Decentralized Sequential Hypothesis Testing in Discrete Time

Georgios Fellouris[1] and George Moustakides[2]

[1] Department of Statistics, Columbia University,
New York, NY 10027, USA
gaf2106@columbia.edu

[2] Department of Electrical and Computer Engineering,
University of Patras,
26500 Patras, GREECE
moustaki@patras.gr

Abstract In this work we deal with the problem of decentralized sequential hypothesis testing in discrete time in the case that the sensors have full local memory. We adopt the scheme called Decentralized Sequential Probability Ratio Test (D-SPRT), which entails asynchronous communication of the sensors with the fusion center at random times. We prove that the D-SPRT is asymptotically optimal and we show that in a certain sense this asymptotic optimality can be of order-2, i.e. for small type-I and type-II error probabilities the expected time for a decision of the D-SPRT differs from that of the optimal centralized SPRT by a constant. These results have important implications on the design of the suggested scheme. Simulation experiments reveal that D-SPRT is efficient and outperforms existing asymptotically optimal schemes of the literature proposed for the same problem.

1 Introduction

The problem of sequential hypothesis testing is one of the most classical and well-studied problems of sequential analysis (see for example [3]). In the last two decades, there has been an intense interest in the *decentralized* formulation of the problem, where the sequentially acquired information for decision-making is distributed across a number of sensors and is transmitted to a global decision-maker (fusion center) which is responsible for making the decision. Moreover, cost, reliability issues as well as, communication bandwidth constraints require that the sensor observations be *quantized* before sent to the fusion center, i.e. the fusion center must send messages that belong to a *finite alphabet*. For more details, see [4].

Depending on the *local memory* that the sensors possess and whether there is *feedback* from the fusion center, there are different configurations of the above sensor-network. Here, we consider the case of *full-local memory*, i.e. we assume that at each time-instant each sensor has access to all its previous observations and can use them in order to quantize the current observation. Mei [2] recently suggested an asymptotically optimal scheme for this problem in a Bayesian setting.

In this work we assume that the alphabet that the sensors have in their disposal is binary and that each time a sensor communicates with the fusion center it must send a one-bit signal.

2 Problem Formulation

Consider the existence of a global digital clock that counts the discrete time instances $\{n\}$ with $n \in \mathbf{N}$. Assume also the existence of K sensors which acquire digital signals $\{\xi_{n,i}\}_{n=1}^{\infty}, i = 1, \ldots, K$ in a synchronized way. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space on which the K random sequences $\{\xi_{n,i}\}$ are *independent* and each sequence has *i.i.d.* samples. We assume that sensor i observes *sequentially* the sequence $\{\xi_{n,i}\}$ whose common distribution we denote by $\mathbb{P}_i$.

We would like to choose between the following two *simple* hypotheses; $\mathbb{H}_0 : \mathbb{P} = \mathbb{P}_0$, $\mathbb{H}_1 : \mathbb{P} = \mathbb{P}_1$, where $\mathbb{P}_0, \mathbb{P}_1$ are two probability measures on $(\Omega, \mathcal{F})$. The distribution $\mathbb{P}_i$ of $\xi_{n,i}$, is equal to $\mathbb{P}_{0,i}$ under $\mathbb{H}_0$ and $\mathbb{P}_{1,i}$ under $\mathbb{H}_1$, where $\mathbb{P}_{0,i}, \mathbb{P}_{1,i}$ are *known* Borel probability measures. Moreover, we set $\xi_n = (\xi_{n,1}, \ldots, \xi_{n,K}),\ n \geq 1$ and we denote by $\mathbb{P}$ the distribution of the random vectors ξ_n, therefore from the independence of observations accross sensors we obtain: $\mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_K$. We also denote by $\{\mathcal{F}_{n,i}\}(\{\mathcal{F}_n\})$ the filtration generated by the process $\{\xi_{n,i}\}$ $(\{\xi_n\})$ with $\mathcal{F}_{0,i}(\mathcal{F}_0)$ denoting the trivial σ-algebra.

