$$

(c) *If* $m \sim N^{1-\beta}$ *for* $\beta > 1 - \zeta$, *then*

$$\frac{\log D_N^{(m)}(T_S(p_0))}{\log N} \to \beta + \zeta - 1.$$

REMARKS. Instead of (2.3), we can model γ growing slowly with N by assuming that

$$\gamma/\log N \to \infty, \qquad \log\gamma = o(N^{\varepsilon}) \qquad \text{for all } \varepsilon > 0. \tag{3.5}$$

Consider the stopping rule $T_S(p_0)$ with $p_0 = cN^{-\frac{1}{2}}$ for some $c > 0$. Under (2.4) and (3.5), the asymptotic (3.4) holds with $\zeta = 0$, and the stopping rule is optimal in view of (2.7).

We shall provide some intuition here on the detectability score transformation g. Consider an i.i.d. sample $Z_1, \dots, Z_N$ that is distributed as $\mathrm{N}(0, 1)$ under the null hypothesis H_0. If $w_N \to \infty$ with $w_N = o(\sqrt{\log N})$, then $\#\{n : Z_n \geq w_N\}/N$ is asymptotically normal with mean α_N and variance α_N/N, where $\alpha_N = P_0\{Z_n \geq w_n\} = \int_{w_N}^{\infty} (2\pi)^{-1/2} e^{-z^2/2}\, dz$.

Therefore, under any alternative hypothesis H_1, $\sqrt{\alpha_N/N}$ is the minimum deviation of $P_1\{Z_n \geq w_N\}$ from α_N that is detectable. Since α_N is essentially $e^{-w_N^2/2}$ (up to logarithmic terms), the minimum detectable deviation is $e^{-w_N^2/4}/\sqrt{N}$. That is, a mixture of $\mathrm{N}(0, 1)$ and a small $p_0 = cN^{-\frac{1}{2}}$ fraction of $\mathrm{N}(0, 2)$ is at the threshold of detectability. The detectability score transformation g is essentially the likelihood ratio between the mixture with p_0 fraction $\mathrm{N}(0, 2)$, and the null distribution. The factor $(\log\gamma)^{1/2}$, in the optimal choice of p_0 in the statement of Theorem 2, adjusts for the additional difficulty of each detection due to the multiple comparison effects of large γ.

It is straightforward to check that the detectability score $\sum_{n=1}^{N} g(Z_{nst}^+)$ in (3.2) is indeed the log likelihood ratio for testing $Z_{1st}^+, \dots, Z_{Nst}^+$ i.i.d. $\mathrm{N}(0, 1)^+$ [the distribution of Z^+ when $Z \sim \mathrm{N}(0, 1)$] against the alternative that $Z_{1st}^+, \dots, Z_{Nst}^+$ are i.i.d.:

$$(1 - p_0)\mathrm{N}(0, 1)^+ + p_0\left[\frac{\lambda}{\sqrt{2}}\mathrm{HN}(0, 2) + \left(1 - \frac{\lambda}{\sqrt{2}}\right)\delta_0\right],$$

where δ_0 denotes a point mass at zero and HN(0, 2) the half-normal distribution with density $\pi^{-1/2}e^{-z^2/4}$ on $z > 0$. The value $\lambda = 2(\sqrt{2} - 1)$ is chosen for convenience, so that g is continuous at 0. The optimality of $T_S(p_0)$ in Theorem 2 does not require the selection of this specific λ.

4. Numerical study. In addition to (3.1), Xie and Siegmund introduce the stopping rule:

$$T_{\mathrm{LR}}(p_0) = \inf\left\{t : \max_{k=t-s+1 \in \mathcal{K}} \sum_{n=1}^{N} (\mu_0 S_{nst} - k\mu_0^2/2 + \log p_0)^+ \ge b\right\}. \tag{4.1}$$

This like (3.1) is motivated by the most powerful likelihood ratio test, but with μ substituted by a pre-determined μ_0 rather than S_{nst}^+/k. It bears resemblance to Mei's stopping rule:

$$T_{\mathrm{Mei}} = \inf\left\{t : \sum_{n=1}^{N} \max_{0<s\le t} (\mu_0 S_{nst} - k\mu_0^2/2)^+ \ge b\right\}, \tag{4.2}$$

with the important difference of an additional $\log p_0$ term in (4.1) that suppresses the contributions of low scoring data streams.

Another key difference is that the sum lies outside the max in (4.2) whereas in T_{LR} (and T_{XS}, T_S), the sum lies inside the max. This confers advantage to Mei's stopping rule when the change-point ν (or ν_n) differs across data streams. We investigate this in Section 5 where we also propose an extension of Mei's stopping rule, denoted by $T_{\mathrm{Mei}}(p_0)$, that like (4.1) weighs down the contributions from nonsignal data streams.

In our numerical study, we benchmark the detectability score stopping rule against the above stopping rules and the max rule:

$$T_{\max} = \inf\left\{t : \max_{0<s\le t} \max_{1\le n\le N} (Z_{nst}^+)^2/2 \ge b\right\}. \tag{4.3}$$

As in [19], we select $N = 100$, $\mu = 1$ and $\#\mathcal{N}$ ranging from 1 to 100. The thresholds b are calibrated to average run length 5000. The set of window sizes chosen is $\mathcal{K} = \{1, \dots, 200\}$, and for Mei's stopping rule and T_{LR} we select $\mu_0 = 1$.

We consider $p_0 = 0.1 (= N^{-\frac{1}{2}})$ for the detectability score stopping rule T_S, corresponding to the optimal choice under (3.5). Another selection is $p_0 = 0.3\{\doteq [(\log\gamma)/N]^{1/2}\}$, which is optimal under (2.3). It is interesting that in [19], the "optimal" $p_0 = N^{-\frac{1}{2}}$ is chosen for T_{XS} and T_{LR} in the numerical study.

We conduct 500 Monte Carlo trials for the estimation of each average run length and detection delay. The thresholds for the stopping rules are in Table 1, the detection delays in Table 2. In Table 2, the simulation outcomes for Mei(p_0) and S(p_0) are new, the other outcomes are reproduced from [19], Table 5.

We see that with a few understandable exceptions, the detectability score stopping rules $T_S(0.1)$ and $T_S(0.3)$ have smaller detection delays compared to their

TABLE 1
Thresholds b for stopping rules calibrated to ARL $\doteq$ 5000. The upper bounds of the thresholds, as given in the statement of Theorems 2 and 3, are in brackets

	$N = 100$		$N = 10^4$	
Test	***b***	**ARL**	***b***	**ARL**
max	12.8	5041	15.9	4930
Mei	88.5 (106.8)	4997	5640 (8722)	4909
Mei($N^{-\frac{1}{2}}$)	3.48 (9.81)	4994	3.03 (8.93)	4973
Mei($3N^{-\frac{1}{2}}$)	5.02 (9.61)	4976	2.31 (6.97)	5017
$S(N^{-\frac{1}{2}})$	4.25 (18.42)	5066	14.49 (18.42)	5121
$S(3N^{-\frac{1}{2}})$	6.30 (18.42)	5195	17.21 (18.42)	4986

competitors over the full range of $\#\mathcal{N}$. This justifies the application of the detectability score stopping rules for a relatively small $N = 100$.

Following the recommendation of a referee, we conduct a second numerical exercise for a larger $N = 10^4$, with $\#\mathcal{N}$ ranging from 1 to 10^4. As in the earlier simulation study, we select $\mu = \mu_0 = 1$, ARL $= 5000$ and $\mathcal{K} = \{1, \dots, 200\}$. The detection thresholds are in Table 1, the detection delays in Table 3. We see again that except for $\#\mathcal{N} = 1$ when $T_{\max}$ is superior, the detection score stopping rules $T_S(p_0)$ for $p_0 = 0.01 (= N^{-\frac{1}{2}})$ and 0.03 $\{\doteq [(\log\gamma)/N]^{1/2}\}$ have the smallest detection delays.

TABLE 2
Detection delays when $\#\mathcal{N}$ (out of $N = 100$) data streams undergo distribution changes. Entries in the last row are standard error upper bounds

	$\#\mathcal{N}$						
Test	**1**	**3**	**5**	**10**	**30**	**50**	**100**
max	25.5	18.1	15.5	12.6	9.6	8.6	7.2
XS(1)	52.3	18.7	12.2	6.7	3.0	2.3	2.0
XS(0.1)	31.6	14.2	10.4	6.7	3.5	2.8	2.0
LR(0.1)	29.1	13.4	9.8	7.1	4.6	4.0	3.4
LR(1)	82.0	27.2	15.5	6.8	3.0	2.3	2.0
Mei	53.2	23.0	15.7	9.6	4.9	3.8	3.0
Mei(0.1)	26.4	14.6	10.8	7.7	4.5	3.4	2.3
Mei(0.3)	34.3	15.9	11.8	7.6	4.1	3.1	2.0
$S(0.1)$	26.8	13.4	9.6	6.4	2.8	2.0	1.1
$S(0.3)$	32.6	14.0	9.5	5.6	2.3	1.5	1.0
s.e.	0.9	0.3	0.1	0.1	0.1	0.1	0.1

TABLE 3
Detection delays when $\#\mathcal{N}$ (out of $N = 10^4$) data streams undergo distribution changes. Entries in the last row are standard error upper bounds

	$\#\mathcal{N}$				
Test	**1**	**10**	**10^2**	**10^3**	**10^4**
max	32.7	18.6	13.9	11.1	9.4
Mei	246.5	46.7	12.0	4.0	1.0
Mei(0.01)	39.7	16.7	8.8	4.0	2.0
Mei(0.03)	53.7	18.6	9.0	4.0	2.0
S(0.01)	37.7	13.3	4.5	1.0	1.0
S(0.03)	49.3	13.7	3.9	1.0	1.0
s.e.	4.0	0.3	0.1	0.1	0.1

5. Detectability of Mei's stopping rule. As mentioned earlier, there is no implicit assumption that the distribution changes occur simultaneously when applying Mei's stopping rule (4.2). Another advantage is the efficient recursive computation of the stopping rule. However, this recursive computation comes with the price of information loss. In this section, we improve Mei's stopping rule by applying a detectability score transformation on each CUSUM score. Due to the information loss, optimality is possible only for specific μ_0.

