Decentralized sequential decision making with asynchronous communication

Georgios Fellouris

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2010

ABSTRACT

Decentralized sequential decision making with asynchronous communication

Georgios Fellouris

We consider three statistical problems – hypothesis testing, change detection and parameter estimation– when the relevant information is acquired sequentially by remote sensors; all sensors transmit *quantized* versions of their observations to a central processor, which is called fusion center and is responsible for making the final decision. Under this *decentralized* setup, the challenge is to choose a quantization rule at the sensors and a fusion center policy that will rely only on the transmitted quantized messages.

We suggest that the sensors transmit messages at stopping times of their observed filtrations, inducing in that way asynchronous communication between sensors and fusion center. Based on such communication schemes, we propose fusion center policies that mimic the corresponding optimal centralized policies. We prove that the resulting decentralized schemes are asymptotically optimal under different statistical models for the observations. These asymptotic optimality properties require moderate, or even rare, communication between sensors and fusion center, which is a very desirable characteristic in applications.

Table of Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank Prof. George Moustakides for his constant and active support, guidance and help. The first two of the three problems in this thesis are based on our joint work and have been motivated by his ideas, whereas his feedback on the third problem has been very important. This dissertation would never have been possible without him.

However, it would be wrong to limit to this dissertation the influence that Prof. Moustakides had on me. Our collaboration generated a lot more projects and opened my research horizons. Despite the geographical distance, I had the opportunity to learn so many things from him. The depth of his knowledge, his dedication to the field of sequential detection, the originality of his thought and the elegance of his work have been and will continue to be great sources of inspiration for me.

I also want to thank my advisor, Prof. Ioannis Karatzas, who believed in me since I was a student in Greece, encouraged me to continue my Ph.D. studies at Columbia University and provided precious feedback to my research efforts. My interaction with him and his lectures have been extremely precious experiences. His enthusiasm, deep knowledge, hard work and extraordinary lecturer skills have inspired and influenced me in a profound way.

I am indebted to Profs. Venugopal V. Veeravalli, Victor De la Peña and David Madigan for their participation in my dissertation committee. Their comments before and during my defense, together with those of Profs. Karatzas and Moustakides, have improved significantly the essence and the presentation of this thesis.

I need to thank Columbia University and the Department of Statistics for their moral and financial support and for providing such a stimulating learning environment. I had the privilege to be taught by inspiring and wonderful teachers, who helped me appreciate and understand in depth probability and statistics. Special thanks should go to Prof. Zhiliang Ying, who was always available to listen my thoughts and answer my questions, Prof. Andrew Gelman, whose approach to applied statistics and the teaching of statistics has influenced me a lot, Prof. David Madigan, who taught me a lot while I was at the statistical consulting office. Moreover, I should add that it was a great pleasure to attend the classes of Profs. Mark Broadie, Paul Glasserman and Ioannis Kontoyiannis.

I also want to thank all my friends in the Statistics Department (and not only), with whom I had stimulating discussions about probability and statistics (and not only) and whose company made my life as a student nicer. Special thanks should go to Dood, Anthony and Faiza for always being so helpful and also to Olympia Hadjiliadis, whose energy and passion inspired me to choose to do research in the field of sequential analysis.

Finally, I want to thank my parents, Anargyro and Eleni, and my brother, Aggelo, for their encouragement and immense emotional support, which has been crucial at bad and good times. Last but not least, I need to thank my girlfriend, Alexandra Chronopoulou, for her patience, understanding, support and friendship all these years.

To my parents

Chapter 1

Introduction

1.1 The decentralized sequential setup

The study of sequential statistical problems started with the pioneering work of A.Wald [63], which was motivated by the need for efficient sampling schemes during World War II. Since then, there has been an enormous literature in sequential methods and sequential analysis has become one of the most mature areas in theoretical and applied statistics ([22], [42], [53]).

The main characteristic of a sequential problem is that the horizon of observations is not fixed in advance. Thus, apart from a *decision rule* for the underlying statistical problem, the designer of a sequential scheme must also choose a *stopping rule* which will determine when he should stop collecting observations.