We assume that $\mathbb{P}_{1,i}, \mathbb{P}_{0,i}$ are mutually absolutely continuous, therefore the Radon-Nikodym derivative $\frac{d\mathbb{P}_{1,i}}{d\mathbb{P}_{0,i}}$ and its logarithm are well-defined. From the i.i.d. assumption within the sensors and the independence assumption accross the sensors, we can define the log-likelihood ratio process locally at sensor i and globally in the sensor-network as:

$$u_{n,i} = \sum_{j=1}^{n} \ell_{j,i}, \; u_n = \sum_{i=1}^{K} u_{n,i},$$

respectively, where $u_{0,i} = 0$ and $\ell_{j,i} = \log \frac{d\mathbb{P}_{1,i}}{d\mathbb{P}_{0,i}} (\xi_{j,i})$. In other words, $\ell_{j,i}$ is the log-likelihood ratio of the jth observation in the ith sensor.

We also define the Kullback-Leibler Divergence $I_{1,i} = \mathbb{E}_1[\ell_{n,i}]$, $I_{0,i} = -\mathbb{E}_0[\ell_{n,i}]$ of $\mathbb{P}_{1,i}$ versus $\mathbb{P}_{0,i}$ and $\mathbb{P}_{0,i}$ versus $\mathbb{P}_{1,i}$ respectively which we assume that are finite in every sensor i. Let also $I_1 = \sum_{i=1}^{K} I_{1,i}$, $I_0 = \sum_{i=1}^{K} I_{0,i}$.

In classical sequential hypothesis testing, the goal is to choose between the hypotheses $\mathbb{H}_0$ and $\mathbb{H}_1$ using a *sequential test*, i.e. a pair $(\mathcal{T}, d_{\mathcal{T}})$, where $\mathcal{T}$ is an $\{\mathcal{F}_n\}$-adapted stopping time and $d_{\mathcal{T}}$ is an $\mathcal{F}_{\mathcal{T}}$-measurable r.v. with values in $\{0,1\}$.

In the decentralized version of the problem we must choose: (1) an increasing sequence of $\{\mathcal{F}_{n,i}\}$-adapted stopping times $\{\tau_k^i\}_{k=1}^{\infty}$ at each sensor i at which times to communicate with the fusion center (sampling strategy), (2) a sequence of quantized signals $\{z_k^i\}_{k=1}^{\infty}$ to transmit to the fusion center from sensor i at the times $\{\tau_k^i\}$ (quantization strategy) and (3) a sequential test that is based on the available information at the fusion center, i.e. the quantized signals $\{z_k^i\}$ from all sensors and the corresponding stopping times $\{\tau_k^i\}$.

The above formulation of the decentralized problem generalizes the typical mathematical setup in the literature (see e.g. [2], [4]), where only steps (2) and (3) are included and the sensors are assumed to communicate, synchronously, with the fusion center at *every* time instant n. It should be noted that in the proposed approach communication between sensors and fusion center is asynchronous and sparse.

2.1 Performance Criteria and the Optimal Centralized Tests

We use Wald's approach [5] to formulate the sequential hypothesis-testing problem. We start by introducing the discrete-time version of the Sequential Probability Ratio Test (SPRT), which is defined as follows:

$$\mathcal{N} = \inf\{n \geq 1 : u_n \notin (-A, B)\}\,, \; d_{\mathcal{N}} = \quad_{\{u_{\mathcal{N}} \geq B\}}\,, \tag{1}$$

where $A, B > 0$ are two constant thresholds. The SPRT was shown by Wald and Wolfowitz in [6] to be optimal in the sense that it solves the following optimization problem:

$$\inf_{(\mathcal{T}, d_{\mathcal{T}})} \mathbb{E}_j[\mathcal{T}]; \;\; \text{subject to } \mathbb{P}_0[d_{\mathcal{T}} = 1] \leq \alpha \text{ and } \mathbb{P}_1[d_{\mathcal{T}} = 0] \leq \beta, \tag{2}$$

