Nonparametric Distributed Sequential Detection via Universal Source Coding

Jithin K. Sreedharan and Vinod Sharma
Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore 560012, India
Email: {jithin,vinod}@ece.iisc.ernet.in

***Abstract*—We consider nonparametric or universal sequential hypothesis testing when the distribution under the null hypothesis is fully known but the alternate hypothesis corresponds to some other unknown distribution. These algorithms are primarily motivated from spectrum sensing in Cognitive Radios and intruder detection in wireless sensor networks. We use easily implementable universal lossless source codes to propose simple algorithms for such a setup. The algorithms are first proposed for discrete alphabet. Their performance and asymptotic properties are studied theoretically. Later these are extended to continuous alphabets. Their performance with two well known universal source codes, Lempel-Ziv code and KT-estimator with Arithmetic Encoder are compared. These algorithms are also compared with the tests using various other nonparametric estimators. Finally a decentralized version utilizing spatial diversity is also proposed and analysed.**

***Index Terms*—Sequential Hypothesis Testing, Universal Testing, Universal Source Codes, Distributed Detection.**

I. Introduction

Distributed detection ([21]) has been quite popular recently due to its relevance to distributed radar, sensor networks, distributed databases and cooperative spectrum sensing in Cognitive radios. It can be either decentralized or centralized. Unlike the centralized framework, where the information received by the sensors are transmitted directly to the fusion center (FC) to decide upon the hypothesis, in decentralized detection each sensor sends a summarized or quantized information. Thus decentralized detection although suboptimal, is more bandwidth and energy efficient.

Two of the important formulations of distributed detection problem are based on the number of samples required for making a decision: fixed sample size and sequential detection ([23]). In the sequential case, the observations are sampled sequentially at the local nodes until a stopping rule is satisfied. The decision and stopping rules are designed with the aim of reducing the number of samples for decision making with reliability constraints. More precisely, sequential detectors can detect change in the underlying hypothesis or test the hypothesis ([16]). In this paper we focus on decentralized sequential hypothesis testing. It is well known that in case of a single node, Sequential Probability Ratio Test (SPRT) outperforms other sequential or fixed sample size detectors ([16]). But optimal solutions for the decentralized setup are not available ([22]). In the parametric case, when there is a full knowledge about the distributions, [4] and [18] propose asymptotically optimal decentralized sequential tests when the communication channel between local nodes and the FC is perfect and noisy MAC respectively.

The sequential methods in case of uncertainties are surveyed in [13] for a parametric family of distributions. For nonparametric sequential methods, [14] provides separate algorithms for different setups like changes in mean, changes in variance etc. In this paper we propose a unified simple universal sequential hypothesis testing algorithm where the unknown alternate distribution can be anything which satisfies a constraint on the Kullback-Leibler divergence ([2]) with the null distribution.

An optimal fixed sample size universal test for finite alphabets is derived in [7]. Statistical inference with universal source codes, started in [15] when classification of finite alphabet sources is studied in the fixed sample size setup. [8] considers the universal hypothesis testing problem in the sequential framework using universal source coding. It derives asymptotically optimal one sided sequential hypothesis tests and sequential change detection algorithms for countable alphabet. In one sided tests one assumes null hypothesis as the default hypothesis and has to wait a long time to confirm whether it is the true hypothesis (it is the true hypothesis only when the test never stops). In many practical applications where it is important to make a quick decision.

In this paper, we consider universal source coding framework for binary hypothesis sequential testing with continuous alphabets. Section II describes the model. Section III provides our algorithm for a single node with the finite alphabet. We prove almost sure finiteness of the stopping time. Asymptotic properties of probability of error and moment convergence of expected stopping times are also studied. Section IV extends the test to continuous alphabet. Algorithms based on two easily implementable universal codes, Lempel-Ziv tree-structured (LZ78) ([26]) codes and Krichevsky-Tofimov (KT) estimator with Arithmetic Encoder ([3]) are studied in Section V. Performance of these tests are compared in Section VI. In Section VII we extend our algorithm to the decentralized scenario. In our distributed algorithm each local node sends a local decision to the FC at asynchronous times leading to considerable saving in communication cost. Section VIII concludes the chapter.

II. Model for single node

We consider the following hypothesis testing problem: Given i.i.d. observations $X_1, X_2, \ldots,$ we want to know whether these observations came from the distribution P_0 (hypothesis H_0) or from another distribution P_1 (hypothesis H_1). We will assume that P_0 is known but P_1 is unknown.

Our problem is motivated from the Cognitive Radio spectrum sensing ([6]) and wireless sensor networks intruder detection scenario ([19]). Then usually P_0 is fully known (e.g., when licensed user is not transmitting in Cognitive Radios). However, under H_1, P_1 will usually not be completely known to the local node (e.g., with unknown licensed user transmission parameters).

We first discuss the problem for a single node and then generalize to decentralized setting. Initially we study the case when X_k take values in a finite alphabet. We will be mainly concerned with continuous alphabet observations because receiver almost always has Gaussian noise. This will be taken up in Section IV

III. Finite Alphabet

In this section, we consider finite alphabet for the distributions P_0 and P_1. A sequential test is usually defined by a stopping time N and a decision rule δ. For SPRT ([16]),

$$N \triangleq \inf\{n : W_n \notin (-\gamma_0, \gamma_1)\}, \quad \gamma_0 > 0, \gamma_1 > 0 \qquad (1)$$

where,

$$W_n = \sum_{k=1}^{n} \log \frac{P_1(X_k)}{P_0(X_k)}. \qquad (2)$$

At time N, the decision rule δ decides H_1 if $W_N \geq \gamma_1$ and H_0 if $W_N \leq -\gamma_0$.