The choice of a sequential policy is characterized by the following trade-off; a quick decision allows any necessary action to be taken in time, however a longer horizon of observations provides more information and leads to a more reliable decision. Thus, the main goal in any sequential problem is to solve optimally this trade-off.

In this thesis we study three statistical problems – hypothesis testing, change detection and parameter estimation– when the available information is acquired sequentially *by remote sensors*; all sensors communicate with a central processor (fusion center), which is responsible for making the final decision.

When the sensors transmit *all* their observations to the fusion center, as it is typically assumed in the classical theory of sequential analysis, then *all* the relevant information is available to the decision maker and we say that we are in a *centralized* setup.

However, a centralized setup is often non-realistic due to practical considerations, such as the the need for data compression, smaller communication bandwidth and robustness of the sensor network. These are crucial issues in application areas such as signal processing, mobile and wireless communication, multisensor data fusion, internet security, robot network and others [57]. In all these applications, the fusion center receives only *partially* the sensor observations. For that reason, we assume that each sensor needs to *quantize* its observations before transmitting them to the fusion center. In other words, each sensor must send to the fusion center mesagges that take values in a small-size alphabet. In this case, we say that we are in a *decentralized* setup.

Therefore, apart from a sequential scheme at the fusion center, the decision maker –under a decentralized setup– must also choose a *communication scheme* at the sensors. A communication scheme consists of a *sampling rule* and a *quantization rule.* The first one determines the times at which the sensors communicate with the fusion center, whereas the second specifies how the sensors quantize their observations.

It is important to stress that the overall goal is not to optimize (in some sense) the communication between sensors and fusion center, but to solve the underlying sequential statistical problem under the above communication constraints. However, it is clear that the fusion center policy depends heavily on the chosen communication scheme, since the decision maker can use only the available information at the fusion center in order to determine its stopping rule and decision rule.

Ideally, we would like to find a decentralized sequential scheme that minimizes a certain performance measure over all admissible communication schemes *and* fusion center policies. This is very difficult (essentially impossible) at this level of

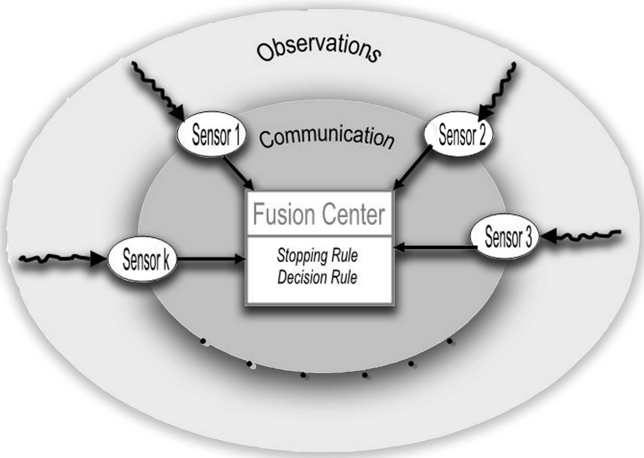


Figure 1.1: The decentralized setup

generality, therefore we resort to *asymptotically* optimal decentralized sequential schemes.

Thus, if $\mathcal{S}$ is the optimal centralized sequential scheme with respect to a certain performance measure $\mathcal{J}$, then a sequential scheme $\tilde{\mathcal{S}}$ will be asymptotically optimal, if $\mathcal{J}[\mathcal{S}]/\mathcal{J}[\tilde{\mathcal{S}}] \to 1$ as the horizon of observations goes to infinity. However, in our framework, the designer of the scheme can control the flow of information at the fusion center in many ways. Thus, not only he decides how long the sensors keep collecting data (*horizon of observations*), but also how often they communicate with the fusion center (*communication frequency*), how often they sample their underlying continuous-time processes (*sampling frequency*) or even the *number of sensors* K which are used. Therefore, it is not only legitimate but also useful to consider asymptotic properties when we have high-frequency sampling, frequent or rare communication or large number of sensors.

