A New Algorithm for Distributed Nonparametric Sequential Detection

Shouvik Ganguly, K. R. Sahasranand, Vinod Sharma
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore, India

Email: mechperson@gmail.com, {sanandkr, vinod}@ece.iisc.ernet.in

Abstract

We consider nonparametric sequential hypothesis testing problem when the distribution under the null hypothesis is fully known but the alternate hypothesis corresponds to some other unknown distribution with some loose constraints. We propose a simple algorithm to address the problem. This is also generalized to the case when the distribution under the null hypothesis is not fully known. These problems are primarily motivated from wireless sensor networks and spectrum sensing in Cognitive Radios. A decentralized version utilizing spatial diversity is also proposed. Its performance is analyzed and asymptotic properties are proved. The simulated and analyzed performance of the algorithm are shown to be better than an earlier algorithm addressing the same problem with similar assumptions. We also modify the algorithm for optimising performance when information about the prior probabilities of occurrence of the two hypotheses are known.

Key words: Sequential Detection, Distributed detection, Asymptotic performance.

I. INTRODUCTION

Presently there is a scarcity of spectrum due to the proliferation of wireless services. However, it has been observed that much of the licensed spectrum remains underutilised for most of the time. Cognitive Radios (CRs) are proposed as a solution to this problem (1). These are designed to exploit the unutilised spectrum for their communication, without causing interference to the primary users. This is achieved through spectrum sensing by the CRs, to gain knowledge about spectrum usage by the primary users.

For CRs, spectrum sensing needs to be achieved at very low SNR in a wireless channel (1). Distributed detection, which can mitigate the time-varying, fading, shadowing and electromagnetic interference is very well-suited for this application. Thus distributed detection has been a highly-studied topic recently (2, 1). This has also found applications in sensor networks (3, 4).

Distributed detection problems can be looked upon in centralised or decentralised framework (5). In a centralised algorithm, the information collected by the local nodes are transmitted directly to the fusion centre which then uses it as a usual detection problem. In a decentralised algorithm, the local nodes transmit certain quantized values (or local decisions) to the fusion node. This has the advantage of requiring less power and bandwidth in transmission, but is suboptimal since the fusion centre has to take a decision based on less information.

Distributed detection problem can also be classified as fixed sample size or sequential (6, 7). In a fixed sample framework, the decision has to be made based on a fixed number of samples, and a likelihood ratio test turns out to be optimal for a simple binary hypothesis problem. In a sequential framework, samples are taken until some conditions are fulfilled, and once the process of taking samples has stopped, a decision is arrived at. It is known that in case of a single node, Sequential Probability Ratio Test (SPRT) outperforms other sequential or fixed sample size detectors in the simple binary hypothesis problem (8). But optimal solutions in the decentralised setup are not available (9). In the parametric case, with full knowledge about the distributions, (10) proposes an asymptotically optimal decentralised sequential test when the communication channel between the local nodes and FC is perfect.

For nonparametric sequential setup, (11) has provided separate algorithms for different problems like changes in mean, variance, etc.

(12) and (13) have studied the distributed decentralised detection problem in a sequential framework, with a noisy reporting MAC. The algorithm in (12) requires complete knowledge of the probability distributions involved, and is thus parametric in nature. The approach in (13) is non-parametric in the sense that it assumes very little knowledge of one of the distributions. In this paper, we have presented a simpler algorithm to address the problem studied in (13). Our algorithm has the added advantage of better performance in most cases, as borne out by simulations and analysis.

The paper is organized as follows. Section II presents the system model and the algorithm. Section III provides theoretical analysis of the algorithm for a single node case. Section IV provides an approximate analysis of the distributed algorithm. The asymptotics of the distributed algorithm are studied in Section V. Section VI compares our algorithm to KTSLRT in (13). Section VII provides a generalization of our algorithm along with an explanation in the CR setup.

II. SYSTEM MODEL

There are L nodes and one fusion centre. Node l makes observation X_{k,l} at time k. We assume \{X_{k,l}, k \geq 0\} are i.i.d. (independent identically distributed). We also assume that the observations received by different nodes are independent of each other. The distribution of the observations at each node is either P_0 or P_1. Each local node makes a decision based on the observations it receives and conveys the decision to the fusion node.
node. The fusion node makes the final decision based on the local decisions it receives. The decision to be made is,

\[ H_0: \text{if the probability distribution is } P_0 \text{ and } H_1: \text{if the probability distribution is } P_1. \]

We assume that \( P_0 \) is known, but that \( P_1 \) belongs to the family \( \{ P : D(P||P_0) \geq \lambda \} \), where \( D(P||P_0) \) is the entropy of \( P \), or differential entropy of the distribution \( P \). The distribution \( P_1 \) can be different for different nodes, allowing for different fading gains.