where $j = 0, 1$ and $\alpha, \beta > 0$ are such that $\alpha + \beta < 1$. The boundaries A, B are chosen so that the error probability constraints in (2) are satisfied with equalities. It is well-known that under appropriate conditions on the process $\{u_n\}$, such as existence and finiteness of the moment-generating function (see [3]), we have that as $\alpha, \beta \to 0$:

$$A = \mathcal{O}(|\log \beta|)\;,\; I_1 \mathbb{E}_1[\mathcal{N}] = |\log \alpha|(1 + o(1))$$
$$B = \mathcal{O}(|\log \alpha|)\;,\; I_0 \mathbb{E}_0[\mathcal{N}] = |\log \beta|(1 + o(1)).$$

2.2 Suggested Decentralized Test

Sampling & Quantization Strategy. Following [1] we suggest that sensor i sends a quantized signal to the fusion center at the stopping times $\{\tau_k^i\}$, which are defined recursively as follows:

$$\tau_k^i = \inf\{n \geq \tau_{k-1}^i : u_{n,i} - u_{\tau_{k-1}^i, i} \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \tag{3}$$

where $\underline{\Delta}_i, \overline{\Delta}_i > 0$ are the *sampling thresholds* in sensor i and $\tau_0^i = 0$. The signal that the ith sensor sends at time τ_k^i , i.e. at the kth time it communicates with the fusion center, will be:

$$z_k^i = \ {}_{\{u_{\tau_k^i,i} - u_{\tau_{k-1}^i,i} \geq \overline{\Delta}_i\}}. \tag{4}$$

We denote by η_k^i the overshoot that occurs in the kth sample from the ith sensor, i.e.

$$\eta_k^i = (u_{\tau_k^i,i} - u_{\tau_{k-1}^i,i} - \overline{\Delta}_i)^+ + (u_{\tau_k^i,i} - u_{\tau_{k-1}^i,i} - \underline{\Delta}_i)^-.$$

We also denote by $\tilde{\ell}_{k,i}$ the log-likelihood ratio of z_k^i. For any given sensor i, $\{\xi_{n,i}\}$ is a sequence of i.i.d. r.v's under both hypotheses, thus $\{z_k^i\}$ is a sequence of i.i.d. Bernoulli r.v's with parameter $1 - \pi_{1,i}$ under $\mathbb{H}_1$ and $\pi_{0,i}$ under $\mathbb{H}_0$, where

$$\pi_{1,i} = \mathbb{P}_1[z_k^i = 0] \, , \, \pi_{0,i} = \mathbb{P}_0[z_k^i = 1],$$

where we note that $\pi_{0,i}, \pi_{1,i} < 0.5$. This suggests that:

$$\tilde{\ell}_{k,i} = \overline{\lambda}_i z_k^i - \underline{\lambda}_i (1 - z_k^i), \quad \text{where} \quad \overline{\lambda}_i = \log\left(\frac{1 - \pi_{1,i}}{\pi_{0,i}}\right) \, , \, \underline{\lambda}_i = \log\left(\frac{1 - \pi_{0,i}}{\pi_{1,i}}\right).$$

Sequential test at the fusion center. In order to define the suggested decentralized sequential test we introduce the following notation, which suppresses the dependence on the sensor: we denote by τ_k the time that the jth signal arrived to the fusion center *independently of the sensor who sent it*. Since it is possible to have signals from different sensors sent at the same time to the fusion center, we order them in an arbitrary way, for example in alphabetic order. Thus, if for example sensors i and m both send a signal at time n=1, with $i < m$, then we set: $\tau_1 = \tau_{1,i} = 1, \tau_2 = \tau_{1,m} = 1$. Similarly, we denote by z_k the signal that arrived at the fusion center at time τ_k, $\tilde{\ell}_k$ the log-likelihood ratio of z_k and η_k the corresponding overshoot. Moreover, we denote by δ_k the identity of the sensor which sent the signal z_k, i.e. $\delta_k = i$, if the kth signal was sent from sensor i, $i = 1, \ldots, K$. Finally we denote by $\{\mathcal{C}_k\}_{k=0}^{\infty}$ the flow of information at the fusion center, i.e.

$$\mathcal{C}_k = \sigma\left\{(z_s, \delta_s), 1 \leq s \leq k\right\}.$$

Clearly, $\mathcal{C}_k \subset \mathcal{F}_{\tau_k}$.

