Spectrum Sensing 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 sequential hypothesis testing when the distribution under null hypothesis is fully known and the alternate hypothesis corresponds to some other unknown distribution. We use easily implementable universal lossless source codes to propose simple algorithms for such a setup. These algorithms are motivated from spectrum sensing application in Cognitive Radios. Universal sequential hypothesis testing using Lempel Ziv codes and Krichevsky-Trofimov estimator with Arithmetic Encoder are considered and compared for different distributions. Cooperative spectrum sensing with multiple Cognitive Radios using universal codes is also considered.**

Keywords- Cognitive Radio, Spectrum Sensing, Sequential Hypothesis Testing, Universal Testing, Universal Source Codes

I. Introduction

Cognitive Radio (CR) technology has been proposed as a working solution for shortage of the spectrum due to the increase in wireless services. Cognitive users are the users who transmit in a frequency band licensed to other (primary) users in such a way that the primary users are not affected. It is the responsibility of the secondary users (Cognitive Radios) to identify the primary (licensed) user's spectrum usage via spectrum sensing. Given the noise, interference and channel uncertainties it is difficult for the secondary users to identify such opportunities reliably. Hence spectrum sensing has become one of the main challenges faced by CRs.

Spectrum sensing problem can be formulated in different ways, one of them being reducing the number of samples taken for deciding if a primary is transmitting or not. In this setting sequential hypothesis testing using Sequential Probability Ratio Test (SPRT) minimizes the mean number of samples used ([17], [20]). In the other formulation, one may decide to know when a primary turns ON and when it turns OFF via detection of change algorithms. Spectrum sensing algorithms based on sequential hypothesis testing are explored in [10], [27] and sequential change detection is studied in [2], [14].

Here we consider sequential hypothesis testing as it is useful when the status of the primary is known to change very slowly, e.g., detecting usage of idle TV bands, which is targeted as the primary application of Cognitive Radios. We explore the case where the noise statistics under no primary transmission is fully known to the secondary node, but the channel gains, modulation schemes etc. of primary transmissions is not available to the secondary user. This is the most common scenario in the CR setup.

The sequential methods in case of uncertainties are studied in [15], [27] for parametric family of distributions. For nonparametric sequential methods, [17] 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 suitable for spectrum sensing where the unknown alternate distribution can be anything which satisfies a constraint on the Kullback-Leibler divergence ([4]) with the noise distribution.

Universal hypothesis fixed sample size tests are considered in [16] from error exponents point of view and in [22] using mismatched divergence. The initial work on statistical inference with the help of universal codes, started in [18], [25], which study classification of finite alphabet sources using universal coding in fixed sample size setup. [9] 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 finite and countable alphabets. In practical applications two sided sequential tests are desirable and often the distributions under the two hypothesis have continuous alphabet. [19] considers both discrete and continuous alphabet for a fixed sample size. For continuous alphabet this paper considers partitions of the real alphabet and proves that with a bound on Type I error, Type II error tends to zero as the sample size goes to infinity.

In this paper we consider sequential universal source coding framework for binary hypothesis with continuous alphabets. This framework captures SNR uncertainty and fading scenarios. Our algorithms also find applications in intruder detection in sensor networks. Our previous work on sequential hypothesis tests using universal codes ([10]) studied tests using Lempel-Ziv (LZ) ([26]) codes and compared it with the composite hypothesis tests. Cooperative setup is also considered there. In this work, we provide some theoretical results of sequential tests using universal codes and propose another universal test using Krichevsky-Tofimov (KT) estimator with Arithmetic Encoder ([5]). We compare both of these tests for different scenarios and find the new work outperforms previous one.

We also extend our algorithm to cooperative spectrum sensing setup in which different CRs interact with each other to provide spectrum sensing, mitigating the effects of multipath fading, shadowing and hidden node problem in single node spectrum sensing methods. It also improves probability of false alarm and probability of miss-detection by making use of spatial diversity. Previous work in cooperative framework ([1]) does not consider the universal setup, to the best of our knowledge.

