Quick Search for Rare Events

Ali Tajer, Member, IEEE, and H. Vincent Poor, Fellow, IEEE

Abstract—Rare events can potentially occur in many applications. When manifested as opportunities to be exploited, risks to be ameliorated, or certain features to be extracted, such events become of paramount significance. Due to their sporadic nature, the information-bearing signals associated with rare events often lie in a large set of irrelevant signals and are not easily accessible. This paper provides a statistical framework for detecting such events so that an optimal balance between detection reliability and agility, as two opposing performance measures, is established. The core component of this framework is a sampling procedure that adaptively and quickly focuses the information-gathering resources on the segments of the dataset that bear the information pertinent to the rare events. Particular focus is placed on Gaussian signals with the aim of detecting signals with rare mean and variance values.

Index Terms—Agility, detection, quick, rare, search.

I. INTRODUCTION

THE problem of searching for scarce and, at the same time, significant events that have certain statistical behavior in a pool of information streams is a classical one and has attracted attention in a wide variety of fields over the past few decades. Such events can broadly model at least three categories of problems. One concentrates on seeking opportunities in a range of domains including trading in finance [1]–[3] and spectrum sensing in telecommunication [4], [5]. The second category pertains to minimizing risks and avoiding catastrophes in applications such as risk analysis in econometrics [6]–[11], blackout cascade avoidance in energy systems [12], [13], intrusion identification in network security [14]–[19], fraud detection [20], [21], and seismology. And the third category deals with extracting certain data segments with prespecified probabilistic features with applications in image and video analysis, medical diagnosis, neuroscience, and remote sensing (seismic, sonar, radar, biomedical) [22]–[24].

While detecting the rare events in some of these applications is time insensitive, in some other applications, time is of the essence and it is important to devise timely and reliable decision-making mechanisms. Such time sensitivity can be due to the transient nature of the opportunities that are attractive only when detected quickly, or due to the substantial costs that risks can incur if not detected and managed swiftly.

Each of these applications can be modeled as a collection of events that can be broadly categorized into two groups. One group constitutes the majority of the events, which are deemed normal and occur most of the time, and the other group consists of the events that occur rarely but are of extreme significance to the observer. In this paper, we assume that the group of all normal events share the same statistical behavior and the group of rare events also share identical statistical behavior, albeit different from that of the normal events. This dichotomous model is adopted to mainly focus the attention on the discrepancy between the rare and the normal events and can be easily generalized to models that involve multiple statistical behaviors for each group. Hence, we assume that the observer has access to a collection of $n$ sequences of random observations $\mathcal{X}^1, \ldots, \mathcal{X}^n$, each modeling one event. Each sequence $\mathcal{X}^i$ consists of independent and identically distributed (i.i.d.) measurements $\mathcal{X}^i = \{X^i_1, X^i_2, \ldots\}$ taking values in the set $\Omega = \mathbb{R}$ endowed with a $\sigma$-field $\mathcal{F}$ of events, obeying one of the two hypotheses:

$$H_0 : X^i_j \sim F_0, \quad j = 1, 2, \ldots$$

$$H_1 : X^i_j \sim F_1, \quad j = 1, 2, \ldots$$

where $F_0$ and $F_1$ denote the cumulative distribution functions (cdfs) of two distinct distributions on $(\Omega, \mathcal{F})$. The distribution $F_0$ models the statistical behavior of the normal events and the distribution $F_1$ models the statistical behavior of the rare events.

For convenience, we assume that $F_0$ and $F_1$ have probability density functions (pdfs) $f_0$ and $f_1$, respectively. Each sequence $\mathcal{X}^i$ is generated by $F_0$ or $F_1$ independently of the rest of the sequences and we assume that hypothesis $H_1$ (a rare event) occurs with prior probability $\epsilon_n$. In order to incorporate the rareness of the sequences generated by $F_1$, we assume that $\epsilon_n = o(1)$ as $n$ grows.

The goal of quick search is to identify one or more rare events among all $n$ given events through 1) designing an information-gathering process for collecting information from the sequences $\mathcal{X}^1, \ldots, \mathcal{X}^n$, and 2) delineating optimal decision rules. Designing a quick search process involves a tension between two performance measures, one being the aggregate amount of information accumulated (i.e., the number of observations made) and the other being the reliability (or cost) of the decision. In this paper, we design an optimal information-gathering process that maximizes the decision reliability subject to a hard constraint on the aggregate number of observations we are allowed to make from $\mathcal{X}^1, \ldots, \mathcal{X}^n$.

\footnote{By a hard constraint, we mean that the aggregate number of observations made cannot exceed a specified level.}

Manuscript received November 17, 2011; revised September 30, 2012 and February 05, 2013; accepted February 20, 2013. Date of current version June 12, 2013. This work was supported in part by the National Science Foundation under Grant DMS-11-18605.

A. Tajer is with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202 USA (e-mail: tajer@wayne.edu).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by G. V. Moustakides, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2253351
A. Related Literature

The quick search problem is closely related to the sequential detection literature, with two major differences. First, the existing approaches in sequential detection that can be applied to the quick search problem at hand often optimize a balance between decision reliability and a soft constraint on the number of observations. This is in contrast with the quick search setting that enforces a hard constraint on the sensing resources. Second, sequential detection often aims to identify all rare events or only one rare event, whereas our setting offers the flexibility to identify one or more rare events. The most relevant sequential detection solutions for identifying all and only one rare event (generated by \( F_1 \)) are the sequential probability ratio test (SPRT) [25] and the cumulative sum (CUSUM) test [26], respectively. More specifically, by denoting the true hypothesis and a decision about sequence \( X_t \) by \( T_i \in \{ H_0, H_1 \} \) and \( D_i \in \{ H_0, H_1 \} \), respectively, the Type-I and Type-II detection error probabilities corresponding to sequence \( X_t \) are

\[
P_{f_1} = P(D_i = H_1, T_i = H_0),
\]

and

\[
P_{f_2} = P(D_i = H_0, T_i = H_1).
\]

When the objective is to identify all rare events (generated by \( F_1 \)), minimizing the average number of observations made with constraints on \( P_{f_1} \) and \( P_{f_2} \) can be decomposed into minimizing the average number of observations necessary for deciding between \( H_0 \) and \( H_1 \) for each individual sequence with the same reliability constraints [27]. The optimal test for each, which is the test that requires the smallest number of observations and satisfies the reliability constraints, is the SPRT [25]. On the other hand, when the objective is to identify only one rare event (generated by \( F_1 \)) there exist, broadly, two classes of problems. In one class, \( n \) is finite and it is known \textit{a priori} that there exists only one rare event (this is called scanning problem). The optimal detector that identifies the rare event for the Brownian motion setting is given in [28] and [29]. In the other class, \( n = \infty \) so that it is ensured that almost surely there exists a rare event. For this class, the optimal test, which is the test that minimizes the average delay subject to a constraint on Type-II error probability, is the CUSUM test [26].

We also remark briefly on the connection between the quick search problem and the problem of sparse support recovery in noisy environments. These two problems become equivalent (in settings and not objectives) when 1) the events space is mapped to the sparse signal space, 2) the rare events are mapped to the constituents of the support of the sparse signal, and 3) the stochastic observations of the normal and rare events follow the distributions of noisy observations of the normal and sparse components of the sparse signal. A recent line of research in sparse recovery relevant to the quick search problem is the notion of adaptive sampling (sensing) in high-dimensional noisy data with sparsity structures, in which the objective is to estimate the support of the signal [30], [31]. The works in [30]–[32] propose adaptive sampling procedures that effectively focus the sensing resources on the segments of the data that have higher likelihood of containing the support. Despite the similarities in the settings, there is one major discrepancy in the objectives of sparse recovery and quick search. By mapping the events space to a signal space, the goal of quick search can be interpreted as identifying one subset of the support with a prescribed cardinality, whereas the goal of sparse recovery is to identify the entire support of the signal. Another major difference is related to the methodology used in this paper and those of [30] and [31] and their effects on the robustness of the detection mechanisms against uncertainties about the underlying distributions. Specifically, our analysis relies on sorting the events based on the observations and making decisions based on these orders. A benefit of such ordering is that the detection procedures become robust against uncertainties about the underlying distribution parameters. This is in contrast with the methods of [30] and [31] that the detection procedure is sensitive to the knowledge of the distribution median.

It is noteworthy that such notion of fractional support recovery is also studied in the context of compressive sensing through nonadaptive data acquisition [33]. The research in [33] considers a sparse signal with known support size (unlike our setting that the number of rare events is stochastic) and takes one set of low-dimensional observation and analyzes the interplay among the sampling rate, the fraction of the support to be recovered, and the necessary and sufficient conditions on signal power for recovering the support reliably.

B. Contributions

The objective of quick search in this paper is to identify a fraction of the rare events (as opposed to the existing literature which targets at identifying either all or only one rare events). This less-investigated objective aims to fill the gap between the two well-studied extreme (all or one) objectives. This shift of objective allows the detector to tolerate a higher level of Type-I errors, i.e., the detector can afford to miss some of the rare events, in favor of quickly detecting the fraction of interest of the rare events (i.e., sequences generated by \( F_1 \)). In general, the number of rare events is a random variable between 0 and \( n \) and is unknown \textit{a priori}.

Besides the objective, another major distinction from the existing literature is the constraint enforced on the sampling resources. This paper considers the less-investigated scenario in which there exists a hard constraint on the sampling resources. This is in contrast with the existing literature that often incorporates the statistical average of the sampling budget in the analysis. We remark that to the best of our knowledge, the general problem of designing the optimal sampling strategy that distributes a limited sampling budget among the sequences under observation in order to yield the highest level of reliability in detecting one, a fraction, or all of the sequences generated by \( F_1 \) is an open problem. In this paper, we focus on the asymptotic performance when the number of sequences \( n \) is sufficiently large and provide the asymptotically optimal sampling procedure when the distributions \( F_0 \) and \( F_1 \) are Gaussian with either different mean or different variance values.

C. Summary of Results

To solve the quick search problem with the aforementioned objective and constraints, a sequential and data-adaptive infor-
mation-gathering procedure is designed that dynamically makes one of the following three data-driven decisions at each time: 1) it decides to postpone making decision and collects more data (observation) as it lacks sufficient evidence to identify the rare events, or 2) it reaches some confidence to eliminate roughly \( \alpha \in (0,1) \) fraction of the events that are deemed to be the weakest candidates for being rare events, but yet does not have enough evidence to make the final decision (refinement), or 3) it has accumulated enough information to identify the rare events of interest (detection). As one major result of this paper, we characterize the asymptotically optimal (in the asymptote of large number of events \( n \)) allocation of the sampling resources among the sequences and the pertinent sampling procedure, which consists of consecutive rounds of coarse observations and refinement actions, followed by consecutive cycles of fine observations. The number of each these cycles (refinement and observation) is also determined as a function of the number of available sampling budget.

It is confirmed in different contexts that in order to recover the rare events (or the support of the sparse signal) reliably, the power of the observed data should be scaling with the data size \( n \) (cf., [30]–[32]). We apply the proposed adaptive sampling procedure for the settings that the normal and rare events are generated according to Gaussian distributions with either different mean or different variance values and characterize necessary and conditions on scaling rates of the mean and variance values for identifying the desired number of rare events. These scaling rates of the means and variances are functions of the data dimension \( n \), the available amount of sampling budget, and the frequency of the rare events. More specifically, we define \( \varepsilon_n \) as the likelihood that an event is a rare event, and define \( \varepsilon_n \in (0,1) \) as

\[
\varepsilon_n = \frac{\ln n \varepsilon_n}{\ln n}.
\]

By assuming that the normal events are distributed as \( \mathcal{N}(\mu_0,1) \) and the rare events are distributed as \( \mathcal{N}(\mu_1,1) \), a necessary and sufficient condition for successfully identifying a \emph{small fraction} of the rare events via the adaptive procedure is that in the asymptote of large \( n \)

\[
\frac{(\mu_0 - \mu_1)^2}{\log n} > \frac{(1 - \varepsilon_n)^2}{\varepsilon_n}
\]

where \( \varepsilon_n \) is a constant independent of \( n \) and determined by the available sampling budget and the number of refinement cycles. We also show that when the normal and rare events are distributed as \( \mathcal{N}(0,A_0) \) and \( \mathcal{N}(0,A_1) \), respectively, the counterpart necessary and sufficient condition is

\[
\frac{\ln A_0/A_1}{\ln n} > \frac{2(1 - \varepsilon_n)}{\varepsilon_n}.
\]

\( \varepsilon_n \) is noteworthy that as long as the objective is to identify a small fraction of the rare events, the scaling behaviors delineated by (5) and (6) do not depend on the exact number of rare events to be identified. This is due to the asymptotic nature of the results, which is dominantly shaped by the regime of interest (small fraction) and the precise number of the rare events has a vanishing effect as \( n \) grows. More discussions on this observation are provided in Section III-A.