Our motivation for this setup is the Cognitive Radio (CR) system. A CR node has to detect if a channel is free (the primary node is not transmitting) or not. When the channel is free then the observations \( X_{k,l} \) is the receiver noise with distribution \( P_0 \). This will often be known (even Gaussian) and hence it is reasonable to assume that \( P_0 \) is known (see, however, the generalization in Section VII). But when the primary is transmitting, it could be using adaptive modulation and coding, unknown to the secondary node, and even the fading of the wireless channel from the primary transmitter to the local CR node may be time-varying and not known to the receiver. This leads to an unknown distribution \( P_1 \) under \( H_1 \). We will elaborate in Section VII how this scenario can lead to the above class of distributions qualified for \( H_1 \).

We have chosen sequential framework for detection at the local nodes as well as at the FC because it provides faster decisions on the average, for any probabilities of errors.

Our detection algorithm works as follows. Each local node \( l \), on receiving \( X_{k,l} \) at time \( k \), computes \( W_{k,l} \) as

\[ W_{k,l} = W_{k-1,l} - \log P_0(X_{k,l}) - H(P_0) - \frac{\lambda}{2} W_{0,l} = 0 \quad (1) \]

If \( W_{k,l} \geq -\log \alpha \), node \( l \) decides \( H_1 \); if \( W_{k,l} \leq -\log \beta \), it decides \( H_0 \); otherwise it waits for the next observation. If at time \( k \), node \( l \) has decided \( H_0 \), it transmits this decision to the FC by transmitting \( b_0 \); if it has decided \( H_1 \) it transmits \( b_1 \); otherwise it sends nothing (i.e., 0). In the algorithm, \( \alpha, \beta, b_0 \) and \( b_1 \) are constants appropriately chosen so as to provide good system performance.

Let \( Y_{k,l} \) be the transmission from node \( l \) to the FC at time \( k \). The FC receives from local nodes at time \( k \),

\[ Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k, \]

where \( Z_k \) is the FC receiver noise. We will assume \( \{ Z_k \} \) to be i.i.d. Because of \( Z_k \), the FC does not directly know the local decisions of the nodes. Thus it cannot use the majority rule, AND rule, etc. usually used in literature. An advantage of allowing all nodes to transmit at the same time is that it reduces transmission delays from the nodes to the FC.

The nodes keep transmitting to the FC till the FC makes its decision. Once the FC makes the decision, it will broadcast a message to all the local nodes to stop transmission. The local nodes make their decisions at random times. Thus \( \{ Y_k \} \) received by the FC are not i.i.d. However, inspired by sequential detection algorithms (e.g. SPRT as in [8]) it uses the following algorithm to make decisions. At time \( k \), it computes

\[ F_k = F_{k-1} + \log \frac{g_{\mu_1}(Y_k)}{g_{\mu_0}(Y_k)} F_0 = 0 \]

After that, it decides \( H_0 \) if \( F_k \leq -\log \beta \); \( H_1 \) if \( F_k \geq -\log \alpha \) and waits for the next observation otherwise. Here the distributions \( g_{\mu_0} \) and \( g_{\mu_1} \) are appropriately decided. We assume that the distribution of noise \( \{ Z_k \} \) is known. Then \( g_{\mu_0} \) is the distribution of \( \mu_0 + Z_k \), and \( g_{\mu_1} \) is the distribution of \( \mu_1 + Z_k \), where \( \mu_0 \) and \( \mu_1 \) are constants. By choosing \( \mu_0 \) and \( \mu_1 \) appropriately, we can ensure that FC makes a decision which is (say) close to a majority decision of local nodes.

The overall algorithm is summarized as

- i Node \( l \) receives \( X_{k,l} \) at time \( k \geq 1 \) and computes \( W_{k,l} \)
- ii Node \( l \) transmits \( Y_{k,l} = b_1 I\{W_{k,l} \geq -\log \alpha_l\} + b_0 I\{W_{k,l} \leq -\log \beta_l\} \)
- iii Fusion node receives at time \( k \),
  \[ Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k \]
- iv Fusion node computes,
  \[ F_k = F_{k-1} + \log \frac{g_{\mu_1}(Y_k)}{g_{\mu_0}(Y_k)} F_0 = 0 \]
- v Fusion node decides \( H_0 \) if \( F_k \leq -\log \beta \), \( H_1 \) if \( F_k \geq -\log \alpha \); otherwise it waits for the next observation.

In the rest of the paper we analyze the performance of this algorithm. First we analyze the performance for a single (local) node in Section III and then for the decentralised algorithm in Section IV. Proofs of the lemmas and theorems provided are similar to those in [12] and [13] and are skipped in this paper for lack of space. We will show that the present algorithm provides a better performance than [13]. Also, unlike in [13], it is simpler to implement because it does not require a universal source coder and it also does not require quantization of observations.