SPRT requires full knowledge of P_1. Now we propose our test when P_1 is unknown by replacing the log likelihood ratio process W_n in (2) by

$$\widehat{W}_n = -L_n(X_1^n) - \log P_0(X_1^n) - n\frac{\lambda}{2}, \quad \lambda > 0, \qquad (3)$$

where $\lambda > 0$ is an appropriately chosen constant and $L_n(X_1^n)$ is the length of the source code for the data $X_1^n \triangleq X_1, \ldots, X_n$ when source coding is done via a universal lossless source code which does not require the distribution of X_k.

The following discussion provides motivation for our test.

1) By Shannon-Macmillan Theorem ([2]) for any stationary, ergodic source $\lim_{n\to\infty} n^{-1} \log P(X_1^n) = -\overline{H}(X)$ a.s. where $\overline{H}(X)$ is the entropy rate. We consider universal lossless codes whose codelength function L_n satisfies $\lim_{n\to\infty} n^{-1} L_n = \overline{H}(X)$ a.s., at least for i.i.d sources. Algorithms like LZ78 ([26]) satisfy this convergence even for stationary, ergodic sources. Thus, for such universal codes,

$$\frac{1}{n}(L_n(X_1^n) + \log P(X_1^n)) \to 0 \text{ w.p.1.} \qquad (4)$$

2) Under hypothesis H_1, $E_1[-\log P_0(X_1^n)]$ is approximately $nH_1(X) + nD(P_1||P_0)$ and for large n, $L(X_1^n)$ is approximately $nH_1(X)$ where $H_1(X)$ is the entropy under H_1 and $D(P_1||P_0)$ is the Kullback-Leibler divergence. This gives the average drift under H_1 as $D(P_1||P_0) - \lambda/2$ and under H_0 as $-\lambda/2$. To get some performance guarantees (average drift under H_1 greater than $\lambda/2$), we limit P_1 to a class of distributions,

$$\mathcal{C} = \{P_1 : D(P_1||P_0) \geq \lambda\}. \qquad (5)$$

Thus our test is to use $\widehat{W}_n$ in (1) when P_0 is known and P_1 can be any distribution in class $\mathcal{C}$ defined in (5). Our test is useful for stationary and ergodic sources also.

The following proposition proves the almost sure finiteness of the stopping time of the proposed test. This proposition holds if $\{X_k\}$ are stationary, ergodic and the universal code satisfies a weak pointwise universality. Let $\overline{H}_i$ be the entropy rate of $\{X_1, X_2, \ldots\}$ under $H_i, i = 0, 1$. The proof of this proposition as well as of Theorems 1 and 2 are available in [9]

Proposition 1. *Let $L_n(X_1^n)/n \to \overline{H}_i$ in probability for $i = 0, 1$. Then*

(a) $P_0(N < \infty) = 1$, *and*

(b) $P_1(N < \infty) = 1$.

Remark 1. The assumption $L_n(X_1^n)/n \to \overline{H}_i$ in probability, which is equivalent to the pointwise universality of the universal code in (4), has been shown to be true for i.i.d. sequences for the two universal source codes LZ78 ([11]) and KT-estimator with Arithmetic encoder ([25] with the redundancy property of Arithmetic Encoder [2]) considered later in this paper.

A stronger version of pointwise universality is

$$\max_{x_1^n \in \mathcal{X}^n} \Big(L_n(x_1^n) + \log P_1(x_1^n) \Big) \sim o(n), \qquad (6)$$

$\mathcal{X}$ being the source alphabet. This property is satisfied by the two universal codes used in this paper for i.i.d. sources: KT-estimator with Arithmetic encoder ([3, Chapter 6]) and LZ78 ([11], [26]).

The following theorem gives a bound for P_{FA} and an asymptotic result for P_{MD}.

Theorem 1. *(1) $P_{FA} \triangleq P_0(\widehat{W}_N \geq \gamma_1) \leq \exp(-\gamma_1)$.*
(2) If the observations $X_1, X_2, \ldots, X_n$ are i.i.d. and (6) is satisfied then

$$P_{MD} \triangleq P_1(\widehat{W}_N \leq -\gamma_0) = \mathcal{O}(\exp(-\gamma_0 s)),$$

where s is the solution of $E_1\left[e^{-s\left(\log\frac{P_1(X_1)}{P_0(X_1)} - \frac{\lambda}{2} - \epsilon\right)}\right] = 1$ for $0 < \epsilon < \lambda/2$ and $s > 0$. ∎

We also have the following.

Theorem 2. *(a) Under H_0, $\lim_{\gamma_1,\gamma_0\to\infty} \frac{N}{\gamma_0} = \frac{2}{\lambda}$ a.s. If (6) is satisfied and $E_0[(\log P_0(X_1))^{p+1}] < \infty$ for some $p \geq 1$, then also,*

$$\lim_{\gamma_1,\gamma_0\to\infty} \frac{E_0[N^q]}{{\gamma_0}^q} = \lim_{\gamma_1,\gamma_0\to\infty} \frac{E_0[(N_0)^q]}{{\gamma_0}^q} = \left(\frac{2}{\lambda}\right)^q,$$

for all $0 < q \le p$.