In our distributed cooperative spectrum sensing algorithm each CR sends a summary static or a local decision to the fusion center instead of full observations. This saves the communication cost in the CR network considerably.

This paper is organized as follows. Section II describes the model. Section III provides our algorithm for a single node. This section starts with a general description of the universal test. Later we provide tests based on Lempel-Ziv codes and Krichovsky-Trofimov estimator with Arithmetic Encoder. In Section IV we compare the two algorithms via simulations. Section V focusses on decentralized setup. Section VI concludes the paper.

II. Model for single CR

We consider the following hypothesis testing problem: Given i.i.d. observations $X_1, X_2, \dots,$ 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). When the observations come from a source with continuous alphabet, we assume P_0 and P_1 have densities f_0 and f_1 with respect to some probability measure. We will assume that P_0 is known but P_1 is unknown. Of course if f_1 belongs to an exponential family with an unknown parameter θ then we can use asymptotically optimal tests mentioned in [15].

We first discuss the problem for a single CR and then generalize to cooperative setting. We will be mainly concerned with continuous alphabet observations because receiver almost always has Gaussian noise.

Our problem is motivated from the Cognitive Radio scenario. In a CR setup, a CR user checks to see if a frequency band is being used by a primary (hypothesis H_1) or not (hypothesis H_0). Under H_0 the CR receiver only senses noise. Usually noise is Gaussian with mean zero and its variance can often be estimated. However, under H_1 the primary is transmitting. The transmit power, its modulation and channel gain may be time varying and not known to the CR node. Thus P_1 (f_1) will usually not be completely known to the CR node.

III. Universal sequential hypothesis testing using universal source codes

We first consider discrete alphabet for the distributions P_0 and P_1. This test is motivated from the universal one sided sequential test for discrete alphabet in [9]. In one sided tests one assumes H_0 as the default hypothesis and has to wait a long time to confirm whether it is the true hypothesis (H_0 is the true hypothesis only when the test never stops) and in spectrum sensing this is not desirable because it is important to make a quick decision. Hence we switch our attention to two sided tests which have a finite decision time under both H_0 and H_1.

A sequential test is usually defined by a stopping time N and a decision rule δ. For SPRT ([20]),

$$N \triangleq \inf\{n : W_n \notin (\log\beta, -\log\alpha)\}, 0 < \alpha, \beta < 1, \quad (1)$$

where,

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

At time N, the decision rule δ decides H_1 if $W_N \geq -\log\alpha$ and H_0 if $W_N \leq \log\beta$, where α and β are defined to satisfy targeted Probability of False Alarm, $P_{FA} = P_0[W_N \geq -\log\alpha]$ and Probability of Miss-detection, $P_{MD} = P_1[W_N \leq \log\beta]$.

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 $\widehat{W}_n$, where

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

$L_n(X_1^n)$ is the codelength function of a universal lossless source code for the data X_1^n and λ is an appropriately chosen constant (see discussion below).

The following discussion provides motivation for our test.

1) The test follows from the pointwise universality of the universal lossless codes:
$$\frac{1}{n}(L_n(X_1^n) + \log P(X_1^n)) \to 0 \text{ w.p.1.} \quad (4)$$
This can be argued in the following way. By Shannon-Macmillan Theorem ([4]) for any stationary ergodic source $\lim_{n\to\infty} n^{-1} \log P(X_1^n) = -\bar{H}(X)$ a.s. where $\bar{H}(X)$ is the entropy rate. We consider universal lossless codes whose codelength function satisfies $\lim_{n\to\infty} n^{-1} L_n = \bar{H}(X)$ a.s., at least for i.i.d sources. Algorithms like LZ78 ([26]) satisfy this convergence even for stationary ergodic sources. From the above two expressions (4) follows.
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 E_i denotes expectation when H_i is the true hypothesis, $i = \{0, 1\}$, $H_1(X)$ is the entropy under H_1 and $D(P_1||P_0)$ is the Kullback-Leibler divergence between distributions P_1 and P_0. 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,
$$\mathcal{C} = \{P_1 : D(P_1||P_0) \geq \lambda\}. \quad (5)$$
3) When considering universal hypothesis testing in Neyman-Pearson framework (fixed sample size) the existing work considers error exponents ([16]),
$$\max_{\delta} \liminf_{n\to\infty} -\log P_{MD},$$
$$\text{such that } \liminf_{n\to\infty} -\log P_{FA} \geq \hat{\alpha}.$$
But in sequential detection framework the aim is to
$$\min_{(N,\delta)} E_1[N], \quad \min_{(N,\delta)} E_0[N],$$