Suppose now that $\{\tilde{u}_k\}$ is the log-likelihood ratio process of the messages $\{z_k\}$ that arrive at the fusion center from *any* sensor. Then, from the independece *across* and *within* sensors and since the fusion center knows which sentor sent each signal, we have the following representation:

$$\tilde{u}_k = \sum_{m=1}^{k} \tilde{\ell}_m = \sum_{i=1}^{K} \sum_{m=1}^{k_i} \tilde{\ell}_{m,i}, \text{ where } k_i = \sum_{m=1}^{k} \ {}_{\{\delta_m = i\}}, i = 1, \ldots, K.$$

We note that index $\{k\}$ counts the number of samples received at the fusion center and not global time. Reference to global time is achieved by using the sequence of communication times $\{\tau_k\}$ since the kth sample received by the center corresponds to the global time τ_k.

The suggested sequential test for the problem in (2) will then be:

$$\tilde{\mathcal{N}} = \tau_{\mathcal{K}}, \text{ where } \mathcal{K} = \inf\{k \geq 1 : \tilde{u}_k \notin (-\tilde{A}, \tilde{B})\} \, , \, d_{\tilde{\mathcal{N}}} = \ {}_{\{\tilde{u}_{\mathcal{K}} \geq \tilde{B}\}}, \tag{5}$$

where $\tilde{A}, \tilde{B}$ are chosen so that $\mathbb{P}_0[d_{\tilde{\mathcal{N}}} = 1] = \alpha$ and $\mathbb{P}_1[d_{\tilde{\mathcal{N}}} = 0] = \beta$.

Notice that $\mathcal{K}$ is a stopping time with respect to filtration of the fusion center, i.e $\mathcal{K}$ is a $\{\mathcal{C}_k\}$-adapted stopping time. Moreover, $\mathcal{C}_{\mathcal{K}} \subset \mathcal{F}_{\tilde{\mathcal{N}}}$, since $\tilde{\mathcal{N}} = \tau_{\mathcal{K}}$. Finally, by the definition of the likelihood ratio process we have the following relationship:

$$\mathbb{E}_0[e^{u_{\tilde{\mathcal{N}}}} | \mathcal{C}_{\mathcal{K}}] = e^{\tilde{u}_{\mathcal{K}}}.$$

This is true, since the probability measures projected onto the space generated by the data transmitted to the fusion center involves only Bernoulli random variables. Consequently the log-likelihood ratio at the kth sample is simply $\tilde{u}_k$.

We now state some Lemmas, which are useful for proving the main results of this paper, however we omit most of the proofs, due to space constraints.

Lemma 1. $\tilde{A} \leq |\log \beta|$, $\tilde{B} \leq |\log \alpha|$.

Lemma 2. *For large values of* $\overline{\Delta}_i, \underline{\Delta}_i$ *we have:*

$$\mathbb{E}_1[|\eta_{1,i}|] = \mathcal{O}(1) \ , \ \overline{\lambda}_i = \overline{\Delta}_i + \overline{\rho}_i + o(1) \ , \ \overline{\Delta}_i = \mathcal{O}(|\log(\pi_{0,i})|)$$
$$\mathbb{E}_0[|\eta_{1,i}|] = \mathcal{O}(1) \ , \ \underline{\lambda}_i = \underline{\Delta}_i + \underline{\rho}_i + o(1) \ , \ \underline{\Delta}_i = \mathcal{O}(|\log(\pi_{1,i})|)$$

where $\overline{\rho}_i, \underline{\rho}_i$ *are positive constants which do not depend on* $\overline{\Delta}_i, \underline{\Delta}_i$.