(b) Under H_1, $\lim_{\gamma_1,\gamma_0\to\infty} \frac{N}{\gamma_1} = \frac{1}{D(P_1||P_0)-\lambda/2}$ *a.s.*

If (6) is satisfied, $E_1[(\log P_1(X_1))^{p+1}] < \infty$ *and* $E_1[(\log P_0(X_1))^{p+1}] < \infty$ *for some* $p \ge 1$*, then also,*

$$\lim_{\gamma_1,\gamma_0\to\infty} \frac{E_1[N^q]}{{\gamma_1}^q} = \lim_{\gamma_1,\gamma_0\to\infty} \frac{E_1[(N_1)^q]}{{\gamma_1}^q} = \left(\frac{1}{D(P_1||P_0)-\frac{\lambda}{2}}\right)^q,$$

for all $0 < q \le p$. ∎

IV. Continuous Alphabet

The above test can be extended to continuous alphabet sources. Let f_i be density of P_i, $i = 0, 1$ with respect to a common probability measure. Now, in (2) P_i is replaced by f_i, $i = 0, 1$. Since we do not know f_1, we would need an estimate of $Z_n \triangleq \sum_{k=1}^{n} \log f_1(X_k)$. If $E[\log f_1(X_1)] < \infty$, then by strong law of large numbers, Z_n/n is close to $E[\log f_1(X_1)]$ for all large n with a high probability. Thus, if we have an estimate of $E[\log f_1(X_1)]$ we will be able to replace Z_n as in (2). In the following we get a universal estimate of $E[\log f_1(X_1)] \triangleq -h(X_1)$, where $h(X_1)$ is the differential entropy of X_1, via the universal data compression algorithms.

First we quantize X_i via a uniform quantizer with a quantization step $\Delta > 0$. Let the quantized observations be X_i^{Δ} and the quantized vector $X_1^{\Delta}, \dots, X_n^{\Delta}$ be $X_{1:n}^{\Delta}$. We know that $H(X_1^{\Delta}) + \log\Delta \to h(X_1)$ as $\Delta \to 0$ ([2]). Given i.i.d. observations $X_1^{\Delta}, X_2^{\Delta}, \dots, X_n^{\Delta}$, its code length for a good universal lossless coding algorithm approximates $nH(X_1^{\Delta})$ as n increases. This idea gives rise to the following modification to (3),

$$\widetilde{W}_n = -L_n(X_{1:n}^{\Delta}) - n\log\Delta - \sum_{k=1}^{n} \log f_0(X_k) - n\frac{\lambda}{2} \quad (7)$$

and as for the finite alphabet case, to get some performance guarantee, we restrict f_1 to a class of densities,

$$\mathcal{C} = \{f_1 : D(f_1||f_0) \ge \lambda\}. \quad (8)$$

Let the divergence after quantization be $D(f_1^{\Delta}||f_0^{\Delta})$, f_i^{Δ} being the probability mass function after quantizing f_i. Then by data-processing inequality ([2]) $D(f_1||f_0) \ge D(f_1^{\Delta}||f_0^{\Delta})$. When $\Delta \to 0$ the lower bound is asymptotically tight and this suggests choosing the class (8)

V. Universal Source Codes

In this section we present two universal source codes which we will use in our algorithms.

A. LZSLRT (Lempel-Ziv Sequential Likelihood Ratio Test)

In the following in (7) we use Lempel-Ziv incremental parsing technique LZ78 ([26]), which is a well known efficient universal source coding algorithm. We call this algorithm LZSLRT.

Thus the test statistic $\widetilde{W}_n^{LZ}$, is

$$\widetilde{W}_n^{LZ} = -\sum_{i=1}^{t} \lceil \log i|A| \rceil - C\left(\frac{1}{\log n} + \frac{\log\log n}{n} + \frac{\log\log n}{\log n}\right) - n\log\Delta - \sum_{k=1}^{n} \log f_0(X_k) - n\frac{\lambda}{2}.$$

The the second term corresponds to the correction to take care of the low redundancy rate of LZ78 ([10]). Here C is a constant which depends on the size of the quantized alphabet and t is the number of phrases after parsing $X_1^{\Delta}, \dots, X_n^{\Delta}$ in LZ78 encoder and $|A|$ is the alphabet size of the quantized alphabet.

B. KTSLRT (Krichevsky-Trofimov Sequential Likelihood Ratio Test)

In this section we propose KTSLRT for i.i.d. sources. The codelength function L_n in (7) now comes from the combined use of KT (Krichevsky-Trofimov [12]) estimator of the distribution of quantized source and the Arithmetic Encoder ([2]). We will denote this combined encoder by KT+AE.

KT-estimator for a finite alphabet source is defined as,

$$P_c(x_1^n) = \prod_{t=1}^{n} \frac{v(x_t/x_1^{t-1}) + \frac{1}{2}}{t - 1 + \frac{|A|}{2}}, \quad (9)$$

where $v(i/x_1^{t-1})$ denotes the number of occurrences of the symbol i in x_1^{t-1}. It is known ([2]) that the coding redundancy of the Arithmetic Encoder is smaller than 2 bits, i.e., if $P_c(x_1^n)$ is the coding distribution used in the Arithmetic Encoder then $L_n(x_1^n) < -\log P_c(x_1^n) + 2$. In our test we actually use $-\log P_c(x_1^n) + 2$ as the code length function and do not need to implement the Arithmetic Encoder. This is an advantage over the scheme LZSLRT presented above.

It is proved in [3] that universal code defined by the KT+AE is nearly optimal for i.i.d. finite alphabet sources.

Writing (9) recursively, (7) can be modified as

$$\widetilde{W}_n^{KT} = \widetilde{W}_{n-1}^{KT} + \log\left(\frac{v(X_n^{\Delta}/X_1^{\Delta n-1}) + \frac{1}{2} + S}{t - 1 + \frac{|A|}{2}}\right) - \log\Delta - \log f_0(X_n) - \frac{\lambda}{2},$$

where f_0^{Δ} is the probability mass function after quantizing f_0 and S is a scalar constant whose value greatly influences the performance. The default value of S is zero.