Moreover, in order to be able to distinguish between asymptotically optimal schemes with qualitatively different behavior, we introduce different *orders* of asymptotic optimality. Thus, as long as $\mathcal{J}[\mathcal{S}] \to \infty$, we say that a (decentralized) sequential scheme $\tilde{\mathcal{S}}$ is *asymptotically optimal* of order-2, if $\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = \mathcal{O}(1)$, and of order-3, if $\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = o(1)$, where by $\mathcal{O}(1)$ we denote an asymptotically bounded term and by $o(1)$ an asymptotically vanishing term. It is clear that, since $\mathcal{J}[\mathcal{S}] \to \infty$, order-3 asymptotic optimality implies order-2 asymptotic optimality and this implies order-1 asymptotic optimality.

1.2 Communication schemes

Let us first of all denote by $\{\mathcal{F}_t^i\}$ the observed filtration locally at sensor i and by $\{\mathcal{F}_t\}$ the global filtration in the sensor network. Thus, if $\{\xi_t^i\}$ is the observed process at sensor i, then

$$\mathcal{F}_t^i = \sigma(\xi_s^i : 0 \leq s \leq t) \quad , \quad \mathcal{F}_t = \sigma(\xi_s^i : 0 \leq s \leq t \, , \; i = 1, \ldots, K) \tag{1.1}$$

Moreover, we denote by $\{0, \ldots, d-1\}$ the available alphabet at all sensors.

We say that (τ_n^i, z_n^i) is a *communication scheme*, if each $\{\tau_n^i\}$ is an increasing sequence of finite $\{\mathcal{F}_t^i\}$-stopping times and each z_n^i is an $\mathcal{F}_{\tau_n^i}^i$-measurable random variable that takes values in $\{0, \ldots, d-1\}$, so that sensor i transmits at each time τ_n^i the message z_n^i to the fusion center. Then, a stopping rule at the fusion center is an $\{\tilde{\mathcal{F}}_t\}$-stopping time, where

$$\tilde{\mathcal{F}}_t = \sigma((\tau_n^i, z_n^i) : \tau_n^i \leq t \, , \, i = 1, \ldots, K), \quad t \geq 0 \tag{1.2}$$

According to this definition of a communication scheme, each sensor may need to remember at any given time all its previous observations in order to determine its next communication time and message to the fusion center (*full* local memory). However, this may not be a realistic assumption, since the memory requirements that it imposes on the sensors are very large and may not be affordable in practice.

For that reason, we modify the definition of a communication scheme so that each sensor does not keep any information regarding its observations before the most recent communication (*limited* local memory). More specifically, we require τ_n^i to be an $\{\mathcal{F}_{\tau_{n-1}^i, t}^i\}$-stopping time and each z_n^i an $\mathcal{F}_{\tau_{n-1}^i, \tau_n^i}^i$-measurable random variable, where

$$\mathcal{F}_{\tau_n^i, t}^i = \sigma(\xi_s^i : \tau_n^i \leq s \leq t), \quad t \geq \tau_n^i, \quad n \in \mathbb{N}. \tag{1.3}$$