Lemma 3. *For* $j = 0, 1,$ *we have the following inequalities:*

$$\mathbb{E}_j\left[\sum_{k=1}^{\mathcal{K}} |\eta_k|\right] \leq \mathbb{E}_j[\mathcal{K}] \sum_{i=1}^{K} \mathbb{E}_j[|\eta_{1,i}|] \ \ , \ \ \mathbb{E}_j\left[\sum_{k=1}^{\mathcal{K}} |\tilde{\ell}_k - \mathbb{E}_j[\tilde{\ell}_k]|\right] \leq \mathbb{E}_j[\mathcal{K}] \sum_{i=1}^{K} \mathbb{E}_j[|\tilde{\ell}_{1,i} - \mathbb{E}_j[\tilde{\ell}_{1,i}]|].$$

The performance of the suggested scheme is characterized by the following inequalities:

Proposition 1.

$$I_1 \frac{\mathbb{E}_1[\tilde{\mathcal{N}}]}{|\log \alpha|} \leq \left(1 + \frac{\sum_{i=1}^K \overline{\lambda}_i}{|\log \alpha|}\right)\left(1 + \frac{\sum_{i=1}^K \mathbb{E}_1[|\eta_{1,i}|]}{\zeta_1}\right), \tag{6}$$

$$I_0 \frac{\mathbb{E}_0[\tilde{\mathcal{N}}]}{|\log \beta|} \leq \left(1 + \frac{\sum_{i=1}^K \underline{\lambda}_i}{|\log \beta|}\right)\left(1 + \frac{\sum_{i=1}^K \mathbb{E}_0[|\eta_{1,i}|]}{\zeta_0}\right), \tag{7}$$

where $\zeta_1 \equiv \min_i \mathbb{E}_1[\tilde{\ell}_{1,i}] - \sum_{i=1}^K \sqrt{\mathbb{V}_1[\tilde{\ell}_{1,i}]}$, $\zeta_0 \equiv \min_i \mathbb{E}_0[\tilde{\ell}_{1,i}] - \sum_{i=1}^K \sqrt{\mathbb{V}_0[\tilde{\ell}_{1,i}]}$ *and* $\mathbb{V}_j$ *denotes variance.*

Proof. We will work under $\mathbb{H}_1$ and prove (6), we can prove (7) in the same way. We observe that $\{u_n\}$ is a random walk and $\tilde{\mathcal{N}}$ an integrable stopping time with respect to the filtration $\{\mathcal{F}_n\}$. Therefore, we can apply Wald's identity and have: $I_1\mathbb{E}_1[\tilde{\mathcal{N}}] = \mathbb{E}_1[u_{\tilde{\mathcal{N}}}] = \mathbb{E}_1[u_{\tilde{\mathcal{N}}} - \tilde{u}_{\mathcal{K}}] + \mathbb{E}_1[\tilde{u}_{\mathcal{K}}]$. From the definition of the overshoots $\{\eta_k\}$ we have: $u_{\tilde{\mathcal{N}}} - \tilde{u}_{\mathcal{K}} = \sum_{k=1}^{\mathcal{K}} \eta_k \leq \sum_{k=1}^{\mathcal{K}} |\eta_k|$, thus from Lemma 3 we obtain: $\mathbb{E}_1[u_{\tilde{\mathcal{N}}} - \tilde{u}_{\mathcal{K}}] \leq \mathbb{E}_1[\mathcal{K}] \sum_{i=1}^K \mathbb{E}_1[|\eta_{1,i}|]$. Moreover, $\tilde{u}_{\mathcal{K}} = \sum_{k=1}^{\mathcal{K}} (\tilde{\ell}_k - \mathbb{E}_1[\tilde{\ell}_k]) + \sum_{k=1}^{\mathcal{K}} \mathbb{E}_1[\tilde{\ell}_k] \geq -\sum_{k=1}^{\mathcal{K}} |\tilde{\ell}_k - \mathbb{E}_1[\tilde{\ell}_k]| + \mathcal{K} \min_i \mathbb{E}_1[\tilde{\ell}_{1,i}]$. Using Lemma 3, the fact that the L_2 norm is larger than the L_1, and assuming $\zeta_1 > 0$ yields $I_1\mathbb{E}_1[\tilde{\mathcal{N}}] \leq (1 + \sum_{i=1}^K \mathbb{E}_1[|\eta_{1,i}|]/\zeta_1)\mathbb{E}_1[u_{\tilde{\mathcal{N}}}]$. Our proof is completed by observing that from the definition of the sequential test in (5) and Lemma 1, we have: $\tilde{u}_{\mathcal{K}} \leq \tilde{B} + \sum_{i=1}^K \overline{\lambda}_i \leq |\log \alpha| + \sum_{i=1}^K \overline{\lambda}_i$, since $\sum_{i=1}^K \overline{\lambda}_i$ is the maximum possible overshoot $\tilde{u}_{\mathcal{K}} - \tilde{B}$ on the event $\{\tilde{u}_{\mathcal{K}} \geq \tilde{B}\}$. Thus: $\mathbb{E}_1[\tilde{u}_{\mathcal{K}}] \leq |\log \alpha| + \sum_{i=1}^K \overline{\lambda}_i$.