such that $P_{FA} \leq \alpha$ and $P_{MD} \leq \beta$.

In case of the universal sequential detection framework, the objective can be to obtain a test satisfying $P_{FA} \leq \alpha$ and $P_{MD} \leq \beta$ with

$$E_1[N] \rightarrow E_1^S[N] = \frac{|\log \alpha|}{D(P_1||P_0)}, \tag{6}$$

$$E_0[N] \rightarrow E_0^S[N] = \frac{|\log \beta|}{D(P_0||P_1)}, \tag{7}$$

as $\alpha + \beta \rightarrow 0$,

where $E_i^S(N)$ is the expected value of N under H_i for SPRT, when $i = 0, 1$.

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 non i.i.d sources also as most of the properties given above are valid for stationary and ergodic sources for certain universal lossless codes. Note that our test is more generally applicable than "robust" sequential tests available which are usually insensitive only against small deviations from the assumed statistical model ([7]).

Proposition. *For our test the following properties holds. Let* $\mathcal{G}_n^1 = W_n - n\lambda/2$ *and* $\mathcal{G}_n^0 = -n\lambda/2$. *If* $n^{-1}|\widehat{W}_n - \mathcal{G}_n^i| \rightarrow 0$ *in probability under hypothesis* H_i *and* $\{X_n\}$ *are i.i.d. then,*

(a) $P_0(N < \infty) = 1$.
(b) $P_1(N < \infty) = 1$.
(c) $P_{FA} \triangleq P_0(\widehat{W}_N \geq -\log \alpha) \leq \alpha$.

Proof: Let $N_1 = \inf\{n : \widehat{W}_n > -\log \alpha\}$ and $N_0 = \inf\{n : \widehat{W}_n < \log \beta\}$. Then $N = \min(N_0, N_1)$. For $i = 0, 1$ and any $\delta > 0$ and ϵ_i,

$$P_i\big[|\frac{\widehat{W}_n}{n} - \epsilon_i| > \delta\big] \leq P_i\big[|\frac{\widehat{W}_n}{n} - \frac{\mathcal{G}_n^i}{n}| > \frac{\delta}{2}\big] + P_i\big[|\frac{\mathcal{G}_n^i}{n} - \epsilon_i| > \frac{\delta}{2}\big]. \tag{8}$$

(a) Since $P_0(N < \infty) \geq P_0(N_0 < \infty)$, it is sufficient to prove $P_0(N_0 < \infty) = 1$. Consider (8). Let $\epsilon_0 = -\lambda/2$. As $n \rightarrow \infty$ the first term in the R.H.S. of (8) goes to zero because of the assumption and obviously the second term also goes to zero. Thus, $\widehat{W}_n/n \rightarrow \epsilon_0$ in probability. Therefore,

$$\begin{aligned} \lim_{n\rightarrow\infty} P_0[N_0 \leq n] &\geq \lim_{n\rightarrow\infty} P_0[\widehat{W}_n < \log \beta] \\ &= \lim_{n\rightarrow\infty} P_0[\frac{\widehat{W}_n}{n} < \frac{\log \beta}{n}] = 1. \end{aligned}$$

We have,

$$P_0[N_0 < \infty] = \lim_{n\rightarrow\infty} \sum_{k=1}^{n} P_0[N_0 = k] = \lim_{n\rightarrow\infty} P_0[N \leq n]. \tag{9}$$

This implies $P_0[N_0 < \infty] = 1$.