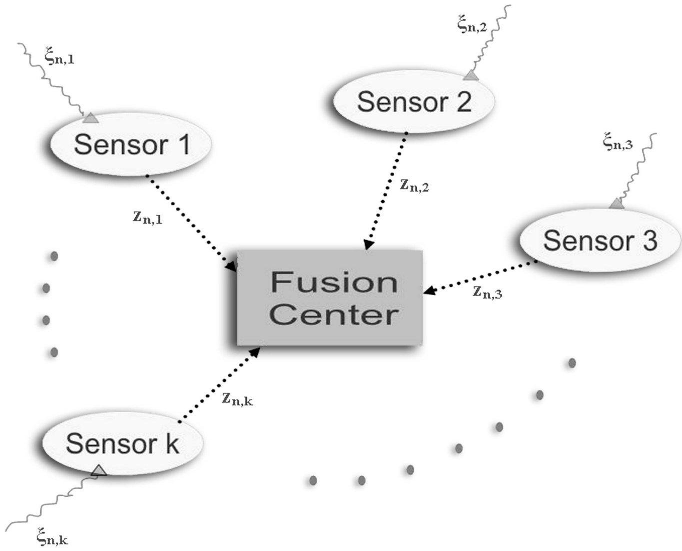


Figure 1.2: Sensor Network

We can think of limited local memory as an intermediate case between full local memory and no local memory, where each sensor does not keep any information regarding its previous observations. Moreover, notice that this is a different notion of limited local memory than the one used typically in the literature ([35], [62]), where each sensor remembers all the quantized messages that it has transmitted to the fusion center.

In the decentralized literature, it is usually implicitly assumed that all sensors communicate at common, deterministic, equidistant times [62]. However, this is not always a realistic assumption, since in practice it may be hard to force distant sensors to transmit messages concurrently. In this thesis, inspired by the works in ([2] , [44], [45]), we suggest that the communication times should be non-trivial stopping times with respect to the sensor observations (*event-triggered sampling*). This will not only induce asynchronous communication between sensors and fusion center, but – as we will prove– will lead to more efficient fusion center policies.

In particular, we focus on a special case of event-triggered sampling, which is called level-triggered (or delta or Lebesgue) and can be expressed as follows:

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : \zeta_t^i - \zeta_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}$$
$$z_n^i = \begin{cases} 1, & \text{if } \zeta_{\tau_n^i}^i - \zeta_{\tau_{n-1}^i}^i \geq \overline{\Delta}_i \\ 0, & \text{if } \zeta_{\tau_n^i}^i - \zeta_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i \end{cases} \tag{1.4}$$

where each $\{\zeta_t^i\}$ is an $\{\mathcal{F}_t^i\}$-adapted process, i.e. locally observed at sensor i, and $\overline{\Delta}_i, \underline{\Delta}_i$ are positive constants.

Based on such communication schemes and mimicking the corresponding optimal centralized schemes, we propose and analyze novel decentralized sequential schemes for the problems of hypothesis testing, quickest detection and parameter estimation for a variety of probabilistic models for the sensor observations. In the remaining part of the introduction we state our main findings and relate them to the literature.

1.3 Sequential hypothesis testing

In sequential hypothesis testing the goal is to distinguish as soon as possible between two simple hypotheses for the sequentially acquired observations. This problem was introduced by Wald in [63], where he defined and proposed the celebrated Sequential Probability Ratio Test (SPRT).

In the case of independent and identically distributed observations under both hypotheses, Wald and Wolfowitz [64] proved that the SPRT minimizes the expected time for a decision under each hypothesis, while controlling the probabilities of both types of error. Kiefer et.al [5] conjectured that the SPRT enjoys a similar optimality property in the case of continuous-time processes with stationary and independent increments. Shiryaev [51] proved this claim in the Brownian case using the theory of optimal stopping and Irle and Schmitz [18] provided a rigorous proof in the general case.

Liptser and Shiryayev [25] generalized the optimality of the SPRT in the case of continuously observed Itô processes. In particular, they proved that the SPRT minimizes the required *information (in a Kullback-Leibler sense)* for decision under both hypotheses, while controlling the probabilities of both types of error. Irle [17] generalized this optimality for general continuous-path processes and more general criteria.

In this thesis, we focus on decentralized sequential testing. Most of the relevant work in this setup assumes that under each hypothesis the sensors take discrete-time, independent and identically distributed observations and communicate with the fusion center whenever they take an observation.

In this framework, Veeravalli et. al. [61] proposed five different formulations for the decentralized testing problem, depending on the *local memory* that the sensors possess and the *feedback* they may receive from the fusion center. Moreover, the authors explained that all network configurations generate information structures that are intractable to dynamic programming arguments with the only exception the case of *full feedback and local memory restricted to past decisions*, where each sensor remembers only its previous messages to the fusion center but it also has access to the previous messages of all other sensors. In this case, the authors found the optimal Bayesian decentralized test.