We can now show that the suggested scheme is asymptotically optimal if we let the thresholds $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ *appropriately* as $\alpha, \beta \to 0$. Before we do that, we state the following Lemma:

Lemma 4. $\zeta_1, \zeta_0 \to \infty$ *as* $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$.

Proposition 2. *If* $\alpha, \beta \to 0$ *and* $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ *so that:*

$$\overline{\Delta}_i = o(|\log \alpha|) \ , \ \underline{\Delta}_i = o(|\log \beta|), \tag{8}$$

then $\frac{\mathbb{E}_j[\tilde{\mathcal{N}}]}{\mathbb{E}_j[\mathcal{N}]} \to 1$, $j = 0, 1$, *i.e. the suggested scheme* $(\tilde{\mathcal{N}}, d_{\tilde{\mathcal{N}}})$ *is asymptotically optimal of order-1.*

2.3 Optimal Rate for the Sampling Thresholds

It is very interesting now to determine the optimal divergence rate of the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ as a function of the error probabilities α, β.

Proposition 3. *If $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ then the optimal divergence rate for the sampling thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ as $\alpha, \beta \to 0$ is:*

$$\overline{\Delta}_i = \mathcal{O}(\sqrt{|\log\alpha|})\,,\ \underline{\Delta}_i = \mathcal{O}(\sqrt{|\log\beta|}).$$

Under this selection, as $\alpha, \beta \to 0$, we have that:

$$\mathbb{E}_1[\tilde{\mathcal{N}}] \leq \frac{|\log\alpha| + \textit{const.}\sqrt{|\log\alpha|} + \textit{const.}}{I_1}\ \ ,\ \ \mathbb{E}_0[\tilde{\mathcal{N}}] \leq \frac{|\log\beta| + \textit{const.}\sqrt{|\log\beta|} + \textit{const.}}{I_0},$$

and

$$\mathbb{E}_1[\tilde{\mathcal{N}}] - \mathbb{E}_1[\mathcal{N}] = \mathcal{O}(\sqrt{|\log\alpha|})\ \ ,\ \ \mathbb{E}_0[\tilde{\mathcal{N}}] - \mathbb{E}_0[\mathcal{N}] = \mathcal{O}(\sqrt{|\log\beta|}).$$

2.4 Oversampling and Asymptotic Optimality of order 2

Suppose now that, at each sensor, we have the possibility to modify the first absolute moment $\mathbb{E}_j[|\xi_{n,i}|]$ of the acquired samples. In a real sensor network system where samples are obtained by sampling continuous-time signals this can be realized by changing the sampling rate. We can then show that provided that the second moment is sufficiently small we obtain asymptotic optimality *of order-2* for the suggested scheme *even with fixed sampling thresholds* $\overline{\Delta}_i, \underline{\Delta}_i$. This is the content of the following proposition.