VI. Performance Comparison

In this section, we compare the performance of LZSLRT and KTSLRT with some other estimators available in literature via simulations. Due to the difference in the expected drift of likelihood ratio process under H_1 and H_0, some algorithms perform better under one hypothesis and worse under the other hypothesis. Hence instead of plotting $E_1[N]$ versus P_{MD} and $E_0[N]$ versus P_{FA} separately, we plot $E_{DD} \triangleq 0.5E_1[N] + 0.5E_0[N]$ versus $P_E \triangleq 0.5P_{FA} + 0.5P_{MD}$. We use an eight bit uniform quantizer.

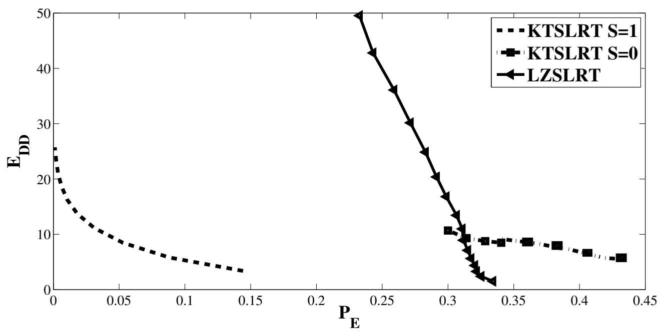

Fig. 1. Comparison between KTSLRT and LZSLRT for Gaussian Distribution.

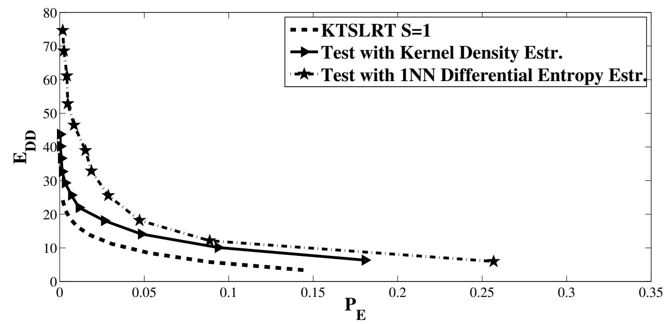

Fig. 2. Comparison among KTSLRT, universal sequential tests using 1NN differential entropy estimator and that using Kernel density estimator.

We compare the performance of LZSLRT to that of SPRT and GLR-Lai ([13]), which is nearly optimal for exponential families, in Table I. We use $f_0 \sim \mathcal{P}(10,2)$ and $f_1 \sim \mathcal{P}(3,2)$, where $\mathcal{P}(K, X_m)$ is the Pareto density function with K and X_m as the shape and scale parameter of the distribution and $\Delta = 0.3125$. We observe that LZSLRT performs better than GLR-Lai for the Pareto Distribution.

E_{DD}	$P_E = 0.05$	$P_E = 0.01$	$P_E = 0.005$
SPRT	7.45	10.86	18.23
GLR-Lai	18.21	29.65	33.42
LZSLRT	16.96	28.31	31.48

TABLE I
COMPARISON AMONG SPRT, GLR-LAI AND LZSLRT FOR PARETO DISTRBUTION.

Performance of KTSLRT is compared with LZSLRT in Figure 1. We take $f_1 \sim \mathcal{N}(0,5)$ and $f_0 \sim \mathcal{N}(0,1)$ where $\mathcal{N}(a,b)$ denotes Gaussian distribution with mean a and variance b. We observe that LZSLRT and KTSLRT with $S = 0$ (the default case) are not able to provide P_E less than 0.3 and 0.23 respectively, although KTSLRT with $S = 1$ provides much better performance. We have found in our simulations with other data also that KTSLRT with $S = 0$ performs much worse than with $S = 1$. Thus in the following we will take KTSLRT with $S = 1$.

The superior performance of KTSLRT over LZSLRT attributes to the pointwise redundancy rate $n^{-1}(L_n(X_1^n) + \log P(X_1^n)) = \mathcal{O}(\log n/n)$ of KT+AE ([25]) as compared to $\mathcal{O}(1/\log n)$ of LZ78 ([11]).

In Figure 2 we compare KTSLRT with sequential tests in which $-n\hat{h}_n$ replaces $\sum_{k=1}^{n} \log f_1(X_k)$ where $\hat{h}_n$ is an estimate of the differential entropy and with a test defined by replacing f_1 by a density estimator $\hat{f}_n$.

It is shown in [24] that 1NN (1st Nearest Neighbourhood) differential entropy estimator performs better than other differential entropy estimators where 1-NN differential entropy estimator is

$$\hat{h}_n = \frac{1}{n}\sum_{i=1}^{n} \log \rho(i) + \log(n-1) + \gamma + 1,$$

and $\rho(i) \triangleq \min_{j:1\le j\le n, j\ne i} ||X_i - X_j||$ and γ is the Euler-Mascheroni constant (=0.5772...).

There are many density estimators available ([17]). We use the Gaussian example in Figure 1 for comparison. For Gaussian distributions, a Kernel density estimator is a good choice as optimal expressions are available for the parameters in the Kernel density estimators ([17]). The Kernel density estimator at a point z is

$$\hat{f}_n(z) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z - X_i}{h}\right),$$

where K is the kernel and $h > 0$ is a smoothing parameter called the bandwidth. If Gaussian kernel is used and the underlying density being estimated is Gaussian then it can be shown that the optimal choice for h is ([17]) $\left(4\hat{\sigma}^5/3n\right)^{1/5}$, where $\hat{\sigma}$ is the standard deviation of the samples.

We provide the comparison of KTSLRT with the above two schemes in Figure 2. We find that KTSLRT with $S = 1$ performs the best.