Let R_{nt} be the CUSUM score of the nth detector at time t, satisfying

$$R_{n0} = 0, \qquad R_{nt} = \left(R_{n,t-1} + \mu_0 X_{nt} - \mu_0^2/2\right)^+, \qquad t \geq 1. \tag{5.1}$$

Define

$$T_{\mathrm{Mei}}(p_0) = \inf\left\{t : \sum_{n=1}^{N} g_M(R_{nt}) \geq b\right\}, \tag{5.2}$$

with the detectability score transformation:

$$g_M(x) = \log\left[1 + p_0\left(\lambda_M e^{x/2} - 1\right)\right], \qquad \lambda_M > 0. \tag{5.3}$$

This is an extension of Mei's test, for $T_{\mathrm{Mei}}(1)$ is equivalent to T_{Mei}. Let $\xi = \lim_{t\to\infty} E_\infty e^{R_{nt}/2}$ and define

$$D_{N,k}(T) = \sup_{k \leq \nu < \infty} E_\nu(T - \nu + 1 | T \geq \nu).$$

THEOREM 3. *Consider stopping rule $T_{\mathrm{Mei}}(p_0)$, $0 < p_0 \leq 1$. Let $u = \log[1 + p_0(\lambda_M \xi - 1)]$. If $\mathrm{ARL}(T_{\mathrm{Mei}}(p_0)) = \gamma$, then the stopping rule threshold $b \leq Nu + \log(4\gamma)$. In addition, if (2.3), (2.4) hold and $p_0 = c[(\log\gamma)/N]^{1/2}$ for some $c > 0$, then the following hold as $N \to \infty$:*

(a) *If* $\frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}$ *and* $\mu_0 = 2\mu$, *then*

$$\frac{D_{N,K_N}(T_{\mathrm{Mei}}(p_0))}{\log N} \to 2\mu^{-2}\rho(\beta,\zeta), \tag{5.4}$$

for $K_N = 2\mu^{-2}(1-\zeta-\beta)\log N$.

(b) *If* $\frac{3(1-\zeta)}{4} < \beta < 1-\zeta$ *and* $\mu_0 = \mu\sqrt{\frac{1-\zeta}{\rho(\beta,\zeta)}}$, *then* (5.4) *holds for* $K_N = 2\mu^{-2}\rho(\beta,\zeta)\log N$.

REMARKS. 1. In Theorem 3, "optimality" occurring when $\mu_0 > \mu$ is a consequence of a small subset of $\mathcal{N}$ dominating the score contributions.

2. Notice the weaker (5.4) instead of (3.4). The extra initial delay is needed for the CUSUM scores R_{nT} for $n \notin \mathcal{N}$ to reach their stationary values and not pull down the total score. In that sense, the detection delay criterion may be disadvantageous to the extended Mei's stopping rule (and hence Mei's test stopping rule itself) since in practice we seldom expect the change-point ν to be that close to 0.

To highlight the unique characteristics of the extended Mei's stopping rule (5.2) in dealing with staggered change-points, we conduct a numerical study with $\mu_{nt} = \mu\mathbf{I}_{\{t \geq n\}}$ in place of (2.1). That is the nth data stream undergoes a distribution change at time n. As in Section 4, the stopping rules are calibrated to average run length 5000, for $N = 100$ detectors, and with $\mu_0 = 1$. The thresholds b for $T_{\mathrm{Mei}}(p_0)$ are in Table 1 (Section 4), the detection delays in Section 4. We select $\lambda_M = 0.64$, this will be explained later. By detection delay, we shall mean the expected stopping time when $\mu_{nt} = \mu\mathbf{I}_{\{t \geq n\}}$.

We see from Tables 2 (Section 4) and 4 that $T_{\mathrm{Mei}}(0.1)$ and $T_{\mathrm{Mei}}(0.3)$ have smaller detection delays compared to T_{Mei}, almost uniformly over $\#\mathcal{N}$ and μ. In Table 3 (for $N = 10^4$), $T_{\mathrm{Mei}}(0.01)$ and $T_{\mathrm{Mei}}(0.03)$ are superior to T_{Mei} for $\#\mathcal{N} \leq 100$. Hence, applying detectability score transformations on the CUSUM scores improves Mei's stopping rule in general, the noise suppression on data streams that do not undergo distribution change is indeed effective. In Table 4, we see that in general $T_S(p_0)$ performs better than $T_{\mathrm{Mei}}(p_0)$ when $\mu \geq 1$ but the

TABLE 4
Detection delays for staggered distribution changes. The standard errors are not more than 0.2

μ	Mei	Mei(0.1)	Mei(0.3)	S(0.1)	S(0.3)
0.5	20.7	21.2	20.6	23.0	20.7
0.7	15.5	15.4	15.1	16.0	14.9
1.0	11.9	10.9	11.1	10.9	10.4
1.3	10.0	8.7	9.0	8.0	7.9

reverse is true when $\mu < 1$. This is consistent with the prediction in Theorem 3 of $T_{\mathrm{Mei}}(p_0)$ performing better for $\mu < \mu_0$.

We end this section with explanations of the choice of the detectability score transformation (5.3) and choice of λ_M. It follows from renewal theory; see, for example, Siegmund [16], equation (8.49), that

$$\lim_{t\to\infty} P_\infty\{R_{nt} \geq x\} \sim \alpha e^{-x} \qquad \text{as } x \to \infty, \tag{5.5}$$

for $\alpha = 2\mu_0^{-2}\exp[-2\sum_{j=1}^\infty j^{-1}\Phi(-\mu_0\sqrt{j}/2)]$. Therefore, the tails of R_{nt} under P_∞ are like that of an i.i.d. sample from $G_1 := (1-\alpha)\delta_0 + \alpha\,\mathrm{Exp}(1)$, where δ_0 denotes a point mass at 0 and $\mathrm{Exp}(\theta)$ the exponential distribution with mean θ.

For large x (smaller than $\log N$) and t, $\#\{n : R_{nt} \geq x\}/N$ is asymptotically normal with mean αe^{-x} and variance $\alpha e^{-x}/N$. Hence, the minimum detectable difference of $P\{R_{nt} \geq x\}$ is $e^{-x/2}/\sqrt{N}$. The distribution at the limit of detectability is therefore $G^* := (1-p_0)G_1 + p_0 G_2$, where $G_2 = (1-\omega)\delta_0 + \omega\,\mathrm{Exp}(2)$ for some $0 < \omega < 1$, and p_0 is of order $N^{-\frac{1}{2}}$. The detectability score transformation g_M [see (5.3)], with $\lambda_M = \frac{1}{1+\alpha}(= 0.64 \text{ for } \mu_0 = 1)$, is the log likelihood ratio between G^* and G_1, with ω selected so that g_M is continuous at 0. We emphasize however that this is for convenience, optimality in Theorem 3 is not restricted to this choice of λ_M.

6. Proof of Theorem 1. To help the reader, we summarize below the definitions of the probability measures used in the proofs of Theorems 1–3 in this and the next two sections:

1. P_s (E_s): This is the probability measure (expectation) under which an arbitrarily chosen data stream has probability $(1-p)$ that all observations are (i.i.d.) N(0, 1), and probability p that observations are N(0, 1) before time s, N(μ, 1) at and after time s. In particular, if
   (a) $s = \infty$, then with probability 1 all observations are N(0, 1).
   (b) $s = 1$, then an arbitrarily chosen data stream has probability $(1-p)$ that all observations are N(0, 1), and probability p that all observations are N(μ, 1).
2. P (E): This is the probability measure (expectation) under which $Y, Y_1, Y_2, \dots$ are i.i.d. N(0, 1) random variables.

We preface the proof of Theorem 1 with the following lemmas. Lemma 1 is well known; see, for example, (3.3) of Lai [9].

LEMMA 1. *Let $k \geq 1$. If T is a stopping rule such that $E_\infty T \geq \gamma$, then $P_\infty\{T \geq s+k | T \geq s\} \geq 1 - k/\gamma$ for some $s \geq 1$.*

Recall the sum $S_{nst} = \sum_{i=s}^t X_{ni}$ and the log likelihood ratio:

$$\ell_{\bullet st} = \sum_{n=1}^N \ell_{nst} \qquad \text{where } \ell_{nst} = \log\bigl(1 - p + pe^{\mu S_{nst} - k\mu^2/2}\bigr), k = t - s + 1.$$

LEMMA 2. *If we can find $b(=b_N)$ and $k(=k_N)$, such that*

$$P_\infty\{\ell_{\bullet 1k} \geq b\}\bigl(= P_\infty\{\ell_{\bullet st} \geq b\}\bigr) \geq k/\gamma, \tag{6.1}$$

$$P_1\{\ell_{\bullet 1k} \geq b\}\bigl(= P_s\{\ell_{\bullet st} \geq b\}\bigr) \to 0, \tag{6.2}$$

then $D_N(T) \geq [1+o(1)]k$ for any stopping rule T satisfying $E_\infty T \geq \gamma$.

