Novel Algorithms for Distributed Sequential Hypothesis Testing

Jithin K S and Vinod Sharma

Department of Electrical Communication Engineering

Indian Institute of Science

Bangalore 560012, India

Email: {jithin,vinod}@ece.iisc.ernet.in

Abstract—This paper considers sequential hypothesis testing in a decentralized framework. We start with two simple decentralized sequential hypothesis testing algorithms. One of which is later proved to be asymptotically Bayes optimal. We also consider composite versions of decentralized sequential hypothesis testing. A novel nonparametric version for decentralized sequential hypothesis testing using universal source coding theory is developed. Finally we design a simple decentralized multihypothesis sequential detection algorithm.

Keywords- Distributed Detection, Sequential Hypothesis Testing, Universal Source Coding, Multihypothesis testing.

I. INTRODUCTION

Distributed detection has been quite popular recently due to its relevance to distributed radar, sensor networks [3], distributed databases and cooperative spectrum sensing in Cognitive radios ([11], [22]).

Distributed detection can be either decentralized or centralized. In the centralized framework, the information received by the sensors are transmitted directly, without doing any processing, to the fusion center. In decentralized detection each sensor sends a summarized or quantized information, which can be a local decision also, to the fusion center. Fusion center ultimately decides upon which hypothesis is true. The later one is more efficient in practical applications as we usually have bandwidth and power constraints at each local node [23]. A drawback of a decentralized scheme is that the fusion center's decision is based on less information. Hence the main challenge of decentralized detection algorithms is to provide a reliable decision with this information. Detection in sensor networks, in particular, is usually based on local node and fusion node detection policies and the type of feedback from the fusion node to the sensors. The main resource constraints for decentralized detection schemes include number of nodes, finite alphabet constraint on output of each sensor, limited spectral bandwidth, total cost of the system and stringent power requirements.

Static or dynamic stopping time is an issue of interest in distributed detection. Static stopping is based on fixed sample size (FSS) detection. In such scenarios, the likelihood ratio test on the received data will minimise the probability of error at the fusion center for a binary hypothesis testing problem. Hence the real problem in a FSS case is to decide the type of information each sensor should send to the fusion center. Interestingly likelihood tests at the sensor nodes are optimal whenever the observations are conditionally independent, given the hypothesis.

Dynamic or sequential case focusses on decentralized schemes where information comes sequentially to the sensors. Sequential detectors can detect change or test hypothesis. It is well known that in case of a single node, Wald's Sequential Probability Ratio Test (SPRT) outperforms other sequential or fixed sample size detectors [7]. In sequential decentralized framework, optimisation needs to be performed jointly over sensors and fusion center policies as well as over time. Decentralized sequential change detection has been discussed extensively in ([1], [20]). In this paper we focus on decentralized sequential hypothesis testing. Unfortunately, this problem is intractable for most of the sensor configurations ([15], [21]). Specifically there is no optimal solution available for sensor configurations with no feedback from fusion center and limited local memory, which is more relevant in practical situations. Recentaly [15] and [6] proposed asymptotically optimal (order 1 and order 2 respectively) decentralized sequential hypothesis tests for such systems with full local memory. But these models do not consider noise at the fusion center and assume a perfect communication channel between the sensor nodes and the fusion center.

We propose a decentralized sequential hypothesis testing algorithm where SPRT is used at both the secondary nodes and the fusion center. This algorithm is called DualSPRT. We prove DualSPRT is asymptotically Bayes. Although DualSPRT performs asymptotically, it is not optimal to use SPRT at the fusion center. Thus we improve over DualSPRT with a modification at the fusion center motivated by CUSUM [11]. Furthermore we introduce a new way of quantizing SPRT decisions of local nodes. We call this algorithm SPRT-CSPRT.

Composite sequential hypothesis testing where there is an uncertainty in some parameters of the assumed distribution is surveyed in ([7], [14]). A unified asymptotically optimal solution, applicable to both cases of with and without indifference zone separating one-sided hypothesis, is provided for exponential family in [13]. We use it at the local nodes and SPRT or CSPRT (fusion center policy of SPRT-CSPRT) at the fusion center. We show that this modification (GLR-SPRT or GLR-CSPRT) works well.

Nonparametric sequential problems for location testing are well documented in [7]. But we focus on universal hypothesis testing, where the distribution under null hypothesis is known, but not under the alternate hypothesis. [25] studies classification of finite alphabet sources using universal coding. [9] considers universal hypothesis testing problem in 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. But in practical applications often the distributions under the two hypothesis have continuous alphabet (e.g., Gaussian noise is ubiquitous). [16] 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. But the author did not consider how to partition the alphabet and proved the result on the assumption of existence of partitions. In this paper we use universal source coding on the problem of distributed sequential hypothesis testing when the alphabet is continuous. For this we use uniform scalar quantizer with universal source coding algorithms (e.g., Lempel Ziv [26]) at the local nodes to approximate the likelihood ratio under alternate (unknown) hypothesis. We show that this universal test works quite favourably compared to other tests.

In summary, this paper makes the following contributions in decentralized hypothesis testing problem. First it summarises our recent asymptotically optimal tests when there is noise at the fusion center. Next it develops new universal distributed algorithms using universal source coding. Finally it develops a simple distributed sequential multihypothesis test from existing algorithms to provide an improved practical algorithm.

This paper is organised as follows. Section II describes the problem. Section III considers parametric decentralized sequential hypothesis testing algorithms. This section starts with the DualSPRT algorithm. Then we provide SPRT-CSPRT. We compare their performance with some asymptotically optimal tests. Later we prove the asymptotic optimality of DualSPRT. Next we consider parametric uncertainty. In Section IV we introduce nonparametric decentralized sequential hypothesis testing algorithms using universal source coding. Section V focusses on multihypothesis case. Section VI concludes the paper.

II. MODEL

Consider a sensor network with one fusion center and L sensors (nodes). The L sensors sense the environment to detect the hypothesis if a signal is present or not. The local decisions made by the sensors are transmitted to a fusion node via a multiple access channel for it to make the final decision. There is no feedback from the fusion center. We allow the possibility that the local nodes may do some processing on the data and transmit to the fusion node without necessarily make a decision.