Next we provide comparison with the asymptotically optimal (in terms of error exponents) universal fixed sample size test for finite alphabet sources. This test is called Hoeffding test ([7], [20]). The decision rule of Hoeffding test, $\delta_{FSS} = \mathbb{I}\{D(\Gamma^n||P_0) \ge \eta\}$, where $\Gamma^n(x)$ is the type of $X_1, \ldots, X_n, = \{\frac{1}{n}\sum_{i=1}^{N} \mathbb{I}\{X_i = x\}, x \in \mathcal{X}\}$, where N is the cardinality of source alphabet $\mathcal{X}$ and $\eta > 0$ is an appropriate threshold. From [20, Theorem III.2], under P_0

$$nD(\Gamma^n||P_0) \xrightarrow[n\to\infty]{d} \frac{1}{2}\chi^2_{N-1},$$

under P_1,

$$\sqrt{n}\big(D(\Gamma^n||P_0) - D(P_1||P_0)\big) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma_1^2).$$

where $\sigma_1^2 = Var_{P_1}\left[\log \frac{P_1(X_1)}{P_0(X_1)}\right]$ and χ^2_{N-1} is the Chi-Squared distribution with $N-1$ degrees of freedom. From the above two approximations, number of samples, n to achieve P_{FA} and P_{MD} can be computed theoretically as a solution of

$$2nD(P_1||P_0) + 2\sqrt{n}F_{\mathcal{N}}^{-1}(P_{MD}) - F_{\chi}^{-1}(1 - P_{FA}) = 0,$$

where $F_{\mathcal{N}}^{-1}$ and F_{χ}^{-1} denote inverse cdf's of the above Gaussian and Chi-Squared distributions.

Since this is a discrete alphabet case, we use (3) with KT+AE as the universal code. Figure 3 provides the comparison when $P_0 \sim B(8, 0.2)$ and $P_1 \sim B(8, 0.5)$, where

$B(n,p)$ represents the Binomial distribution with n trials and p as the success probability in each trial. It can be seen that our test outperforms Hoeffding test in terms of average number of samples.

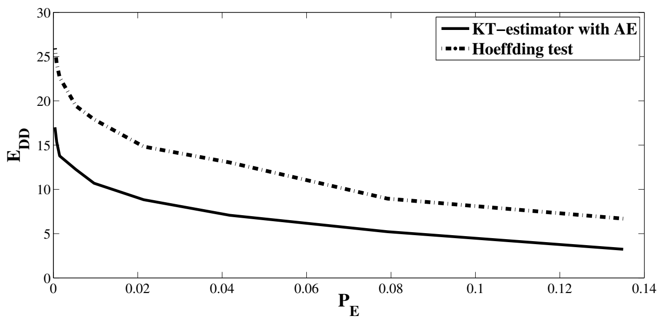

Fig. 3. Comparison between Hoeffding test and our discrete alphabet test (3) for Binomial distribution.

VII. Decentralized Detection

A. Algorithm

Motivated by the satisfactory performance of a single node case, we extend LZSLRT and KTSLRT to the decentralized setup in [18]. Let $X_{k,l}$ be the observation made at node l at time k. We assume that $\{X_{k,l}, k \geq 1\}$ are i.i.d. and that the observations are independent across nodes. We will denote by $f_{1,l}$ and $f_{0,l}$ the densities of $X_{k,l}$ under H_1 and H_0 respectively. Using a detection algorithm on $\{X_{n,l}, n \leq k\}$ the local node l transmits $Y_{k,l}$ to the fusion node at time k. We assume a multiple-access channel (MAC) between nodes and FC in which the FC receives Y_k, a coherent superposition of the node transmissions: $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k$, where $\{Z_k\}$ is i.i.d., zero mean Gaussian receiver noise with variance σ^2 (for our algorithms Gaussian assumption is not required, but its distribution is assumed to be known). FC observes Y_k, runs a decision rule and decides upon the hypothesis.

Now our assumptions are that at local nodes, $f_{0,l}$ is known but $f_{1,l}$ is not known. The variance σ^2 of Z_k is known to the FC. Thus we use LZSLRT at each local node and Wald's SPRT at FC (we call it LZSLRT-SPRT). Similarly we can use KTSLRT at each node and SPRT at FC and call it KTSLRT-SPRT. In both the cases whenever at a local node, the stopping time is reached, it transmits b_1 if its decision is H_1 and transmits b_0 if the decision is H_0. At the FC we have SPRT for the binary hypothesis testing of two densities g_1 (density of $Z_k + \mu_1$) and g_0 (density of $Z_k - \mu_0$), where μ_0 and μ_1 are design parameters. At the FC, the Log Likelihood Ratio Process (LLR) crosses the upper threshold under H_1 with a high probability when a sufficient number of local nodes (denoted by I, to be specified appropriately) transmit b_1. Thus $\mu_1 = b_1 I$ and similarly $\mu_0 = b_0 I$.

In the following we compare the performance of LZSLRT-SPRT, KTSLRT-SPRT and DualSPRT developed in [18] which requires knowledge of $f_{1,l}$ at CR l. Asymptotically, DualSPRT is shown to achieve the performance of the optimal centralized test, which does not consider fusion center noise. We choose $b_1 = 1$, $b_0 = -1$, $I = 2$, $L = 5$ and $Z_k \sim \mathcal{N}(0,1)$. We use an eight bit quantizer in all these experiments. In Figure 4 $f_{0,l} \sim \mathcal{N}(0,1)$ and $f_{1,l} \sim \mathcal{N}(0,5)$, for $1 \leq l \leq L$. We observe that KTSLRT-SPRT performs much better than LZSLRT-SPRT. It also performs better than DualSPRT for higher values of P_E.