In the following, \( E_s[\cdot] \) and \( P_s(\cdot) \) denote the expectation and probability, respectively, under \( H_s \), \( i = 0, 1 \).

### III. PERFORMANCE ANALYSIS FOR A SINGLE NODE

In this section we provide performance of the test for a single node. We will omit the node index \( i \) in this section. Let

\[ N_1 \triangleq \inf \{ n : W_n > -\log \alpha \} \]

\[ N_0 \triangleq \inf \{ n : W_n < \log \beta \} \]

\[ N \triangleq \min(N_0, N_1) \]

**Lemma 3.1.** \( P(N < \infty) = 1 \) under \( H_0 \) and \( H_1 \).

We will use the notation that \( P_{FA} \triangleq P_0(\text{decide } H_1) \) and \( P_{MD} \triangleq P_1(\text{decide } H_0) \). Also, \( f(x) = O(g(x)) \) denotes that \( \lim_{x \to \infty} \frac{f(x)}{g(x)} < \infty \).

**Theorem 3.2.**

a) \( P_{FA} = O(\alpha^s) \) where \( s \) is a solution of \( E_0[\frac{1}{e^{-\epsilon(H(P_0) + \log P_0(X_n) + \frac{1}{2} - \epsilon)}}] = 1 \) where \( 0 < \epsilon < \frac{1}{2} \) and \( s > 0 \).

b) \( P_{MD} = O(\beta^s) \) where \( s^* \) is a solution of

\[ E_1[\frac{1}{e^{-s^*(H(P_0) + \log P_0(X_n) - \frac{1}{2} + \epsilon)}}] = 1 \]

\[ 0 < \epsilon < D(P_1||P_0) + H(P_1) - H(P_0) - \frac{1}{2} \] and \( s^* > 0 \).

Let,

\[ N_0^*(\epsilon) \triangleq \sup \{ n \geq 1 : -\log P_0(x_n^*) - nH(P_0) > n\epsilon \}, \]
\(N^*_\epsilon(e) \triangleq \sup\{n \geq 1 : |-\log P_0(x^*_\epsilon) - nH(P_1) - nD(P_i||P_0)| > \epsilon n\}\)

Theorem 3.3: a) Under \(H_0\), 
\[
\lim_{\alpha, \beta \to 0} \frac{N}{|\log \beta|^\alpha} = \frac{2}{\lambda} \text{ a.s.}
\]
If in addition, \(E_0(N^*_{\epsilon}(e)^p) < \infty\) and \(E_0(|\log P_0(X)|^{p+1}) < \infty\) for all \(\epsilon > 0\) and for some \(p \geq 1\), then
\[
\lim_{\alpha, \beta \to 0} \frac{E_0[N^\epsilon]}{|\log \beta|^\alpha} = \left(\frac{2}{\lambda}\right)^p
\]
for all \(0 < \epsilon < p\).

b) Under \(H_1\),
\[
\lim_{\alpha, \beta \to 0} \frac{E_1[N^\epsilon]}{|\log \beta|^\alpha} = \left(D(P_1||P_0) + H(P_1) - H(P_0) - \frac{\lambda}{2}\right)^{-\alpha}
\]
for all \(0 < \epsilon < p\).

From Lemma 3.1 and Theorem 3.2 we see that the asymptotic behaviour of our algorithm is comparable to that in [13] and also to dualSPRT ([12]). However, we will see via simulations that it substantially outperforms KTSRLRT in [13].

IV. APPROXIMATE PERFORMANCE OF DECENTRALISED ALGORITHM

In the following, we take, for convenience, \(\alpha = \beta = 1 \forall l\), \(a = b, b_1 = -b_0 = b\), and \(\mu_1 = -\mu_0 = \mu = 1.b\), for some \(1 \leq I \leq L\).

Roughly speaking, this ensures that the FC makes decisions \(H_1\) when \(I\) more nodes decide \(H_1\) compared to the nodes deciding \(H_0\). Similarly for \(H_0\). In the following, \(N^i_l\) corresponds to \(N^i_l\) at node \(l\) and \(N_i\) corresponds to \(N\). Similarly, \(N^1, N^0\) and \(N\) represent the corresponding terms for the FC.

**Lemma 4.1.** For \(i = 0, 1\),
\[
P_i(N_l = N^i_l) \to 1 \text{ as } \alpha_i, \beta_i \to 0
\]
\[
P_i(N = N^i) \to 1 \text{ as } \alpha, \beta \to 0
\]

Note: In general, when \(\alpha_i \neq \beta_i\), the results of Lemma 4.1 under \(H_0\) demand that \(\beta\) and/or \(\beta_i\) \(\to 0\), and the results under \(H_1\) demand that \(\alpha\) and/or \(\alpha_i\) \(\to 0\). Analogous comments will hold true for the subsequent results as well.