\( \text{Remark 2:} \) By comparing the scaling laws in (5) and (6) by their counterparts corresponding to full recovery of the rare events, it is evident that full recovery imposes more stringent conditions on the scaling of \( (\mu_0 - \mu_1)^2 \) and \( \frac{\mu_0}{\mu_1} \). Specifically, when the normal events are distributed as \( \mathcal{N}(\mu_0,1) \) and the rare events as \( \mathcal{N}(\mu_1,1) \), a necessary and sufficient condition for full recovery of the rare events is

\[
\frac{(\mu_1 - \mu_0)^2}{\log n} > \frac{1 - \varepsilon_n}{\varepsilon_n}.
\]

By noting that \( \varepsilon_n < 1 \), recovering all the rare events clearly imposes a harder condition on \( (\mu_1 - \mu_0)^2 \) than recovering a fraction of them. The differences between such scaling behaviors in different regimes are further discussed in Section IV.

\( \text{Remark 3:} \) Finally, we remark that constant \( \varepsilon \) increases exponentially with the number of refinement cycles. This translates into less stringent conditions on the scaling of \( (\mu_0 - \mu_1)^2 \) and \( A_0/A_1 \) as the number of refinement cycles increases.

Throughout this paper, a sequence \( \{a_m\} \) is said to be asymptotically dominated by \( \{b_m\} \), denoted by \( a_m = o(b_m) \), when \( \lim_{m \to \infty} \frac{a_m}{b_m} = 0 \). Also, \( \{a_m\} \) is said to be asymptotically dominating \( \{b_m\} \), denoted by \( a_m = o(b_m) \), when \( b_m - o(a_m) \).

II. PROBLEM STATEMENT

A. Sampling Model

With the ultimate objective of identifying \( T_n = o(n^{\varepsilon_n}) \) rare events, i.e., a \emph{small} fraction of the rare events, the proposed sampling procedure is initiated by making observations from \emph{all} sequences \( X_1, \ldots, X^n \). Based on these rough observations, a fraction of the sequences that are least-likely generated by \( F_1 \) are discarded and the rest are retained for further and more accurate scrutiny. Repeating this procedure successively refines the search support and progressively focuses the observations on the more promising sequences. More specifically, at each time, the sampling procedure selects a subset of the sequences \( X_1, \ldots, X^n \) and takes one sample from each of these sequences. Upon collecting these samples, it takes one of the following actions:

- **A1 (Detection)** stops further sampling and identifies \( T_n \) sequences that have the highest likelihood of being rare events (generated by \( F_1 \));
- **A2 (Observation)** continues to further observe the same set of sequences in order to gather more information about their statistical behavior; or
- **A3 (Refinement)** discards a portion of the sequences and declares that they are most likely normal events (generated by \( F_0 \)). When a sequence is discarded it will be deemed a weak candidate for being a rare event (generated by \( F_1 \)) and will not be observed anymore, while the remaining sequences are retained for more scrutiny. By denoting the number of sequences retained prior to a refinement action by \( \ell \), the number of sequences that this action discards is \( (1 - \alpha)(\ell - T_n) \) for some \( \alpha \in (0,1) \). Discarding the
sequences at this rate ensures that at least $T_n$ sequences will be retained for the final detection action (action $A_1$). We denote the set of the indices of the sequences observed at time $t \in \mathbb{N}$ by $\mathcal{L}_t$. As we initialize the information-gathering procedure by including all sequences for observation, we have $\mathcal{L}_1 = \{1, \ldots, n\}$. Also, we denote the stopping time of the procedure, i.e., the time after which detection (action $A_1$) is performed, by $\tau$. Furthermore, we define the switching function $\psi : \{1, \ldots, \tau\} \rightarrow \{0, 1\}$ to model actions $A_2$ (observation) and $A_3$ (refinement). At each time $1 \leq t \leq \tau - 1$, we set $\psi(t) = 0$ if we decide in favor of performing observation, while $\psi(t) = 1$ indicates a decision in favor of performing refinement, i.e., $\forall t \in \{1, \ldots, \tau - 1\}$:

$$
\psi(t) = \begin{cases} 
0, & \text{action } A_2 \text{ and } \mathcal{L}_{t+1} = \mathcal{L}_t \\
1, & \text{action } A_3 \text{ and } \mathcal{L}_{t+1} \subset \mathcal{L}_t.
\end{cases}
$$

Let $X^i_t$ denote the observations made from sequence $i \in \mathcal{L}_t$ at time $t$ and denote the $\sigma$-algebra generated by the observation $\{X^1_t, \ldots, X^n_t\}$ by

$$
\mathcal{F}_t = \sigma(X^1_t, \ldots, X^n_t).
$$

Given $\mathcal{F}_t$, we denote the posterior probability that the sequence $X^i_t$, for $i \in \mathcal{L}_t$, is generated by $\theta_i$ by

$$
\pi^i_t = \frac{1}{1 + \epsilon_n \sum_{u=1}^{t} \frac{f_0(X^i_u)}{f_1(X^i_u)}}.
$$

By defining the likelihood ratio

$$
\Lambda^i_t \triangleq \prod_{u=1}^{t} \frac{f_0(X^i_u)}{f_1(X^i_u)}
$$

we have

$$
\pi^i_t = \left( 1 + \frac{1 - \epsilon_n}{\epsilon_n} \Lambda^i_t \right)^{-1}
$$

and the actions $A_1$, $A_2$, and $A_3$ can be formalized as follows.

$A_1$: At the stopping time $\tau$, identify the set $\mathcal{U} \subset \mathcal{L}_\tau$ as the detector’s decision according to

$$
\mathcal{U} = \arg \max_{\mathcal{U} \subset \mathcal{L}_\tau: \mathcal{U} = \{\tau\}} \mathbb{P} \left\{ \bigvee_{i \in \mathcal{U}} \left( T_i = H_1 \mid \mathcal{F}_\tau \right) \right\} = \arg \max_{\mathcal{U} \subset \mathcal{L}_\tau: \mathcal{U} = \{\tau\}} \prod_{i \in \mathcal{U}} \mathbb{P} \left( T_i = H_1 \mid \mathcal{F}_\tau \right)
$$

Hence, $\mathcal{U}$ contains the indices of the $T_n$ smallest elements of the set $\{A^i_\tau: i \in \mathcal{L}_\tau\}$.

$A_2$: At time $t$, the decision is to further measure the same set of sequences (i.e., $\mathcal{L}_{t+1} = \mathcal{L}_t$), and set

$$
\Lambda^i_{t+1} = \Lambda^i_t \cdot \frac{f_0(X^i_{t+1})}{f_1(X^i_{t+1})}
$$

$A_3$: At time $t$, the decision is to refine $\mathcal{L}_t$ and the set $\mathcal{L}_{t+1}$ is obtained as

$$
\mathcal{L}_{t+1} = \arg \max_{\mathcal{L} \subset \mathcal{L}_t} \mathbb{P} \left\{ \bigvee_{i \in \mathcal{L}} \left( T_i = H_1 \mid \mathcal{F}_\tau \right) \right\} = \{\mathcal{L} \setminus \{T_n\}\} + T_n.
$$

B. Quality of Sampling

Characterizing the experimental design and decision rules for performing quick search involves an interplay between two performance measures, one being the aggregate number of observations made and the other being the frequency of erroneous detection. Any improvement in either of these performance measures penalizes the other one and the optimal design of such procedures typically involves optimizing a tradeoff between them. Furthermore, once the optimal aggregate number of observations is determined, we also need to ascertain how these observations should be split and allotted to the refinement and detection actions. For a given stopping time $\tau$ and a given sequence of switching functions $\bar{\psi}(\tau) = (\psi(1), \psi(2), \ldots, \psi(\tau - 1))$, the probability of erroneous detection, i.e., the probability that the detected sequences include a normal event (generated by $F_0$), is

$$
P_n(\tau, \bar{\psi}(\tau)) \triangleq \mathbb{P} \left\{ \{ i \in \mathcal{U}: T_i = H_0 \} \neq \emptyset \right\}.
$$

Our goal is to minimize this above detection error probability over all possible stopping times $\tau$, all switching rules $\bar{\psi}$, and all possible allocations of the observations to refinement and detection actions subject to two hard constraints. One constraint incorporates the impact of the aggregate number of observations made, and the other one captures the cost of the refinement actions. Specifically, for the first constraint, the aggregate number of observations normalized by the number of sequences is constrained to be less than a prespecified value $S$. As the second constraint, the number of refinement actions must also be smaller than a prespecified value $K$. This latter constraint captures the cost incurred by each refinement action, which is the permanent loss of the sequences discarded after the refinement actions. This optimization problem can be formalized as

$$
P_n(S, K) \triangleq \inf_{\tau, \bar{\psi}(\tau)} \frac{1}{\sum_{\tau=1}^{\tau-1} \bar{\psi}(\tau)} \mathbb{P} \left\{ \{ i \in \mathcal{U}: T_i = H_0 \} \neq \emptyset \right\}.
$$

We solve this problem in the asymptote of large $n$ for two hypothesis testing problems that test the mean and the variance of Gaussian observations, i.e., for $m \in \{0, 1\}$

1) Mean:

$$
H_m: X^i_j \sim \mathcal{N}(\mu_m, 1).
$$

2) Variance:

$$
H_m: X^i_j \sim \mathcal{N}(0, A_m).
$$
In this paper, for the asymptotically large values of $n$, we characterize:

1) the optimal stopping time and sampling process (through designing the optimal switching sequence) that maximize the detection reliability subject to the hard constraints on the sampling resources and refinement actions;

2) the asymptotic detection error $P_n(S, K)$ defined in (18) for any given $S$ and $K$;

3) the minimum distance between the unknown pair of parameters in each test such that the two hypotheses are guaranteed to be distinguished perfectly, i.e., $P_n(S, K) \to 0$;

4) efficiencies of the proposed sequential and adaptive sampling procedure with respect to that of the nonadaptive procedure that does not involve any refinement action and performs detection directly, i.e., $K = 0$; and

5) comparisons with a few relevant tests.

C. Methodology

Note that among the three actions, action $A_2$ (observation) concentrates on accumulating further evidence about the statistical behavior of the sequences under observation. Action $A_3$ (refinement) is intended to purify the set of the sequences on which we ultimately perform the detection action. Each iteration of the refinement actions monitors the events retained by the previous iterations and eliminates those deemed least-likely to be rare events. The core idea of the refinement action is that it is relatively easy to identify sequences drawn from $H_0$ with low-quality measurements as the events of interest occur rarely ($\epsilon_n$ is small). After iteratively performing actions $A_2$ and $A_3$, we will have a more condensed proportion of the desired sequences to the undesired ones. The sequences retained after the last action $A_3$ are further observed and finally fed to the detector in order to identify $T_n$ sequences of interest.

Characterizing the detection error rate $P(\tau, \bar{\psi}(\tau))$, as a result, depends on 1) the evolution of the proportion of the number of the normal and rare events throughout the refinement actions, and 2) the quality of the detector. For any time $t$, let us partition the set $L_t$, which includes the indices of the sequences observed at time $t$, into two disjoint sets $L^o_t$ and $L^r_t$ as

$$L^o_t \triangleq \{ i \in L_t : T_i = m \}, \text{ for } m \in \{ 0, 1 \}.$$ 

Also let us define $n^o_t = |L^o_t|$ and $n^r_t = |L^r_t|$ for all $t \in \{ 1, \ldots, \tau \}$.

As will be made clear in the subsequent sections, implementing action $A_1$ (detection) necessitates obtaining the low-order statistics of the sets $\{ A^o_t : i \in L^o_t \}$ and $\{ A^r_t : i \in L^r_t \}$ with sample sizes $n^o_t$ and $n^r_t$, respectively. Evaluating the dynamics of the refinement action $A_3$ additionally requires analyzing the high-order statistics of the sets $\{ A^o_t : i \in L^o_t \}$ and the low-order statistics of the sets $\{ A^r_t : i \in L^r_t \}$ for all $t \in \{ 1, \ldots, \tau \}$. Moreover, $n^o_t$ and $n^r_t$ are also random variables that change (reduce) after each refinement action. Analyzing the evolution of the set sizes $n^o_t$ and $n^r_t$ as well as the necessary order statistics depends on the distributions $P_0$ and $P_1$. The necessary tools and results from the order statistics literature that will be used frequently throughout the analysis are summarized in Appendix A.

III. OPTIMAL SAMPLING

In this section, we determine the optimal stopping time $\tau$ and the associated optimal switching sequence $\bar{\psi}(\tau) = \{ \psi(1), \ldots, \psi(\tau - 1) \}$. For this purpose, we initially characterize the error probability $P(\tau, \bar{\psi}(\tau))$ defined in (17) for any given stopping time $\tau$ and switching sequence $\bar{\psi}(\tau)$ and then optimize it over the valid choices of $\tau$ and $\bar{\psi}(\tau)$ such that the constraints on the aggregate sampling budget and the number of refinement actions are satisfied. As stated earlier, characterizing the detection performance $P(\tau, \bar{\psi}(\tau))$ for a given stopping time $\tau$ and a switching sequence $\bar{\psi}(\tau)$ involves analyzing 1) the evolution of the proportion of the number of sequences generated by $F_0$ and $F_1$ throughout the refinement actions, and 2) the quality of the detection action. Analyzing both aspects heavily depends on the underlying distributions $F_0$ and $F_1$ and are treated separately for two the problems given in (19).