PROOF. Let T satisfies $E_\infty T \geq \gamma$, and let b, k satisfy (6.1) and (6.2). By Lemma 1, we can find s satisfying

$$P_\infty\{T \geq s+k | T \geq s\} \geq 1 - k/\gamma. \tag{6.3}$$

Let $P_\infty^*\{\cdot\} = P_\infty\{\cdot | T \geq s\}$ and $P_s^*\{\cdot\} = P_s\{\cdot | T \geq s\}$.

Let $t = s+k-1$, and consider the test, conditioned on $T \geq s$, of

$$\begin{aligned}
H_0: \quad & X_{nu} \sim \mathrm{N}(0,1) \qquad \text{for } 1 \leq n \leq N, 1 \leq u \leq t,\\
\text{vs } H_s: \quad & X_{nu} \sim \mathrm{N}(\mu \mathbf{I}_{\{u \geq s, n \in \mathcal{N}\}}, 1) \qquad \text{for } 1 \leq n \leq N, 1 \leq u \leq t,\\
& \text{with } \mathbf{I}_{\{n \in \mathcal{N}\}} \sim \text{ Bernoulli}(p).
\end{aligned}$$

By (6.3) the test "reject H_0 if $T < s+k$, accept H_0 otherwise" has Type I error probability not exceeding k/γ. By (6.1), the likelihood ratio test rejecting H_0 when $\ell_{\bullet st}$ exceeds b has Type I error probability at least k/γ, and hence by the Neyman–Pearson lemma, it is at least as powerful as the test based on T. That is,

$$P_s^*\{\ell_{\bullet st} \geq b\} \geq P_s^*\{T < s+k\}. \tag{6.4}$$

A key observation here is that the conditioning on $\{T \geq s\}$ does not affect the distribution of X_{nu} for $u \geq s$ under either H_0 or H_s. Therefore, by (6.4),

$$\begin{aligned}
D_N(T) &\geq E_s(T-s+1 | T \geq s) \geq k P_s^*\{T \geq s+k\}\\
&\geq k P_s^*\{\ell_{\bullet st} < b\} = k P_s\{\ell_{\bullet st} < b\},
\end{aligned}$$

and we conclude $D_N(T) \geq [1+o(1)]k$ from (6.2). $\square$

LEMMA 3. *If k is such that $\log k = o(N^\zeta)$ and*

$$P_1\{\ell_{\bullet 1k} \geq 2N^\zeta/3\} \to 0, \tag{6.5}$$

then (6.1) and (6.2) follow from selecting b satisfying

$$P_1\{2N^\zeta/3 \geq \ell_{\bullet 1k} \geq b\} = \exp(-N^\zeta/4). \tag{6.6}$$

PROOF. It follows from (6.5) and (6.6) that (6.2) holds. Moreover, since $\ell_{\bullet 1k}$ is the log change of measure between P_1 and P_∞ at time k,

$$\begin{aligned}
&P_\infty\{\ell_{\bullet 1k} \geq b\}\\
&\quad \geq P_\infty\{2N^\zeta/3 \geq \ell_{\bullet 1k} \geq b\}\\
&\quad = E_1\bigl(e^{-\ell_{\bullet 1k}} \mathbf{I}_{\{2N^\zeta/3 \geq \ell_{\bullet 1k} \geq b\}}\bigr) \geq \exp(-2N^\zeta/3) P_1\{2N^\zeta/3 \geq \ell_{\bullet 1k} \geq b\},
\end{aligned}$$

and (6.1) follows from (6.6) since $\log(\gamma/k) \sim N^\zeta$. $\square$

In view of Lemmas 2 and 3, to prove Theorem 1 it suffices to check (6.5) for

$$k = \lfloor (1-\delta)2\mu^{-2}\rho(\beta,\zeta)\log N \rfloor \qquad \text{if } \frac{1-\zeta}{2} < \beta < 1-\zeta, \tag{6.7}$$

with $\delta > 0$ small. Motivations behind the above choice of k are given in Appendix B.

Let $Z_{nk} = S_{n1k}/\sqrt{k}$ and

$$\ell_{nk}(=\ell_{n1k}) = \log\bigl(1 - p + pe^{Z_{nk}\mu\sqrt{k}-k\mu^2/2}\bigr). \tag{6.8}$$

Note that Z_{nk}, $1 \le n \le N$, are i.i.d. N(0, 1) under P_∞, and i.i.d. $(1-p)$N(0, 1) + pN$(\mu\sqrt{k}, 1)$ under P_1. More specifically, Z_{nk} has the distribution of $Y \sim$ N(0, 1) if $n \notin \mathcal{N}$, and the distribution of $Y + \mu\sqrt{k}$ if $n \in \mathcal{N}$. Hence, conditioned on $n \notin \mathcal{N}$, ℓ_{nk} has the distribution of

$$\ell_0 = \log\bigl(1 - p + pe^{Y\mu\sqrt{k}-k\mu^2/2}\bigr), \tag{6.9}$$

whereas conditioned on $n \in \mathcal{N}$, ℓ_{nk} has the distribution of

$$\ell_1 = \log\bigl(1 - p + pe^{Y\mu\sqrt{k}+k\mu^2/2}\bigr). \tag{6.10}$$

Consider $\frac{1-\zeta}{2} < \beta < 1-\zeta$ and let $\widetilde{\ell}_{nk} = \ell_{nk}\mathbf{I}_{\{Z_{nk}\le\omega_N\}}$, where

$$\omega_N = \sqrt{2(1-\zeta)\log N + 2\log\log N}.$$

We shall check on two cases that

$$\widetilde{\mu} := E_1\widetilde{\ell}_{nk} = o\bigl(N^{\zeta-1}\bigr), \tag{6.11}$$

$$\sup_{1\le n\le N} \widetilde{\ell}^{+}_{nk} = O(1), \tag{6.12}$$

$$E_1\widetilde{\ell}^{2}_{nk} = o\bigl(N^{\zeta-1}\bigr), \tag{6.13}$$

$$P_1\{Z_{nk} > \omega_N\} = o\bigl(N^{\zeta-1}/\log N\bigr). \tag{6.14}$$

Since $\max_{1\le n\le N} Z_{nk} = O_p(\sqrt{\log N})$, it follows from (6.7) and (6.8) that $\max_{1\le n\le N} \ell_{nk} = O_p(\log N)$. Hence, by (6.14), $\sum_{n=1}^N \ell_{nk}\mathbf{I}_{\{Z_{nk}>\omega_N\}} = o_p(N^\zeta)$ and, therefore, there exists $\alpha_N \to 0$ such that

$$\sum_{n=1}^N \ell_{nk}\mathbf{I}_{\{Z_{nk}>\omega_N\}} = O_p\bigl(\alpha_N N^\zeta\bigr). \tag{6.15}$$

Recall that $\ell_{\bullet 1k} = \sum_{n=1}^N \ell_{nk}$ and let $\widetilde{\ell}_{\bullet 1k} = \sum_{n=1}^N \widetilde{\ell}_{nk}$. By Chebyshev's inequality and (6.13),

$$\begin{aligned} P_1\bigl\{\widetilde{\ell}_{\bullet 1k} - N\widetilde{\mu} \ge N^{\zeta/2}\bigr\} &\le N^{-\zeta}E_1(\widetilde{\ell}_{\bullet 1k} - N\widetilde{\mu})^2 \\ &= N^{-\zeta+1}E_1(\widetilde{\ell}_{nk} - \widetilde{\mu})^2 \le N^{-\zeta+1}E_1\widetilde{\ell}^{2}_{nk} \to 0. \end{aligned} \tag{6.16}$$

By (6.15), noting that $\ell_{\bullet 1k} - \widetilde{\ell}_{\bullet 1k} = \sum_{n=1}^N \ell_{nk} \mathbf{I}_{\{Z_{nk} > \omega_N\}}$,

$$P_1\{\ell_{\bullet 1k} - \widetilde{\ell}_{\bullet 1k} \geq \sqrt{\alpha_N} N^\zeta\} \to 0. \tag{6.17}$$

It follows from (6.16) and (6.17) that $P_1\{\ell_{\bullet 1k} \geq \widehat{b}\} \to 0$ for $\widehat{b} = N\widetilde{\mu} + N^{\zeta/2} + \sqrt{\alpha_N} N^\zeta [= o(N^\zeta)$ by (6.11)], hence (6.5) holds.

Checking (6.11)–(6.14):

(a) $\frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}$ and $\rho(\beta, \zeta) = \beta - \frac{1-\zeta}{2}$. By Jensen's inequality, $E\ell_0 \leq \log E e^{\ell_0} = 0$, therefore, to show (6.11), it suffices to show that

$$pE\ell_1^+ = o(N^{\zeta-1}). \tag{6.18}$$

Indeed, as $\log(1+x) \leq x$, by (6.7),

$$pE\ell_1^+ \leq p^2 E e^{Y\mu\sqrt{k} + k\mu^2/2} = p^2 e^{k\mu^2} = O(N^{-2\beta + (1-\delta)(2\beta - 1 + \zeta)}), \tag{6.19}$$

and (6.18) holds.

To show (6.12), note that

$$\begin{aligned} \sup_{1 \leq n \leq N} \widetilde{\ell}_{nk}^+ &\leq p e^{\omega_N \mu\sqrt{k} - \mu k^2/2} = p e^{\omega_N^2/2 - (\omega_N - \mu\sqrt{k})^2/2} \\ &\sim N^{-\beta + 1 - \zeta} e^{-(w_N - \mu\sqrt{k})^2/2} \log N. \end{aligned} \tag{6.20}$$

Express $\beta = \frac{1-\zeta}{2} + \alpha(1-\zeta)$ for some $0 < \alpha < \frac{1}{4}$. Since $\rho(\beta, \zeta) = \alpha(1-\zeta)$ and $w_N \geq \sqrt{2(1-\zeta)\log N}$, by (6.7) there exists $\varepsilon > 0$ small such that

$$\begin{aligned} \frac{(w_N - \mu\sqrt{k})^2}{2\log N} &\geq (1-\zeta)(1-\sqrt{\alpha})^2 + \varepsilon \\ &= (1-\zeta)\left[\frac{(1-2\sqrt{\alpha})^2}{2} + \frac{1}{2} - \alpha\right] + \varepsilon \\ &\geq (1-\zeta)\left(\frac{1}{2} - \alpha\right) + \varepsilon \\ &= 1 - \zeta - \beta + \varepsilon. \end{aligned} \tag{6.21}$$

Substituting (6.21) into (6.20) shows (6.12).