The case of no feedback and no local memory was treated in [58], while the case of full local memory with no feedback in [12],[60] (also under a Bayesian setting). However, in the last two cases no exactly optimal decentralized test has been discovered (see [62] for a review). In the case of full local memory, the first asymptotically optimal test was suggested by Mei [31]. According to this scheme, each sensor performs its local SPRT and at every communication with the fusion center it transmits its decision. The fusion center policy follows a consensus rule, according to which the fusion center stops the first time that all sensors agree.

As we mentioned before, all the previous schemes require each sensor to communicate every time it acquires an observation (although Mei's scheme has an asynchronous flavor). This requirement was dropped in the decentralized sequential test proposed by Hussain [16], where level-triggered communication is combined with an "asynchronous" SPRT at the fusion center. Hussain used the term D-SPRT (*decentralized SPRT*) to describe his scheme, however he did not support it with a strong theoretical justification.

Here, we prove that the D-SPRT is asymptotically optimal in the discrete-time setup and we propose a continuous-time version of this decentralized test, which we also prove to be asymptotically optimal. It should be underlined that while the communication scheme for the D-SPRT is the same both in discrete and continuous time, the fusion center policy that we suggest in continuous-time is essentially *model-free*, since it relies on the path-continuity of the observed sensor process and not on their dynamics. Thus, we are able to extend the properties of the D-SPRT for more general communication schemes and observation models.

In order to state more precisely our results, we introduce the generic parameters γ and Δ, which control the horizon of observations and the frequency of communication respectively, so that $\gamma \to \infty$ implies a long horizon of observations, whereas $\Delta \to \infty (\Delta \to 0)$ implies a rare (frequent) communication between sensors and fusion center.

Thus, when the sensor processes are independent and each sensor observes a sequence of independent and identically distributed random variables,

- we prove that the D-SPRT is order-1 asymptotically optimal as $\gamma \to \infty$ and $\Delta \to \infty$ so that $\Delta = o(\gamma)$ with the optimal rate being $\Delta = \mathcal{O}(\sqrt{\gamma})$.
- we show that if the sensors observe independent Brownian motions only at discrete times with a common sampling period h, then – for any fixed Δ– the D-SPRT is order-2 asymptotically optimal as $\gamma \to \infty$ and $h \to 0$ so that $h^{1/4} = \mathcal{O}(\gamma^{-1})$.
- we present simulation experiments which verify the above findings and show that the D-SPRT performs better than other decentralized schemes in the literature.

When the underlying sensor processes are independent Itô processes or correlated Brownian motions,

- we prove that the continuous-time D-SPRT is order-2 asymptotically optimal as $\gamma \to \infty$ for *any* fixed Δ.

- we present simulation experiments which show that the continuous-time D-SPRT performs better than then discrete-time SPRT when the underlying sensor processes are independent Brownian motions.

1.4 Quickest detection

The goal in statistical quality control is to monitor and improve the productivity of industrial process. This area of statistics has dominated applications since the introduction of the Shewhart charts [50] in the 1920s and has contributed significantly to technological innovations and the increase in productivity since then. The main method in statistical quality control is quickest detection or change-detection. There, it is assumed that the distribution of the observed process changes at some unknown time and the goal is to detect the change as soon as possible using the sequentially acquired observations. For reviews on statistical control theory and quickest detection we refer to [54] , [13] , [23], [4] , [42].

In this thesis, we focus on the CUSUM (Cumulative Sums) test, which has been one of the most popular detection rules, both in theory and in application, since its introduction by Page [39]. A strong theoretical argument in favor of the CUSUM test was provided by Moustakides [33], who proved that the CUSUM is the optimal detection rule according to the minimax criterion suggested by Lorden [28] in the case of independent and identically distributed observations before and after the change. In particular, it was shown in [33] that the CUSUM rule minimizes the worst-case conditional expected delay given the worst possible history up to the time of the change among detection rules with a period of false alarms larger than a pre-specified constant.