Let $X_{k,l}$ be the observation made at sensor l at time k. We assume that $\{X_{k,l}, \ k \geq 1\}$ are independent and identically distributed (i.i.d.) and that the observations are independent across sensors. Using the detection algorithm based on $\{X_{n,l}, \ n \leq k\}$ the sensor l transmits $Y_{k,l}$ to the fusion node at time k. We assume that the sensors are synchronised so that the fusion node receives $Y_k = \sum_{l=1}^{L} Y_{k,l} + Z_k$, where $\{Z_k\}$ is i.i.d. zero mean Gaussian receiver noise with variance σ^2 (for our algorithms Gaussian assumption is not required). The fusion center observes $\{Y_k\}$ and decides upon the hypothesis.

The observations $\{X_{k,l}\}$ depend on whether the true hypothesis is hypothesis H_1 or hypothesis H_0:

$$X_{k,l} = \begin{cases} Z_{k,l}, & k = 1, 2, \ldots, \text{ under } H_0, \\ h_l S_k + Z_{k,l}, & k = 1, 2, \ldots, \text{ under } H_1, \end{cases} \quad (1)$$

where h_l is the channel gain of the lth sensor, S_k is the signal and $Z_{k,l}$ is the noise at the lth sensor at time k. We assume $\{Z_{k,l}, k \geq 1\}$ are i.i.d. We will denote by $f_{1,l}$ and $f_{0,l}$ the densities of $X_{k,l}$ under H_1 and H_0 respectively. The fusion center makes a decision at a random time N. We assume that N is much less than the coherence time of the channel so that the slow fading assumption is valid. This means that h_l is random but remains constant during the sensing duration.

The general problem is to develop a distributed algorithm in the above setup which solves the problem:

$$\min E_{DD} \triangleq E[N|H_i] , \quad (2)$$

subject to $P_{FA} \leq \alpha$ and $P_{MD} \leq \beta$

where H_i is the true hypothesis, $i = 0, 1$ and P_{FA} and P_{MD} are the probability of false alarm and the probability of miss detection respectively, i.e., probability of making a wrong decision under H_0 and under H_1.

It is well known that for a single node case ($L = 1$) Wald's SPRT performs optimally for i.i.d. observations in terms of reducing $E[N|H_1]$ and $E[N|H_0]$ for a given P_{FA} and P_{MD}. If there is no communication cost or energy cost in transmitting data from the local nodes to the fusion node, then again we can reliably send data sensed by the local nodes to the fusion node and run SPRT at the fusion center. Otherwise, the optimal algorithm is not known. Asymptotically optimal algorithms for i.i.d. observations are recently proposed. But they do not take into account the fusion center noise and the uncertainties in the parameters of the distributions under H_0 and H_1. Motivated by the good performance of DualCUSUM (a decentralized sequential change detection algorithm which runs CUSUM at the local nodes and at the fusion center) in [1] and the optimality of SPRT for a single node, we propose DualSPRT (a decentralized sequential hypothesis testing algorithm which runs SPRT at local nodes and at the fusion center). Later on we will present an algorithm when there is uncertainties in the distribution.

III. PARAMETRIC DECENTRALIZED SEQUENTIAL HYPOTHESIS TESTING ALGORITHMS

We first present DualSPRT [10].

A. DualSPRT Algorithm

1) Secondary node, l, runs SPRT,

$$W_{0,l} = 0$$
$$W_{k,l} = W_{k-1,l} + \log\left[\frac{f_{1,l}(X_{k,l})}{f_{0,l}(X_{k,l})}\right], k \geq 1. \qquad (3)$$

2) Secondary node l transmits a constant b_1 at time k if $W_{k,l} \geq \gamma_{1,l}$ or transmits b_0 when $W_{k,l} \leq \gamma_{0,l}$, i.e.,

$$Y_{k,l} = b_1 1_{\{W_{k,l} \geq \gamma_{1,l}\}} + b_0 1_{\{W_{k,l} \leq \gamma_{0,l}\}}$$

where $\gamma_{0,l} < 0 < \gamma_{1,l}$ and 1_A denotes the indicator function of set A. Parameters $b_1, b_0, \gamma_1, \gamma_0$ are chosen appropriately.

3) Finally, fusion center runs SPRT:

$$F_k = F_{k-1} + \log\left[g_1(Y_k)/g_0(Y_k)\right], \quad F_0 = 0, \qquad (4)$$

where g_0 is the density of $Z_k + \mu_0$ and g_1 is the density of $Z_k + \mu_1$, μ_0 and μ_1 being design parameters such that F_k has a positive drift when atleast $L/2$ nodes transmit b_1 and negative drift if atleast $L/2$ nodes transmit b_0.

4) The fusion center decides about the hypothesis at time N where

$$N = \inf\{k : F_k \geq \beta_1 \text{ or } F_k \leq \beta_0\}$$

and $\beta_0 < 0 < \beta_1$. The decision at time N is H_1 if $F_N \geq \beta_1$; otherwise H_0.

Physical layer fusion reduces transmission time, but requires synchronisation of different local nodes. If synchronisation is not possible, then some other algorithm, e.g., TDMA can be used. This algorithm in the context of spectrum sensing in Cognitive Radio has been studied in our previous work [10]. Performance analysis and parameter uncertainty (unknown SNR and fading) in distributions at the local nodes were handled in the same paper.

B. SPRT-CSPRT Algorithm

In DualSPRT given above, observations to the fusion center are not always identically distributed. Till the first transmission from secondary nodes, these observations are i.i.d. $\sim \mathcal{N}(0, \sigma^2)$ where $\mathcal{N}(a, b)$ is the Gaussian pdf with mean a and variance b. But after the transmission from the first local node and till the transmission from the second node, they are i.i.d. Gaussian with another mean and but same variance σ^2. Thus the observations at the fusion center are no longer i.i.d. Since the optimality of SPRT is known for i.i.d. observations ([7]), DualSPRT is not optimal.