A. Gaussian Mean

The hypothesis-testing problem in this case is

$$H_0 : X^0_j \sim N(\mu_0, 1), \quad j = 1, 2, \ldots, 0.$$ 

$$H_1 : X^1_j \sim N(\mu_1, 1), \quad j = 1, 2, \ldots, 1.$$ 

(20)

where without loss of generality $\mu_1 > \mu_0$ and $\mu_0, \mu_1 \in \mathbb{R}$. Under this setting, the likelihood ratio at time $t$ for the sequences $i \in L_t$ is

$$\Lambda^r_t = \prod_{u=1}^{t} \frac{f_0(X^0_u)}{f_1(X^1_u)} = \exp \left\{ \frac{\mu_0^2 - \mu_1^2}{2} \right\}. \quad (21)$$

By defining

$$\forall t \in \{ 1, \ldots, \tau \} \text{ and } \forall i \in L_t : \quad Z^r_t \triangleq \sum_{u=1}^{t} X^r_u$$

(22)

detection and refinement actions formalized in (13) and (15), respectively, are equivalently given by

$$U = \text{indices of the } T_n \text{ smallest elements of } \{ Z^r_t : i \in L_t \}$$

(23)

$$L_{t+1} = \text{indices of the } n_t \text{ smallest elements of } \{ Z^r_t : i \in L_t \}.$$ 

(24)

On the other hand, by invoking the distribution of $X^0_1$ from (20), we immediately have $\forall t \in \{ 1, \ldots, \tau \} \text{ and } \forall i \in L_t$:

$$Z^0_t \mid H_m \sim N\left( \mu_m \cdot t, t \right).$$

(25)

Furthermore, let us for each $t \in \{ 1, \ldots, \tau \}$ define

$$U^r_j \triangleq \text{the } j\text{th smallest element of } \{ Z^r_t : i \in L_t \} \quad (26)$$

$$U_t^o \triangleq \text{the } j\text{th smallest element of } \{ Z^r_t : i \in L_t \}.$$ 

(27)
Given these definitions, the probability $P(\tau, \tilde{\psi}(\tau))$, which is the probability that at least one member of $\mathcal{U}$ is distributed according to $F_{\nu}$, can be equivalently written as

$$P_n(\tau, \tilde{\psi}(\tau)) = P \left( \mathcal{U} \cap \mathcal{L}^T_{\nu} \geq 0 \right) = P \left( U^T_{\nu} > \bar{U}^T_{\nu} \right). \quad (28)$$

Assessing this performance measure involves finding the cardinalities and the distributions of the first and the $T_n^{th}$ order statistics of the sets of random variables $\{Z^i : i \in \mathcal{L}^1_{\nu}\}$ and $\{Z^i : i \in \mathcal{L}^L_{\nu}\}$, respectively. In order to proceed with analyzing $P_n(\tau, \tilde{\psi}(\tau))$, we first assess the variations of the number of normal and rare events, denoted by $n_\epsilon$ and $\tilde{n}_\epsilon$, respectively, and their associated ratio throughout the refinement cycles. The following lemma sheds light on the tendency of the refinement action $(\mathcal{A}_q)$ toward retaining almost all the rare events and discarding almost $\alpha$ proportion of the normal events.

**Lemma 1 (Refinement Performance):** Let $\tilde{n}_\epsilon = \mathcal{L}^0_{\nu}$ and $n_\epsilon = \mathcal{L}^L_{\nu}$ denote the number of sequences generated by $F_{\nu}$ and $F_{\nu}$ that are retained up to time $t$. For any arbitrary $\delta \in (0, 1]$ and for sufficiently large $n$, the event

$$n_\epsilon \geq (1 - \delta)n_\epsilon \quad (29)$$

holds almost surely if

$$(\mu_0 - \mu_1)^2 = \omega \left( n^{-\epsilon n} \right) \quad (30)$$

where $\epsilon_n$ is defined as

$$\epsilon_n = \frac{\Delta \ln n \epsilon n}{\ln n}. \quad (31)$$

**Proof:** See Appendix C.  

Therefore, when the condition in (30) is satisfied, the refinement actions almost surely discard no more than a fraction $\delta$ of the rare events, for any arbitrary $\delta \in (0, 1]$. Therefore, the final ratio of the number of rare events to that of the normal events increases dramatically throughout the refinement actions. Besides the performance of the refinement actions, the overall detection reliability also depends on the performance of the detection action $(\mathcal{A}_q)$. The next lemma describes a necessary and sufficient condition that guarantees asymptotically error-free detection. Note that this lemma does not restrict itself to any specific performance for the outcome of the refinement actions and applies to any arbitrary sequence of refinement and observation actions captured by $\tilde{\psi}(\tau)$.

**Lemma 2 (Detection Performance):** For a given stopping time $\tau$ and switching sequence $\tilde{\psi}(\tau)$, the detection error probability $P_n(\tau, \tilde{\psi}(\tau))$ tends to zero in the asymptote of large $n$ if and only if

$$\tau_m > \frac{(1 - \sqrt{\epsilon_n})^2}{\tau} \quad (32)$$

where we have defined

$$\tau_m = \frac{(\mu_0 - \mu_1)^2}{2 \ln n}. \quad (33)$$

**Proof:** See Appendix D.  

Therefore, this lemma, conditionally on the value of the stopping time which is stochastic, provides a necessary and sufficient condition on the distance between distributions $F_{\nu}$ and $F_{\nu}$ (captured by $(\mu_0 - \mu_1)$) for achieving asymptotically optimal detection performance. In the next lemma, we show that the stochastic stopping time is upper bounded by a constant.

**Lemma 3:** The stopping time $\tau$ is upper bounded by $S/\alpha^K$ for sufficiently large $n$.

**Proof:** See Appendix E.

Combining the results of Lemmas 2 and 3 and replacing the stopping time $\tau$ in (32) with its upper bound from Lemma 3 provides that

$$\frac{(\mu_0 - \mu_1)^2}{2 \ln n} > \frac{(1 - \sqrt{\epsilon_n})^2}{S/\alpha^K}$$

is a necessary condition for ensuring asymptotically error-free detection. This clearly imposes a more stringent condition on the distance $(\mu_0 - \mu_1)^2$ than (30), which is a sufficient condition for retaining at least a fraction $(1 - \delta)$ of the rare events throughout the refinement cycles. As a result, irrespective of the optimal value of the stopping time $\tau$, ensuring asymptotically error-free detection requires the refinement process to retain at least $(1 - \delta)$ fraction of the rare events.

Finally, we would like to remark that while the exact characterization of the error probability $P_n(\tau, \tilde{\psi}(\tau))$ does depend on $T_n$ (the number of events to be returned by the detector), the scaling rate given in (32) does not depend on $T_n$. The intuitive reason is that in analyzing such scaling laws, in the asymptote of large $n$, only the dominant (leading) terms survive and the impacts of the nonleading terms vanish as $n$ grows. $T_n$ appears to impact only the nonleading terms and its impact is not observed on the ultimate scaling laws. In a more technical sense, the reason lies at the core of the analysis on order statistics. Note that the focus of this paper is on the regime $T_n = o(n\epsilon_n)$. The analysis reveals that while the error probabilities vary for different choices of $T_n$, the scaling laws of the mean values are identical for all values of $T_n$ in the regime $T_n = o(n\epsilon_n)$. The underlying reason for this observation is that according to the definition of the detection action, the ultimate $T_n$ sequences selected are the sequences with the smallest likelihood ratios and, hence, constitute the $T_n$ smallest order statistics of the set comprised of the likelihood ratios of the retained sequences. When $T_n = o(n\epsilon_n)$ all these order statistics are of low order (Definition 3) and have identical asymptotic impacts on the scaling rates.

**B. Gaussian Variance**

In this section, we analyze the performance of detection and refinement in the Gaussian variance hypothesis testing problem. The presentation of the results follows the same flow as the problem of testing the mean. The proofs, however, are different due to different statistical behavior of the likelihood ratio and its pertinent sufficient statistic. The problem of interest can be posed as

$$H_0: \quad X^j \sim \mathcal{N}(0, A_0), \quad j = 1, 2, \ldots$$

$$H_1: \quad X^j \sim \mathcal{N}(0, A_1), \quad j = 1, 2, \ldots \quad (34)$$
where $A_C > A_1$ and $A_D, A_3 \in \mathbb{R}^+$. Hence, the likelihood ratio at time $t$ for the sequences $i \in \mathcal{L}_t$ is

$$
A_i^t = \prod_{u=1}^{t} \frac{f_{0}(X_u^i)}{f_{1}(X_u^i)} = \exp \left\{ \frac{1}{2} \left( \frac{1}{A_1} - \frac{1}{A_0} \right) \sum_{u=1}^{t} (X_u^i)^2 \right\}.
$$

Similarly to (22), we define

$$
\forall t \in \{1, \ldots, \tau\} \quad \text{and} \quad \forall i \in \mathcal{L}_t : \quad Z_i^t = \sum_{u=1}^{t} (X_u^i)^2.
$$

Therefore, the indices of the sequences yielded by the detection and refinement actions defined in (13) and (15), respectively, are given by

$$
\mathcal{U} = \text{indices of the } T_n \text{ smallest elements of } \{Z_i^t : i \in \mathcal{L}_t\}
$$

$$
\mathcal{L}_{t+1} = \text{indices of the } \bar{n}_t \text{ smallest elements of } \{Z_i^t : i \in \mathcal{L}_t\}
$$

where $\bar{n}_t$ is defined in (16). By recalling the distribution of $X_1^t$ given in (34), we have $\forall t \in \{1, \ldots, \tau\}$ and $\forall i \in \mathcal{L}_t$:

$$
Z_i^t \mid H_m \sim A_m \cdot \chi^2(t)
$$

where $\chi^2(k)$ denotes a chi-squared distribution with $k$ degrees of freedom. Next, for each $t \in \{1, \ldots, \tau\}$, let us define

$$
\bar{V}_j^t = \text{the } j^{\text{th}} \text{ smallest element of } \{Z_i^t : i \in \mathcal{L}_t\}
$$

$$
\tilde{V}_j^t = \text{the } j^{\text{th}} \text{ smallest element of } \{\tilde{Z}_i^t : i \in \mathcal{L}_t\}
$$

As shown in (28), the probability $P(\tau, \tilde{\psi}(\tau))$ is given by

$$
P(\tau, \tilde{\psi}(\tau)) = P \left( \left\{ \mathcal{U} \cap \mathcal{L}_t^c \mid \frac{\ln n}{\tau} \right\} \right.
$$

$$
= P \left( \left\{ \frac{\ln n}{\tau} > \tilde{V}_1^t \right\} \right).
$$

As assessing this probability requires knowing the distributions of the first and the $T_n$th order statistics of the sets of random variables $\{Z_i^t : i \in \mathcal{L}_t^c\}$ and $\{\tilde{Z}_i^t : i \in \mathcal{L}_t^c\}$, respectively, which can be found based on the results on the minimal domain of attraction of chi-squared distributions provided in Lemma 7. Given these distributions, the next lemma provides the performance of the refinement action in retaining the sequences generated by $F_0$ and $F_1$.

**Lemma 4 (Refinement Performance):** Let $\bar{n}_t = |\mathcal{L}_t^c|$ and $n_t = |\mathcal{L}_t|$ denote the number of sequences generated by $F_0$ and $F_1$, respectively, that are retained up to time $t$. For sufficiently large $n_t$, the event

$$
n_t = n_1
$$

holds almost surely if

$$
\frac{A_0}{A_1} = \omega (\varepsilon_n \ln n)
$$

where $\varepsilon_n$ is defined as

$$
\varepsilon_n = \frac{\ln n}{n_1 \ln n}.
$$

**Proof:** See Appendix F.

Therefore, when the scaling law in (44) is satisfied, the proportion of the sequences generated by $F_1$ to those generated by $F_0$ increases after the refinement actions. In the next lemma, we provide a necessary and sufficient condition on the scaling of $A_0/A_1$ that ensures perfect identification of the $T_n$ sequences generated by $F_1$ (rare events).

**Lemma 5 (Detection Performance):** For a given stopping time $\tau$ and switching sequence $\bar{\psi}(\tau)$, and conditionally on $n_t$, and $\bar{n}_t$, the detection error probability $P_n(\tau, \bar{\psi}(\tau))$ tends to zero in the asymptote of large $n$ if and only if

$$
\xi_n > \frac{2(1 - \varepsilon_n)}{\tau}
$$

where we have defined

$$
\xi_n = \frac{\ln n_1 \ln n}{\ln n}.
$$

**Proof:** See Appendix G.

By comparing the scaling laws offered by Lemmas 4 and 5, we find that the scaling law necessary for making a reliable detection, irrespective of $\tau$ and $\bar{\psi}(\tau)$, dominates the one that is sufficient for maintaining $n_t = n_1$ almost surely. In other words, in order to perform reliable detection, the refinement action (A2) retains all the rare events almost surely.

**C. Optimal Stopping Time**

Given the performance of the refinement action offered by Lemmas 1 and 4, and the detection action given by Lemmas 2 and 5, in this section, we provide the optimal choices of the stopping time and the switching sequence. Given the discussions at the end of Sections III-A and III-B, irrespective of the discrepancies in the analysis and the ensuring scaling laws in the mean and variance settings, these two settings conform in the fact that targeting at error-free detection forces the refinement process to retain almost all of the rare events. More specifically, the refinement process is guaranteed (probabilistically) to discard at most a fraction $\delta$ of the rare events in testing the mean and retain almost all such events in testing the variance.