To show (6.13), note that by (6.19),

$$E\ell_0^2 = O(p^2 e^{2Y\mu\sqrt{k} - k\mu^2}) = O(p^2 e^{k\mu^2}) = o(N^{\zeta-1}). \tag{6.22}$$

Since $\beta > \frac{1-\zeta}{2}$,

$$(\widetilde{\ell}_{nk}^-)^2 = O(p^2) = o(N^{\zeta-1}), \tag{6.23}$$

and (6.13) follows from (6.12), (6.18) and (6.22).

Finally, to show (6.14), note that $P\{Y > \omega_N\} = o(N^{\zeta-1}/\log N)$, and that by (6.21),

$$
\begin{aligned}
pP\{Y+\mu\sqrt{k} > \omega_N\} &= O\big(N^{-\beta}e^{-(\omega_N-\mu\sqrt{k})^2/2}\big) \\
&= o\big(N^{\zeta-1}/\log N\big).
\end{aligned}
\tag{6.24}
$$

(b) $\frac{3(1-\zeta)}{4} < \beta < 1-\zeta$ and $\rho(x,y) = (x-y)^2$, where $x = \sqrt{1-\zeta}$, $y = \sqrt{1-\zeta-\beta}$. By $\log(1+v) \le v$,

$$
\begin{aligned}
pE\big(\ell_1^+ \mathbf{I}_{\{Y+\mu\sqrt{k}\le\omega_N\}}\big) &\le p^2 \int_{-\infty}^{\omega_N-\mu\sqrt{k}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2+z\mu\sqrt{k}+k\mu^2/2}\,dz \\
&= p^2 e^{k\mu^2}\Phi(\omega_N - 2\mu\sqrt{k}).
\end{aligned}
\tag{6.25}
$$

Since $\omega_N \sim x\sqrt{2\log N}$, $\mu\sqrt{k} = (1-\delta)(x-y)\sqrt{2\log N} + O(1)$ and $x > 2y$, it follows that $\omega_N < 2\mu\sqrt{k}$ for $\delta > 0$ small and, therefore,

$$
\begin{aligned}
p^2 e^{k\mu^2}\Phi(\omega_N - 2\mu\sqrt{k}) &= O\big(p^2 e^{k\mu^2-(\omega_N-2\mu\sqrt{k})^2/2}\big) \\
&= O\big(p^2 e^{\omega_N^2/2-(\omega_N-\mu\sqrt{k})^2}\big) \\
&= O\big(N^{-2\beta+x^2-2y^2-\varepsilon}\big) = O\big(N^{\zeta-1-\varepsilon}\big)
\end{aligned}
\tag{6.26}
$$

for some $\varepsilon > 0$, since $-2\beta + x^2 - 2y^2 = \zeta - 1$, and since $E\ell_0 \le 0$, (6.11) follows from (6.25) and (6.26).

By the first line of (6.20),

$$
\widetilde{\ell}_{nk} \le p e^{\omega_N\mu\sqrt{k}-\mu k^2/2} = O\big(N^{-\beta+x^2-y^2-\varepsilon}\big)
$$

for some $\varepsilon > 0$, therefore, (6.12) holds.

Note that by (6.9) and $\log(1+v) \le v$,

$$
\begin{aligned}
E\big(\ell_0^2 \mathbf{I}_{\{Y\le\omega_N\}}\big) &= O\left(p^2 \int_{-\infty}^{\omega_N} \frac{1}{\sqrt{2\pi}} e^{-z^2/2+2z\mu\sqrt{k}-k\mu^2}\,dz\right) \\
&= O\big(p^2 e^{k\mu^2}\Phi(\omega_N - 2\mu\sqrt{k})\big),
\end{aligned}
$$

and (6.13) follows from (6.12), (6.23), (6.25) and (6.26). It is easy to check that (6.24), and hence (6.14), holds in this case.

7. Proof of Theorem 2. The following lemma provides an upper bound for the threshold of the detectability score stopping rule.

LEMMA 4. *Consider stopping rule $T_S(p_0)$, $0 < p_0 \le 1$, with arbitrary window-sizes $\mathcal{K}$. If the stopping rule threshold $b = \log(4\gamma^2 + 2\gamma)$, then $E_\infty T_S(p_0) \ge \gamma$.*

PROOF. It suffices to show that

$$P_\infty\{T_S(p_0) < 2\gamma\} \le \frac{1}{2}. \tag{7.1}$$

Let $Z_{nst} = S_{nst}/\sqrt{k}$, $k = t - s + 1$. Since $V_{st} := \sum_{n=1}^N g(Z_{nst}^+)$ is a log likelihood ratio against $Z_{1st}, \dots, Z_{Nst}$ i.i.d. N(0, 1), it follows from a change of measure argument that

$$P_\infty\{V_{st} \ge b\} \le e^{-b} = (4\gamma^2 + 2\gamma)^{-1}.$$

By Bonferroni's inequality,

$$P_\infty\{T_S < 2\gamma\} \le \sum_{(s,t):1\le s\le t<2\gamma} P_\infty\{V_{st} \ge b\} \le \binom{\lfloor 2\gamma + 1\rfloor}{2}(4\gamma^2 + 2\gamma)^{-1},$$

and (7.1) follows. □

Assume (2.3), (2.4) and let the threshold b [$\le \log(4\gamma^2 + 2\gamma)$ by Lemma 4] be such that $E_\infty T_S(p_0) = \gamma$. Let $\eta = \min_{m\in J_N} P_1\{\sum_{n=1}^N g(Z_{n1k}^+) \ge b|\#\mathcal{N} = m\}$, where

$$J_N = \begin{cases} \left\{m : m \ge \frac{1}{2}Np\right\}, & \text{if } \beta < 1 - \zeta, \\ \{m\} \text{ for some } m \sim N^{1-\beta}, & \text{if } \beta > 1 - \zeta. \end{cases} \tag{7.2}$$

We show in Appendix C, using large deviations theory, that for N large

$$P_1\left\{\#\mathcal{N} < \frac{1}{2}Np\right\} \le \exp\left(-\frac{1}{8}Np\right). \tag{7.3}$$

We shall show in various cases below that $\eta \to 1$ when

$$k = \begin{cases} 1, & \text{if } \beta < \frac{1-\zeta}{2}, \\ \lfloor (1+\delta)2\mu^{-2}\rho(\beta,\zeta)\log N\rfloor, & \text{if } \frac{1-\zeta}{2} < \beta < 1 - \zeta, \\ \lfloor MN^{\beta+\zeta-1}\rfloor, & \text{if } \beta > 1 - \zeta, \end{cases} \tag{7.4}$$

for all $\delta > 0$, and M large. For $j \ge 1$ and $m \in J_N$, $P_1\{T_S(p_0) \ge jk + 1|\#\mathcal{N} = m\} \le (1-\eta)^j$. Hence, if $\beta < 1 - \zeta$, then $Np \sim N^{1-\beta} \gg N^\zeta$ and it follows from (7.3) that

$$D_N(T_S(p_0)) \le k\sum_{j=0}^\infty (1-\eta)^j + \gamma P_1\{\#\mathcal{N} \notin J_N\} \sim k. \tag{7.5}$$

By similar arguments, if $m \sim N^{1-\beta}$ for $\beta > 1 - \zeta$, then $D_N^{(m)}(T_S(p_0)) \le [1 + o(1)]k$. The proof of Theorem 2 is thus complete.

Let $V_N = \sum_{n=1}^N g(Y_n^+)$ for $Y_1, \dots, Y_N$ i.i.d. N(0, 1).

LEMMA 5. *If* $p_0 \sim cN^{(\zeta-1)/2}$, *then* $P\{V_N \ge -N^\zeta\} \to 1$.