The following heuristic arguments provide the motivation of the proposed modifications to DualSPRT. If the SPRT sum defined in (4) goes below zero it delays in crossing the positive threshold β_1. Hence if we keep SPRT sum at zero whenever it goes below zero, it reduces E_{DD}. This happens in CUSUM [7]. Similarly one can use a CUSUM type algorithm under H_0. These arguments were verified via simulations and theory in [11]. Thus we obtain the following algorithm :

Steps (1)-(2) are same as in DualSPRT. The steps (3) and (4) are replaced by

3) Fusion center runs two algorithms:

$$F_k^1 = (F_{k-1}^1 + \log[g_1(Y_k)/g_0(Y_k)])^+ \qquad (5)$$
$$F_k^0 = (F_{k-1}^0 + \log[g_1(Y_k)/g_0(Y_k)])^-, \qquad (6)$$

$F_0^1 = 0, F_0^0 = 0$, where $(x)^+ = \max(0, x)$ and $(x)^- = \min(0, x)$.

4) The fusion center decides about the hypothesis at time N where

$$N = \inf\{k : F_k^1 \geq \beta_1 \text{ or } F_k^0 \leq \beta_0\}$$

and $\beta_0 < 0 < \beta_1$. The decision at time N is H_1 if $F_N^1 \geq \beta_1$; otherwise H_0.

Under H_1, (5) has a positive drift and hence it approaches the threshold β_1 quickly, but under H_0, (5) will most probably be hovering around zero. Similarly under H_0, (6) moves towards β_0, but under H_1 will be mostly around zero. This means that P_{FA} for this algorithm is expected to be less compared to DualSPRT.

We consider one more improvement. When a local SPRT sum crosses its threshold, it transmits b_1/b_0. This node transmits till the fusion center SPRT crosses the threshold. If it is not a false alarm, then its SPRT sum keeps on increasing (decreasing). But if it is a false alarm, then the sum will eventually move towards the other threshold. Hence instead of transmitting b_1/b_0 the local node can transmit a higher/lower value in an intelligent fashion. This should improve the performance. Thus we modify the step (2) in DualSPRT as follows. Secondary node l transmits a constant from $\{b_1^1, b_2^1, b_3^1, b_4^1\}$ at time k if $W_{k,l} \geq \gamma_1$ or transmits from $\{b_1^0, b_2^0, b_3^0, b_4^0\}$ when $W_{k,l} \leq \gamma_0$, as follows :

$$Y_{k,l} = \begin{cases} b_1^1 \text{ if } W_{k,l} \in [\gamma_1, \gamma_1 + \Delta_1), \\ b_2^1 \text{ if } W_{k,l} \in [\gamma_1 + \Delta_1, \gamma_1 + 2\Delta_1), \\ b_3^1 \text{ if } W_{k,l} \in [\gamma_1 + 2\Delta_1, \gamma_1 + 3\Delta_1), \\ b_4^1 \text{ if } W_{k,l} \in [\gamma_1 + 3\Delta_1, \infty), \\ b_1^0 \text{ if } W_{k,l} \in [\gamma_0, \gamma_0 - \Delta_0), \\ b_2^0 \text{ if } W_{k,l} \in [\gamma_0 - \Delta_0, \gamma_0 - 2\Delta_0), \\ b_3^0 \text{ if } W_{k,l} \in [\gamma_0 - 2\Delta_0, \gamma_0 - 3\Delta_0), \\ b_4^0 \text{ if } W_{k,l} \in [\gamma_0 - 3\Delta_0, -\infty). \end{cases} \qquad (7)$$

where Δ_1 and Δ_0 are the parameters to be tuned. The expected drift under H_1 (H_0) is a good choice for Δ_1 (Δ_0).

We call the algorithm with the above two modifications as SPRT-CSPRT (with 'C' as an indication about the motivation from CUSUM).

The theoretical analysis and extensive numerical experiments for SPRT-CSPRT are provided in [11].

C. Performance Comparison

In this section we compare the performance of DualSPRT, SPRT-CSPRT and the asymptotically optimal algorithms in [6] and [15]. The algorithm in [15] has been shown to be first order optimal and the algorithm in [6] is second order optimal. The simulation example shows a difference in their

performance at finite parameter values. The scenario in this example is motivated from Cognitive Radio systems ([10], [11]).

Throughout the paper we use $\gamma_1 = -\gamma_0 = \gamma$, $\beta_1 = -\beta_0 = \beta$ and $\mu_1 = -\mu_0 = \mu$ for simplicity. Parameters used for simulation are as follows: There are 5 nodes ($L = 5$). We take $f_{0,l} \sim \mathcal{N}(0,1)$, $f_{1,l} \sim \mathcal{N}(1,1)$ for $1 \leq l \leq L$ and $Z_k \sim \mathcal{N}(0,1)$. We also take $\{b_1^1, b_2^1, b_3^1, b_4^1\} = \{1, 2, 3, 4\}$, $\{b_1^0, b_2^0, b_3^0, b_4^0\} = \{-1, -2, -3, -4\}$ and $b_1 = -b_0 = 1$ (for Dual-SPRT). Parameters γ and β are chosen from a range of values to achieve a particular P_{FA}. Performance comparisons with the asymptotically optimal decentralized sequential algorithms which do not consider fusion center noise (DSPRT [6], Mei's SPRT [15]) are given in Figure 1. Note that DualSPRT and SPRT-CSPRT include fusion center noise. Here We find that the SPRT-CSPRT's performance is close to that of DSPRT and is better than DualSPRT and Mei's SPRT. Also, SPRT-CSPRT performs better than DualSPRT. This has happened even when we did not use fusion node noise with the algorithms [6] and [15].