(b) Since $P_1(N < \infty) \geq P_1(N_1 < \infty)$, it is sufficient to prove $P_1(N_1 < \infty) = 1$. Here $\epsilon_1 = \lim_{n\rightarrow\infty} n^{-1}\mathcal{G}_n^1 > 0$ a.s. . When $n \rightarrow \infty$ the first term in the R.H.S. of (8) approaches zero because of the assumption and the second term also goes to zero by strong law of large numbers. Thus,

$$\begin{aligned} \lim_{n\rightarrow\infty} P_1[N_1 \leq n] &\geq \lim_{n\rightarrow\infty} P_1[\widehat{W}_n > -\log \alpha] \\ &= \lim_{n\rightarrow\infty} P_1[\frac{\widehat{W}_n}{n} > \frac{-\log \alpha}{n}] = 1. \end{aligned}$$

This gives $P_1[N_1 < \infty] = 1$ as in (9).

(c) We have,

$$\begin{aligned} P_{FA} &= P_0(N_1 < N_0) \leq P_0(N_1 < \infty), \\ &\stackrel{(a)}{\leq} \frac{\alpha}{2^{\lambda/2} - 1} \leq \alpha, \end{aligned} \tag{10}$$

where (a) is given by the [9, Lemma 2]. ∎

The assumption $n^{-1}|\widehat{W}_n - \mathcal{G}_n^i| \rightarrow 0$ in probability has been shown to be true for i.i.d. sequences for the two universal source codes LZ78 ([12]) and KT-estimator with Arithmetic encoder ([24] with the redundancy property of Arithmetic Encoder [4]) considered later in this section. However, another desirable property $P_{MD} \triangleq P_1[\widehat{W}_N \leq \log \beta] \leq \beta$ seems to require extra conditions. Indeed we have seen via simulations that it holds for KT-estimator with the Arithmetic Encoder but not for LZ78. Also we have observed from our simulations that (6) and (7) hold for both the encoders with appropriate modification in the denominator of R.H.S. as $D(P_1||P_0) - \lambda/2$ and $\lambda/2$ respectively (this corresponds to the absolute value of the expected drift under H_1 and H_0). These results are not reported due to lack of space.

Design parameter α is chosen so as to meet a P_{FA} requirement for both the aforementioned universal source codes and β as P_{MD} to be achieved for KT-estimator with Arithmetic Encoder. By (5), λ can be chosen as the minimum Kullback-Leibler divergence, which is related to the minimum SNR under consideration.

A. Continuous Alphabet

The above test can be extended to continuous alphabet sources. 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, it is a.s. close to Z_n/n for all large n. 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 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 from X_1^Δ to X_n^Δ be $X_{1:n}^\Delta$. We know that $H(X_1^\Delta) + \log \Delta \rightarrow h(X_1)$ as $\Delta \rightarrow 0$ ([4]). Given i.i.d. observations $X_1^\Delta, X_2^\Delta, \ldots, X_n^\Delta$, 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),

$$\widetilde{W}_n = -L_n(X_{1:n}^\Delta) - n \log \Delta - \sum_{k=1}^{n} \log f_0(X_k) - n\frac{\lambda}{2} \tag{11}$$

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

$$\mathcal{C} = \{f_1 : D(f_1||f_0) \geq \lambda\}. \tag{12}$$

The following comments justify the above quantization.

1) It is known that uniform scalar quantization with variable-length coding of n successive quantizer outputs achieves

the optimal operational distortion rate function for quantization at high rates and even for low rates ([6]). This further justifies the development of our algorithm.