PROOF. Let $\bar\Phi(z) = \int_z^\infty \phi(y)\,dy$ where $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$, and $\widetilde{g}(z) = g(z)\mathbf{I}_{\{z \le w_N\}}$ where $w_N = \sqrt{2(1-\zeta)\log N - (\log\log N)/2}$. By $\log(1+x) \sim x$ and $p_0 e^{w_N^2/4} \to 0$,

$$
\begin{aligned}
E\widetilde{g}(Y_1) &\sim cN^{(\zeta-1)/2} \int_{-\infty}^{w_N} (\lambda e^{z_+^2/4} - 1)\phi(z)\,dz \\
&= cN^{(\zeta-1)/2}\left[\frac{\lambda}{2} + \lambda\sqrt{2}\int_0^{w_N} \frac{1}{\sqrt{4\pi}} e^{-z^2/4}\,dz - \Phi(w_N)\right] \qquad (7.6)\\
&= cN^{(\zeta-1)/2}\left\{\left[\frac{\lambda}{2} - \Phi(w_N)\right] + \lambda\sqrt{2}\left[\frac{1}{2} - \bar\Phi\left(\frac{w_N}{\sqrt{2}}\right)\right]\right\}.
\end{aligned}
$$

Since $\lambda = 2(\sqrt{2}-1)$ solves $\frac{\lambda}{2} + \frac{\lambda}{\sqrt{2}} = 1$ and $\bar\Phi(w_N) \le \bar\Phi(\frac{w_N}{\sqrt{2}}) = o(N^{(\zeta-1)/2})$, by (7.6),

$$
|E\widetilde{g}(Y_1)| = o(N^{\zeta-1}). \qquad (7.7)
$$

Since

$$
E\widetilde{g}^2(Y_1) \sim c^2 N^{\zeta-1} \int_{-\infty}^{w_N} (\lambda e^{z_+^2/4} - 1)^2 \phi(z)\,dz = O\left(N^{\zeta-1}\sqrt{\log N}\right),
$$

and $g \ge \widetilde{g}$, we conclude Lemma 5 from (7.7) and Chebyshev's inequality. $\square$

Let $h(z) = g((z+\mu\sqrt{k})^+) - g(z^+)(\ge 0)$ and $H_N = \sum_{n\in\mathcal{N}} h(Y_n)$. Then

$$
\eta = \min_{m\in J_N} P\{V_N + H_N \ge b \mid \#\mathcal{N} = m\}.
$$

In view of Lemmas 4 and 5, to show $\eta \to 1$ and hence (7.5), it suffices to show that

$$
\min_{m\in J_N} P\{H_N \ge 4N^\zeta \mid \#\mathcal{N} = m\} \to 1. \qquad (7.8)
$$

We shall check (7.8) in three cases below. For notational simplicity, we shall let C denote a generic positive constant.

Case 0: $\beta < \frac{1-\zeta}{2}$ and $k = 1$. Since $\log(1+x) \sim x$ as $x \to 0$,

$$
h(z) \sim c\lambda N^{(\zeta-1)/2}\left(e^{(z+\mu)^2/4} - e^{z^2/4}\right) \ge c\lambda N^{(\zeta-1)/2}\left(e^{\mu^2/4} - 1\right),
$$

uniformly over $0 \le z \le 1$. Hence, by LLN,

$$
H_N \ge [C + o_p(1)]N^{1-\beta+(\zeta-1)/2},
$$

and (7.8) holds because $1 - \beta + \frac{\zeta-1}{2} > \zeta$.

Case 1: $\frac{1-\zeta}{2} < \beta < 1 - \zeta$ and $k = \lfloor (1+\delta)2\mu^{-2}\rho(\beta,\zeta)\log N \rfloor$. We shall show (7.8) for the following sub-cases:

(a) $\frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}$ and $\rho(\beta,\zeta) = \beta - \frac{1-\zeta}{2}$. For $\delta > 0$ small and $z \geq \mu\sqrt{k}$,

$$
\begin{aligned}
h(z) &\sim \log\bigl[1 + cN^{(\zeta-1)/2}\bigl(\lambda e^{(z+\mu\sqrt{k})^2/4} - 1\bigr)\bigr] \\
&\quad - \log\bigl[1 + cN^{(\zeta-1)/2}\bigl(\lambda e^{z^2/4} - 1\bigr)\bigr] \qquad (7.9)\\
&\geq [c\lambda + o(1)]N^{(\zeta-1)/2}e^{\mu^2 k}.
\end{aligned}
$$

Since $P\{Y_n \geq \mu\sqrt{k}\} \geq Ce^{-\mu^2 k/2}/\sqrt{\log N}$, by (7.9) and LLN,

$$
\begin{aligned}
H_N &\geq [C + o_p(1)]N^{1-\beta+(\zeta-1)/2}e^{\mu^2 k/2}/\sqrt{\log N} \\
&\geq [C + o_p(1)]N^{1-\beta+(\zeta-1)/2+\rho(\beta,\zeta)+\varepsilon}/\sqrt{\log N}
\end{aligned}
$$

for some $\varepsilon > 0$, and (7.8) holds because $1 - \beta + \frac{\zeta-1}{2} + \rho(\beta,\zeta) = \zeta$.

(b) $\frac{3(1-\zeta)}{4} < \beta < 1 - \zeta$ and $\rho(\beta,\zeta) = (x-y)^2$, where $x = \sqrt{1-\zeta}$, $y = \sqrt{1-\zeta-\beta}$. Since $\mu^2 k = 2(1+\delta)(x-y)^2 \log N + O(1)$, by the first relation in (7.9), $h(z) \geq C$ for $z \geq \sqrt{2(y^2-\varepsilon)\log N}$ with $\varepsilon > 0$ small. Since $P\{Y_n \geq \sqrt{2(y^2-\varepsilon)\log N}\} \geq CN^{-y^2+\varepsilon}/\sqrt{\log N}$, by LLN,

$$
H_N \geq [C + o_p(1)]N^{1-\beta-y^2+\varepsilon}/\sqrt{\log N},
$$

and (7.8) holds because $1 - \beta - y^2 = \zeta$.

Case 2: $\beta > 1 - \zeta$ and $k = \lfloor MN^{\beta+\zeta-1}\rfloor$. By the first relation in (7.9), for $z \geq 0$,

$$
h(z) \geq \frac{k\mu^2}{4} + O(\log N) = \left[\frac{M\mu^2}{4} + o(1)\right]N^{\beta+\zeta-1}.
$$

By LLN, $H_N \geq [\frac{M\mu^2}{8} + o_p(1)]N^{\zeta}$, and (7.8) holds for $M > 32\mu^{-2}$.

8. Proof of Theorem 3. In Lemma 6 below, we provide an upper bound of the detection threshold of the extended Mei's stopping rule, and follow this with conditions under which this bound is exceeded under P_ν. We complete the proof by checking these conditions for various cases. Let

$$
\begin{aligned}
g_0(x) = g_M(x) - u \qquad &\text{where } g_M(x) = \log\bigl[1 + p_0\bigl(\lambda_M e^{x/2} - 1\bigr)\bigr], \\
&u = \log\bigl[1 + p_0(\lambda_M \xi - 1)\bigr],
\end{aligned}
$$

and $\xi = \lim_{t\to\infty} Ee^{R_{nt}/2}$.

LEMMA 6. *Consider stopping rule $T_{\mathrm{Mei}}(p_0)$, $0 < p_0 \leq 1$. If the stopping rule threshold $b = Nu + \log(4\gamma)$, then $E_\infty T_{\mathrm{Mei}}(p_0) \geq \gamma$.*

PROOF. If $b = Nu + \log(4\gamma)$, then

$$T_{\text{Mei}}(p_0) = \inf\left\{t : \sum_{n=1}^{N} g_0(R_{nt}) \geq \log(4\gamma)\right\}.$$

Let $S_j = \sum_{i=1}^{j} Y_i$ with Y_i i.i.d. N(0, 1), and let

$$R = \sup_{j \geq 0}\bigl(\mu_0 S_j - j\mu_0^2/2\bigr).$$

Let $R_1, \dots, R_N$ be an i.i.d. sample with the distribution of R. Let $1 \leq t < 2\gamma$. Since R_{nt} is bounded stochastically by R_n, it follows from $Ee^{g_0(R_n)} = 1$, a change of measure argument and g_0 monotone that

$$P_\infty\left\{\sum_{n=1}^{N} g_0(R_{nt}) \geq \log(4\gamma)\right\} \leq P\left\{\sum_{n=1}^{N} g_0(R_n) \geq \log(4\gamma)\right\} \leq (4\gamma)^{-1}.$$

Therefore, $\{T_{\text{Mei}}(p_0) < 2\gamma\}$ is a union of no more than 2γ events, each with probability bounded by $(4\gamma)^{-1}$ under P_∞. We conclude that $P_\infty\{T_{\text{Mei}}(p_0) < 2\gamma\} \leq \frac{1}{2}$. Hence, $E_\infty T_{\text{Mei}}(p_0) \geq \gamma$. $\square$

Let $\nu \geq K_n$ and $t = \nu + k - 1$, where $k = \lfloor (1+\delta) 2\mu^{-2}\rho(\beta, \zeta)\log N\rfloor$ for $\delta > 0$ small. Let $U_n = \mu_0 \sum_{i=\nu}^{t} X_{ni} - k\mu_0^2/2$ ($\leq R_{nt}$). Under P_ν, $U_n \sim \mathrm{N}(k\mu_0(\mu - \frac{\mu_0}{2}), k\mu_0^2)$ when $n \in \mathcal{N}$. Theorem 3 follows from

$$\inf_{m \geq Np/2} P_\nu\left\{\sum_{n \notin \mathcal{N}} g_0(R_{nt}) \geq -N^{\zeta+\varepsilon} \,|\, \#\mathcal{N} = m\right\} \to 1, \tag{8.1}$$

$$\inf_{m \geq Np/2} J_N P_\nu\left\{\sum_{n \in \mathcal{N}} g_0\bigl(U_n^+\bigr) \geq 2N^{\zeta+\varepsilon} \,|\, \#\mathcal{N} = m\right\} \to 1, \tag{8.2}$$

for some $\varepsilon > 0$, see (7.3).

The following lemma provides the framework for showing (8.1) and (8.2). Let $\widetilde{g}_0(y) = g_0(y)\mathbf{I}_{\{y \leq \upsilon_N\}}$, where

$$\upsilon_N = (1-\zeta)\log N - \log\log N. \tag{8.3}$$

LEMMA 7. *If* $t \geq 4\mu_0^{-2}(1-\zeta)\log N$ *and* $n \notin \mathcal{N}$, *then for all* $\varepsilon > 0$,

$$E_\infty\bigl(e^{R_{nt}/2}\mathbf{I}_{\{R_{nt} \leq \upsilon_N\}}\bigr) = \xi + o\bigl(N^{(\zeta-1)/2+\varepsilon}\bigr). \tag{8.4}$$

Moreover, if $p_0 \sim cN^{(\zeta-1)/2}$ *with* $c > 0$, *then*

$$\Bigl[\inf_{y \geq 0} \widetilde{g}_0(y)\Bigr]^2 = O\bigl(N^{\zeta-1}\bigr), \tag{8.5}$$

$$\sup_{y \geq 0} \widetilde{g}_0(y) = O(1). \tag{8.6}$$

PROOF. The relation (8.5) follows from

$$\left|\inf_{y\geq 0}\widetilde{g}_0(y)\right| = |g_0(0)| = O(p_0) = O\big(N^{(\zeta-1)/2}\big),$$

whereas (8.6) follows from $\sup_{y\geq 0}\widetilde{g}_0(y) = g_0(v_N) = O(1)$.