We compared these algorithms with other system parameters and similar conclusions were drawn.
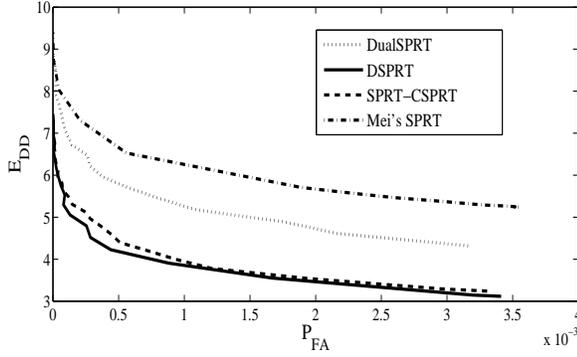

Fig. 1. Comparison among DualSPRT, SPRT-CSPRT, Mei's SPRT and DSPRT under H_1

D. Asymptotic optimality of DualSPRT

We have observed above that DualSPRT and SPRT-CSPRT perform close to or better than asymptotically optimal algorithms available which do not model fusion center noise. Now we prove asymptotic optimality of DualSPRT with the Bayes setup following the arguments in [15].

The two hypotheses H_0 and H_1 are assumed to have known prior probabilities π and $1 - \pi$ respectively. A cost c (≥ 0) is assigned to each time step taken for decision. Let $W_i > 0$, $i = 0, 1$ be the cost of falsely rejecting H_i. Then Bayes risk of DualSPRT (the stopping time is denoted by $N(c)$ and the Bayes test is denoted by δ) is

$$\mathcal{R}_c(\delta) = \pi[cE_0(N) + W_0 P_0\{reject\ H_0\}]$$
$$+ (1 - \pi)[cE_1(N) + W_1 P_1\{reject H_1\}] \quad (8)$$

where E_i denotes expectation and P_i denotes probability under H_i, $i = 0, 1$. We also use the following notation:

$$I(f_{1,l}, f_{0,l}) = \int \log\left(\frac{f_{1,l}(x)}{f_{0,l}(x)}\right) f_{1,l}(x)dx, \quad (9)$$

is the Kullback-Leibler (K-L) Divergence.

$$I_{tot} = \sum_{l=1}^{L} I(f_{0,l}, f_{1,l}), J_{tot} = \sum_{l=1}^{L} I(f_{1,l}, f_{0,l}),$$

$$r_l = \frac{I(f_{0,l}, f_{1,l})}{I_{tot}}, \rho_l = \frac{I(f_{1,l}, f_{0,l})}{J_{tot}}.$$

Theorem 1: Assume,
1) $\{X_{k,l}, k \geq 0\}$ is i.i.d. and independent of $\{X_{k,j}, k \geq 0\}$ for all $l \neq j$.
2) the following hold

$$\int \left(\log\left(\frac{f_{1,l}(x)}{f_{0,l}(x)}\right)\right)^2 f_{1,l}(x)dx < \infty$$

and

$$\int \left(\log\left(\frac{f_{0,l}(x)}{f_{1,l}(x)}\right)\right)^2 f_{0,l}(x)dx < \infty.$$

Then DualSPRT with local node thresholds $\gamma_{0,l} = -r_l|\log c|, \gamma_{1,l} = \rho_l|\log c|$ and fusion center thresholds $\beta_0 = -|\log c|, \beta_1 = |\log c|$, is asymptotically Bayes, i.e., $\lim_{c \to 0} \mathcal{R}_c(\delta^*)/\mathcal{R}_c(\delta) = 1$, where δ^* is the Bayes solution and δ indicates DualSPRT

Proof: See the Appendix. ∎

Remarks 1: As the cost c decreases, essentially we are allowing more samples for detection, which is captured in the modified expressions of $\gamma_{0,l}$, $\gamma_{1,l}$, β_0 and β_1.

E. Unknown Parameters

In this section we consider the setup when the distributions $f_{0,l}$ and $f_{1,l}$ belong to a parametric family with some uncertainty in the parameters.

In ([13], [14]) Lai has proposed a one sided sequential composite hypothesis test, which is asymptotically Bayes and also nearly optimal from the frequentist view point for testing one sided composite hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$, $\theta_0 < \theta_1$. The distributions under H_0 and H_1 belong to the exponential family of distributions and θ denotes a parameter for the family. We propose the following test in our distributed setup. The local nodes use Lai's algorithm while the fusion node runs the SPRT. We have made this modification for DualSPRT as well as for SPRT-CSPRT. We call the modified algorithm as GLR-SPRT and GLR-CSPRT respectively.

The sensor's hypothesis testing problem, likelihood ratio sum, stopping criteria and decision criteria are as follows. We consider

$$H_0 : \theta = \theta_0 \ ; \ H_1 : \theta \geq \theta_1 \ . \quad (10)$$

where $\theta_0 = 0$ and θ_1 is appropriately chosen. Then,

$$W_{n,l} = max\left[\sum_{k=1}^{n} \log \frac{f_{\hat{\theta}_n}(X_k)}{f_{\theta_0}(X_k)}, \sum_{k=1}^{n} \log \frac{f_{\hat{\theta}_n}(X_k)}{f_{\theta_1}(X_k)}\right] \ , \quad (11)$$

$$N = \inf\{n : W_{n,l} \geq g(cn)\}, \quad (12)$$

where $g(cn)$ is a time varying threshold and c is the cost assigned for each observation. An approximate expression for g is given in [13]. Also for Gaussian f_0 and f_1, $\theta \in [a_1, a_2]$, $S_n = \sum_{k=1}^n X_{k,l}$, $\hat{\theta}_n = max\{a_1, min[S_n/n, a_2]\}$. At time N decide upon H_0 or H_1 according as $\hat{\theta}_N \leq \theta^*$ or $\hat{\theta}_N \geq \theta^*$, where θ^* is obtained by solving $I(\theta^*, \theta_0) = I(\theta^*, \theta_1)$, and $I(\theta, \lambda)$ is the Kullback-Leibler information number, which is the K-L Divergence $I(f_\theta, f_\lambda)$ in (9). Here, as the threshold $g(cn)$ is a time varying and decreasing function, the quantisation (7) is changed in the following way: if $\hat{\theta}_N \geq \theta^*$

$$Y_{k,l} = \begin{cases} b_1^1 \text{ if } W_{k,l} \in [g(kc), g(kc3\Delta)), \\ b_2^1 \text{ if } W_{k,l} \in [g(kc3\Delta), g(kc2\Delta)), \\ b_3^1 \text{ if } W_{k,l} \in [g(kc2\Delta), g(kc\Delta)), \\ b_4^1 \text{ if } W_{k,l} \in [g(kc\Delta), \infty). \end{cases} \quad (13)$$

If $\hat{\theta}_N \leq \theta^*$ the local node will transmits from $\{b_1^0, b_2^0, b_3^0, b_4^0\}$ under the same conditions. Here Δ is a tuning parameter and $0 \leq 3\Delta \leq 1$. The choice of θ_1 in (10) affects the performance of $E[N|H_0]$ and $E[N|H_1]$ for the algorithm (11)-(12). As θ_1 increases, $E[N|H_0]$ decreases and $E[N|H_1]$ increases.