In discrete time, the (exact) Lorden-optimality of the CUSUM rule was generalized by Moustakides in [37] for a certain class of processes with dependent observations. Lai [24] established the asymptotic optimality of the CUSUM test for general observation models. In continuous time, Shiryaev [52] proved the Lorden-optimality of the CUSUM rule in the Brownian case and Moustakides [34] extended

–under a modified Lorden's criterion– the CUSUM optimality in the case of Itô processes.

Unlike the SPRT which is also optimal under the Bayesian formulation of sequential testing, the CUSUM is not optimal under the Bayesian formulation of change-detection [51], where it is assumed that the change-point is a geometric random variable independent of the observations. Actually, the CUSUM does not optimize neither the other popular minimax criterion in the literature due to Pollack [30]. For an insightful discussion regarding these criteria we refer to [36].

As it is the case for decentralized sequential testing, the decentralized change-detection problem is typically set in a discrete-time setup and it is assumed that the sensors communicate with the fusion center every time they take an observation. Again, we can have different formulations depending on the local memory and the feedback that is available at the sensors [62].

Veeravalli [62] solved the change detection problem in a Bayesian setting in the case of full feedback and local memory restricted to past decisions. The corresponding results under a non-Bayesian formulation are given by Moustakides in [35], where the fusion center policies are CUSUM statistics. In the case of no-local memory and no feedback from the fusion center, Tartakovsky and Veeravalli [56] proposed and studied decentralized schemes that use threshold quantization at the sensors and CUSUM detection rules at the fusion center.

In the case of full-local memory, Mei [32] proposed an asymptotically optimal detection rule, according to which each sensor performs its local CUSUM rule and at every communication with the fusion center it transmits its decision. The fusion center decision rule is a consensus rule, according to which the fusion center raises an alarm when all sensors agree that the change has occurred. Comparisons of this scheme with other decentralized and centralized alternatives are reported in [55], where it is shown that Mei's scheme may perform worse in practice than asymptotically sub-optimal schemes.

In this thesis, we propose a novel decentralized detection rule, both in a continuous-time and a discrete-time framework. The suggested scheme has the same communication scheme as the D-SPRT but uses a CUSUM rule at the fusion center, thus we call it D-CUSUM (decentralized CUSUM).

In order to state more precisely our results, we recall the generic parameters γ and Δ, where $\gamma \to \infty$ implies a large horizon of observations and $\Delta \to \infty (\Delta \to 0)$ implies very rare (frequent) communication between sensors and fusion center. Then, when the sensor processes are independent and each sensor observes a sequence of independent and identically distributed random variables both before and after the change, we prove that the D-CUSUM is order-1 asymptotically optimal as $\gamma \to \infty$ and $\Delta \to \infty$ so that $\Delta = o(\gamma)$, with the optimal rate being $\Delta = \mathcal{O}(\sqrt{\gamma})$. Moreover, we show that if the sensors observe independent Brownian motions only at discrete times with a common sampling period h , then the D-CUSUM is order-2 asymptotically optimal as $\gamma \to \infty$ and $h \to 0$ so that $h^{1/4} = o(\gamma^{-1})$, while Δ is fixed.

Moreover, when the underlying sensor process are independent Itô processes or correlated Brownian motions both before and after the change, we prove that the continuous-time D-CUSUM is order-2 asymptotically optimal as $\gamma \to \infty$ for *any* fixed Δ and order-1 asymptotically optimal as $\gamma \to \infty$ and $\Delta \to \infty$ so that $\Delta = o(\gamma)$.

1.5 Sequential parameter estimation

The statistical inference of continuous-time stochastic processes has drawn the attention of many probabilists and statisticians ([20], [43]). When one observes independent Brownian motions with the same unknown drift, it is well-known that the Maximum Likelihood Estimator (MLE) of the drift is unbiased, Gaussian and has the smallest possible mean square error. However, these properties are not preserved when the drifts are random.

In particular, when all drifts are linear with respect to an unknown and common parameter, the MLE of this parameter can be computed explicitly, but is no longer unbiased, Gaussian or optimal –in a mean square error sense– for finite horizons.