By (5.1), we can express

$$R_{nt} = \sup_{1\leq s\leq t}[\mu_0 S_{nst} - (t-s+1)\mu_0^2/2]^+.$$

Extend $\{X_{nu} : u \geq 1\}$ to $\{X_{nu} : -\infty < u < \infty\}$ by letting X_{nu} i.i.d. N(0, 1) under P_∞ for $u \leq 0$. Fix t and let

$$R_n^* = \sup_{-\infty<s\leq t}\big[\mu_0 S_{nst} - (t-s+1)\mu_0^2/2\big]^+,$$

extending the definition of $S_{nst} = \sum_{i=s}^t X_{ni}$ to $s \leq 0$.

Since $\xi = \lim_{t\to\infty} E_\infty e^{R_{nt}/2}$, to show (8.4), it suffices to show that

$$E_\infty\big(e^{R_{nt}/2}\mathbf{I}_{\{R_{nt}>v_N\}}\big) = o\big(N^{(\zeta-1)/2+\varepsilon}\big), \tag{8.7}$$

$$E_\infty\big(e^{R_n^*/2}\mathbf{I}_{\{R_n^*>R_{nt}\}}\big) = o\big(N^{(\zeta-1)/2+\varepsilon}\big). \tag{8.8}$$

We conclude (8.7) from (5.5) and (8.3). Let $Q = \sup_{j\geq t}(\mu_0 S_j - j\mu_0^2/2)$ and $R' = \sup_{j\geq 0}(\omega S_j - j\omega^2\mu_0^2/2)$ for some $\omega > \frac{1}{2}$. By (5.5), for $x \geq 0$,

$$\begin{aligned}P_\infty\{R_n^* > R_{nt}, R_n^* \geq x\} &\leq P\{Q \geq x\} \leq P\{R' \geq \omega x + t(\omega-\omega^2)\mu_0^2/2\}\\ &= O\big(e^{-\omega x - t(\omega-\omega^2)\mu_0^2/2}\big).\end{aligned}$$

Hence, by selecting ω close enough to $\frac{1}{2}$, it follows that

$$E_\infty\big(e^{R_n^*/2}\mathbf{I}_{\{R_n^*>R_{nt}\}}\big) = \int_{-\infty}^\infty \frac{1}{2}e^{x/2}P_\infty\{R_n^* > R_{nt}, R_n^* \geq x\}\,dx = O\big(e^{-t\mu_0^2/8+\varepsilon}\big),$$

and (8.8) holds for $t \geq 4\mu_0^{-2}(1-\zeta)\log N$. $\square$

We conclude (8.1) from (8.4), $p_0 \sim cN^{(\zeta-1)/2}$ and LLN. We note that indeed $t(=\nu+k-1) \geq 4\mu_0^{-2}(1-\zeta)\log N$ when

(a) $\frac{1-\zeta}{2} < \beta \leq \frac{3(1-\zeta)}{4}$, $\rho(\beta,\zeta) = \beta - \frac{1-\zeta}{2}$, $\mu_0 = 2\mu$, $\nu \geq 2\mu^{-2}(1-\zeta-\beta)\log N$,

(b) $\frac{3(1-\zeta)}{4} < \beta < 1-\zeta$, $\mu_0 = \mu\sqrt{\frac{1-\zeta}{\rho(\beta,\zeta)}}$, $\nu \geq 2\mu^{-2}\rho(\beta,\zeta)\log N$.

It remains for us to check (8.2) on:

(a) $\frac{1-\zeta}{2} < \beta < \frac{3(1-\zeta)}{4}$. Since $\mu_0 = 2\mu$, we have $U_n \sim \mathrm{N}(0, k\mu_0^2)$ when $n \in \mathcal{N}$. Hence,

$$\begin{aligned}&E_\nu\big[\widetilde{g}_0\big(U_n^+\big)|n\in\mathcal{N}\big]\\ &\quad\sim p_0\big[E_\nu\big(e^{U_n^+/2}\mathbf{I}_{\{U_n\leq v_N\}}|n\in\mathcal{N}\big) - \xi\big]\end{aligned} \tag{8.9}$$

$$\sim cN^{(\zeta-1)/2}e^{k\mu_0^2/8}\int_{-\infty}^{v_N}\frac{1}{\sqrt{2k\pi\mu_0^2}}e^{-(y-k\mu_0^2/2)^2/(2k\mu_0^2)}\,dy$$
$$= cN^{(\zeta-1)/2}e^{k\mu_0^2/8}\Phi\left(\frac{v_N-k\mu_0^2/2}{\mu_0\sqrt{k}}\right).$$

Check that $e^{k\mu_0^2/8} = N^{[1+\delta+o(1)]\rho(\beta,\zeta)} \geq N^{\beta+(\zeta-1)/2+2\varepsilon}$ for $\varepsilon > 0$ small and N large. Moreover, $\rho(\beta,\zeta) < \frac{1-\zeta}{4}$, therefore, $v_N > \frac{k\mu_0^2}{2}$ $[\sim 4(1+\delta)\rho(\beta,\zeta)\log N]$ for $\delta > 0$ small. Hence, by (8.9),

$$E_\nu[\widetilde{g}_0(U_n^+)|n \in \mathcal{N}] \sim [c+o(1)]N^{\beta+\zeta-1+2\varepsilon}. \tag{8.10}$$

By (8.5), (8.6) and (8.10), we can conclude

$$E_\nu[\widetilde{g}_0^2(U_n^+)|n \in \mathcal{N}] = O(|E_\nu[\widetilde{g}_0(U_n^+)|n \in \mathcal{N}]|),$$

and (8.2) then follows from (8.10), Chebyshev's inequality and $g_0 \geq \widetilde{g}_0$.

(b) $\frac{3(1-\zeta)}{4} \leq \beta < 1-\zeta$. For $n \in \mathcal{N}$, express $U_n = k\mu_0(\mu - \frac{\mu_0}{2}) + \sqrt{k}\mu_0 Y_n$, with $Y_n \sim \mathrm{N}(0,1)$. Let $\mathcal{N}_1 = \{n \in \mathcal{N} : Y_n \geq \sqrt{2(1-\zeta-\beta-2\varepsilon)\log N}\}$ for $\varepsilon > 0$ satisfying

$$1-\zeta-\beta-2\varepsilon \geq (1-\zeta-\beta)/(1+\delta). \tag{8.11}$$

By LLN, if $\#\mathcal{N} \geq \frac{1}{2}Np$, then

$$\begin{aligned}\#\mathcal{N}_1 &= (\#\mathcal{N})[C+o_p(1)]N^{\zeta+\beta-1+2\varepsilon}/\sqrt{\log N}\\ &\geq \left[\frac{C}{2}+o_p(1)\right]N^{\zeta+2\varepsilon}/\sqrt{\log N}.\end{aligned} \tag{8.12}$$

Let $r = \mu_0/\mu$ $(= \sqrt{\frac{1-\zeta}{\rho(\beta,\zeta)}} \leq 2)$. Since $r-1 = \sqrt{\frac{1-\zeta-\beta}{\rho(\beta,\zeta)}}$, by (7.4) and (8.11), for $n \in \mathcal{N}_1$ with N large,

$$\begin{aligned}U_n &\geq k\mu^2\left(r-\frac{r^2}{2}\right)+\mu r(r-1)\sqrt{\frac{2k\rho(\beta,\zeta)\log N}{1+\delta}}\\ &= 2\rho(\beta,\zeta)\log N\left[(1+\delta)\left(r-\frac{r^2}{2}\right)+r^2-r\right]+O(1)\\ &\geq r^2\rho(\beta,\zeta)\log N = (1-\zeta)\log N,\end{aligned}$$
$$g_0(U_n^+) \geq \log[1+p_0(N^{(\zeta-1)/2}-1)] - \log[1+p_0(\xi-1)] \to \log(1+c).$$

Hence, by (8.12),

$$\sum_{n\in\mathcal{N}_1} g_0(U_n^+) \geq \left[\frac{C}{2}+o_p(1)\right]N^{\zeta+2\varepsilon}\Big/\sqrt{\log N}.$$

This, combined with

$$\sum_{n \in \mathcal{N} \setminus \mathcal{N}_1} g_0(U_n^+) \geq -[C + o_p(1)] N^{1-\beta} \log p_0$$

$$\sim -[C + o_p(1)] N^{1-\beta+(\zeta-1)/2},$$

and noting that $1 - \beta + \frac{\zeta - 1}{2} \leq 1 - \frac{5}{4}(1 - \zeta) < \zeta$, shows (8.2).