**Lemma 4.2.** Under \(H_i\),

a) \(|N_l - N^i_l| \to 0\) a.s. as \(\alpha_i, \beta_i \to 0\) and
\[
\lim_{\alpha_i \to 0} \frac{N_l}{|\log \alpha_i|} = \lim_{\alpha \to 0} \frac{N^i_l}{|\log \alpha_i|} = \frac{1}{[\delta^i_l]} \alpha_i \text{ a.s. and in } L^i.
\]

b) \(|N - N^i| \to 0\) a.s. and
\[
\lim_{\alpha, \beta \to 0} \frac{N}{|\log \alpha|} = \lim_{\alpha \to 0} \frac{N^i}{|\log \alpha|} \text{ a.s. and in } L^i.
\]
as \(\alpha_i, \beta_i \to 0\) and \(\alpha, \beta \to 0\).

Lemmas 4.1 and 4.2 show that the local nodes make the right decisions as the thresholds \(|\log \alpha_i|\) and \(|\log \beta_i|\) tend to infinity. Then the FC also makes the right decisions when its own thresholds increase. We need to set the thresholds such that the probabilities of errors are small.

We will use the following notation:
\[
\delta^i_l, FC \triangleq \text{ mean drift of the fusion centre process } \{F_k\} \text{ under } H_i, \quad \text{where } j \text{ local nodes are transmitting}.
\]
\[
t^i_l \triangleq \text{ time at which the mean drift of } \{F_k\} \text{ changes from } \delta^i_l, FC \text{ to } \delta^i_l, FC.
\]
\[
\tilde{F}_j \triangleq E[F_{t^i_l}]
\]
Now, it is seen that under \(H_i\), \(\tilde{F}_j = \tilde{F}_{j-1} + \delta^i_l, FC(E(t_j) - E(t_{j-1})), \tilde{F}_0 = 0\).

**Lemma 4.3:** \(P_i(\text{decision of local node } l \text{ at time } t_k \text{ is } H_i \text{ and } t_k \text{ is the } k\text{th statistics of } \{N^1_l, \ldots, N^i_l\}) \to 1 \text{ as } \alpha_i \to 0 \forall l\).

**Proof:** See Theorem 5.1, Chapter 3 in [14].

Let \(E_{DD} \triangleq E_i[N], i = 0, 1\). In the following we provide an approximation for this for both \(i = 0, 1\).

When \(\alpha_i\) and \(\alpha\) are small, probabilities of error are small, as proved in the above lemmas. Hence in such a scenario, for approximation, we assume that local nodes are making correct decisions.

Let
\[
l^i_t \triangleq \min\{j : \delta^i_l, FC > 0 \text{ and } \frac{\pm |\log \alpha| - t^i_l}{\delta^i_l, FC} < E(t_{j+1}) - E(t_j)\},
\]
where the ‘+’ sign occurs under \(H_1\) and − sign under \(H_0\).

**Proof:** See Theorem 5.1, Chapter 3 in [14].

The first term in approximation (2) corresponds to the mean time till the mean drift of \{\(F_k\)\} becomes positive (for \(H_1\)) or negative (for \(H_0\)), and the second term corresponds to the mean time from then on till it crosses the threshold. Using the Gaussian approximation of Lemma 4.4, the \(t_k\)’s are the order statistics of i.i.d. Gaussian random variables and hence, the \(\tilde{F}_k\)’s can be computed. See, for example, [15].

In the following, we compute approximate expressions for \(P_{FA} \triangleq P[\text{decision is } H_1|H_0] \) and \(P_{MD} \triangleq P[\text{decision is } H_0|H_1]\).

Under the same setup of small \(\alpha_i\) and \(\alpha\), for \(P_{FA}\) analysis, we assume that all local nodes are making correct decisions. Then for false alarm, the dominant event is \(\{N^1 < t_1\}\). Also, for reasonable performance, \(P_0(N^0 < t_1)\) should be small. Then, the probability of false alarm, \(P_{FA}\), can be approximated as
\[
P_{FA} = P_0(N^1 < N^0) \geq P_0(N^1 < t_1, N^0 > t_1) \approx P_0(N^1 < t_1).
\]

(3)
Also,

$$P_0(N^1 < N^0) \leq P_0(N^1 < \infty)$$

$$= P_0(N^1 < t_1) + P_0(t_1 \leq N^1 < t_2) + \cdots$$ \hspace{1cm} (4)

The first term in the RHS of (4) should be the dominant term since after $t_1$, the drift of $F_k$ will have the desired sign (will at least be in the favourable direction) with a high probability, if the local nodes make correct decisions.

Equations (3) and (4) suggest that $P_0(N^1 < t_1)$ should serve as a good approximation for $P_{FA}$. Similar arguments show that $P_1(N^0 < t_1)$ should serve as a good approximation for $P_{MD}$.