Proposition 4. *Assume that as $\alpha, \beta \to 0$ we have the ability to force $\mathbb{E}_j[|\xi_{n,i}|] \to 0,\ j = 0, 1,\ i = 1, \ldots, K$. If for every sensor i we keep the sampling thresholds $\underline{\Delta}_i, \overline{\Delta}_i$ fixed and select the rates as follows*

$$|\log\alpha| \cdot \mathbb{E}_1[|\xi_{1,i}|] \to 0\ ,\ |\log\beta| \cdot \mathbb{E}_0[|\xi_{1,i}|] \to 0, \tag{9}$$

then $\mathbb{E}_j[\tilde{\mathcal{N}}] - \mathbb{E}_j[\mathcal{N}] = \mathcal{O}(1), j = 0, 1$, i.e. the suggested scheme $(\tilde{\mathcal{N}}, d_{\tilde{\mathcal{N}}})$ is asymptotically optimal of order-2 under both $\mathbb{H}_0$ and $\mathbb{H}_1$ for the problem in (2).

Example: Suppose that the i.i.d. sequence of observations in each sensor is obtained from canonical deterministic sampling of a continuous-time process $\{\xi_{t,i}\}_{t\geq 0}$ at the discrete times $t = nh,\ n \in \mathbf{N}$, where each $\{\xi_{t,i}\}$ is a Brownian Motion with drift 0 under $\mathbb{H}_0$ and μ_i under $\mathbb{H}_1$. Each μ_i is a real non-zero constant and $h > 0$ is the common sampling period for all sensors. We then have the followning hypothesis testing problem:

$$\mathbb{H}_0 : \{\xi_{nh,i} - \xi_{(n-1)h,i}\} \sim^{iid} \mathcal{N}(0, h)\ ,\ \mathbb{H}_1 : \{\xi_{nh,i} - \xi_{(n-1)h,i}\} \sim^{iid} \mathcal{N}(\mu_i h, h) \tag{10}$$

Letting $h \to 0$ makes the discrete problem converge to the continuous problem, thus: $\lim_{h\to 0}\mathbb{E}_j[|\xi_{1,i}|] = 0,\ j = 0, 1$. Therefore, letting $h \to 0$ and $\alpha, \beta \to 0$ in such a way that condition (9) is satisfied, leads to *order-2* asymptotic optimality of the D-SPRT. Of course the question is how dense the sampling must be in order to have performance which is comparable to the optimum. As the next simulation example reveals, even crude sampling is sufficient to guarantee a very satisfactory performance.