2) An adaptive uniform quantizer, which is changing at each time step, makes the scalar quantized observations dependent (learning from the available data at that time) and non-identically distributed. This makes the universal codelength function unable to learn the underlying distribution.
3) Non-uniform partitions with width Δ_j at j^{th} bin and with probability mass p_j require knowledge of p_j which is unknown under H_1.
4) Assuming we have i.i.d observations, uniform quantization has another advantage. (11) can be written as

$$-L_n(X_{1:n}^{\Delta}) - \sum_{k=1}^{n} \log(f_0(X_k)\Delta) - n\frac{\lambda}{2}.$$

Under the high rate assumption, $f_0(X_i)\Delta \approx p_0(X_i^{\Delta})$ (p_0 is the probability mass function after quantizing f_0). Thus, $\widetilde{W}_n = -L_n(X_{1:n}^{\Delta}) - \sum_{j=1}^{n} \log p_0(X_j^{\Delta}) - n\lambda/2$. This test entirely depends upon the quantized observations which is not possible for non-uniform quantization.
5) The range of the quantization can be fixed by considering only those f_1's whose tail probabilities are less than a small specific value at a fixed boundary and use these boundaries as range.

We could possibly approximate differential entropy $h(X_1)$ by universal lossy coding algorithms ([3], [8]). But these algorithms require a large number of samples (more than 1000) to provide a reasonable approximation. In our application we are interested in minimising the expected number of samples in a sequential setup. Thus, we found the algorithms in [3] and [8] inappropriate for our applications.

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

In the following we use Lempel-Ziv incremental parsing technique LZ78 ([26]), which is a well known efficient algorithm, for universal source coding in (3). This algorithm parses the input string into phrases, where each phrase is the shortest phrase not seen earlier and code each phrase by giving the location of the prefix of the phrase and the value of the latest symbol. We call this algorithm LZSLRT. [10] contains extensive simulation results of this algorithm for different distributions. Comparisons with a nearly optimal sequential GLR test ([15]) and extension to decentralized case are also considered in the same paper. It has been observed that the test is better than the GLR test for some classes of distributions.

At low n, which is of interest in sequential detection, the approximation for the log likelihood function via LZSLRT is usually poor as universal coding requires a few samples to learn the source. Hence we add a correction term $n\epsilon_n$, in the likelihood sum in (11), where ϵ_n is the redundancy for universal lossless codelength function. It is shown in [11], that

$$L_n(X_{1:n}^{\Delta}) \leq n\tilde{H}_n(X_1^{\Delta}) + n\epsilon_n, \tag{13}$$

where

$$\epsilon_n = C\left(\frac{1}{\log n} + \frac{\log\log n}{n} + \frac{\log\log n}{\log n}\right). \tag{14}$$

Here C is a constant which depends on the size of the quantized alphabet and $\tilde{H}_n(X_1^{\Delta})$ is the empirical entropy, which is the entropy calculated using the empirical distribution of samples upto time n.

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

In this section we propose KTSLRT for i.i.d. sources. The codelength function in KTSLRT comes from the combined use of KT (Krichevsky-Trofimov) estimator ([13]) and the Arithmetic Encoder ([4]). We will show that the test obtained via this universal code often substantially outperforms LZSLRT.

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

$$P_c(x_1^n) = \prod_{t=1}^{n} \frac{v(x_t/x_1^{t-1}) + \frac{1}{2}}{t - 1 + \frac{|A|}{2}}, \tag{15}$$

where $v(i/x_1^{t-1})$ denotes the number of occurrences of symbol i in x_1^{t-1} and $|A|$ is the alphabet size. It is known ([4]) 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 $L_n(x_1^n) < -\log P_c(x_1^n) + 2$. It is proved in [5] that universal codes defined by the KT-estimator with the Arithmetic Encoder are nearly optimal in the sense that the worst case maximum redundancy of this code achieves the lower bound. Writing (15) in a sequentially updating fashion, (11) can be modified as

$$\widetilde{W}_n^{KT} = \widetilde{W}_{n-1}^{KT} + \log\left(\frac{v(X_n^{\Delta}/X_1^{\Delta n-1}) + \frac{1}{2} + S}{t - 1 + \frac{|A|}{2}}\right) - \log p_0(X_n^{\Delta}), \tag{16}$$

where S is a scalar constant whose value greatly changes the performance. The default value of S is zero.