In the following, we provide approximations for these.

Let $\xi_k \triangleq \log \frac{g_{\mu_k}(Y_k)}{g_{\mu_0}(Y_k)}$ before $t_1$ have mean 0 and probability distribution symmetric about 0. Then, from the Markov property of the random walk $\{F_k\}$, before $t_1$,

$$P_0(N^1 < t_1) \approx \sum_{k=1}^{\infty} P_0(\{F_k \geq -\log \alpha\} \cap \{F_n < -\log \alpha\} \cap \{F_0 > k\})$$

$$= \sum_{k=1}^{\infty} P_0(\{F_k \geq -\log \alpha\} \cap \{F_n < -\log \alpha\}). P_0(\{F_0 > k\})$$

$$= \sum_{k=1}^{\infty} P_0(F_k \geq -\log \alpha). P_0(\sup_{1 \leq n \leq k-1} F_n < -\log \alpha). |1 - \Phi_{t_1}(k)|$$

$$= \sum_{k=1}^{\infty} \left[ \int_{u=0}^{\infty} P_0(\xi_k > u) f_{F_{k-1}}(-\log \alpha - u) du \right]. P_0(\sup_{1 \leq n \leq k-1} F_n < -\log \alpha). |1 - \Phi_{t_1}(k)|$$

$$= \sum_{k=1}^{\infty} \left[ \int_{u=0}^{\infty} P_0(\xi_k > u) f_{F_{k-1}}(-\log \alpha - u) du \right]. |1 - \Phi_{t_1}(k)|$$

Similarly, $P_{MD}$ can be approximated as

$$P_{MD} \approx \sum_{k=1}^{\infty} \left[ \int_{u=0}^{\infty} P_1(\xi_k < u) f_{F_{k-1}}(\log \beta + u) du \right]. |1 - \Phi_{t_1}(k)|$$

We will show in Section VI that these approximate results compare well with simulations.

V. ASYMPTOTIC RESULTS

In this section, we assume

1. $E_i[N_{1,i}^1(\epsilon)] < \infty$, where these quantities are as defined in Lemma 3.3, but for local node $l$.

2. $E_i[N_{1,i}^1(\epsilon)] < \infty$, where $N_{1,i}^1(\epsilon)$ is the drift of the test statistic at the local nodes.

3. $E_i[N_{1,i}^1(\epsilon)] < \infty$, where $N_{1,i}^1(\epsilon)$ is the variance of $V_{k,l}$.

4. $\rho^2_{k,l} < \infty$, where $\rho^2_{k,l}$ is the variance of $V_{k,l}$.

In this section, we also use the notation

1. $\theta_i = E_i(\xi_k^*)$

2. $A^i = \{\omega \in \Omega : \text{ all local nodes transmit correct decisions (}b_i\text{) under } H_1\}$

3. $\Delta(A^i) = \text{Drift of fusion centre LLR under } A^i$, i.e. $E_i[\xi_k|A^i]$

4. $D_{tot}^i = \frac{\lambda^i}{L}$

5. $D_{tot}^i = \sum_{i=1}^{N} [D(f_{1,i}||f_0,i) + H(f_1,i) - H(f_0,i) - \frac{\lambda}{2}]$

6. $r_1 = \frac{1}{2}$

7. $\rho_i = \frac{D(f_{1,i}||f_0,i) + H(f_1,i) - H(f_0,i) - \frac{\lambda}{2}}{D_{tot}^i}$

8. $\Lambda_i(\alpha) = \sup[\alpha |\log g_i(\lambda)|$

9. $g_i = \text{m.g.f. of } |\xi_k^*|$

10. $\alpha_i^* = \text{ess sup } |\xi_k^*|$

Furthermore, local node thresholds are $-r_1|\log c|$ and $\rho_i |\log c|$, where $c$ is a constant, and the fusion centre thresholds are $-|\log c|$ and $|\log c|$.

Theorem 5.1. Under $H_1$,

$$\lim_{c \to 0} \frac{N}{\log c} \leq \frac{1}{D_{tot}^i} + \frac{C_i}{\Delta(A^i)}$$

a.s. and in $L^1$,

where $C_0 = -(1 + \frac{\theta_0}{D_{tot}^i})$ and $C_1 = 1 + \frac{\theta_1}{D_{tot}^i}$.