Primarily based on the similar behavior of the refinement process in both settings, the optimal choices of the stopping time and the switching sequence turn out to be exactly the same in both settings. The scheme of the proofs in this section are as follows. By using the result of Lemma 3, which states that the stopping time $\tau$ is upper bounded by a constant, it is concluded that regardless of the choice of the stopping time and the switching sequence, the conditions that Lemmas 2 and 5 impose on $(\mu_0 - \mu_1)^2$ and $A_0/A_1$ dominate those imposed by Lemmas 1 and 4, respectively. This immediately indicates that the refinement process in both settings retains almost all rare events almost surely. Based on this property of the refinement process, we obtain the choices of the stopping time and switching sequence that minimize the error probability.

**Theorem 1 (Stopping Time):** The optimal switching sequence for achieving $P_n(S, K) \to 0$ satisfies

$$
\forall t \in \{1, \ldots, K^*\} : \psi(1) = 1, \quad \text{and} \quad \forall t > K^* : \psi(t) = 0
$$

where

$$
K^* = \begin{cases} 
K, & \text{if } \alpha \leq 1 - \frac{1}{8} \\
0, & \text{if } \alpha > 1 - \frac{1}{8} 
\end{cases}
$$

**Proof:** See Appendix G.
Also the optimal stopping time is
\[ r^* = \begin{cases} K + s(K), & \text{if } \alpha \leq 1 - \frac{1}{2\beta(K)} \\ S, & \text{if } \alpha > 1 - \frac{1}{2\beta(K)} \end{cases} \] (50)
where
\[ s(K) \triangleq \left[ S \cdot \alpha^{-K} + \frac{1 - \alpha}{1 - \alpha} \right]. \] (51)

**Proof:** See Appendix H.

This theorem demonstrates that when $\alpha$ is large (close to 1), the optimal sampling does not involve any refinement action and adaptive sampling does not offer any gain over the nonadaptive sampling procedure. However, for sufficiently large $S$, only for limited choices of $\alpha$ does adaptation in sampling offer no gain, and for a wide range of $\alpha$, the optimal sampling procedure involves refinement actions and becomes adaptive. Throughout the remainder of this paper, we focus on the regime $\alpha \leq 1 - \frac{1}{\delta}$.

### IV. ADAPTATION GAINS

The main component of the proposed sampling procedure is the inclusion of the refinement actions, which makes the detection procedure adaptive to the data. In this section, we show the gains attained by the inclusion of the refinement process. These gains can be viewed as the gain of adapting the detection process to the observed data and can be interpreted in two ways, namely in the forms of agility and scaling gains defined in the sequel. These gains essentially evaluate the contribution of the refinement actions by comparing the performance of the proposed sampling procedure against the same procedure without any refinement action, i.e., $K = 0$.

**Definition 1 (Agility Gain):** The agility gain, denoted by $G_\alpha$, is the ratio of the minimum sampling budget required by the nonadaptive procedure ($K = 0$) to that required by the adaptive procedure with $K > 0$ refinement actions such that both achieve asymptotically error-free detection while enjoying identical scaling for $(\mu_0 - \mu_1)$ in the mean setting or identical scaling for $A_0/A_1$ in the variance setting.

In order to quantify this gain, we consider a nonadaptive detection procedure with the aggregate sampling budget controlled by $S_0$ and obtain the required scaling law for $(\mu_0 - \mu_1)^2$ that guarantees asymptotically error-free detection. For the same scaling law in an adaptive procedure with $K$ refinement actions, we assess the minimum sampling budget $S$ that ensures asymptotically error-free detection by the adaptive procedure as well. Then, we find the agility gain as the ratio $S_0/S$. In order to analyze the agility gain for the variance setting, we repeat the same steps by replacing all the arguments on the scaling of $(\mu_0 - \mu_1)^2$ by those on the scaling of $A_0/A_1$.

**Definition 2 (Scaling Gain):** The scaling gain, denoted by $G_\alpha$, is the ratio of the smallest scaling law required by the nonadaptive procedure to that required by the adaptive procedure when both target at achieving asymptotically error-free detection while enjoying identical sampling budgets.

In order to analyze the scaling gain of the adaptive procedure with $K$ refinement actions in the Gaussian mean setting, we assume that both adaptive and nonadaptive procedures are allocated the aggregate sampling budget $S \cdot n$ and aim at identifying the smallest values of $r_m$ and $r_m^0$ such that the scaling laws $(\mu_0 - \mu_1)^2 = 2 r_m n$ and $(\mu_0 - \mu_1)^2 = 2 r_m^0 n$ guarantee asymptotically error-free detection for the adaptive and the nonadaptive procedures, respectively. The scaling gain can be found as $r_m^2$. For computing the scaling gain in the variance setting, we follow the same logic, and instead of finding $r_m$ and $r_m^0$, we consider the scaling laws $A_0/A_1 = n\varepsilon^2$ and $A_0/A_1 = n\varepsilon^2$ for the adaptive and nonadaptive procedures and find $\varepsilon$ and $\varepsilon^0$, respectively. In this case, the scaling gain is defined as $\varepsilon^2/\varepsilon^2$.

Analyzing the scaling and agility gains strongly relies on the connection between the available sampling budget $S$ and the scaling laws that lead to reliable detection of the sequences of interest. The following two theorems establish this connection.

**Theorem 2 (Mean Scaling):** When $\alpha < 1 - \frac{1}{S}$, a necessary and sufficient condition for $P_n(S, K) \xrightarrow{n \to \infty} 0$ is that
\[ r_m > \frac{(1 - \sqrt{\varepsilon n})^2}{K + s(K)} \] (52)
where $r_m$ is defined in (33) and $s(K)$ was defined as
\[ s(K) \triangleq \left[ S \cdot \alpha^{-K} + \frac{1 - \alpha}{1 - \alpha} \right]. \] (53)
where $\alpha$ controls what fraction of the events are discarded at each refinement cycle, $S$ is the aggregate sampling budget normalized by $n$, and $K$ is the maximum allowable number of refinement actions.

**Proof:** The proof can be established by combining the results from Lemma 2 and the optimal stopping time given in Theorem 1. As shown in the proof of Lemma 2, a necessary and sufficient condition for having asymptotically error-free detection is that $B_\omega = \omega(1)$, where $B_\omega$ is defined in (93). Moreover, we have also proved that a necessary and sufficient condition for $B_\omega = \omega(1)$ is that
\[ r_m > \frac{(1 - \sqrt{\varepsilon n})^2}{\tau}. \] (54)
On the other hand, in the proof of Theorem 1, we showed that the detection error probability is minimized when $\tau$ is maximized. In other words, the smallest necessary scaling law for $(\mu_0 - \mu_1)^2$ is obtained when $\tau$ is maximized and is equal to $K + s(K)$. By substituting this value into the equation above, we obtain the desired result.

For a given sampling budget $S \cdot n$ and the number of refinement actions $K$, this theorem delineates the asymptotic performance of the proposed detector in the $(r_m, \varepsilon)$ plane. It shows that when $(\mu_0 - \mu_1)^2$ scales as $2 r m n$, if $r > r_m$, the proposed sequential detection procedure is guaranteed to make error-free decisions for identifying $T_n$ sequences generated by $F_1$. On the other hand, when $r \leq r_m$, the probability of erroneous detection is bounded away from zero. Therefore, $r_m$ defines a sharp threshold for identifying the sequences of interest under the objective and constraints in (18).

On a related context, note the relevant results provided by [34] and [35] when the objective is to identify all sequences generated by $F_1$ through observing each sequence only once,
i.e., $S = 1$ and $K = 0$. The works in [34] and [35] show that for $\varepsilon_n \in (0, \frac{1}{2})$, the false-discovery and nondiscovery proportions\(^4\) tend to zero if and only if $(\mu_0 - \mu_1)^2$ scales as $2 r \ln n$ and $r > 1 - \varepsilon_n$. This is clearly a more stringent condition than the requirement $r > r_m = (1 - \sqrt{\varepsilon_n})^2$ corresponding to the setting $S = 1$ and $K = 0$ in our sampling procedure. This discrepancy demonstrates the tradeoff between partial versus full recovery of the sequences of interest on one hand, and the required scaling law on $(\mu_0 - \mu_1)^2$ on the other hand.

In the following lemma, we focus on the variance setting, and by using the result of Lemma 5 and Theorem 1, we offer a necessary and sufficient condition on the scaling of $A_n/A_1$ in order to guarantee asymptotically error-free detection.

**Theorem 3 (Variance Scaling):** When $\varepsilon_n = o(1), n \varepsilon_n = \omega(1)$, and $\alpha \leq 1 - \frac{1}{2}$, a necessary and sufficient condition for $P_n(S, K) \sim \infty 0$ is that

$$\xi_v = \frac{2(1 - \varepsilon_n)}{K + s(K)}$$

where $\xi_v$ is defined in (47) and $s(K)$ was defined as

$$s(K) \triangleq \frac{S \cdot r - K + 1 - \alpha^{-K}}{1 - \alpha}$$

where $\alpha$ controls what fraction of the events are discarded at each refinement cycle, $S$ is the aggregate sampling budget normalized by $n$, and $K$ is the maximum allowable number of refinement actions.

**Proof:** This results follows from the same line of argument as in the proof of Theorem 2.

Fig. 1 compares the regions in the $(r_m, \varepsilon_n)$ plane over which the adaptive and the nonadaptive procedures are guaranteed to make error-free detections. Specifically, the diagonally shaded region is the region in which both schemes succeed to detect the $T_n$ sequences of interest. In the vertically dashed region, however, only the adaptive procedure succeeds and the nonadaptive procedure makes an erroneous decision almost surely, and finally both schemes fail in the horizontally shaded region. It is observed that, depending on the choice of $K$, the detectability region corresponding to the adaptive procedure can be substantially larger than the one corresponding to the nonadaptive procedure. Fig. 2 depicts the counterpart regions in the variance setting in the $(\xi_v, \varepsilon_n)$ plane. Given the necessary and sufficient conditions on the scaling law for performing error-free detection, in the following corollaries, we determine the agility and scaling gains.

**Corollary 1 (Agility Gain):** When $\varepsilon_n = o(1), n \varepsilon_n = \omega(1)$, and $\alpha \leq 1 - \frac{1}{2}$ the agility gain satisfies

$$1 - \alpha^K \leq S_0(\nu_a^{-1} - \alpha^K) \leq \frac{1 - \alpha K + 1}{1 - \alpha}$$

and in the asymptote of large $K$, the upper and lower bounds on the agility gain meet and are equal to $S_0(1 - \alpha)$.

**Proof:** See Appendix I.

This result indicates that for a given set of sequences, the proposed sequential detection procedure can make a detection decision with substantially fewer measurements compared with the nonadaptive sampling procedure. Specifically, the sampling budget required by the adaptive procedure reduces exponentially with the number of refinement actions $K$. It is noteworthy that while the number of refinement actions $K$ can be made arbitrarily large (but fixed as a function of $n$), increasing it beyond some point increases the agility gain marginally. More specifically, for large $K$, both upper and lower bounds on the agility gain tend to $S_0(1 - \alpha)$, which is a constant. This observation sheds light on the fundamental limit of the agility gain yielded by adaptive sampling.

In Fig. 3, we provide a numerical comparison between the proposed adaptive sampling scheme and the CUSUM test repeated $T_n$ times. With the target error probability $P_n(S, K) = 10^{-9}$, the plot depicts the necessary average
number of samples by the CUSUM and the necessary number of samples by the adaptive sampling schemes for different number of refinement actions. We consider events and set the prior probability and aim to identify rare events. The normal events are distributed as and the rare events as with the variance values satisfying . It is observed that for the setting that distribution occurs very rarely (i.e., smaller values of ), the adaptive sampling strategy is quicker than the repeated CUSUM tests. On the other hand, by increasing , the repeated CUSUM test outperforms the adaptive sampling strategy.

Corollary 2 (Scaling Gain): When , the scaling gain satisfies

and in the asymptote of large , the upper and lower bounds on meet and are equal to .

Proof: The proof follows the same line of argument as the proof of Corollary 1. 

This corollary indicates that the scaling gain grows exponentially with the number of refinement actions. This result also indicates that the adaptive procedure can detect signals with much smaller and values than those detectable by the nonadaptive procedure. More specifically, by noting that is substantially larger than 1, the mean scaling requirement in the adaptive scenario becomes substantially less stringent than its counterpart in the nonadaptive scenario. As a result, there are scenarios in which nonadaptive schemes fail to successfully identify sequences of interest, while the adaptive scheme succeeds.

V. CONCLUSION

In this paper, we have presented an adaptive sampling methodology for quickly searching over finitely many events with the objective of identifying multiple events that occur sparsely and are distributed according to a given distribution of interest. The main idea of the sampling procedure is to successively and gradually adjust the measurement process using information gleaned from the previous measurements. Compared to corresponding nonadaptive procedures, dramatic gains in terms of reliability and agility are achieved.

APPENDIX A

ORDER STATISTICS ANALYSIS

In this appendix, we provide some definitions, notation, and basic results on asymptotic statistical behavior of order statistics that we frequently use throughout the rest of this paper. Given random variables , we denote the corresponding sequence of order statistics by , where is the -order statistic. When the ’s are independent and drawn from the same parent distribution with cdf , the cdf of , denoted by , satisfies

for some nondegenerate cdfs and . The random variables and , in general, have completely different statistical behavior in low-, central-, and high-order statistics. In the analysis in this paper, we need the distributions of all kind of orders, which are formally defined next.