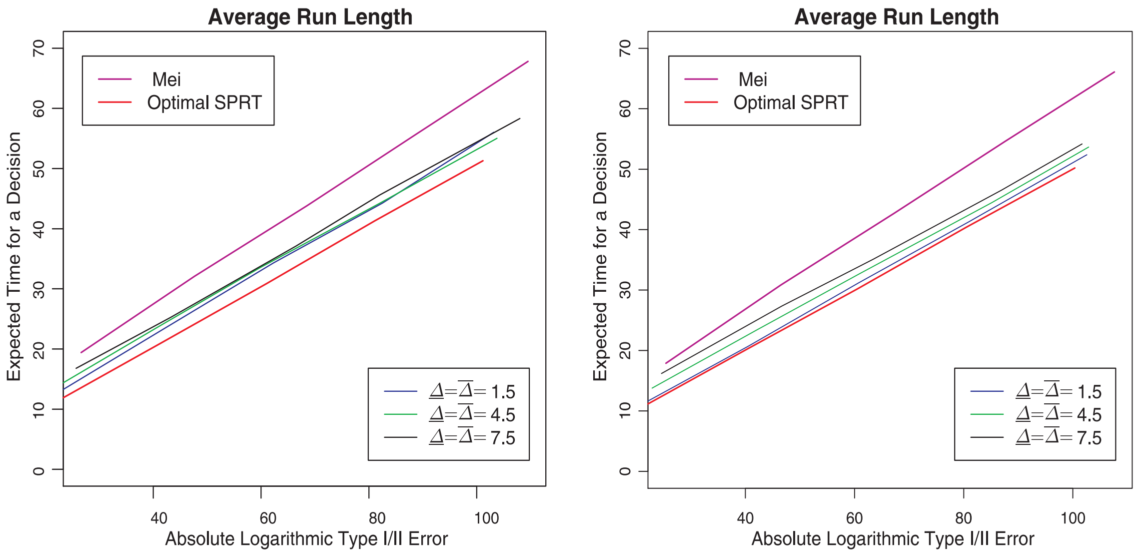
3 Design and Simulation Experiments

The main challenge in the implementation of the D-SPRT is the choice of the sampling thresholds $\overline{\Delta}_i$ and $\underline{\Delta}_i$. Small values of $\overline{\Delta}_i$, $\underline{\Delta}_i$'s entail more frequent communication between the sensors and the fusion center, but make the scheme more vulnerable to the overshoot effect while overly large values of the same parameters result in larger detection delays. Thus, we should choose the sampling thresholds to be large enough in order to stabilize the overshoot effect, but not too large to affect the frequency of communication between sensors and fusion center.

In Propositions 2 and 3 we deal exactly with this problem and provide the (optimal) rate of divergence for the sampling thresholds that allow the scheme to be asymptotically optimal. However, the scheme will in practice be implemented with constant predetermined sampling thresholds; thus there is still a lot of flexibility in the specification of the sampling thresholds, since Proposition 3 determines only the optimal divergence rate of the $\overline{\Delta}_i, \underline{\Delta}_i$'s with respect to the error probabilities α, β.

In Proposition 4, we take a different approach and consider the sampling frequency h of the continuous-time signal at the sensors as the control parameter, instead of the sampling thresholds. The results of these propositions imply that *oversampling* at the sensor-level improves dramatically the efficiency of the D-SPRT by minimizing the overshoot effect. In that case, we expect smaller $\underline{\Delta}_i, \overline{\Delta}_i$'s to lead to better performances for the D-SPRT.

We illustrate these ideas by performing two simulation experiments in the context of problem (10). We set $K = 4$ and $\mu_1 = \ldots = \mu_4 = 1$. We compare our scheme defined in (5) with the optimal centralized SPRT (1) and also with the test suggested by Mei in [2], which is also asymptotically optimal. We consider two cases $h = 1$ and 0.1 while the sampling thresholds take the values $\overline{\Delta}_i = \underline{\Delta}_i = 1.5, 4.5, 7.5$. We plot the resulting average-length-run (ARL) curves and compare them to the corresponding curves of the optimal SPRT and Mei's test. The horizontal axis represents $|\log \alpha| (= |\log \beta|)$, since in our example we consider $\alpha = \beta$ and the vertical axis represents the expected time for a decision.



Both graphs show the D-SPRT has a substantially better performance than Mei's test and is also very close to the optimal performance. This is true for all three choices of the sampling thresholds. Moreover, in the graph to the right which corresponds to $h = 0.1$, the D-SPRT (with the same choices for the sampling thresholds) is much closer to the optimal centralized test than in the left case where $h = 1$. In addition to that, when $h = 0.1$ there is a clear ordering in the curves that correspond to the different $\overline{\Delta}_i = \underline{\Delta}_i$'s with smaller sampling thresholds leading to better performances. These graphs are consistent with our results and seem to advocate the use of our scheme especially in combination with oversampling at the sensor level.

References

Hussain, A.M. (1994). Multisensor Distributed Sequential Detection. In: *IEEE Tran. Aer. Elect. Syst.* Vol. 30, **3**, 698–708.

Mei, Y. (2008). Asymptotic optimality theory for sequential hypothesis testing in sensor networks. In: *IEEE Tran. Inf. Th.*, Vol. 54, **5**, 2072–2089.

Siegmund,D. (1985). *Sequential Analysis, Tests and Confidence Intervals.* Springer-Verlag, New York.

Veeravalli V.V (1999). Sequential Decision Fusion: Theory and Applications In: *J. Franklin Inst.*, Vol. 336, 301–322.

Wald, A. (1945) *Sequential Analysis.* Wiley, New York.

Wald A. and Wolfowitz J. (1948) Optimum character of the sequential probability ratio test, In: *Ann. Math. Statist*, Vol. 19, 326–339.