We have used this algorithm in energy detector setup ([10], [11]) to take care of the following scenarios: the received SNR at the local nodes and/or the channel gains h_l are not known.

Numerical results and our observations for GLR-SPRT and GLR-CSPRT are presented in ([10], [11]) and are shown to provide good performance.

IV. USING UNIVERSAL SOURCE CODING FOR NONPARAMETRIC TESTING

In this section we consider a completely nonparametric setup for f_1. For this we use universal data compression algorithms. There is comparatively little literature available on using universal data compression algorithms for the detection problem; especially sequential detection algorithms. Thus we first discuss the problem for a single node and then generalize to the decentralized setting. We will be mainly concerned with continuous alphabet observations because the receiver almost always has Gaussian noise. Of course, if our observations have discrete alphabets, the following algorithm will simplify.

A. Single node case

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 distribution P_0 (hypothesis H_0) or from distribution P_1 (hypothesis H_1). Let P_0 and P_1 have densities f_0 and f_1 with respect to some probability measure. We will assume that f_0 is known but f_1 is unknown. Of course if f_1 belongs to a parametric family with an unknown parameter θ then we can use Lai's GLR [13] which is optimal under some conditions. However, if we use nonparametric setup for f_1 then in the following we propose to use universal data compression algorithms.

The idea for the new algorithm is as follows. If we know f_1 we use SPRT:

$$N \triangleq \inf\{n : W_n = \sum_{k=1}^n \log \frac{f_1(X_k)}{f_0(X_k)} \notin (-\beta_0, \beta_1)\} \quad (14)$$

and if $W_N > \beta_1$ decide H_1; otherwise H_0. 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 in (14). 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_1 via a uniform quantizer with a quantization step $\Delta > 0$:

$$X_1^\Delta = [X_1/\Delta]\Delta, \text{ where } [X] \text{ is the integer part of } X \quad (15)$$

We know that $H(X_1^\Delta) + \log \Delta \to h(X_1)$ as $\Delta \to 0$ [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. In the following we use Lempel-Ziv incremental parsing algorithm LZ78 [26], which is a well known efficient algorithm. But any universal lossless codelength function satisfying $\lim_{n \to \infty} n^{-1} L_n = H(X_1^\Delta)$ a.s. can be used (L_n is the code length of first n quantized samples). Examples of such algorithms include Context Tree Weighting algorithm, Kolmogrov complexity and Rissanen stochastic complexity [24]. Specifically, for Lempel-Ziv algorithm (LZ78) $n^{-1} L_n \to H(X^\Delta)$ a.s for any stationary ergodic source $\{X_1^\Delta, X_2^\Delta, \ldots\}$ with a finite alphabet.

In (14) W_n has an expected drift of $I(f_1, f_0)$ under H_1 and $-I(f_0, f_1)$ under H_0. However if we replace Z_n in (14) by the approximation $-L_n - n \log \Delta$, then the drift of W_n under H_0 becomes zero. In order to make it negative, we add an extra term in the expression of W_n and define (14) by,

$$\hat{W}_n = -L_n - n \log \Delta - \sum_{k=1}^n \log f_0(X_k) - n\frac{\lambda}{2} \quad (16)$$

where λ is a positive constant. To get some performance guarantees, we limit to a class of densities f_1:

$$\mathcal{C} = \{f_1 : I(f_1, f_0) \geq \lambda\} \quad (17)$$

Then under H_0, $\hat{W}_n$ has a negative drift of $\lambda/2$ and under H_1, in the worst case $\hat{W}_n$ has a positive drift of $\lambda/2$. Instead of $\lambda/2$, we can be less conservative and choose a larger fraction of λ. We call this algorithm as LZSLRT (Lempel-Ziv Sequential Likelihood Ratio Test).

We compare LZSLRT obtained with SPRT and GLR-Lai [13] for different f_1 and f_0 in Table I and Table II for Gaussian and Pareto distributions respectively. The experimental set up for Table I is, $f_0 \sim \mathcal{N}(0, 5)$ and $f_1 \sim \mathcal{N}(3, 5)$. $\Delta = 0.3125$. The setup for Table II is, $f_0 \sim \mathcal{P}(10, 2)$ 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. $\Delta = 0.3125$. Here P_E represents P_{FA} when the actual hypothesis is H_0 and P_{MD} when it is H_1. We observe that although LZSLRT performs worse for Gaussian distribution, it works better than GLR-Lai for the Pareto Distribution. We

have made comparisons for other distributions also and found that LZSLRT compares favourably with GLR-Lai.

This motivates us to use LZSLRT for the decentralized setup in the next section.

We make a few observations on LZSLRT. Although LZSLRT is universal, we need to fix the parameter Δ and upper and lower truncation of support of f_1 and f_0 to obtain a finite alphabet. At low n, which is of interest in sequential detection, the approximation for the likelihood function is usually bad as universal lossless coding requires a few samples to learn the source (as is of course needed by any learning algorithm including the test in [13]). Hence we add $n\epsilon_n$ in (16), where ϵ_n is the redundancy for the universal lossless code length function. It can be shown that for Lempel-Ziv coding [12],

$$L_n \leq nH(X_1^\Delta) + n\epsilon_n \tag{18}$$

where

$$\epsilon_n = C\left(\frac{1}{\log n} + \frac{\log\log n}{n} + \frac{\log\log n}{\log n}\right) \tag{19}$$

Here C is a constant which depends on the size of quantized alphabet. The dominating term in the ϵ_n is the third term. Hence we use $L_n - C(\log\log n)(\log n)^{-1}$ instead of L_n in (16).