Definition 3: The order statistic is said to be of low order and the order statistic is said to be of high order iff

Definition 4: The order statistic is said to be of central order iff

such that

Definition 5 (Domain of Attraction): A given parent cdf for is said to belong to the minimal domain of attraction of a cdf if there exists at least one pair of sequences and such that the sequence of random variables defined in satisfies . Similarly, we say that belongs to the maximal domain of attraction of a cdf if for at least one pair of sequences and , the sequence of random variables defined in satisfies .

Finding the possible cdfs and corresponding to any given parent distribution has a rich literature; cf. [36] or [37]. The well-known result is that only one parametric family is possible for and only one family for .

Theorem 4 (Von-Mises Family of Distribution): The only nondegenerate family of distributions satisfying is of the form

where and the only nondegenerate family of distributions satisfying is of the form

When tends to infinity, the limit distributions and become degenerate as

and

A common approach to avoid degeneracy is to identify two appropriate sets of affine transformations of , , that

and

for some nondegenerate cdfs and . The random variables and , in general, have completely different statistical behavior in low-, central-, and high-order statistics. In the analysis in this paper, we need the distributions of all kind of orders, which are formally defined next.

Definition 3: The order statistic is said to be of low order and the order statistic is said to be of high order iff

Definition 4: The order statistic is said to be of central order iff

such that

Definition 5 (Domain of Attraction): A given parent cdf for is said to belong to the minimal domain of attraction of a cdf if there exists at least one pair of sequences and such that the sequence of random variables defined in satisfies . Similarly, we say that belongs to the maximal domain of attraction of a cdf if for at least one pair of sequences and , the sequence of random variables defined in satisfies .

Finding the possible cdfs and corresponding to any given parent distribution has a rich literature; cf. [36] or [37]. The well-known result is that only one parametric family is possible for and only one family for .
where

\[ 1 - \kappa \left( \frac{w - \lambda}{\sigma} \right) \geq 0. \]

**Proof:** See [36] and [37].

Besides the limit distributions of minima and maxima, in this paper, we also need to find the asymptotic nondegenerate distributions corresponding to low- and central-order statistics. The following theorems characterize the asymptotic nondegenerate distributions for these order statistics as functions of the limit distributions of minima.

**Theorem 5 (Low-Order Statistics):** Let \( G \) be a continuous cdf that belongs to the minimal domain of attraction of a cdf \( L \) with the associated pair of sequences \( \{a_m\}_{m=1}^\infty \) and \( \{b_m\}_{m=1}^\infty \). Then, for \( r = o(m) \) and \( w \in \mathbb{R} \), we have

\[
\lim_{m \to \infty} Q_{r,m}(w; m) = 1 - I_0 \{ L(w) < 1 \} \cdot \left[ 1 - I_0(w) \right] \sum_{i=0}^{r} \frac{[-\ln(1 - L(w))]^i}{i!}. \tag{63}
\]

**Proof:** See [38, Th. 8.4.1].

**Theorem 6 (Central-Order Statistics):** Let a continuous cdf \( G \) with associated pdf \( g \) be the distribution of \( m \) i.i.d. random variables. Let \( \zeta \) be a real number in the interval \([0,1]\) such that \( g(G^{-1}(\zeta)) \neq 0 \). Then, central-order statistic \( r = \lceil m \zeta \rceil \) is distributed as

\[ Y_{r,m} \sim \mathcal{N} \left( G^{-1}(\zeta); \frac{\zeta(1 - \zeta)}{m [g(G^{-1}(\zeta))]^2} \right). \]

**Proof:** See [38, Th. 8.5.1].

**APPENDIX B**

**DOMAINS OF ATTRACTION**

Analyzing the performance of the refinement actions as well as the final detection action strongly rely on analyzing the domain of attraction of Gaussian and \( \chi^2 \) distributions. In this appendix, we provide the necessary domains of attractions for these distributions. The following lemma determines to what minimal domain of attraction the standard Gaussian distribution belongs.

**A) Gaussian Minimum:**

**Lemma 6 (Gaussian Minimum):** The standard Gaussian distribution belongs to the minimal domain of attraction of

\[ L(w) = 1 - \exp \left( -\exp(w - \ln(2\sqrt{\pi})) \right), \quad \forall w \in \mathbb{R} \]

corresponding to (61) for \( \kappa \to 0 \). The associated affine transformation is characterized by the sequences \( \{a_m\} \) and \( \{b_m\} \) given by

\[ a_m = h(m), \quad \text{and} \quad b_m = \sqrt{h(m)}, \quad \forall m \in \{2, 3, \ldots\} \]

where \( h : \{x \in \mathbb{R} : x > 1\} \to \mathbb{R}^+ \) is defined as

\[ h(x) = 2 \ln x - \ln \ln x. \tag{64} \]

**Proof:** Let us define \( Y_1, \ldots, Y_m \) as i.i.d. standard Gaussian random variables with cdf \( G \) and pdf \( g \) and define the function \( h : \{x \in \mathbb{R} : x > 1\} \to \mathbb{R}^+ \) as in (64).

By defining

\[ W_i = a_m + b_m Y_i \tag{65} \]

and setting

\[ a_m = h(m), \quad \text{and} \quad b_m = \sqrt{h(m)} \tag{66} \]

for the cdf of \( W_{1:m} \), denoted by \( Q_{1:m}(\cdot; m) \), we have

\[
1 - Q_{1:m}(w; m) = 1 - P \left( W_{1:m} \leq w \right) \tag{65};
\]

\[
= 1 - P \left( Y_m \leq \frac{w - a_m}{b_m} \right) \tag{66};
\]

\[
= 1 - G \left( \frac{w}{\sqrt{h(m)}} - \sqrt{h(m)} \right)^m. \tag{67}
\]

We next show that \( \forall w \in \mathbb{R} \)

\[ \lim_{x \to \pm \infty} x \cdot \log \left[ 1 - G \left( \frac{w}{\sqrt{h(x)}} - \sqrt{h(x)} \right) \right] = -\frac{e^w}{2\sqrt{\pi}} \tag{68} \]

by using L'Hôpital's rule, we obtain the equations given at the bottom of the page.
Equations (67) and (68) in conjunction with the continuity of \( \ln(\cdot) \) establish that
\[
\lim_{m \to \infty} Q_{1:m}(w; m) = 1 - \exp \left( - \frac{e^w}{2\sqrt{m}} \right) \tag{69}
\]
which is the desired result.

B) \( \chi^2 \) Minimum:

Lemma 7 (\( \chi^2 \) Minimum): The chi-squared distribution with \( k \) degrees of freedom belongs to the minimal domain of attraction of
\[
I(w) = 1 - \exp \left( -w^{k/2} \right), \quad \forall w \in \mathbb{R}
\]
which is corresponding to (61) for \( \kappa \to 0 \). The associated affine transformation is characterized by the sequences \( \{a_m\} \) and \( \{b_m\} \) for \( m \in \{2, 3, \ldots\} \) given by
\[
a_m = 0, \quad \text{and} \quad b_m = \frac{1}{2} \left[ \frac{m}{\Gamma(k/2 + 1)} \right]^{2/k}
\]

Proof: Let us define \( Y_1, \ldots, Y_m \) as i.i.d. random variables distributed according to \( \chi^2(k) \). For the cdf of \( \chi^2(k) \) random variables, we have
\[
F_Y(y) = \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^y t^{k/2-1} \exp(-t/2) \, dt \tag{70}
\]
As for \( t \in [0, y] \) we have \( \exp(-y/2) \leq \exp(-t/2) \leq 1 \), we obtain the following bounds on \( F_Y(y) \):
\[
\begin{align*}
F_Y(y) & \geq \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^y t^{k/2-1} \, dt \\
F_Y(y) & \leq \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^y t^{k/2-1} \, dt
\end{align*}
\tag{71}
\]
which subsequently provide
\[
\frac{\exp(-y/2)y^{k/2}}{2^{k/2}\Gamma(k/2 + 1)} \, dt \leq F_Y(y) \leq \frac{y^{k/2}}{2^{k/2}\Gamma(k/2)} 
\tag{72}
\]
Next, by setting
\[
a_m = 0, \quad \text{and} \quad b_m = \frac{1}{2} \left[ \frac{m}{\Gamma(k/2 + 1)} \right]^{2/k}
\]
for \( m \in \{2, 3, \ldots\} \) and defining
\[
W_i = a_m + b_m Y_i
\]
for the cdf of \( W_i \), denoted by \( F_W(w) \), we have \( F_W(w) = \mathcal{P}(W_i \leq w) = F_Y(w/b_m) \). By invoking the inequalities in (73), we find the following bounds on \( F_W(w) \):
\[
\exp \left( -w \left[ \frac{\Gamma(k/2 + 1)}{m} \right]^{2/k} \right) \leq F_W(w) \leq \frac{w^{k/2}}{m} \tag{74}
\]
which implies that in the asymptote of large \( m \), we have
\[
F_W(w) = \frac{w^{k/2}}{m} (1 + o(1)).
\]
Therefore, for the cdf of \( W_{1:m} \) (the first-order statistic of \( \{W_1, \ldots, W_m\} \), denoted by \( Q_{1:m}(\cdot; m) \), in the asymptote of large \( m \), we have
\[
Q_{1:m}(\cdot; m) = 1 - \left[ 1 - \frac{w^{k/2}}{m} (1 + o(1)) \right]^m = 1 - \exp \left( -w^{k/2} \right).
\]

Also, as will be shown later, for analyzing the performance of the detection action, we also need to determine to what maximal domain of attraction a \( \chi^2 \) distribution belongs.

C) \( \chi^2 \) Maximum:

Lemma 8 (\( \chi^2 \) Maximum): The chi-squared distribution with \( k \) degrees of freedom belongs to the maximal domain of attraction of
\[
H(w) = \exp \left( \frac{1}{\Gamma(k/2)} \cdot \exp(-w) \right), \quad \forall w \in \mathbb{R}
\]
which is corresponding to (62) for \( \kappa \to 0 \). The associated affine transformation is characterized by the sequences \( \{a_m\} \) and \( \{b_m\} \) for \( m \in \{2, 3, \ldots\} \) given by
\[
c_m = - \left( \ln m + \frac{K - 1}{2} \ln \ln m \right), \quad \text{and} \quad d_m = \frac{1}{2}
\]

Proof: Let us define \( Y_1, \ldots, Y_m \) as i.i.d. random variables distributed according to \( \chi^2(k) \) with cdf \( F_Y(y) \) and define the function \( g: \{x \in \mathbb{R} : x > 1\} \to \mathbb{R}^+ \) as
\[
g(x) = \frac{x^n}{\exp(\frac{K - 1}{2} \ln \ln x)}.
\]
By defining
\[
W_i = c_m + d_m Y_i \tag{75}
\]
and setting
\[
c_m = u(m), \quad \text{and} \quad b_m = \frac{1}{2}
\tag{76}
\]
for the cdf of \( W_{m:m} \), denoted by \( Q_{m,m}(\cdot; m) \), we have
\[
Q_{m,m}(w; m) = \mathcal{P}(W_{m:m} \leq w) \tag{78}
\]
\[
= \left[ \mathcal{P} \left( Y_{m:m} \leq \frac{w - c_m}{d_m} \right) \right]^m \tag{77}
\]
We next show that
\[
\lim_{x \to \infty} x \cdot \log \left[ F_Y(2(w - u(x))) \right] = -\frac{\exp(-w)}{\Gamma(k/2)}, \quad \forall w \in \mathbb{R}
\tag{78}
This is shown by using L’Hôpital’s rule in (79) given at the bottom of the page.

Equations (77) and (78) demonstrate that

\[
\lim_{m \to \infty} Q_{m:m}(w; m) = \exp \left( -\frac{1}{\Gamma(k/2)} \exp(-w) \right)
\]

which is the desired result.

**APPENDIX C**

**PROOF OF LEMMA 1**

From the definitions of \( n_t \) and \( n_{t+1} \) for all \( t \in \{1, \ldots, \tau \} \), we have \( L_t = n_t + \bar{\eta}_t \). Therefore, for any time \( t \in \{1, \ldots, \tau - 1\} \) that the refinement action is taken (i.e., \( \psi(t) = 1 \)), the condition

\[
(1 - \bar{\delta})n_t \leq n_{t+1}
\]

for some \( \bar{\delta} \in (0, 1) \) is equivalent to having

\[
\bar{\eta}_{t+1} \leq |L_{t+1}| - (1 - \bar{\delta})n_t^*.
\]

By taking into account the distributions of the order statistics \( U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \) and \( U^t\{1, \ldots, \bar{\eta}_t\} \) given in (26) and (27), respectively, the event that the \( |L_{t+1}| \) retained events after a refinement action at time \( t \) contain \( (1 - \bar{\delta})n_t \) rare events is equivalent to

\[
U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \leq U^t\{1, \ldots, \bar{\eta}_t\} - (1 - \bar{\delta})n_t^*.
\]

In order to analyze this probability in the following two lemmas, we show that the order statistics \( U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \) and \( U^t\{1, \ldots, \bar{\eta}_t\} \) are of central orders and specify their distributions.