APPENDIX A: MINIMUM DETECTION DELAY UNDER THE MINIMAX SETTING

Let $E_{\nu,\mathcal{N}}$ denote expectation with respect to $X_{nt} \sim \mathrm{N}(\mu_{nt}, 1)$, with $\mu_{nt} = \mu \mathbf{I}_{\{t \geq \nu, n \in \mathcal{N}\}}$ for some $\mu > 0$. For a given stopping rule T, define

$$D_N^{(m)}(T) = \sup_{1 \leq \nu < \infty} \Big[\max_{\mathcal{N}:\#\mathcal{N}=m} E_{\nu,\mathcal{N}}(T - \nu + 1 | T \geq \nu) \Big].$$

THEOREM 4. *Let T be a stopping rule such that* $\mathrm{ARL}(T) \geq \gamma$, *with* $\log \gamma \sim N^\zeta$ *for some* $\zeta > 0$. *Let* $m \sim N^{1-\beta}$ *for some* $0 < \beta < 1$.

(a) *If* $\frac{1-\zeta}{2} < \beta < 1 - \zeta$, *then*

$$\liminf_{N \to \infty} \frac{D_N^{(m)}(T)}{\log N} \geq 2\mu^{-2} \rho(\beta, \zeta).$$

(b) *If* $\beta > 1 - \zeta$, *then*

$$\liminf_{N \to \infty} \frac{\log D_N^{(m)}(T)}{\log N} \geq \beta + \zeta - 1.$$

PROOF. Let

$$k = \begin{cases} \lfloor (1-\delta) 2\mu^{-2} \rho(\beta, \zeta) \log N \rfloor, & \text{if } \dfrac{1-\zeta}{2} < \beta < 1 - \zeta, \\ \lfloor \delta N^{\beta+\zeta-1} \rfloor, & \text{if } \beta > 1 - \zeta, \end{cases} \tag{A.1}$$

with $\delta > 0$ small. By Lemma 1, we can find $s \geq 1$ such that

$$P_\infty\{T \geq s + k | T \geq s\} \geq 1 - k/\gamma. \tag{A.2}$$

Let $t = s + k - 1$, and consider the test, conditional on $T \geq s$, of

$$\begin{aligned} H_0: \quad & X_{nu} \sim \mathrm{N}(0,1) \qquad \text{for } 1 \leq n \leq N, 1 \leq u \leq t, \\ \text{vs } H_{s,m}: \quad & X_{nu} \sim \mathrm{N}(\mu \mathbf{I}_{\{u \geq s, n \in \mathcal{N}\}}, 1) \qquad \text{for } 1 \leq n \leq N, 1 \leq u \leq t, \\ & \text{with } \mathcal{N} \text{ a random subset of } \{1, \dots, N\} \text{ of size } m. \end{aligned}$$

By (A.2), the test rejecting H_0 when $T < s + k$ has Type I error probability not exceeding k/γ.

Let $\mathcal{A}_j = \{\mathcal{N} : \#\mathcal{N} = j\}$. At time t, the (conditional) likelihood ratio between $H_{s,m}$ and H_0 is $L_m (= L_{mst})$, where

$$L_j = \binom{N}{j}^{-1} \sum_{\mathcal{N} \in \mathcal{A}_j} \Bigl(\prod_{n \in \mathcal{N}} e^{Z_n \mu \sqrt{k} - k\mu^2/2} \Bigr), \qquad Z_n = Z_{nst}.$$

Let $P_{s,m}$ ($E_{s,m}$) denote probability (expectation) with respect to $H_{s,m}$.

We shall check on various cases below that

$$P_{s,m}\{L_m \geq J\} \to 0, \qquad J = \exp(2N^{\zeta}/3). \tag{A.3}$$

Let B be such that $P_{s,m}\{J \geq L_m \geq B\} = \exp(-N^{\zeta}/4)$. It follows from (A.3) that

$$P_{s,m}\{L_m \geq B\} \bigl(= P_{s,m}\{L_m \geq B | T \geq s\}\bigr) \to 0, \tag{A.4}$$

and that for N large,

$$\begin{aligned} P_{\infty}\{L_m \geq B\} &\bigl(= P_{\infty}\{L_m \geq B | T \geq s\}\bigr) \\ &\geq P_{\infty}\{J \geq L_m \geq B\} \\ &= E_{s,m}\bigl(L_m^{-1} \mathbf{I}_{\{J \geq L_m \geq B\}}\bigr) \geq J^{-1} \exp\bigl(-N^{\zeta}/4\bigr) \geq k/\gamma. \end{aligned} \tag{A.5}$$

By (A.2), (A.5) and the Neyman–Pearson lemma, the test rejecting H_0 when $L_m \geq B$ is at least as powerful as the one based on T, that is,

$$P_{s,m}\{T \geq s + k | T \geq s\} \geq P_{s,m}\{L_m < B\}. \tag{A.6}$$

It follows from (A.4) and (A.6) that

$$D_N^{(m)}(T) \geq E_{s,m}(T - s + 1 | T \geq s) \geq k P_{s,m}\{T \geq s + k | T \geq s\} = k\bigl[1 + o(1)\bigr],$$

and the proof of Theorem 4 is complete. $\square$

We shall now proceed to check (A.3). Let $p_1 = 2N^{-\beta}$ and

$$L(p_1) = \prod_{n=1}^{N} \bigl(1 - p_1 + p_1 e^{Z_n \mu \sqrt{k} - k\mu^2/2}\bigr) \left[= \sum_{j=0}^{N} (1 - p_1)^{N-j} p_1^j \binom{N}{j} L_j \right]. \tag{A.7}$$

Since $Z_n \sim \mathrm{N}(\mu\sqrt{k}, 1)$ if $n \in \mathcal{N}$ and $Z_n \sim \mathrm{N}(0, 1)$ if $n \notin \mathcal{N}$, it follows that

$$E e^{Z_n \mu \sqrt{k} - k\mu^2/2} = \begin{cases} e^{k\mu^2}, & \text{if } n \in \mathcal{N}, \\ 1, & \text{if } n \notin \mathcal{N}. \end{cases}$$

Therefore, by (A.7),

$$E_{s,m} L(p_1) = \bigl(1 - p_1 + p_1 e^{k\mu^2}\bigr)^m, \tag{A.8}$$

the exponent m in (A.8) due to $\#\mathcal{N} = m$ for each $\mathcal{N}$ under $H_{s,m}$. By the monotonicity $E_{s,m}L_1 \le \cdots \le E_{s,m}L_N$, and by $P\{W \ge m\} \to 1$ for $W \sim \text{Binomial}(N, p_1)$, it follows from (A.7) that

$$E_{s,m}L(p_1) \ge P\{W \ge m\}E_{s,m}L_m = [1 + o(1)]E_{s,m}L_m. \tag{A.9}$$

By (A.8), (A.9) and Markov's inequality, to show (A.3) it suffices to show that

$$(1 - p_1 + p_1 e^{k\mu^2})^m = o(\exp(2N^\zeta/3)), \tag{A.10}$$

and this can be easily done for the following cases.

Case 1(a): $\frac{1-\zeta}{2} < \beta \le \frac{3(1-\zeta)}{4}$, $k = \lfloor (1-\delta)\mu^{-2}(2\beta + \zeta - 1)\log N \rfloor$. We show (A.10) by applying the inequality:

$$(1 - p_1 + p_1 e^{k\mu^2})^m \le \exp(mp_1 e^{k\mu^2}).$$

Case 2: $\beta > 1 - \zeta$, $k = \lfloor \delta N^{\beta+\zeta-1} \rfloor$, $\delta > 0$ small. We show (A.10) by applying the inequalities (for large N):

$$(1 - p_1 + p_1 e^{k\mu^2})^m \le (2p_1 e^{k\mu^2})^m \le e^{k\mu^2 m}.$$

The final case below is more complicated. Additional truncation arguments are needed to show (A.3).

Case 1(b): $\frac{3(1-\zeta)}{4} < \beta < 1 - \zeta$, $k = \lfloor (1-\delta)2\mu^{-2}(x-y)^2 \log N \rfloor$, where $x = \sqrt{1-\zeta}$ and $y = \sqrt{1 - \zeta - \beta}$. The outline of the arguments needed to show (A.3) is as follows:

1. Let $\widetilde{Z}_n = \min(Z_n, \omega)$, where

$$\omega(= \omega_N) = \sqrt{2(1-\zeta)\log N + 2\log\log N} (\doteq x\sqrt{2\log N}).$$

Let $p_1 = 2N^{-\beta}$ and

$$\widetilde{L}(p_1) = \prod_{n=1}^{N} (1 - p_1 + p_1 e^{\tilde{Z}_n \mu\sqrt{k} - k\mu^2/2}). \tag{A.11}$$

Show that $E_{s,m}\widetilde{L}(p_1) = o(J^{1/2})[= o(\exp(N^\zeta/3))]$.

2. Argue that we have monotonicity $E_{s,m}\widetilde{L}_1 \le \cdots \le E_{s,m}\widetilde{L}_N$, where

$$\widetilde{L}_j = \binom{N}{j}^{-1} \sum_{\mathcal{N} \in \mathcal{A}_j} \Bigl(\prod_{n \in \mathcal{N}} e^{\tilde{Z}_n \mu\sqrt{k} - k\mu^2/2}\Bigr),$$

and conclude that

$$E_{s,m}\widetilde{L}(p_1) \ge P\{W \ge m\}E_{s,m}\widetilde{L}_m = [1 + o(1)]E_{s,m}\widetilde{L}_m, \tag{A.12}$$

where $W \sim \text{Binomial}(N, p_1)$.

3. Let $C > 0$ and $\widehat{L}_m = L_m \mathbf{I}_G$, where $G(= G_N)$ is the event that
$$\max_{1\le n\le N} Z_n \le C\sqrt{\log N}, \qquad F_N := \#\{n : Z_n > \omega\} \le N^{\zeta}/(\log N)^{5/4}.$$
Show that uniformly under G,
$$\max_{\mathcal{N}\in\mathcal{A}_m}\Bigl(\prod_{n\in\mathcal{N}} e^{(Z_n-\tilde{Z}_n)\mu\sqrt{k}}\Bigr) = o(J^{1/2})[= o(\exp(N^{\zeta}/3))],$$
and conclude that $\widehat{L}_m/\widetilde{L}_m = o(J^{1/2})$.
4. Show that for C large, $P_{s,m}(G_N) \to 1$ and so $P_{s,m}\{L_m > \widehat{L}_m\} \to 0$.