IV. Performance Comparison

We now provide the performance of KTSLRT via simulations and compare with LZSLRT. We also compare it with some other estimators available in literature. It has been observed from our initial experiments that 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 $E_{H_1}[N]$ versus P_{MD} and $E_{H_0}[N]$ versus P_{FA} separately, we plot E_{DD} $(0.5E_{H_1}[N] + 0.5E_{H_0}[N])$ versus P_E $(0.5P_{FA} + 0.5P_{MD})$. We use an eight bit uniform quantizer.

Figure 1 shows the Gaussian case when $f_1 \sim \mathcal{N}(0, 5)$ and $f_0 \sim \mathcal{N}(0, 1)$, where $\mathcal{N}(a, b)$ denotes the Gaussian distribution with mean a and variance b. We observe that LZSLRT and KTSLRT with $S = 0$ (the default case) are not able to give 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 that $S = 0$ case performs much worse than $S = 1$ and hence in the following we consider KTSLRT with $S = 1$ only. Next we provide comparison for two heavy tail distributions.

Figure 2 displays the Lognormal distribution comparison when $f_1 \sim ln\mathcal{N}(3, 3)$, $f_0 \sim ln\mathcal{N}(0, 3)$ and $ln\mathcal{N}(a, b)$ indicates the density function of Lognormal distribution with the

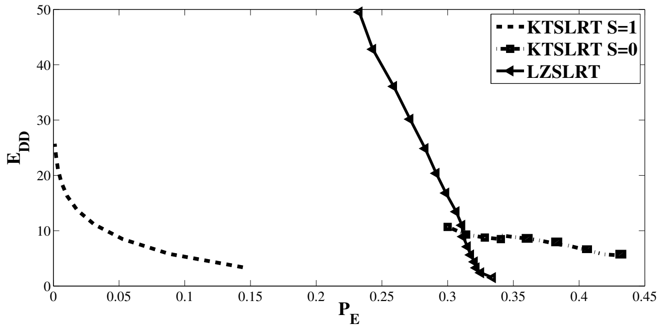

Fig. 1. Comparison between KTSLRT and LZSLRT for Gaussian Distribution

underlying Gaussian distribution $\mathcal{N}(a, b)$. It can be observed that P_E less than 0.1 is not achievable by LZSLRT. KTSLRT with $S = 1$ provides a good performance.

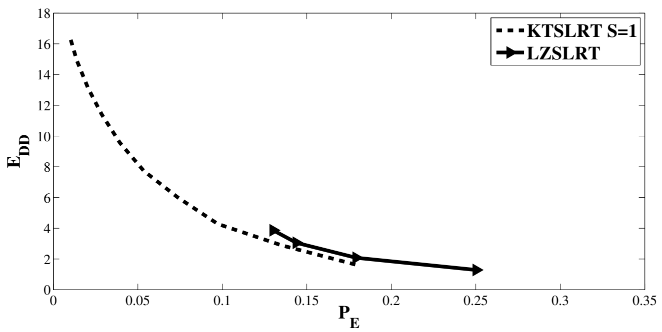

Fig. 2. Comparison between KTSLRT and LZSLRT for Lognormal Distribution

Figure 3 shows the results for Pareto distribution. Here $f_1 \sim \mathcal{P}(3, 2)$, $f_0 \sim \mathcal{P}(10, 2)$ and support set $(2, 10)$, where $\mathcal{P}(K, \alpha_m)$ is the Pareto density function with K and α_m as the shape and scale parameter of the distribution. We observe that KTSLRT with $S = 1$ and LZSLRT have comparable performance.