**Lemma 9**: The order statistic \( U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \) is of central order and is almost surely distributed as \( \mathcal{N}(\tilde{M}_t, \tilde{\sigma}_t^2) \) for

\[
\tilde{M}_t = \mu_0 t + \sqrt{2t} \text{erf}^{-1}(2\bar{\delta} - 1)
\]

and

\[
\tilde{\sigma}_t^2 = \frac{2t}{n_t} \exp(2\text{erf}^{-1}(2\bar{\delta} - 1)).
\]

Proof: Note that in the asymptote of large \( n \), we almost surely have \( n_t = \omega(1) \) and Definition 4 confirms that \( U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \) is an order statistic of central order from a sequence of \( \bar{\eta}_t \) i.i.d. random variables with the parent distribution \( \mathcal{N}(\mu_0, \tilde{\sigma}_t^2) \). By using Theorem 6 in the asymptote of large \( n \), we have \( U^t_{\{(1 - \bar{\delta})n_t^* + 1\}} \sim \mathcal{N}(\tilde{M}_t, \tilde{\sigma}_t^2) \) for \( \tilde{M}_t \) and \( \tilde{\sigma}_t^2 \) defined in (82).

**Lemma 10**: The order statistic \( U^t\{1, \ldots, \bar{\eta}_t\} \) is of central order and is almost surely distributed as \( \mathcal{N}(\tilde{M}_t, \tilde{\sigma}_t^2) \) for

\[
\tilde{M}_t = \mu_0 t + \sqrt{2t} \text{erf}^{-1}(2\alpha - 1)
\]

and

\[
\tilde{\sigma}_t^2 = \frac{2t}{n_t} \alpha \exp(2\text{erf}^{-1}(2\alpha - 1)).
\]

where \( \text{erf}(\cdot) \) denotes the Gauss error function.

Proof: The proofs follows the same line of arguments as in the proof of Lemma 9 and noting that in the asymptote of large \( n \)

\[
\frac{|L_{t+1} - (1 - \bar{\delta})n_t^* + 1|}{\bar{\eta}_t} = \frac{\alpha(n_t + \bar{\eta}_t - T_u)}{n_t} = (1 - \bar{\delta})n_t^* + 1
\]

holds almost surely.

Next, by defining

\[
\tilde{\delta}_t^2 \triangleq \sigma_t^2 \tilde{\sigma}_t^2 + \tilde{\sigma}_t^2
\]

and by using (83)–(82), we obtain after a refinement action at time \( t \) (i.e., \( \psi(t) = 1 \))

\[
\mathbb{P} \left( (1 - \bar{\delta})n_t \leq n_{t+1} | n_t \right) = 1 - \int_{\tilde{M}_t}^{\tilde{M}_t + \tilde{\sigma}_t \sqrt{2\pi}} \frac{\tilde{L}_t}{\tilde{\sigma}_t} \text{erf}(\tilde{L}_t) \exp \left( -2t \right) \text{d}L_t
\]

\[
> 1 - \frac{1}{M_t - \tilde{M}_t} \exp \left( -\frac{1}{2} \left( \frac{M_t - \tilde{M}_t}{\tilde{\sigma}_t} \right)^2 \right)
\]

\[
(85)
\]

\[
(79)
\]
where we have used $\int_0^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \ dt < \frac{1}{\sqrt{2}} \cdot \exp(-x^2/2)$ for $x > 0$. Hence, for the number of rare events discarded by the refinement actions, collectively, we have

$$
\mathbb{P} \left( n_\tau \geq n_1 (1 - \delta)^K n_1 \right)
\geq \prod_{t: \psi(t) = 1} \mathbb{P} \left( n_{\tau+1} \geq n_1 (1 - \delta) n_1 \right)
\geq \prod_{t: \psi(t) = 1} \left[ 1 - \frac{M_t - \bar{M}_t}{\tilde{\sigma}_t}, \exp \left\{ \frac{1}{2} \cdot \left( \frac{M_t - \bar{M}_t}{\tilde{\sigma}_t} \right)^2 \right\} \right]^{-K}
\geq \left[ 1 - \frac{M_0 - \bar{M}_0}{\tilde{\sigma}_0}, \exp \left\{ -\frac{1}{2} \cdot \left( \frac{M_0 - \bar{M}_0}{\tilde{\sigma}_0} \right)^2 \right\} \right]^{-K}
$$

(87)

where the first inequality holds since $\sum_{t: \psi(t) = 1} \leq K$ and the last inequality holds as $\tilde{\sigma}_t$ is increasing in $t$. By setting $\tilde{\sigma}_0 = 1 - (1 - \delta)^1/K$, (86) indicates that a sufficient condition that ensures $n_\tau \geq (1 - \delta)n_1$ almost surely is that

$$
\frac{M_0 - \bar{M}_0}{\tilde{\sigma}_0} n_1 \rightarrow 1.
$$

(88)

Now, from (84), recall that $\tilde{\sigma}_t^2$ has two summands, where one scales with $1/n_1$ and the other one scales with $1/n_1$. By noting that the stopping time can be bounded by a constant (Lemma 3), it can be readily verified that (88) holds if

$$
\mu_0 - \mu_1 = \omega \left( \sqrt{\frac{1}{\bar{n}_1} + \frac{1}{n_1}} \right)
$$

which, in turn, holds almost surely if $\mu_0 - \mu_1 = \omega \left( \bar{n}^{-\tau/2} \right)$.

**APPENDIX D**

**PROOF OF LEMMA 2**

By using (28), we have the connection between the error probability $\mathbb{P}_r(\tau, \tilde{\psi}(\tau))$ for given $\tau$ and $\tilde{\psi}(\tau)$ and low-order statistics of two sequences of sizes $n_\tau$ and $\bar{n}_\tau$ with Gaussian parent distributions with different means. According to (28), we have

$$
\mathbb{P}_r(\tau, \tilde{\psi}(\tau)) = \mathbb{P} \left( U_{n_\tau}^T > \bar{U}_1^T \right)
$$

In order to find the asymptotic distribution and the associated minimal domain of attraction of $U_{n_\tau}^T$, we use the result of Lemma 6. For this purpose, let us define $h : \{ x \in \mathbb{R} : x > 1 \} \rightarrow \mathbb{R}^+$ as

$$
h(x) = 2 \ln x - \ln \ln x.
$$

Based on the definition and distribution of $Z_i^\tau$ for $i \in \mathcal{L}_\tau$ given in (22)–(25), let us define

$$
\forall i \in \mathcal{L}_\tau^c : \quad \bar{W}_i^\tau = h(\bar{n}_\tau) + \sqrt{h(\bar{n}_\tau)} \left( \frac{Z_i^\tau - \mu_0 \cdot \tau}{\sqrt{\tau}} \right) \sim \mathcal{N}(0, 1)
$$

(89)

$$
\forall i \in \mathcal{L}_\tau^c : \quad \bar{W}_i^\tau = h(n_\tau) + \sqrt{h(n_\tau)} \left( \frac{Z_i^\tau - \mu_0 \cdot \tau}{\sqrt{\tau}} \right) \sim \mathcal{N}(0, 1)
$$

(90)

Given the definitions in (89) and (90), we obtain

$$
\mathbb{P}_r(\tau, \tilde{\psi}(\tau)) = \mathbb{P} \left( U_{n_\tau}^T > \bar{U}_1^T \right) = \mathbb{P} \left( \frac{U_{\bar{n}_\tau}^T - \mu_0 \cdot \tau}{\sqrt{\tau}} > \bar{U}_1^T - \frac{\bar{n}_\tau - \mu_0 \cdot \tau}{\sqrt{\tau}} + \sqrt{\tau} (\mu_0 - \mu_1) \right)
$$

(91)

$$
= \mathbb{P} \left( W_{n_\tau, n_1} > A_n \cdot \bar{W}_{n_1, n_1} + B_n \right)
$$

(92)

where we have defined

$$
A_n \triangleq \frac{h(n_\tau)}{h(\bar{n}_\tau)} \quad \text{and} \quad H_n \triangleq \sqrt{h(n_\tau)} - \sqrt{h(\bar{n}_\tau)} + \sqrt{\tau} (\mu_0 - \mu_1)
$$

(93)

In order to assess $\mathbb{P}_r(\tau, \tilde{\psi}(\tau))$ given in (92), we find the distributions of $W_{n_\tau, n_1}$ and $\bar{W}_{n_1, n_1}$. For this purpose, according to Lemma 6, the cdfs of $W_{n_\tau, n_1}$ and $\bar{W}_{n_1, n_1}$, denoted by $Q_{1,n_\tau, n_1}$ and $Q_{1,n_1, n_1}$, respectively, satisfy

$$
\lim_{n_\tau \rightarrow \infty} Q_{1,n_\tau, n_1}(w; \bar{n}_\tau) = 1 - \exp(-\exp(w - \ln 2\sqrt{\tau}))
$$

$$
\lim_{n_\tau \rightarrow \infty} Q_{1,n_\tau, n_1}(w; n_\tau) = 1 - \exp(-\exp(w - \ln 2\sqrt{\tau})).
$$

(94)

Furthermore, given the asymptotic distributions of the first-order statistics by using Theorem 5, we can find the distributions of the $r$th other (low-order) statistics. By denoting the cdfs of $W_{r,n_\tau}$ and $\bar{W}_{n_1, n_1}$ by $Q_{r,n_\tau, n_1}$ and $Q_{r,n_1, n_1}$, respectively, we have

$$
\lim_{n_\tau \rightarrow \infty} Q_{r,n_\tau, n_1}(w + \varsigma; \bar{n}_\tau) = 1 - \sum_{i=0}^{r-1} \frac{\exp(iw - \exp(w))}{i!}
$$

$$
\lim_{n_\tau \rightarrow \infty} Q_{r,n_\tau, n_1}(w + \varsigma; n_\tau) = 1 - \sum_{i=0}^{r-1} \frac{\exp(iw - \exp(w))}{i!}
$$

(95)
where
\[ \zeta \triangleq \ln 2\sqrt{\pi}. \]
The associated pdfs, consequently, are given by
\[ \lim_{n_\tau \to \infty} q_{n,n_\tau}(w + \zeta; n_\tau) = \sum_{i=0}^{\infty} \frac{(\exp(w) - i)^i \exp(iw - \exp(w))}{i!}, \quad \forall w \in \mathbb{R} \]
and
\[ \lim_{n_\tau \to \infty} q_{n,n_\tau}(w + \zeta; n_\tau) = \sum_{i=0}^{\infty} \frac{(\exp(w) - i)^i \exp(iw - \exp(w))}{i!}, \quad \forall w \in \mathbb{H}. \]

By using the distributions given in (94)–(97) for given \( \tau \) and \( \bar{\psi}(\tau) \), we find lower and upper bounds on \( P_n(\tau, \bar{\psi}(\tau)) \). These bounds, in turn, serve as the basis for finding necessary and sufficient conditions on the scaling laws of \((\mu_0 - \mu_1)\) that guarantee
\[ P_n(\tau, \bar{\psi}(\tau)) \to 0. \]

**Lower bound:** By replacing \( W_{T,n_\tau} \) by \( W_{1,n_\tau} \), based on (92), \( P_n(\tau, \bar{\psi}(\tau)) \) can be lower bounded for sufficiently large \( n \) as follows:
\[ P_n(\tau, \bar{\psi}(\tau)) \geq P\left(W_{1,n_\tau} > A_n \cdot W_{1,n_\tau} + B_n\right) \]
\[ = 1 - \int_{-\infty}^{0} Q_{1,n_\tau}(A_n w + B_n) q_{1,n_\tau}(w) \, dw \]
\[ - \int_{0}^{\infty} Q_{1,n_\tau}(w + B_n) q_{1,n_\tau}(w) \, dw \]
\[ \geq 1 - \int_{-\infty}^{0} Q_{1,n_\tau}(B_n) q_{1,n_\tau}(w) \, dw \]
\[ - \int_{0}^{\infty} Q_{1,n_\tau}(w) q_{1,n_\tau}(w) \, dw \]
\[ - 1 - Q_{1,n_\tau}(0) \]
\[ + \int_{0}^{\infty} (1 - Q_{1,n_\tau}(w)) q_{1,n_\tau}(w) \, dw \]
\[ = \exp(-\exp(-\zeta)) \]
\[ \times \left\{ 1 - \exp(-\exp(-\zeta)) + \frac{\exp(-\exp(-\zeta))}{\exp(B_n) + 1} \right\}. \]

This strictly positive lower bound on \( P_n(\tau, \bar{\psi}(\tau)) \) approaches zero if and only if \( B_n = \omega(1) \).

**Upper Bound:** From (28) for sufficiently large \( n \), we have
\[ P_n(\tau, \bar{\psi}(\tau)) = P\left(W_{T,n_\tau}^0 > A_n \cdot W_{1,n_\tau} + B_n\right) \]
\[ = 1 - \int_{-\infty}^{0} Q_{T,n_\tau}(A_n w + B_n) q_{1,n_\tau}(w) \, dw \]
\[ - \int_{0}^{\infty} Q_{T,n_\tau}(w + B_n) q_{1,n_\tau}(w) \, dw \]
\[ < 1 - \int_{-\infty}^{0} Q_{T,n_\tau}(B_n) q_{1,n_\tau}(w) \, dw \]
\[ - \int_{0}^{\infty} Q_{T,n_\tau}(w) q_{1,n_\tau}(w) \, dw \]
\[ - 1 - Q_{T,n_\tau}(0) \]
\[ + \int_{-\infty}^{\infty} \exp(w - \exp(w + B_n) - \exp(w)) \, dw \]

which can be easily shown to approach zero if and only if \( B_n = \omega(1) \).