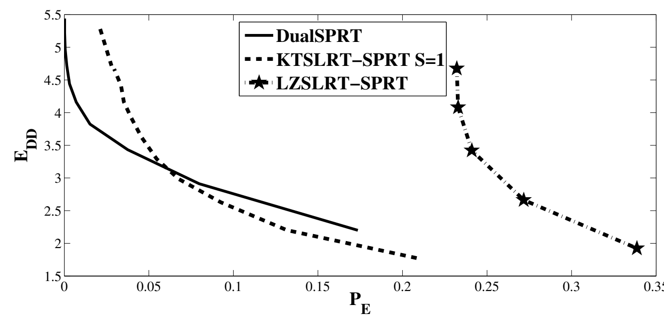

Fig. 4. Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Gaussian Distribution.

B. Performance Analysis

Definition 1 ([5]). A process $\{Z_n, n \geq 1\}$, such that $Z_n = S_n + \xi_n, n \geq 1$, is called a perturbed random walk if $\{S_n, n \geq 1\}$ is a random walk whose increments have positive finite mean, and $\{\xi_n, n \geq 1\}$ is a sequence of random variables with $\xi_n/n \to 0$ a.s. as $n \to \infty$.

$\widehat{W}_n$ in (3) can be written as

$$\widehat{W}_n = \xi_n + S_n, \tag{10}$$

where

$$\xi_n = -L_n(X_1^n) - \log P_1(X_1^n) \text{ and } S_n = \sum_{k=1}^{n} \log \frac{P_1(X_k)}{P_0(X_k)} - \frac{\lambda}{2}.$$

Observe that $\{S_n, n \geq 1\}$ as a random walk and $\xi_n/n \to 0$ a.s. due to the pointwise convergence of universal source codes. Thus $\widehat{W}_n$ is a perturbed random walk.

Since the analysis is almost same under H_1 and H_0 with necessary modifications, we provide only under H_1.

At node l, let $P_{0,l}$ and $P_{1,l}$ denote the distribution under H_0 and H_1 at local node l, and

$$\delta_l = E_1 \left[\log \frac{P_{1,l}(X_{k,l})}{P_{0,l}(X_{k,l})} - \frac{\lambda}{2} \right],$$

$$\rho_l^2 = Var_{H_1} \left[\log \frac{P_{1,l}(X_{k,l})}{P_{0,l}(X_{k,l})} - \frac{\lambda}{2} \right].$$

We will assume δ_l finite throughout this paper. By Jensen's Inequality and (8), $\delta_l > 0$. Let $\widehat{W}_{k,l}$ be the test static at local node l, time k, equivalent to (7). Also let F_k be the FC SPRT test static at time k. $\{\widehat{W}_{k,l}, k \geq 0\}$ is a perturbed random walk with expected drift of the random walk given by δ_l. We use γ_0 and γ_1 as $\log \beta$ and $-\log \alpha$ in Section III. Let

$$N_l = \inf\{k : \widehat{W}_{k,l} \notin (-\gamma_0, \gamma_1)\},$$

$$N_l^1 = \inf\{k : \widehat{W}_{k,l} > \gamma_1\},\ N_l^0 = \inf\{k : \widehat{W}_{k,l} < -\gamma_0\}.$$

Then $N_l = \min\{N_l^0, N_l^1\}$. Let N_d denote the stopping time of FC SPRT: $N_d = \inf\{k : F_k \notin (-\beta_0, \beta_1)\}$. Let $N_d^0 = \inf\{k : F_k \leq -\beta_0\}$ and $N_d^1 = \inf\{k : F_k \geq \beta_1\}$. Then $N_d = \min\{N_d^1, N_d^0\}$.

We choose $\gamma_1 = \gamma_0 = \gamma$, $\beta_1 = \beta_0 = \beta$, $b_1 = -b_0 = b$ and $\mu_1 = \mu_0 = \mu$ for simplicity of notation.

From the perturbed random walk results, by Central Limit Theorem for the first passage time N_l^1 ([5, Theorem 2.3 in Chapter 6]), when

$$\xi_n/\sqrt{n} \to 0 \text{ a.s.}, \tag{11}$$

$$N_l^1 \sim \mathcal{N}\left(\frac{\gamma}{\delta_l}, \frac{\rho_l^2 \gamma}{\delta_l^3}\right). \tag{12}$$

Remark 2. From the redundancy rates given in Section VI, it can be seen that KT+AE satisfies (11), but not LZ78. The following decentralized analysis is applicable for any universal source code which satisfies (11).

1) $E[N_d|H_1]$ Analysis: At the fusion node F_k crosses β first under H_1 with a high probability when a sufficient number of local nodes transmit b_1. The dominant event occurs when the number of local nodes transmitting are such that the mean drift of the random walk F_k will just have turned positive. In the following we find the mean time to this event and then the time to cross β after this.

The following lemmas provide justification for considering only the events $\{N_l^1\}$ and $\{N_d^1\}$ for analysis of $E[N_d|H_1]$.