By steps 1, 2 and Markov's inequality, $P_{s,m}\{\widetilde{L}_m \ge J^{1/2}\} \to 0$. By step 3, we can further conclude that $P_{s,m}\{\widehat{L}_m \ge J\} \to 0$, and (A.3) then follows from step 4. We shall now provide details to the above outline:

1. If $n \notin \mathcal{N}$, then $Ee^{\tilde{Z}_n\mu\sqrt{k}-k\mu^2/2} \le 1$, and if $n \in \mathcal{N}$, then
$$\begin{aligned} Ee^{\tilde{Z}_n\mu\sqrt{k}-k\mu^2/2} &= e^{k\mu^2}\Phi(\omega - 2\mu\sqrt{k}) + [1 - \Phi(\omega - \mu\sqrt{k})]e^{\omega\mu\sqrt{k}-k\mu^2/2} \\ &= o(N^{2(x-y)^2-(2y-x)^2}) + o(N^{-y^2+2x(x-y)-(x-y)^2}) \\ &= o(N^{x^2-2y^2}). \end{aligned}$$
Since $\#\mathcal{N} = m$ for each $\mathcal{N}$ under $H_{s,m}$, by (A.11),
$$E_{s,m}\widetilde{L}(p_1) \le [1 + p_1 o(N^{x^2-2y^2})]^m \le \exp[mp_1 o(N^{x^2-2y^2})] = o(J^{1/2}).$$
2. The monotonicity follows from $\widetilde{Z}_n$ stochastically larger when $n \in \mathcal{N}$ compared to when $n \notin \mathcal{N}$, whereas the inequality in (A.12) follows from the monotonicity and the expansion:
$$\widetilde{L}(p_1) = \sum_{j=0}^{N}(1-p_1)^{N-j}p_1^j\binom{N}{j}\widetilde{L}_j.$$
3. Under G, there exists $\widetilde{C} > 0$ not depending on N such that for all $\mathcal{N} \in \mathcal{A}_m$,
$$\begin{aligned} \prod_{n\in\mathcal{N}} e^{(Z_n-\tilde{Z}_n)\mu\sqrt{k}} &\le \exp(F_N C\sqrt{\log N}\mu\sqrt{k}) \\ &\le \exp\Bigl(\frac{N^{\zeta}}{(\log N)^{5/4}}\cdot \widetilde{C}\log N\Bigr) = o(J^{1/2}). \end{aligned}$$
4. Let $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. We apply Markov's inequality to show $P_{s,m}(G_N) \to 1$ by checking that

$$m\bar{\Phi}(\omega - \mu\sqrt{k}) + (N-m)\bar{\Phi}(\omega) = o\Bigl(\frac{N^{\zeta}}{(\log N)^{5/4}}\Bigr), \tag{A.13}$$

and that for C large,

$$m\bar{\Phi}(C\sqrt{\log N} - \mu\sqrt{k}) + (N-m)\bar{\Phi}(C\sqrt{\log N}) \to 0. \tag{A.14}$$

By Mill's inequality, (A.14) holds for C large and $N\bar{\Phi}(\omega) = o(\frac{N^\zeta}{(\log N)^{5/4}})$. Moreover,

$$\limsup_{N\to\infty} \log_N\bigl[m\bar{\Phi}(\omega - \mu\sqrt{k})\bigr] < 1 - \beta - y^2 = \zeta,$$

and so (A.13) holds as well.

APPENDIX B: MOTIVATIONS BEHIND (6.7) AND (A.1)

In view of the need to satisfy (6.5), we choose k to be the "largest" possible such that

$$E_1\ell_{\bullet 1k} < 2N^\zeta/3. \tag{B.1}$$

Under P_1, $Z_{nk} \sim (1-p)\mathrm{N}(0,1) + p\mathrm{N}(\mu\sqrt{k}, 1)$. Let $Y \sim \mathrm{N}(0,1)$. Since

$$\log(1+x) \le x, \qquad Ee^{Y\mu\sqrt{k}} = e^{k\mu^2/2}, \tag{B.2}$$

it follows that

$$\begin{aligned} E_1\ell_{\bullet 1k} \bigl\{ &= \mathrm{N}(1-p)E\log\bigl[1 + p\bigl(e^{Y\mu\sqrt{k} - k\mu^2/2} - 1\bigr)\bigr] \\ &\quad + NpE\log\bigl[1 + p\bigl(e^{Y\mu\sqrt{k} + k\mu^2/2} - 1\bigr)\bigr]\bigr\} \\ &\le NpE\log\bigl[1 + p\bigl(e^{Y\mu\sqrt{k} + k\mu^2/2} - 1\bigr)\bigr]. \end{aligned} \tag{B.3}$$

Case 1(a): $\frac{1-\zeta}{2} < \beta \le \frac{3(1-\zeta)}{4}$. It follows from applying (B.2) on (B.3) that

$$E_1\ell_{\bullet 1k} \le Np^2\bigl(e^{k\mu^2} - 1\bigr) \sim N^{1-2\beta+(k\mu^2/\log N)}. \tag{B.4}$$

Hence, choosing $k = \lfloor(1-\delta)\mu^{-2}(2\beta + \zeta - 1)\log N\rfloor$ as in (6.7) ensures $E_1\ell_{\bullet 1k} = o(N^\zeta)$, and so (B.1) holds.

Case 1(b): $\frac{3(1-\zeta)}{4} < \beta < 1 - \zeta$. The inequality in (B.4) is further sharpened to allow for larger k satisfying (B.1). Let ω be the root of

$$e^{\omega\mu\sqrt{k} + k\mu^2/2} = N^\beta\bigl(\sim p^{-1}\bigr). \tag{B.5}$$

By (A5), applying the inequalities

$$\log(1+x) \le \begin{cases} x, & \text{if } -1 < x < 1, \\ \log 2 + \log x, & \text{if } x \ge 1, \end{cases} \tag{B.6}$$

on (B.3) results in

$$E_1\ell_{\bullet 1k} \le Np^2 \int_{-\infty}^{\omega} \frac{1}{\sqrt{2\pi}} e^{-z^2/2+z\mu\sqrt{k}+k\mu^2/2}\,dz + O(Npke^{-\omega^2/2})$$

$$= Np^2 e^{k\mu^2}\Phi(\omega-\mu\sqrt{k}) + O(Npke^{-\omega^2/2}) \tag{B.7}$$

$$= O(Npke^{-\omega^2/2}).$$

By (B.5),

$$\omega\mu\sqrt{k} + k\mu^2/2 = \beta \log N \left(\Rightarrow \quad \mu\sqrt{k} = -\omega + \sqrt{\omega^2 + 2\beta \log N}\right), \tag{B.8}$$

and by (B.7), we satisfy (B.1) if

$$(1-\beta)\log N - \omega^2/2 < \zeta \log N \left(\Rightarrow \quad \omega > \sqrt{2(1-\beta-\zeta)\log N}\right). \tag{B.9}$$

Combining (B.8) and (B.9) leads to $k < 2\mu^{-2}(\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2 \log N$. Hence, the choice of $k = \lfloor (1-\delta)2\mu^{-2}(\sqrt{1-\zeta} - \sqrt{1-\zeta-\beta})^2 \log N \rfloor$ in (6.7).

Case 2: $\beta > 1-\zeta$. By (B.1) and (B.6), choosing $k = \lfloor \delta N^{\beta+\zeta-1} \rfloor$ as in (A.1) ensures that

$$E_1\ell_{\bullet 1k} \le [1+o(1)]NpE(Y\mu\sqrt{k} + k\mu^2/2) \sim \delta\mu^2 N^{\zeta}/2,$$

and (B.1) indeed holds for $\delta > 0$ small.

APPENDIX C: PROOF OF (7.3)

Let Q be a probability measure under which $\#\mathcal{N} \sim \text{Binomial}(N, r)$, where $r = \frac{p}{2}$. Then

$$\begin{aligned} P_1\{\#\mathcal{N} < Nr\} &= \sum_{x:x<Nr} \frac{P_1\{\#\mathcal{N} = x\}}{Q\{\#\mathcal{N} = x\}} Q\{\#\mathcal{N} = x\} \\ &\le \left(\frac{p}{r}\right)^{Nr} \left(\frac{1-p}{1-r}\right)^{N(1-r)} Q\{\#\mathcal{N} < Nr\} \\ &\le \exp\{Nr\log 2 - [1+o(1)]Np + [1+o(1)]Nr\} \\ &= \exp\left\{[1+o(1)]Np\left(\frac{\log 2 - 1}{2}\right)\right\}, \end{aligned}$$

and (7.3) holds.

Acknowledgments. We thank an Associate Editor and the referees for their insights and helpful comments. The Appendices are due to questions and suggestions from the referees.

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] CHAN, H. P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37** 3985–4010. MR2572450

[4] CHAN, H. P. and WALTHER, G. (2015). Optimal detection of multi-sample aligned sparse signals. *Ann. Statist.* **43** 1865–1895. MR3375870

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

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

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

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

[9] LAI, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 613–658. MR1354072

[10] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. MR0309251

[11] MEI, Y. (2006). Sequential change-point detection when unknown parameters are present in the pre-change distribution. *Ann. Statist.* **34** 92–122. MR2275236

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

[13] MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387. MR0868306

[14] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. MR0773162

[15] POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779. MR0888438

[16] SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. MR0799155

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

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

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

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