Therefore, a necessary and sufficient condition for having both the lower and upper bounds \( P_n(\tau, \bar{\psi}(\tau)) \) approach zero is that \( B_n = \omega(1) \), which subsequently becomes a necessary and sufficient condition for \( P_n(\tau, \bar{\psi}(\tau)) \to 0 \). According to (93), \( B_n = \omega(1) \) is equivalent to
\[ \frac{B_n}{h(n_\tau) \sqrt{2 \ln n}} \to \frac{1}{h(n_\tau) \sqrt{2 \ln n}} \]
\[ \to \frac{1}{\sqrt{\tau} \cdot \sqrt{r_m} - (1 - \sqrt{\varepsilon_n})} = \omega\left(\frac{1}{h(n_\tau) \sqrt{2 \ln n}}\right) \]

where \( h(\cdot) \) is defined in (64) and it holds in the asymptote of large \( n \) almost surely if and only if
\[ \sqrt{\tau} \cdot \sqrt{r_m} - (1 - \sqrt{\varepsilon_n}) = \omega\left(\frac{1}{h(n_\tau) \sqrt{2 \ln n}}\right) \]
or equivalently
\[ r_m > \frac{(1 - \sqrt{\varepsilon_n})^2}{\tau}. \]

**APPENDIX E**

**PROOF OF LEMMA 3**

From the definition of \( \psi(t) \) given in (8) \( \forall t \in \{1, \ldots, \tau - 1\} \), we have
\[ \mathcal{L}_{t+1} = |\mathcal{L}_{t} \cdot \mathbb{1}_{\psi(t)=0} + \{\alpha(\mathcal{L}_{t} - T_n) + T_n \cdot \mathbb{1}_{\psi(t)=1}\} | \]
\[ \geq |\mathcal{L}_{t} \cdot \mathbb{1}_{\psi(t)=0} + \{\alpha(\mathcal{L}_{t} - 1) \cdot \mathbb{1}_{\psi(t)=1}\} | \]

(102)
where \( \mathbb{1}_{\{1\}} \) is the indicator function. Following this relationship, the number of sequences retained and fed to the detector is related to the initial number of sequences \( n \) according to

\[
|L_t| \geq \alpha \sum_{t=0}^{t} \mathbb{1}_{\{1\}} |L_1| - \sum_{t=0}^{t} \alpha^t
\]

\[
\geq \alpha^K |L_1| - \frac{1 - \alpha^K}{1 - \alpha}
\]

\[
= \alpha^K \cdot n - \frac{1 - \alpha^K}{1 - \alpha}.
\]

(103)

The constraint \( \frac{1}{n} \sum_{t=1}^{T} |L_t| \leq S \) in conjunction with (103) provides

\[
S \cdot n \geq \sum_{t=1}^{T} |L_t| \geq \tau \cdot \left( \alpha^K \cdot n - \frac{1 - \alpha^K}{1 - \alpha} \right)
\]

which shows that in the asymptote of large \( n \), we have \( \tau \leq S/\alpha^K \).

**APPENDIX F**

**PROOF OF LEMMA 4**

From the definitions of \( n_t \) and \( \bar{n}_t \) for all \( t \in \{1, \ldots, \tau\} \), we have \( |L_t| = n_t + \bar{n}_t \). Therefore, for any time \( t \in \{1, \ldots, \tau - 1\} \) that the refinement action is taken (i.e., \( \psi(t) = 1 \)), the condition

\[
n_t = n_{t+1}
\]

is equivalent to having

\[
\bar{n}_{t+1} = |L_{t+1}| - n_t.
\]

(104)

By taking into account the distributions of the order statistics \( V_{n_t}^t \) and \( V_{\bar{n}_t}^t \) given in (40) and (41), respectively, the event that the \( |L_{t+1}| \) retained events after a refinement action at time \( t \) contain \( n_t \) rare events is equivalent to

\[
V_{n_t}^t < V_{|L_{t+1}| - n_{t+1}}^t.
\]

(105)

In order to analyze this probability in the following lemma, we show that the order statistic \( V_{|L_{t+1}| - n_{t+1}}^t \) is of central order and specify their distributions.

**Lemma 11:** The order statistic \( V_{|L_{t+1}| - n_{t+1}}^t \) is of central order and

\[
V_{|L_{t+1}| - n_{t+1}}^t \sim \mathcal{N} \left( \bar{\mu}, \bar{\sigma}^2 \right)
\]

(106)

where

\[
\bar{\mu} \sim \mathcal{N} \left( \bar{\mu}, \bar{\sigma}^2 \right)
\]

(107)

for

\[
\bar{\mu} = \mathcal{G}^{-1}(\alpha), \quad \text{and} \quad \bar{\sigma}^2 = \alpha \left( 1 - \alpha \right) \frac{2}{\mathbb{n}_t \mathbb{g}(\mathcal{G}^{-1}(\alpha))}^2
\]

and \( \mathcal{G} \) and \( \mathbb{g} \) denote the cdf and pdf of \( \chi^2(t) \).

**Proof:** The proof of this follows from the same line of argument as in the proof of Lemma 9.

Therefore, from (106) and (107), we obtain when \( \psi(t) = 1 \)

\[
\mathbb{P} \left( n_{t+1} = n_t | n_t \right) = \int_{-\infty}^{\infty} \mathbb{P} \left( \frac{V_{n_t}^t - \ln n_t}{2A_1} < \frac{A_0}{2A_1} \cdot \alpha - \ln n_t \right) f_A(a) \, da
\]

(108)

where \( f_A \) denotes the pdf of \( \bar{\lambda} \). Based on the definition of \( V_{n_t}^t \) given in (40), \( V_{n_t}^t \) is the highest order statistic of a sequence of \( n_t \) i.i.d. random variables with the parent distribution \( \chi^2(t) \).

Hence, by using Lemma 8, we obtain

\[
\mathbb{P} \left( \frac{V_{n_t}^t}{2A_1} - \ln n_t \leq w \right) = \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp(-w) \right).
\]

(109)

Therefore, from (108) and (109), we get

\[
\mathbb{P} \left( n_{t+1} = n_t \right) = \int_{-\infty}^{\infty} \mathbb{P} \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot a \right) \right) f_A(a) \, da
\]

(110)

\[
\geq \int_{\bar{\mu}/2}^{\infty} \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot a \right) \right) f_A(a) \, da
\]

(111)

\[
\geq \int_{\bar{\mu}/2}^{\infty} \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot \frac{\bar{\mu}}{2} \right) \right) f_A(a) \, da
\]

(112)

\[
= \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot \frac{\bar{\mu}}{2} \right) \right) \times \int_{\bar{\mu}/2}^{\infty} f_A(a) \, da
\]

\[
- \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot \frac{\bar{\mu}}{2} \right) \right) \times \mathbb{P}(\bar{\lambda} - \bar{\mu} \leq \bar{\mu}/2)
\]

\[
- \exp \left( \frac{1}{\Gamma(t/2)} \cdot \exp \left( \ln n_t - \frac{A_0}{2A_1} \cdot \frac{\bar{\mu}}{2} \right) \right) \times \left[ 1 - \left( \frac{2\bar{\sigma}}{\Gamma} \right)^2 \right]
\]

(113)

where (111) holds by narrowing the interval of integral, (112) holds as the integrand in (111) takes its smallest value at \( a = \bar{\mu}/2 \), and (113) holds according to Chebyshev’s inequality. Next, note that in the asymptote of large \( n \) almost surely, we have \( n_t = \omega(1) \), and consequently, \( \left( \frac{2\bar{\sigma}}{\Gamma} \right)^2 \xrightarrow{n \to \infty} 0 \). Hence

\[
\frac{A_0}{2A_1} \cdot \frac{\bar{\mu}}{2} - \ln n_t = \omega(1)
\]

(114)

is a sufficient condition for ensuring \( \mathbb{P} \left( n_{t+1} = n_t \right) \xrightarrow{n \to \infty} 1 \). In turn, this condition is satisfied in the asymptote of large \( n \) almost surely when \( \frac{\bar{\mu}}{\Gamma} = \omega(\ln \epsilon \ln n) \) or equivalently \( \frac{A_0}{2A_1} = \omega(\epsilon \ln n) \).
APPENDIX G
PROOF OF LEMMA 5

According to (42), the detection error probability is

\[ P_n(\tau, \tilde{\psi}(\tau)) = P \left( V^*_\tau > \bar{V}^*_\tau \right). \]

In order to find the asymptotic distribution and the associated minimal domains of attraction of \( V^*_\tau \) and \( \bar{V}^*_\tau \), we use the result of Lemma 7. For this purpose, let us define

\[ \bar{h}(x) = \frac{1}{2} \cdot x^{2/\tau}. \quad (116) \]

Also, let us set

\[ \forall \bar{i} \in L^0 : \quad \bar{W}^{\bar{i}} = \frac{A_0 \cdot \bar{h}(n_{\tau})}{A_0} \quad (117) \]

and \( \forall \bar{i} \in L^1 : \quad W^{\bar{i}} = \frac{h(n_{\tau})}{A_1} \). \quad (118)

Therefore, according to Lemma 7, the cdfs of \( W_{1,n} \) and \( \bar{W}_{1,n} \), denoted by \( Q_{1,n}(w; n_{\tau}) \) and \( \bar{Q}_{1,n}(w; n_{\tau}) \), respectively, satisfy

\[ \lim_{n_{\tau} \to \infty} \bar{Q}_{1,n_{\tau}}(w; n_{\tau}) = 1 - \exp \left( -w^{\tau/2} \right) \]

and \( \lim_{n_{\tau} \to \infty} Q_{1,n_{\tau}}(w; n_{\tau}) = 1 - \exp \left( -w^{\tau/2} \right). \)

Furthermore, by using Theorem 5, we can find the distribution of the \( T_n^{\tau} \)-other (low-order) statistic \( W_{1,n_{\tau}} \) as follows:

\[ \lim_{n_{\tau} \to \infty} Q_{T_n^{\tau},n_{\tau}}(w) = 1 - \exp \left( -w^{\tau/2} \right) \cdot \sum_{i=0}^{T_n^{\tau}-1} \frac{w^{i/2}}{i!}. \quad (119) \]

Given the definitions in (117) and (118), we obtain

\[ P_n(\tau, \tilde{\psi}(\tau)) = P \left( V^*_\tau > \bar{V}^*_\tau \right) = 1 - P \left( W_{1,n_{\tau}} < \frac{A_1}{\bar{h}(n_{\tau})} \cdot \frac{A_0}{A_0} \cdot \bar{h}(n_{\tau}) \right) \\
= 1 - P \left( W_{T_n^{\tau},n_{\tau}} < \frac{A_1}{\bar{h}(n_{\tau})} \cdot \frac{A_0}{A_0} \cdot \bar{h}(n_{\tau}) \right). \]

By setting

\[ \Theta = \frac{A_0}{A_1} \cdot \frac{\bar{h}(n_{\tau})}{\bar{h}(n_{\tau})} \]

we find that \( P_n(\tau, \tilde{\psi}(\tau)) \) can be cast as in (115). By further setting

\[ \Theta^{\tau/2} \]

we find

\[ P_n(\tau, \tilde{\psi}(\tau)) = \sum_{i=0}^{T_n^{\tau}-1} \frac{\Theta^{i/2}}{i!} \cdot \frac{1}{\tau/2 (1 + \Theta^{\tau/2})^{i+1}} \times \int_{0}^{\infty} \exp(-x) x^i \ dx \]

\[ = 1 - \left( \frac{1 - \sum_{i=0}^{T_n^{\tau}-1} \Theta^{i/2}}{1 + \Theta^{\tau/2}} \right)^{T_n^{\tau}} \]

\[ = 1 - \left( \frac{1 - \sum_{i=0}^{T_n^{\tau}-1} \Theta^{i/2}}{1 + \Theta^{\tau/2}} \right)^{T_n^{\tau}} \]

Hence, the requirement \( P_n(\tau, \tilde{\psi}(\tau)) \to \infty \) is equivalent to \( \Theta^{\tau/2} = \omega(1) \), which by taking into account (116) and (120) is, in turn, equivalent to

\[ \frac{A_0}{A_1} = \omega \left( \sqrt{\frac{\bar{h}_{n_{\tau}}}{n_{\tau}}} \right). \]

By taking into account that \( \bar{h}_{n_{\tau}} \approx \omega(n_{\tau}) \) for some constant \( c \in R_+ \), this condition can be equivalently cast as

\[ \xi_{n_{\tau}} = \frac{\ln \frac{A_0}{A_1}}{n_{\tau}} \to 2(1 - \varepsilon_{n_{\tau}}) \]

(123)
APPENDIX H

PROOF OF THEOREM 1

From the definition of $P_n(S, K)$, the optimal stopping time $\tau$ and the optimal switching sequence $\hat{S}(\tau)$ are the minimizers of $P_n(\tau, \hat{S}(\tau))$ within the constraints on the sampling budget and the number of switchings. In the sequel, we first prove that the impacts of the stopping time and the switching sequence are embedded in the term $\tau$ and minimizing $P_n(\tau, \hat{S}(\tau))$ reduces to maximizing $\tau$. We provide separate proofs for the mean and variance cases.