We could possibly approximate differential entropy $h(X_1)$ by universal lossy coding algorithms ([2], [24]). 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 ([2], [24]) inappropriate for our applications. It is known [8] that uniform scalar quantization with variable-length coding of n successive quantizer outputs achieves the optimal operational distortion rate function for quantizer at high rates and even for low rates. This further justifies the development of our algorithm although in our setup rate is not critical.

Hyp	E_{DD}	$P_E = 0.05$	$P_E = 0.01$	$P_E = 0.005$
H_1	SPRT	3.18	4.56	6.34
H_1	GLR-Lai	4.8	8.59	12.17
H_1	LZSLRT	12.3	14.78	18.9
H_0	SPRT	3.23	4.62	6.23
H_0	GLR-Lai	5.18	8.46	13.49
H_0	LZSLRT	13.6	15.6	19.67

TABLE I
COMPARISON AMONG SPRT, GLR-LAI AND USC-SLRT FOR GAUSSIAN DISTRBUTION

Hyp	E_{DD}	$P_E = 0.05$	$P_E = 0.01$	$P_E = 0.005$
H_1	SPRT	3.45	6.86	14.23
H_1	GLR-Lai	10.23	13.78	19.37
H_1	LZSLRT	9.6	11.79	19.31
H_0	SPRT	3.78	5.9	8.92
H_0	GLR-Lai	11.45	14.5	21.56
H_0	LZSLRT	10.8	14.23	19.6

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

B. Decentralized case

Motivated by satisfactory performance of a single node case, we extend LZSLRT to the decentralized setup. Now our assumptions are that at local nodes, $f_{0,l}$ is known but $f_{1,l}$ is not known and the fusion node only observes Gaussian data (this setup is amply justified in sensor network and Cognitive Radio applications). Thus we use LZSLRT at each local node and Wald's SPRT at the fusion center (we call it LZSLRT-SPRT). An advantage of the decentralized (multiple nodes) is the reduction in the detection delay compared to a single node for the same P_{FA} and P_{MD}. This is usually attained by reducing the local node thresholds, which imparts high P_{FA} and P_{MD} in the single node case. The increase in the P_{FA} and P_{MD} in the decentralized case is mitigated at the fusion center. In our case it has been observed that for lower thresholds, we have to use all the three terms in (19) to obtain a good approximation of entropy.

We compare decentralized LZSLRT (LZSLRT-SPRT) with DualSPRT and GLRSPRT in Tables III and IV. The experimental set up for Table III is, $f_{0,l} \sim \mathcal{N}(0,5)$ and $f_{1,l} \sim \mathcal{N}(3,5)$, for $1 \leq l \leq L$, $L = 5$, $Z_k \sim \mathcal{N}(0,1)$. $\Delta = 0.3125$. The setup for Table IV is, $f_{0,l} \sim \mathcal{P}(10,2)$ and $f_{1,l} \sim \mathcal{P}(3,2)$, for $1 \leq l \leq L$, $L = 5$ 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. $\Delta = 0.3125$. Fusion center noise has distribution $\mathcal{N}(0,1)$. Here P_E represents P_{FA} when the actual hypothesis is H_0 and P_{MD} when it is H_1. As for the single node case, LZSLRT-SPRT performs worse than GLR-SPRT for the Gaussian case but better than GLR-SPRT for the Pareto case.

Hyp	E_{DD}	$P_E = 0.01$	$P_E = 0.005$	$P_E = 0.001$
H_1	DualSPRT	2.6	3.93	4.47
H_1	GLR-SPRT	4.38	6.4	10.89
H_1	LZSLRT-SPRT	9.4	11.53	16.03
H_0	DualSPRT	3.1	3.9	4.42
H_0	GLR-SPRT	5.9	5.86	11.78
H_0	LZSLRT-SPRT	10.44	12.67	16.33

TABLE III
COMPARISON AMONG DUALSPRT, GLR-SPRT AND LZSLRT-SPRT FOR GAUSSIAN DISTRBUTION

Hyp	E_{DD}	$P_E = 0.01$	$P_E = 0.005$	$P_E = 0.001$
H_1	DualSPRT	2.3	2.64	3.05
H_1	GLR-SPRT	9.26	11.8	18.24
H_1	LZSLRT-SPRT	8.01	9.37	15.23
H_0	DualSPRT	2.89	3.4	4.49
H_0	GLR-SPRT	10.79	15.88	20.39
H_0	LZSLRT-SPRT	9.38	13.44	16.11

TABLE IV
COMPARISON AMONG DUALSPRT, GLR-SPRT AND LZSLRT-SPRT FOR PARETO DISTRBUTION

V. DECENTRALIZED MULTIHYPOTHESIS SEQUENTIAL HYPOTHESIS TESTING

Consider the problem of decentralized sequential multihypothesis testing with $M > 2$ hypothesis and with no feedback from the fusion center. Let the hypotheses be $H_m : X_{k,l} \sim f_l^m$, $m = 0, \ldots, M-1$ where l is the sensor index and k is the time index,

There has been some work on a single node multihypothesis sequential testing problem both in the Bayesian [5] as well as non-Bayesian ([7], [17], [18]) framework. In [19] decentralized multihypothesis sequential testing problem is considered. The authors use a test at each local node, which is provided in [17] and at the fusion center they use a test loosely based on a method in [18]. We have found through numerical experiments that this distributed test requires a very large number of samples to make a decision. This motivates us to provide a simple and practically relevant distributed algorithm for multihypothesis sequential testing.

We propose a simple modification of *Test-D1* in [17] and for ease of reference we call this modification as *MTest-D1*. The test, *MTest-D1*, is as follows. Define

$$\mathcal{Z}_l^{i,j}(k) = \max\left(\sum_{t=1}^{k} \log \frac{f_l^i(X_{t,l})}{f_l^j(X_{t,l})}, 0\right). \quad (20)$$

The stopping time at local node l is,

$$N_l = \inf\{k : \mathcal{Z}_l^{i,j}(k) > A \text{ for all } j \neq i \text{ and some } i\} \quad (21)$$

where A is an appropriately chosen constant. At time N_l, node l makes the decision $d_l = i$ where i is given in (21).