We introduce the following function,

$$s_i(\eta) \triangleq \begin{cases} \frac{\eta}{\alpha_i^*} & \text{if } \eta \geq \Lambda_i(\alpha_i^*) \\ \frac{\eta}{\alpha_i^*} & \text{if } \eta \in (0, \Lambda_i(\alpha_i^*)) \\ \end{cases}$$

where $g_i(\lambda) = \text{m.g.f. of } |\xi_k^*|$, and $\Lambda_i(\alpha) = \sup[\alpha |\log g_i(\lambda)|$

$$R_i \triangleq \min_{1 \leq L \leq L_1} \{ -t(-\log f_0,l(X_{k,l}) - H(P_0) - \frac{\lambda}{2}) \}.$$}

Theorem 5.2. The following hold:

a) $\lim_{c \to 0} \frac{P_{FA}}{c} = 0$ if for some $0 < \eta < R_0$, $s_0(\eta) > 1$.

b) $\lim_{c \to 0} \frac{P_{MD}}{c} = 0$ if for some $0 < \eta < R_1$, $s_1(\eta) > 1$.

From Theorems 5.1 and 5.2 we see that the asymptotic performance of our algorithm is comparable to SPRT and KTSLRT in [12] and [13].

VI. SIMULATIONS

In this section, we compare the simulated and theoretical performances of the new algorithm with KT-SLRT ([13]). For simulations, we have taken $b_1 = -b_0 = 1$, $L = 5$, $\mu_1 = -\mu_0 = 2$. Also, the FC noise has been taken as zero mean Gaussian with variance $\sigma^2$. Hence in this case, $\Delta(A^1) = -\Delta(A^0) = \frac{20}{\sigma^2} = \theta_0 = \theta_1$. In the following simulations, $E_{DD} \triangleq 0.5[E_0(N) + E_1(N)]$ and $P_e \triangleq 0.5(P_{FA} + P_{MD})$

The observations $X_{k,l}$ are considered with the following distributions:
Pareto Distribution $P$, $P_0 \sim \mathcal{P}(10, 2)$ and $P_1 \sim \mathcal{P}(3, 2)$. Lognormal Distribution $\ln N$, $P_0 \sim \ln N(0, 3)$ and $P_1 \sim \ln N(3, 3)$. Gaussian Distribution $\mathcal{N}$ with $P_0 \sim \mathcal{N}(0, 1)$ and under $P_1 \sim \mathcal{N}(1, 1)$. The channel gains from the primary to the secondary nodes are 1 except in the Gaussian case, where these are taken as 0 dB, -1.5 dB, -2.5 dB, -4 dB and -6 dB.

We plot the results in Figs 1-7. We see that the new algorithm markedly outperforms KT-SLRT. This may be due to the presence of compression in KT-SLRT, due to which redundancy is introduced, leading to inaccuracies in the estimate. Also, the approximations provided in Section IV are much closer to the simulated values than the asymptotics.

We plot the results in Figs 1-7. We see that the new algorithm markedly outperforms KT-SLRT. This may be due to the presence of compression in KT-SLRT, due to which redundancy is introduced, leading to inaccuracies in the estimate. Also, the approximations provided in Section IV are much closer to the simulated values than the asymptotics.

VII. FURTHER GENERALIZATIONS

Let us now consider a generalization of the problem, in which $P_0$ is not exactly known. Specifically, the hypothesis testing problem we now consider is:

$$H_0 : P \in \{P' \in \mathcal{P} : D(P' || P_0) \leq \gamma \lambda \}, \text{ for some } 0 \leq \gamma < 1. \quad (5)$$

$$H_1 : P \in \{P' \in \mathcal{P} : D(P' || P_0) \geq \lambda \text{ and } H(P') > H(P_0),$$

for all $P_0' \in H_0$.

The detection algorithm remains the same except that now we write the test statistic at the local node $l$ as

$$\tilde{W}_{k,l} = \tilde{W}_{k-1,l} - \log \hat{P}_0(X_{k,l}) - H(\hat{P}_0) - \nu \lambda.$$  

For good performance we should pick $\hat{P}_0$ from the class in (5) and choose $\nu$ carefully. We elaborate on this in the following.

Let us try to justify this problem statement from a practical CR standpoint. In a CR setup, $H_0$ actually indicates the presence of only noise, while under $H_1$, the observatios are signal + noise. Due to electromagnetic interference, the receiver noise can be changing with time ($\lambda$). Thus we assume that the noise power $P_N$ is bounded as $\sigma_{N,L}^2 \leq P_N \leq \sigma_{N,H}^2$. Similarly, let the signal power be bounded as $\sigma_{S,L}^2 \leq P_S \leq \sigma_{S,H}^2$. Now we formulate these constraints in the form (5) where we should select appropriate $P_0$, $\lambda$ and $\gamma$. We will compute these assuming we are limiting
Fig. 6: Performance for Gaussian Distribution with different received SNRs
Top: Detection Delay; Bottom: Error Rate

Fig. 5: Performance of KTSLRT for Gaussian Distribution with different received SNRs
Top: Detection Delay; Bottom: Error Rate

Fig. 7: Performance Comparison with KT-SLRT for Gaussian Distribution with different received SNRs