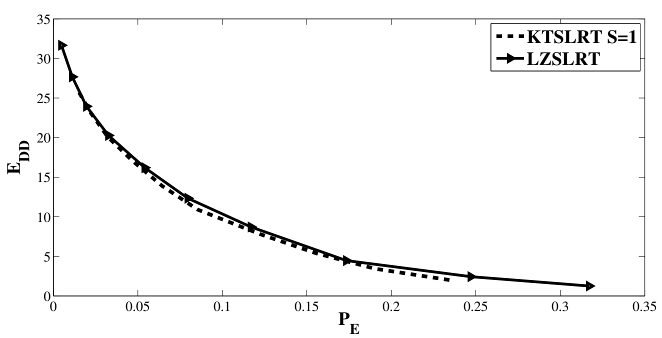

Fig. 3. Comparison between KTSLRT and LZSLRT for Pareto Distribution

It is observed by us that as S increases, till a particular value the performance of KTSLRT improves and afterwards it starts to deteriorate. For all the examples we considered, $S = 1$ provides good performance.

In Figure 4 we compare KTSLRT with the sequential tests defined by replacing $\sum_{k=1}^{n} \log f_1(X_k)$ by $-n\hat{h}_n$ 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 [23] that 1NN (1st Nearest Neighbourhood) differential entropy estimator performs better than other differential entropy estimators where 1-NN differential entropy estimator is

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

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

There are many density estimators available ([21]). 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 ([21]). The Kernel density estimator at a point z is

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

where K is the kernel and w_n is the bandwidth. See [21] for optimal K and w_n. We provide the comparison of KTSLRT with these two schemes in Figure 4. We find that KTSLRT with $S = 1$ performs the best.

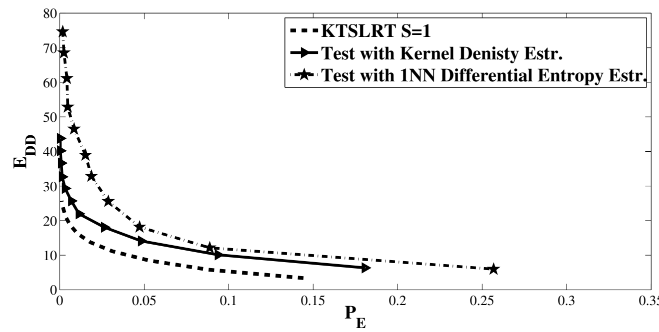

Fig. 4. Comparison among KTSLRT, universal sequential tests using 1NN differential entropy estimator and that using Kernel density estimator

V. Decentralized Detection

Motivated by the satisfactory performance of a single node case, we extend LZSLRT to the decentralized setup in [10]. In this setup we consider a CR network with one fusion center (FC) and L CRs. The CRs use local observations to make local decisions about the presence of a primary and transmit them to the FC. The FC makes the final decision on the local decisions it received.

Let $X_{k,l}$ be the observation make at CR l at time k. We assume $\{X_{k,l}, k \geq 1\}$ are independent and identically distributed (i.i.d.) and that the observations are independent across CRs. 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 the detection algorithm based 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 CRs and FC in which the FC receives Y_k, a coherent superposition of the CR 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). 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. Thus we use LZSLRT at each local node and Wald's SPRT at the fusion center (we call it LZSLRT-SPRT). Similarly we can use KTSLRT at each CR and SPRT at the fusion center and call it KTSLRT-SPRT. In both the cases whenever the CR stopping time is reached, it transmits b_1 if its decision is H_1, otherwise b_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 upper threshold under H_1 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 [10] which runs SPRT at CRs and FC and hence requires knowledge of $f_{1,l}$ at CR l. DualSPRT is known to be asymptotically optimal. We choose $b_1 = 1$, $b_0 = -1$, $I = 2$, $L = 5$ and $Z_k \sim \mathcal{N}(0, 1)$ and assume same SNR for all the CRs to reduce the complexity of simulations. We use eight bit quantizer in all these experiments. Figure 5 is the Gaussian example. The simulation setup has $f_{0,l} \sim \mathcal{N}(0, 1)$ and $f_{1,l} \sim \mathcal{N}(0, 5)$, for $1 \leq l \leq L$. The setup for Figure 6 is $f_{0,l} \sim \mathcal{P}(10, 2)$ and $f_{0,l} \sim \mathcal{P}(3, 2)$, for $1 \leq l \leq L$. FC thresholds are chosen appropriately with the available expressions for SPRT.