Lemma 1. *$P_1(N_l = N_l^1) \to 1$ as $\gamma \to \infty$ and $P_1(N_d = N_d^1) \to 1$ as $\gamma \to \infty$ and $\beta \to \infty$.*

Proof: From (10), $\widehat{W}_n/n \to D(f_1||f_0) - \lambda/2$ a.s. since $\xi_n/n \to 0$ and $S_n/n \to D(f_1||f_0) - \lambda/2$ a.s. Thus by (8), $\widehat{W}_n \to \infty$ a.s. This in turn implies that $\widehat{W}_n$ never crosses some finite negative threshold a.s. This implies that $P_1(N_l^0 < \infty) \to 0$ as $\gamma \to \infty$ but $P_1(N_l^1 < \infty) = 1$ for any $\gamma < \infty$. Thus $P_1(N_l = N_l^1) \to 1$ as $\gamma \to \infty$. This also implies that for large γ, the drift of F_k is positive for H_1 with a high probability and $P_1(N_d = N_d^1) \to 1$ as $\gamma \to \infty$ and $\beta \to \infty$. ■

From Lemma 1 we also get that $|N_l - N_l^1| \to 0$ a.s. as $\gamma \to \infty$ and $|N_d - N_d^1| \to 0$ a.s. as $\gamma \to \infty$ and $\beta \to \infty$. From this fact, along with Theorem 2, we can use the result in (12) for N_l also. The following lemma also holds.

Lemma 2. *Let t_k be the time when k local nodes have made the decision. As $\gamma \to \infty$,*

$$P_1(\textit{Decision at time } t_k \textit{ is } H_1 \textit{ and } t_k \textit{ is the } k^{th} \textit{ order statistics of } N_1^1, N_2^1, \ldots, N_L^1) \to 1.$$

Proof: From Lemma 1,

$$\begin{aligned} &P_1(\text{Decision at time } t_k \text{ is } H_1 \text{ and} \\ &\quad t_k \text{ is the } k^{th} \text{ order statistics of } N_1^1, N_2^1, \ldots, N_L^1) \\ &\qquad \geq P_1(N_l^1 < N_l^0, l = 1, \ldots, L) \to 1, \ \text{ as } \gamma \to \infty. \end{aligned}$$

■

We use Lemma 1-2, Theorem 2 and equation (12) in the following to obtain an approximation for $E[N_d|H_1]$ when γ and β are large. Large γ and β are needed for small probability of error. Then we can assume that the local nodes are making correct decisions. Let δ_{FC}^j be the mean drift of the fusion center SPRT F_k, when j local nodes are transmitting. Then t_j is the point at which the drift of F_k changes from δ_{FC}^{j-1} to δ_{FC}^j and let $\bar{F}_j = E_1[F_{t_j - 1}]$, the mean value of F_k just before transition epoch t_j.

Let

$$l^* = min\{j : \delta_{FC}^j > 0 \text{ and } \frac{\beta - \bar{F}_j}{\delta_{FC}^j} < E[t_{j+1}] - E[t_j]\}.$$

$\bar{F}_j$ can be iteratively calculated as

$$\bar{F}_j = \bar{F}_{j-1} + \delta_{FC}^j \left(E[t_j] - E[t_{j-1}]\right),\ \bar{F}_0 = 0. \tag{13}$$

Note that $\delta_{FC}^j\, 0 \leq j \leq L$ is assumed to be jb and t_j is the j^{th} order statistics of $\{N_l^1,\, 0 \leq l \leq L\}$. The Gaussian approximation (12) can be used to calculate the expected value of the order statistics using the method given in [1]. This implies that $E[t_j]$s and hence $\bar{F}_j s$ are available offline. By using these values $E_1[N_d]$ ($\approx E_1[N_d^1]$) can be approximated as

$$E_1[N_d] \approx E[t_{l^*}] + \frac{\beta - \bar{F}_{l^*}}{\delta_{FC}^{l^*}}, \tag{14}$$

where the first term on R.H.S. is the mean time till the drift becomes positive at the fusion node while the second term indicates the mean time for F_k to cross β from t_{l^*} onward.

In case of continuous alphabet sources which is assumed in our decentralized algorithm, $\widetilde{W}_n$ in (7) can be modified as,

$$\begin{aligned} &= -L_n(X_{1:n}^{\Delta}) - \sum_{k=1}^{n} \log(f_1(X_k)\Delta) + \sum_{k=1}^{n} \log \frac{f_1(X_k)}{f_0(X_k)} - \frac{\lambda}{2} \\ &\overset{(a)}{\approx} -L_n(X_{1:n}^{\Delta}) - \log(f_1^{\Delta}(X_{1:n}^{\Delta})) + \sum_{k=1}^{n} \log \frac{f_1(X_k)}{f_0(X_k)} - \frac{\lambda}{2}. \end{aligned}$$

Here f_1^{Δ} is the probability mass function after quantizing f_1, which is the distribution being learnt by $L_n(X_{1:n}^{\Delta})$. (a) is due to the approximation $f_1(x)\Delta \approx f_1^{\Delta}(x^{\Delta})$ at high rate uniform quantization. By taking $\xi_n = -L_n(X_{1:n}^{\Delta}) - \log(f_1^{\Delta}(X_{1:n}^{\Delta}))$ and $S_n = \sum_{k=1}^{n} \log \frac{f_1(X_k)}{f_0(X_k)} - \frac{\lambda}{2}$, it is clear that $\widetilde{W}_n$ can be approximated as a perturbed random walk since $\{S_n, n \geq 1\}$ is a random walk and $\xi_n/n \to 0$ a.s. from the pointwise convergence of universal source codes.