Let \( f(\sigma) \triangleq \ln \frac{\sigma_0}{\sigma} + \frac{1}{2} \frac{\sigma^2}{\sigma_0} - 1 \). We choose \( \sigma_0 \) such that
\[ f(\sigma_{N,L}) = f(\sigma_{N,H}) \]
This can be achieved for some \( \sigma_0 \in (\sigma_{N,L}, \sigma_{N,H}) \), since \( f \) is convex with a minimum at \( \sigma_0 \). This choice ensures that \( P_0 \) is at some sort of a “centre” of the class of distributions under consideration in \( H_0 \). We now choose \( \gamma \lambda \triangleq f(\sigma_{N,L}) = f(\sigma_{N,H}) \).

For the class of distributions considered under \( H_1, \sigma_{N,L}^2 + \sigma_{N,H}^2 \leq E[X^2] \leq \sigma_{N,L}^2 + \sigma_{N,H}^2 \). We take \( \lambda \triangleq \inf \left( \sigma^2|\sigma_{N,L}^2 + \sigma_{N,H}^2 \right) f(\sigma) = f\left(\sqrt{\sigma_{N,L}^2 + \sigma_{N,H}^2}\right) \).

Next we compute \( \hat{P}_0 \). If the \( X_{k,1} \) has distribution \( P_i \) for \( i = 0, 1 \), then the drift at the local nodes is \( D(P_0||\hat{P}_0) + H(P_0) - H(P_0) - v\lambda \) under \( H_0 \), and \( D(P_i||\hat{P}_0) + H(P_i) - H(P_0) - v\lambda \) under \( H_1 \). This drift is an important parameter in determining the algorithm performance and will decide \( \hat{P}_0 \).

Let \( W_i \) be the cost of rejecting \( H_i \) wrongly, and \( c \) be the cost of taking each observation. Then, Bayes risk for the test is given (17) by
\[ \mathcal{R}_c(\delta) = \sum_{i=0}^{1} \pi^i [W_i \mathbb{P}_i(\text{reject } H_i) + cE_i(N)], \]
where \( \pi^i \) is the prior probability of \( H_i \).

Taking the same thresholds as in Section V and using Theorems 5.1 and 5.2,
\[
\lim_{c \to 0} \frac{\mathcal{R}_c(\delta)}{c \log c} \leq \frac{\pi^0}{L[D(P_0||\hat{P}_0) + H(P_0) + v\lambda]} \left( 1 - \frac{\theta_1}{\Delta(A^1)} \right) + \frac{\pi^1}{L[D(P_1||\hat{P}_0) + H(P_1) - H(P_0) - v\lambda]} \left( 1 + \frac{\theta_1}{\Delta(A^1)} \right) - \frac{\pi_0}{\Delta(A^0)} + \frac{\pi^1}{\Delta(A^1)}.
\]
Following a minimax approach, we first maximize the above expression with respect to \( \hat{P}_0 \) and \( P_i \), and then minimize the resulting maximal risk w.r.t. \( \hat{P}_0 \) and \( v \). As noted before, we achieve this optimization limiting ourselves to only Gaussian family.

The second term in (6) is maximized when \( D(P_1||\hat{P}_0) + h(P_i) \) is minimized. Let us denote the variance of \( \hat{P}_0 \) by \( \Gamma \). Now, the variances of all eligible \( P_i \)'s are greater than \( \Gamma^2 \). Hence, \( D(P_1||\hat{P}_0) + h(P_i) \) is minimized when \( P_i \) has the least possible variance, i.e. \( \sigma_{N,L}^2 + \sigma_{N,H}^2 \). Using \( \mathcal{N}(0, \sigma_{N,L}^2 + \sigma_{N,H}^2) \) in place of \( P_i \), the second term in (6) becomes (after simplification),
\[
\frac{(\pi^1/L)(1 + \theta_1/\Delta(A^1))}{\frac{1}{2}\frac{\sigma_{N,L}^2 + \sigma_{N,H}^2}{\sigma_0^2} - 1 - v\lambda}.
\]
Similarly, to maximize the first term in (6), we have to minimize $D(P_0^0\|P_0) + H(P_0^0)$ w.r.t. $P_0$. After this, the first term becomes
\[
u \lambda - \frac{1}{2} (\frac{\sigma^2_{N,H}}{\nu_{2\nu}^2} - 1)
\]
Taking $x \equiv \frac{1}{\Gamma^2}$, $y \equiv \nu \lambda$, $a = \sigma^2_{N,H}$, $b = \sigma^2_{N,L} + \sigma^2_{S,L}$,
\[A = (\pi^0/L)(1 - \theta_0 \Delta(A^0))\text{ and } B = (\pi^1/L)(1 + \theta_1 \Delta(A^1)),
\]the non-constant part of the above expression can be written as a function of $x$ and $y$ in the form,
\[g(x, y) = \frac{A}{y + x - 1} + \frac{B}{2x - y - 1}\]
Minimizing this w.r.t. $y$ yields,
\[y_{opt} = \frac{1}{2} \sqrt{A} (bx - 1) + \sqrt{B} (ax - 1)
\]
Together with this, we can choose $x \in \left(\frac{1}{\sigma^2_{N,H}}, \frac{1}{\sigma^2_{N,L}}\right)$.