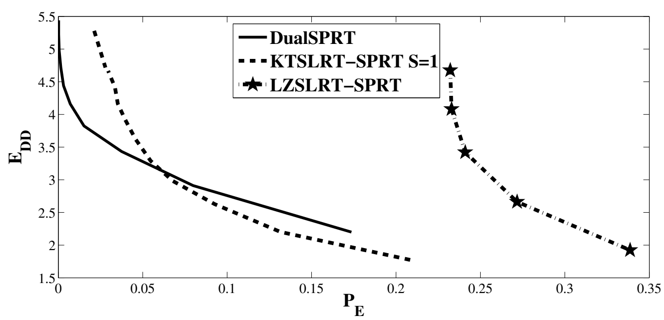

Fig. 5. Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Gaussian Distribution

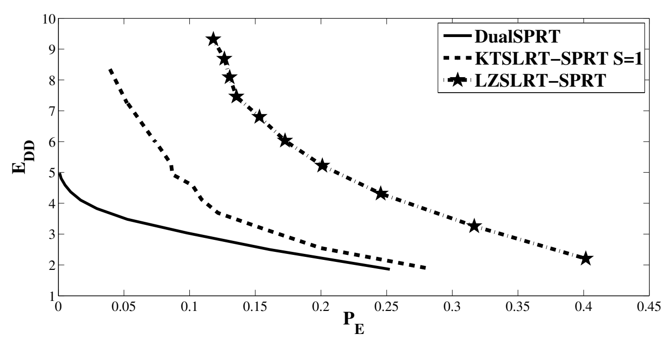

Fig. 6. Comparison among DualSPRT, KTSLRT-SPRT and LZSLRT-SPRT for Pareto Distribution

VI. Conclusions

In this paper we have presented a novel algorithm for spectrum sensing. We start with a universal sequential testing spectrum sensing framework where the CRs do not have any knowledge about the distribution (not even parametric family) when the primary transmits. This setup covers uncertainty in the SNR at CR receivers and fading channels between primary and CR. We propose a simple test using universal lossless codes. Our algorithm can be used for continuous and discrete distributions. We have compared our algorithms when the lossless codes are Lempel-Ziv codes and KT-estimator with Arithmetic Encoder. Finally we have extended these algorithms to distributed cooperative setup.

References

[1] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40 – 62, 2011.

[2] T. Banerjee, V. Sharma, V. Kavitha, and A. K. JayaPrakasam, "Generalized analysis of a distributed energy efficient algorithm for change detection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 91 –101, Jan 2011.

[3] D. Baron and T. Weissman, "An MCMC approach to lossy compression of continuous sources," in *Proc. Data Compression Conference (DCC)*, March 2010, pp. 40 –48.

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

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

[6] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325 –2383, Oct 1998.

[7] P. J. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.

[8] E. hui Yang, Z. Zhang, and T. Berger, "Fixed-slope universal lossy data compression," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1465 –1476, Sep 1997.

[9] 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.

[10] K. S. Jithin and V. Sharma, "Novel algorithms for distributed sequential hypothesis testing," in *Proc. 49th Annual Allerton Conference on Communication, Control, and Computing*, Sep 2011, pp. 1529 –1536.

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

[12] 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.

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

[14] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2008, pp. 1 –5.

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

[16] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2215 – 2229, Aug 2002.

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

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

[19] B. Ryabko, "Applications of universal source coding to statistical analysis of time series," *Selected Topics in Information and Coding Theory, World Scientific Publishing*, pp. 289–338, 2010.

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

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

[22] 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.

[23] 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.

[24] 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.

[25] J. Ziv, "On classification with empirically observed statistics and universal data compression," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 278 –286, Mar 1988.

[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.

[27] Q. Zou, S. Zheng, and A. H. Sayed, "Cooperative sensing via sequential detection," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6266 –6283, Dec 2010.