**Mean:**
As shown in (28) and (91), we have

$$P_n(\tau, \hat{S}(\tau)) = \frac{\theta^\tau}{\theta^{\tau+1}}.$$  \hfill (124)

where $U_\tau$ and $\bar{U}_\tau$ denote the order statistics of the sets $\{Z_i^t : t \in L_i^t\}$ and $\{Z_i^t : t \in L_i^t\}$. Furthermore, as shown in (89) and (90), we also have

$$\forall i \in L_i^t : \frac{Z_i^t - \mu_{01} - \mu_{02}}{\sqrt{\tau}} \sim N(0, 1)$$

and

$$\forall i \in L_i^t : \frac{Z_i^t - \mu_{01} - \mu_{02}}{\sqrt{\tau}} \sim N(0, 1).$$

Hence, the distributions of $\frac{U_\tau - \mu_{01}}{\sqrt{\tau}}$ and $\frac{\bar{U}_\tau - \mu_{02}}{\sqrt{\tau}}$ are independent of $\tau$ and the effect of $\tau$ on $P_n(\tau, \hat{S}(\tau))$ is captured entirely by the term $\sqrt{\tau(\mu_{01} - \mu_{02})}$. Therefore, the error probability is minimized when $\tau$ is maximized.

**Variance:**
By recalling (116), (120), and (121), we have

$$P_n(\tau, \hat{S}(\tau)) = 1 - \left(\frac{\Theta^{\tau/2}}{1 + \Theta^{\tau/2}}\right)^{T_n}.$$  \hfill (125)

where from (116) and (120), we have

$$\Theta^{\tau/2} = \frac{A_2}{A_1} \tau^{\tau/2} \cdot n_\tau.$$  \hfill (126)

Therefore, minimizing $P_n(\tau, \hat{S}(\tau))$ is equivalent to maximizing $\Theta^{\tau/2}$, which happens when $\tau$ is maximized. Next, we obtain the choices of $\tau$ and $\hat{S}(\tau)$ that maximize $\tau$. Let us denote the optimal number of refinement actions by $K^* \leq K$. We argue that the optimal switching sequence that maximizes $\tau$ is of the form

$$\hat{S}(\tau) = \{1, \ldots, 1, 0, \ldots, 0\}. \quad (127)$$

To prove this, we first show that the aggregate number of observations taken corresponding to the switching sequence

$$\hat{S}(\tau) = \{\hat{S}_1, 0, 1, \hat{S}_2\}$$  \hfill (128)

is strictly smaller than that corresponding to the switching sequence

$$\hat{S}(\tau) = \{\hat{S}_1, 1, 0, \hat{S}_2\}$$  \hfill (129)

where $\hat{S}_1$ and $\hat{S}_2$ are switching subsequences. Note that the sequences $\hat{S}(\tau)$ and $\hat{S}(\tau)$ differ only in two switch values. By denoting the lengths of $\hat{S}_1$ and $\hat{S}_2$ by $|\hat{S}_1|$ and $|\hat{S}_2|$, respectively, the number of samples taken according to the switching sequence $\hat{S}(\tau)$ is

$$\sum_{t=1}^{\tau} |L_t| = \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_\hat{S}_1| + \sum_{t=1}^{\tau} |L_\hat{S}_2| + \sum_{t=1}^{\tau} |L_\hat{S}_3| + \sum_{t=1}^{\tau} |L_\hat{S}_4|$$  \hfill (130)

and the number of samples taken according to the switching sequence $\hat{S}(\tau)$ is

$$\sum_{t=1}^{\tau} |L_t| = \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_t| + \sum_{t=1}^{\tau} |L_t|$$  \hfill (131)

By comparing (128) and (129), all the summands of $\sum_{t=1}^{\tau} |L_t|$ are identical except the second terms, where it is strictly smaller in (128). By following the same line of argument, it is concluded that among all switching sequences that contain $K^*$ switches with value 1, the sequence that has its initial $K^*$ switches set to 1 takes the least number of samples. Therefore, among all switching sequences with $K^*$ refinement actions, the sequence of the form (125) leaves the most unused sampling resources, which in turn can be exploited to take further samples and delay (increase) the stopping time.

Next, we need to determine the optimal number of refinement actions $K^*$. By using (16), the following equation delineates the number of sequences of sequences observed at time $t$:

$$L_t \left| = \alpha \sum_{t=1}^{\tau} |\hat{S}(\tau)| (n - T_n) \right| + T_n. \hfill (132)$$

Hence, by invoking that the switching sequence is of form (125), we obtain

$$\sum_{t=1}^{\tau} |L_t| = (\tau - K^*) \left(\alpha K^* (n - T_n) + T_n\right)$$

where $\hat{S}_1$ and $\hat{S}_2$ are switching subsequences.
Furthermore, by taking into account the hard constraint on the sampling resources, i.e., \( \sum_{t=1}^{T} |L_t| \leq S \cdot n \), in the asymptote of large \( n \), we have
\[
\tau = \frac{nS - \sum_{t=1}^{T} K_1 \left( \alpha^{n - T_n} + T_n \right)}{\alpha^K n - T_n + T_n} + K^* \\
\geq \left[ S \cdot K^* + \frac{1 - \alpha}{1 - \alpha} \right].
\] (132)

Also, by noting that \( \alpha \in (0, 1) \), it can be readily shown that for \( x \in \{0, N\} \)
\[
S \cdot \alpha^x + \frac{1 - \alpha}{1 - \alpha} x
\]
is increasing in \( x \) when \( \alpha \leq 1 - \frac{1}{N} \) and decreasing in \( x \) when \( \alpha > 1 - \frac{1}{N} \). As a result
\[
K^* = \begin{cases} 
K_1, & \text{if } \alpha \leq 1 - \frac{1}{N} \\
0, & \text{if } \alpha > 1 - \frac{1}{N}.
\end{cases}
\]

By using (132), the optimal stopping time is
\[
\tau = \begin{cases} 
K + \left[ S \cdot \alpha^{-K} + \frac{1 - \alpha}{1 - \alpha} - S_0 \right], & \text{if } \alpha \leq 1 - \frac{1}{N} \\
S, & \text{if } \alpha > 1 - \frac{1}{N}.
\end{cases}
\]

**APPENDIX I**

**PROOF OF COROLLARY 1**

By following the same line of argument as in the proof of Theorem 1, it can be readily shown that the impact of the average sampling budget on the asymptotic error probability \( P_n(S, K) \) is captured by \( \tau \). Therefore, achieving identical asymptotic performance \( P_n(S, K) \approx P_n(S_0, 0) \) is equivalent to equating the term \( \tau \) under the adaptive and the nonadaptive procedures. By recalling the results of Theorem 1 and (48)–(50), this latter equivalence can be stated as
\[
S \cdot \alpha^{-K} + \frac{\alpha - \alpha^{-K}}{1 - \alpha} \leq S_0 < S \cdot \alpha^{-K} + \frac{\alpha - \alpha^{-K}}{1 - \alpha}.
\] (133)

which, in turn, provides
\[
S \cdot \alpha^{-K} + \frac{\alpha - \alpha^{-K}}{1 - \alpha} < S_0 < S \cdot \alpha^{-K} + \frac{\alpha - \alpha^{-K}}{1 - \alpha}.
\] (134)

After some simple manipulations, we obtain the desired result.

**REFERENCES**

[1] E. Andreou and E. Ghysels, “The impact of sampling frequency and volatility estimators on change-point tests,” *J. Financ. Econ.*, vol. 2, no. 2, pp. 290–318, 2004.

[2] M. Beibl and H. R. Lerche, “A new look at optimal stopping problems related to mathematical finance,” *Stat. Sinica*, vol. 7, pp. 63–108, 1997.

[3] A. N. Shiryaev, “Quickest detection problems in the technical analysis of financial data,” *Math. Financ.*, vol. 427–521, 2002.

[4] H. Li, C. Li, and H. Dai, “Quickest spectrum sensing in cognitive radio,” in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 2008, pp. 203–208.

[5] H. Jiang, L. Lai, R. Fan, and H. V. Poor, “Cognitive radio: How to maximally utilize spectrum opportunities in sequential sensing,” presented at the IEEE Global Commun. Conf., New Orleans, LA, USA, Dec. 2008.

[6] E. Andersson, D. Bock, and M. Frisén, “Detection of turning points in business cycles,” *J. Bus. Cycle Meas. Anal.*, vol. 1, no. 1, pp. 93–108, 2004.

[7] E. Andersson, D. Bock, and M. Frisén, “Some statistical aspects of methods for detection of turning points in business cycles,” *J. Appl. Stat.*, vol. 33, no. 3, pp. 257–278, 2006.

[8] E. Andreou, “The impact of sampling frequency and volatility estimators on change-point tests,” *J. Financ. Econ.*, vol. 2, no. 2, pp. 290–318, 2004.

[9] D. W. K. Andrews, I. Lee, and W. Ploberger, “Optimal change-point tests for normal linear regression,” *J. Econom.*, vol. 70, no. 1, pp. 9–38, 1996.

[10] I. Berkes, E. Gombay, L. Horváth, and P. Kokoszka, “Sequential change-point detection in Garch \( \{ p, q \} \) models,” *Econometric Theory*, vol. 20, no. 6, pp. 1140–1167, 2004.

[11] C. Han, P. Willett, and D. Abraham, “Some methods to evaluate the quality of communication links,” *IEEE Trans. Commun.*, vol. COM-27, no. 4, pp. 682–693, Apr. 1976.

[12] A. A. Cardenas, J. S. Baras, and V. Ramezani, “Distributed change detection for worms, dods and other network attacks,” in *Proc. Amer. Control Conf.*, Jun. 2004, pp. 1008–1013.

[13] R. K. C. Chang, “Defending against flooding-based distributed denial-of-service attacks: A tutorial,” *IEEE Commun. Mag.*, vol. 40, no. 10, pp. 42–51, Oct. 2002.

[14] H. Kim, B. L. Rozovskii, and A. G. Tartakovsky, “A nonparametric multichart CUSUM test for rapid detection of dos attacks in computer networks,” *Int. J. Comput. Inf. Sci.*, vol. 2, no. 3, pp. 149–158, 2004.

[15] P. Papantoni-Kazakos, “Algorithms for monitoring changes in the quality of communication links,” *IEEE Trans. Commun.*, vol. COM-27, no. 4, pp. 682–693, Apr. 1976.

[16] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, and H. Kim, “A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods,” *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3372–3382, Sep. 2006.

[17] M. Thottan and C. Ji, “Anomaly detection in IP networks,” *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2191–2204, Aug. 2003.

[18] E. Andreou, “The impact of sampling frequency and volatility estimators on change-point tests,” *J. Financ. Econ.*, vol. 2, no. 2, pp. 290–318, 2004.

[19] L. D. Broemling and H. Tsurumi, “Properties and use of the Shewhart method and its follow-up,” in *Ann. Math. Statist.*, vol. 57, no. 9, pp. 6222–6235, Sep. 2011.

[20] I. Berkes, E. Gombay, L. Horváth, and P. Kokoszka, “Sequential change-point detection in Garch \( \{ p, q \} \) models,” *Econometric Theory*, vol. 20, no. 6, pp. 1140–1167, 2004.
[34] D. Donoho and J. Jin, “Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data,” Ann. Statist., vol. 34, no. 6, pp. 2980–3018, 2006.

[35] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone, “Adapting to unknown sparsity by controlling the false discovery rate,” Ann. Statist., vol. 34, no. 2, pp. 584–653, 2006.

[36] R. A. Fisher and L. H. C. Tippett, “Limiting forms of the frequency distribution of the largest or smallest member of a sample,” Math. Proc. Cambridge Philosoph. Soc., vol. 24, no. 2, pp. 180–190, 1928.

[37] J. Galambos, The Asymptotic Theory of Extreme Order Statistics, 1st ed. New York, NY, USA: Wiley, 1978.

[38] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. A First Course in Order Statistics. New York, NY, USA: Wiley, 1992.

Ali Tajer (S’05–M’10) received the B.Sc. and M.Sc. degrees in Electrical Engineering from Sharif University of Technology, and the M.A. degree in Statistics and the Ph.D. degree in Electrical Engineering from Columbia University. During 2010–2012 he was a Postdoctoral Research Associate at Princeton University and an Adjunct Assistant Professor at Columbia University. Since August 2012 he has been an Assistant Professor of Electrical and Computer Engineering at Wayne State University. His research interests lie in the general areas of applied statistics, network information theory, and energy systems.

H. Vincent Poor (S’72–M’77–SM’82–F’87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. He has also held visiting appointments at several other institutions, most recently at Imperial College and Stanford. Dr. Poor’s research interests are in the areas of stochastic analysis, statistical signal processing and information theory, and their applications in wireless networks and related fields including social networks and smart grid. Among his publications in these areas are Smart Grid Communications and Networking (Cambridge University Press, 2012) and Principles of Cognitive Radio (Cambridge University Press, 2013).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, and an International Fellow of the Royal Academy of Engineering (U. K.). He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, in 2004–05 as the Editor-in-Chief of these Transactions, and in 2009 as General Co-chair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal for Achievement in Communications, the 2011 IEEE Eric E. Sumner Award, the 2011 IEEE Information Theory Paper Award, and honorary doctorates from Aalborg University, the Hong Kong University of Science and Technology, and the University of Edinburgh.