In the following, we demonstrate the advantage of optimizing the above parameters on the examples considered in Section VI. The bounds on the noise and signal power were chosen in each case such that the distributions specified in Section VI satisfy those constraints. Also, the thresholds were chosen the same as before.

For the following simulations, we have taken $\Gamma^2 = \frac{\sigma^2_{N,L} + \sigma^2_{S,L}}{2}$ and determined $y_{opt}$ in accordance with (8).

For Gaussian distribution, $P_0 \equiv N(0, 1)$, $P_1 \equiv N(0, 5)$

For Lognormal distribution, $P_0 \equiv \log N(0, 3)$, $P_1 \equiv \log N(3, 3)$

For Pareto distribution, $P_0 \equiv P(10, 2)$, $P_1 \equiv P(3, 2)$

We compare the performances in Figs. 8-10. We see that the optimized version performs noticeably better, even for distributions other than Gaussian.

\[\text{Fig. 8: Optimization for Pareto Distribution}\]

\[\text{Fig. 9: Optimization for Lognormal Distribution}\]

\[\text{Fig. 10: Optimization for Gaussian Distribution}\]

REFERENCES

[1] I. F. Akyildiz, B. Lo, and R. Balakrishnan, “Cooperative spectrum sensing in cognitive radio networks: A survey,” Physical Communication, vol. 4, no. 1, pp. 40–62, 2011.
[2] J. Umakrishnan, D. Huang, S. Meyn, A. Surana, and V. Veeravalli, “Universal and composite hypothesis testing via mismatched divergence,” IEEE Transactions on Information Theory, vol. 57, no. 3, pp. 1587–1603, 2011.
[3] J. Chamberland and V. Veeravalli, “Wireless sensors in distributed detection applications,” vol. 24, pp. 16–25, 2007.
[4] P. Varshney and R. Viswanathan, “Distributed detection with multiple sensors,” in IEEE, Proceedings of, vol. 85, pp. 54–63, IEEE, 1997.
[5] K. Liu and A. M. Sayeed, “Optimal distributed detection strategies for wireless sensor networks,” vol. 10, 2017.
[6] E. Geraniotis, “Robust distributed discrete-time block and sequential detection in uncertain environments,” Naval Research Journal.
[7] S. Marano and V. Matta, “Elements of sequential detection with applications to sensor networks.”
[8] D. Siegmund, Sequential Analysis: Tests and Confidence Intervals. Springer, 1985.
[9] V. Veeravalli, T. Basar, and H. V. Poor, “Decentralized sequential detection with a fusion centre performing the sequential test,” IEEE Transactions on Information Theory, vol. 59, no. 2, pp. 433–442, 1993.
[10] G. Fellouris and G. V. Moustakides, “Decentralized sequential hypothesis testing using asynchronous communication,” IEEE Transactions on Information Theory, vol. 57, no. 1, pp. 534–548, 2011.
[11] N. Mukhopadhyay and B. De Silva, Sequential Methods and Their Applications: CRC: Chapman and Hall.
[12] J. K. Sreedharan and V. Sharma, “Spectrum sensing using distributed sequential detection via noisy reporting mac,” No. arXiv:1211.5562v2 [cs.IR], 2013.
[13] J. K. Sreedharan and V. Sharma, “Nonparametric decentralized sequential detection via universal source coding,” No. arXiv:1308.6481v1 [cs.IT], 2013.
[14] A. Gut, Stopped random walks: limit theorems and applications, vol. 5 of Applied probability. New York: Springer-Verlag, 1988.
[15] H. Barakat and Y. Abdelkader, “Computing the moments of order statistics from non-identical random variables,” Statistical Methods and Applications, vol. 13, pp. 15–26, 2004.
[16] P. Billingsley, Probability and Measure. John Wiley and Sons, 1986.
[17] E. Lehmann and G. Casella, Theory of Point Estimation. Springer Texts in Statistics, New York: Springer, 2003.
[18] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley-Interscience, 2 ed.
[19] V. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 2 ed., 1994.
[20] T. Banerjee, V. Sharma, V. Kavitha, and A. Jayaparakash, “Generalized analysis of a distributed energy efficient algorithm for change detection,” IEEE Transactions on Wireless Communication, vol. 10, pp. 91–101, 2011.