2) P_{MD} Analysis: At reasonably larger local node thresholds, according to Lemma 2, with a high probability local nodes are making the right decisions and t_k can be taken as the order statistics assuming that all local nodes make the right decisions. P_{MD} at the fusion node is given by,

$$P_{MD} = P_1(\text{accept } H_0) = P_1(N_d^0 < N_d^1).$$

It can be easily shown that $P_1(N_d^1 < \infty) = 1$ for any $\beta > 0$. Also $P_1(N_d^0 < \infty) \to 0$ as $\beta \to \infty$. We should decide the different thresholds such that $P_1(N_d^1 < t_1)$ is small for reasonable performance. Therefore

$$P_{MD} = P_1(N_d^0 < N_d^1) \geq P_1(N_d^0 < t_1, N_d^1 > t_1) \approx P_1(N_d^0 < t_1). \quad (15)$$

Also,

$$P_1(N_d^0 < N_d^1) \leq P_1(N_d^0 < \infty) = P_1(N_d^0 < t_1) + P_1(t_1 \leq N_d^0 < t_2) + P_1(t_2 \leq N_d^0 < t_3) + \dots \quad (16)$$

The first term in the right hand side is expected to be the dominant term. This is because, from Lemma 2, after t_1, the drift of F_k will be most likely more positive than before t_1 (if P_{MD} at local nodes are reasonably small) and causes fewer errors if the fusion center threshold is chosen appropriately. We have verified this from simulations also. Hence we focus on the first term. Combining this fact with (15), $P_1(N_d^0 < t_1)$ will be a good approximation for $P_1(\text{reject } H_1)$. For calculating $P_1(N_d^0 < t_1)$, we use the bounding technique and approximate expression given in [18, Section III-B2] with the distribution of N_l^1 in (12).

Table II provides comparison between analysis and simulations for continuous distributions. The simulation setup is same as that in Figure 4. It shows that at low P_{MD}, $E_1[N_d]$ from theory approximates the simulated value reasonably well.

$P_{MD} Sim.$	$P_{MD} Anal.$	$E_1[N_d] Sim.$	$E_1[N_d] Anal.$
0.081	0.072	4.12	5.91
0.067	0.059	5.32	6.43
0.034	0.031	6.63	6.35

TABLE II
KTSLRT-SPRT: Comparison of $E_1[N_d]$ and P_{MD} obtained via analysis and simulation.

VIII. Conclusions

The problem of universal sequential hypothesis testing is very relevant in practical applications, e.g., quickest detection with SNR uncertainty in Cognitive Radio systems. We have used universal lossless source codes for learning the underlying distribution. The algorithm is first proposed for discrete alphabet and almost sure finiteness of the stopping time is proved. Asymptotic properties of probability of error and stopping times are also derived. Later on the algorithm is extended to continuous alphabet with the use of uniform quantization. We have used Lempel-Ziv code and KT-estimator with Arithmetic Encoder as universal lossless codes. From the performance comparisons, it is found that KT-estimator with Arithmetic Encoder (KT+AE) always performs the best. We have compared this algorithm with other universal hypothesis testing schemes also and found that KT+AE performs the best. Finally we have extended these algorithms to decentralized setup.

References

[1] H. M. Barakat and Y. H. Abdelkader, "Computing the moments of order statistics from nonidentical random variables," *Statistical Methods & Applications*, vol. 13, no. 1, pp. 15–26, 2004.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[3] I. Csiszár and P. C. Shields, *Information theory and statistics: A tutorial*. Now Publishers Inc, 2004.

[4] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 534–548, 2011.

[5] A. Gut, *Stopped Random Walks: Limit Theorems and Applications*, 2nd ed. Springer, 2009.

[6] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201 – 220, Feb 2005.

[7] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *Annals of Mathematical Statistics*, vol. 36, pp. 369–408, 1965.

[8] T. Jacob and R. K. Bansal, "Sequential change detection based on universal compression algorithms," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 1108 –1112.

[9] K. S. Jithin, "Spectrum sensing in cognitive radios using distributed sequential detection," Master's thesis, Indian Institute of Science, Bangalore, 2012.

[10] J. C. Kieffer, "Lecture notes in source coding," Chapter 5, unpublished.

[11] J. C. Kieffer and E. H. Yang, "A simple technique for bounding the pointwise redundancy of the 1978 lempel-ziv algorithm," in *Proc. Data Compression Conference (DCC)*, Mar 1999, pp. 434 –442.

[12] R. Krichevsky and V. Trofimov, "The performance of universal encoding," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, pp. 199 – 207, Mar 1981.

[13] T. L. Lai, "Sequential Analysis: Some Classical Problems and New Challenges," *Statistica Sinica*, vol. 11, pp. 303–408, 2001.

[14] N. Mukhopadhyay and B. M. De Silva, *Sequential methods and their applications*. Chapman & Hall/CRC, 2009.

[15] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 629 – 636, Jul 1984.

[16] D. Siegmund, *Sequential analysis: tests and confidence intervals*. Springer, 1985.

[17] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed. Chapman and Hall/CRC, 1986.

[18] J. K. Sreedharan and V. Sharma, "Spectrum sensing using distributed sequential detection via noisy MAC," *Arxiv preprint arXiv:1211.5562*, Dec 2012.

[19] A. Swami, Q. Zhao, Y.-W. Hong, and L. Tong, Eds., *Wireless Sensor Networks: Signal Processing and Communications*. Wiley, 2007.

[20] J. Unnikrishnan, D. Huang, S. P. Meyn, A. Surana, and V. V. Veeravalli, "Universal and composite hypothesis testing via mismatched divergence," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1587 –1603, Mar 2011.

[21] P. K. Varshney and C. S. Burrus, *Distributed detection and data fusion*. Springer Verlag, 1997.

[22] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 433 –442, Mar 1993.

[23] A. Wald, *Sequential analysis*. Wiley, New York, 1947.

[24] Q. Wang, S. R. Kulkarni, and S. Verdú, "Universal estimation of information measures for analog sources," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 3, pp. 265–353, 2009.

[25] Q. Xie and A. R. Barron, "Asymptotic minimax regret for data compression, gambling, and prediction," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 431 –445, Mar 2000.

[26] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530 – 536, Sep 1978.