The modification in the decision at node l compared to *Test-D1* is using a reflected random walk in (20) instead of random walk in *Test-D1*. We use this modified test at the local nodes. Local node l transmits a value b_i, when $d_l = i$, to the fusion center. Hence the transmitting values of each local node would be $\{b_0, \ldots, b_{M-1}\}$, where b_i's are appropriately chosen. Using physical layer fusion in the current setup would cause a lot of confusion. Thus the nodes transmit data using TDMA. We assume that the fusion center has i.i.d. zero mean Gaussian noise Z_k with variance σ^2. At the fusion center we run another multihypothesis sequential test of the form (21) with hypothesis $G_m : Y_k \sim f_{FC}^m = \mathcal{N}(b_m, \sigma^2)$, $m = 0, \ldots, M-1$. Define, for $i, j, = 0, \ldots, M-1$

$$\mathcal{Z}_{FC}^{i,j}(k) = \max\left(\sum_{t=1}^{k} \log \frac{f_{FC}^i(Y_t)}{f_{FC}^j(Y_t)}, 0\right).$$

The stopping time at the fusion center is,

$$N = \inf\{k : \mathcal{Z}_{FC}^{i,j}(k) > B \text{ for all } j \neq i \text{ and some } i\} \quad (22)$$

where B is appropriately chosen. At time N, the fusion center selects hypothesis G_i where i is given in (22) and decides H_i in the decentralized setup. The thresholds A and B can be different for different hypothesis to enable different P_{FA} for different hypothesis. We call this decentralized scheme as *DualMTest-D1*.

We demonstrate the effectiveness of the proposed algorithm through a Gaussian mean change example. The number of hypothesis, M is 5 and the number of local nodes L is also 5. Fusion center noise has variance 1. Under hypothesis $H_m, X_{k,l} \sim \mathcal{N}(m, 1)$, $m=0, \ldots, M-1$. As there is not much literature on decentralized multihypothesis sequential testing with no feedback from fusion center (except [19]), we compare our algorithm to the decentralized schemes created by

combinations of existing single node methods. We note that the distributed algorithm in [19] in our setup provides a very large expected detection delay. In the local node test of the distributed algorithm in [19], rejection times of each hypothesis are found by calculating the likelihood ratio of all other hypothesis with respect to the hypothesis under consideration and comparing with a positive upper threshold. The test stops when all but one of the hypothesis are rejected. But it can happen that there is a negative drift in any of the likelihood ratios with respect to more than one hypothesis and this makes the rejection time of more than one hypothesis to be very large. Thus this local node test, though it is theoretically worthy to consider, requires a large average number of samples to stop. We believe that this can be the reason for a large expected delay in the algorithm in [19].

For making combinations for comparing distributed algorithms we have considered the following tests at local nodes and fusion center: *Test-D1* and *Test-D2* of [17], *Test1* and *Test2* of [18], and *MSPRT* [5] (with equal prior probabilities). Among them we found that the combination of *Test-D1* and our *MTest-D1* outperforms other combinations. Hence in Figure 2 we plot only different configurations of *Test-D1* and *MTest-D1*. Here *DualTest-D1* indicates using *Test-D1* at both the local nodes and the fusion center; *MTest-D1:Test-D1* is using *MTest-D1* at local nodes and *Test-D1* at the fusion center and *Test-D1:MTest-D1* uses the other way. P_{FA} is the probability of false alarm in rejecting the true hypothesis. The performance is almost same under different hypothesis. Hence we show the plot of E_{DD} vs P_{FA} only for the true hypothesis taken as H_3. $b_i = i + 1, 0 \leq i \leq M - 1$. We use TDMA. *DualMTest-D1*, using *MTest-D1* at the local nodes as well as the fusion center gives the best performance.
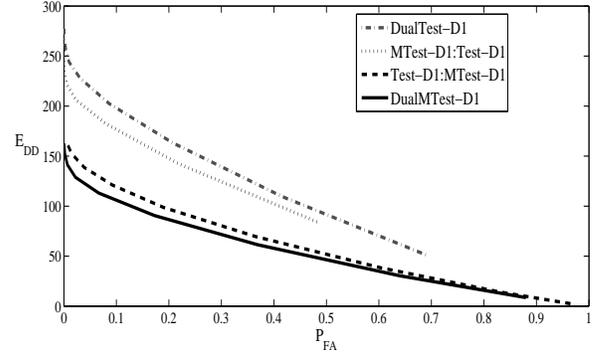

Fig. 2. Comparison among different Multihypothesis schemes

VI. CONCLUSIONS

In this paper we have presented several novel algorithms for decentralized sequential hypothesis testing problem. We start with two algorithms DualSPRT and SPRT-CSPRT for binary hypothesis testing scenario when the densities under both the hypothesis are known. We show that DualSPRT is asymptotically (first order) optimal. SPRT-CSPRT is an improvement

over DualSPRT. Next we discuss GLR-SPRT which can be used when the densities at the local nodes have parametric uncertainties. We next consider the scenario when one of the densities (under H_0) is completely known but the other belongs to a nonparametric family. We use universal source coding to obtain a good algorithm in this scenario. Finally we develop a new decentralized multihypothesis algorithm with no feedback from the fusion center and show its performance via simulations.

APPENDIX
PROOF OF THEOREM 1

The theorem follows from [15, Theorem 3] if we prove Theorem 2 of the same paper for DualSPRT. Let $P_{FA} = \alpha$ and $P_{MD} = \beta$. [15, Theorem 2] gives upper bounds for P_{FA}, P_{MD}, $E_0(N)$ and $E_1(N)$. We prove that these bounds are valid for DualSPRT also.

The stopping time of DualSPRT $N = \min(T_0(c), T_1(c))$ where $T_0(c) = \inf\{k : F_k \leq \beta_0\}$ and $T_1(c) = \inf\{k : F_k \geq \beta_1\}$. Then,

$$
\begin{aligned}
\alpha &= P_0\{\delta \ rejects \ H_0\} \\
&= P_0\{T_1(c) < T_0(c)\} \leq P_0\{T_1(c) < \infty\} \\
&\overset{(A)}{=} E_1[\exp^{-F_{T_1(c)}} 1_{\{T_1(c) < \infty\}}] \\
&\leq E_1[\exp^{-\beta_1} 1_{\{T_1(c) < \infty\}}] \overset{(B)}{=} c
\end{aligned}
$$

where (A) follows from Wald's likelihood ratio identity [7] and (B) follows since $\beta_1 = |\log c|$ and $P_1[T_1(c) < \infty] = 1$. Similarly we can prove that $\beta \leq c$. Let

$$
N_l = \inf\{n \geq 1 : \sum_{i=1}^{n} \log \frac{f_{1,l}(X_{i,l})}{f_{0,l}(X_{i,l})} \geq \rho_l |\log c|, \}
$$

$$
\tau_l(N_l) = \sup\{n \geq 1 : \sum_{i=N_l+1}^{N_l+n} \log \frac{f_{1,l}(X_{i,l})}{f_{0,l}(X_{i,l})} \leq 0\}.
$$

Then,

$$
N \leq \max_{1 \leq l \leq L} (N_l + \tau_l(N_l) + T)
$$

$$
\leq \max_{1 \leq l \leq L} N_l + \sum_{l=1}^{L} \tau_l(N_l) + T \qquad (23)
$$

where T indicates the stopping time of the fusion center SPRT after all sensors start transmitting b_1. It can be shown that the fusion center SPRT eventually will have i.i.d observations that make $E_1(T) < \infty$ [7, Chapter 3]. From [15, p. 2084] for i.i.d. observations $E_1(\tau_l(N_l)) < \infty$. The expectation of first term in (23) is also given as $|\log c|/J_{tot} + \mathcal{O}(\sqrt{|\log c|})$ [15]. Therefore,

$$
E_1(N) \leq \frac{|\log c|}{J_{tot}}(1 + o(1)).
$$

Similarly we can show for $E_0(N)$. This proves [15, Theorem 2].

REFERENCES

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

[2] D. Baron, T. Weissman, "An MCMC Approach to Lossy Compression of Continuous Sources," *Proc. of Data Compression Conference (DCC)*, Mar 2010.

[3] J. F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks", *IEEE Trans. Signal Proces.*, vol.51, issue 2, pp.407-416, Feb 2003.

[4] T. M. Cover and J. Thomas, *Elements of Information Theory*, New York:Wiley, 2006.

[5] V. P. Draglin, A. G. Tartakovsky, V. V. Veeravalli, , "Multihypothesis sequential probability ratio tests .I. Asymptotic optimality," *IEEE Trans. Information Theory*, vol.45, no.7, pp.2448-2461, Nov 1999.

[6] G. Fellouris, G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Information Theory*, vol.57, no.1, pp.534-548, Jan 2011.

[7] B. K. Ghosh and P. K. Sen, *"Handbook of Sequential Analysis"*, Statistics: a Series of Textbooks and Monographs, CRC press, 1991.

[8] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Information Theory*, Vol. 44, pp. 23252384, Oct 1998.

[9] T. Jacob, R. K. Bansal, "Sequential change detection based on universal compression algorithms," *Proc. of IEEE International Symposium on Information Theory*, July 2008.

[10] K. S. Jithin and V. Sharma, "Cooperative distributed sequential spectrum sensing", *Proc. of National Conference on Communication (NCC)*, Jan 2011.

[11] K. S. Jithin and V. Sharma, "Novel algorithms for cooperative distributed sequential spectrum sensing", *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, April 2011.

[12] J. C. Kieffer,*"Lecture notes in source coding"*, chapter 5, unpublished.

[13] T. L. Lai, "Nearly optimal sequential tests of composite hypotheses", *The Annals of Statistics*, vol.16, no.2, pp.856-886, 1988.

[14] T. L. Lai, "Sequential analysis: Some classical problems and new challenges (with discussion)", *Statistica Sinica*, vol.11, pp.303-408, 2001.

[15] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks", *IEEE Trans. Information Theory*, vol.54, no.5, pp.2072-2089, May 2008.

[16] B. Ryabko, "Applications of Universal Source Coding to Statistical Analysis of Time Series", *Selected Topics in Information and Coding Theory*, World Scientific Publishing, 2010.

[17] A. Tartakovsky, "Asymptotic Optimality of certain Multihypothesis Tests: Non-i.i.d. Case", *Statistical Inference for Stochastic Processes*, vol. 1, no. 3, pp. 265-295, 1998.

[18] A. Tartakovsky, "Asymptotically Optimal Sequential Tests for Nonhomogeneous Processes." *Sequential Analysis*, vol. 17, no. 1, pp. 33-61, 1998

[19] A. Tartakovsky, X. Li Rong, "Sequential testing of multiple hypotheses in distributed systems," Information Fusion, *Proceedings of the Third International Conference on FUSION 2000*, July 2000.

[20] A. Tartakovsky and V. V. Veeravalli, "Asymptotically Optimum Quickest Change Detection in Distributed Sensor Systems," *Sequential Analysis*, vol. 27, no. 4, pp. 441-475, 2008.

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

[22] V. V. Veeravalli and J. Unnikrishnan, "Cooperative spectrum sensing for primary detection in cognitive radios," *IEEE journal on Selected Topics in Signal Processing*, pp.18-27, Feb 2008.

[23] R. Viswanathan, P. K. Varshney, "Distributed detection with multiple sensors I. Fundamentals," *Proceedings of the IEEE*, vol.85, no.1, pp.54-63, Jan 1997.

[24] En-hui Yang, Zhen Zhang, T Berger, "Fixed-slope universal lossy data compression," *IEEE Trans. Information Theory*, vol.43, no.5, pp.1465-1476, Sep 1997.

[25] J. Ziv, "On classification with empirically observed statistics and universal data compression," *IEEE Trans. on Information 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. on Information Theory*, vol. IT-24, pp. 530536, 1978.