However, as Liptser and Shiryayev showed in [25], a *sequential* version of the MLE recovers these properties in this more general framework. Thus, even though this parameter estimation problem may not be sequential by nature, a sequential formulation is more natural, because it leads to an elegant, intuitive and efficient solution in contrast to a fixed-horizon formulation.

In this thesis we focus on decentralized parameter estimation in the above framework. In particular, we suggest and analyze a novel decentralized sequential estimator, which is based on level-triggered communication at the sensors and mimics the MLE at the fusion center.

In order to state more precisely the properties of this estimator, which we call decentralized MLE (D-MLE), we recall the generic parameters γ and Δ, where $\gamma \to \infty$ implies a large horizon of observations and $\Delta \to \infty (\Delta \to 0)$ very rare (frequent) communication between sensors and fusion center. Then, we prove that as $\gamma \to \infty$ and $\Delta \to \infty$ the D-MLE is a consistent estimator as $\Delta = o(\gamma)$ and asymptotically normal and efficient (in a mean-square-error sense) as $\gamma \to \infty$ and $\Delta \to \infty$ as $\Delta = o(\sqrt{\gamma})$.

Finally, we suggest a modification of the D-MLE, which relies on between-sensor communication, when the sensors observe correlated diffusion processes. For additional literature on decentralized (non-sequential) parameter estimation we refer to [46], [14], [15], [10].

1.6 Wald's identities

Throughout this thesis we apply the celebrated Wald's identities –as well as generalized versions of the classical identities– in order to analyze the suggested decentralized sequential schemes. For that reason we present here these identities in the forms that we use them both in discrete and continuous time. For a review of Wald's identies and their applications to sequential analysis we refer to [21].

1.6.1 Discrete-time techniques

1.6.1.1 Wald's identities

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space which hosts a sequence of independent and identically distributed random variables $\{X_n\}$ with $\mathsf{E}[|X_1|] < \infty$. We denote by $\{\mathcal{F}_n^X\}$ the filtration generated by $\{X_n\}$, i.e. $\mathcal{F}_n^X = \sigma(X_1, \ldots, X_n)$, and we $S_n = X_1 + \ldots + X_n$ and $\mu = \mathsf{E}[X_1]$. If additionally $\mathsf{E}[X_1^2] < \infty$, we set $\sigma^2 \equiv \mathsf{E}[(X_1 - \mu)^2]$.

If $\mathcal{T}$ is an *integrable* $\{\mathcal{F}_n^X\}$-stopping time, then:

$$\mathsf{E}[S_{\mathcal{T}}] = \mu\, \mathsf{E}[\mathcal{T}] \quad , \quad \mathsf{E}[(S_{\mathcal{T}} - \mu\, \mathcal{T})^2] = \sigma^2\, \mathsf{E}[\mathcal{T}]. \tag{1.5}$$

where the second identity is of course valid as long as $\sigma^2 < \infty$. An example of such an integrable stopping time is the first time the random walk $\{S_n\}$ exits the interval $(-A, B)$, i.e. $\mathcal{T} = \inf\{n \in \mathbb{N} : S_n \notin (-A, B)\}$ The integrability of this stopping time is based on the fact that for every $n \in \mathbb{N}$ there exist constants $C > 0$ and $0 < \rho < 1$ (independent of n) so that $\mathsf{P}(\mathcal{T} > n) \leq C\rho^n$.

1.6.1.2 Change of measure

Let $(\Omega, \mathcal{F})$ be a measurable space which hosts a sequence of random variables $(X_n)_{n \in \mathbb{N}}$. We now consider two probability measures P and $\tilde{\mathsf{P}}$, so that the joint probability density function of the random vector $(X_1, \ldots, X_n)$ is $f_n(x_1, \ldots, x_n)$ under P and $g_n(x_1, \ldots, x_n)$ under $\tilde{\mathsf{P}}$, i.e. for $n \in \mathbb{N}$

$$(X_1, \ldots, X_n) \sim \begin{cases} f_n(x_1, \ldots, x_n) & \text{under } \mathsf{P} \\ g_n(x_1, \ldots, x_n) & \text{under } \tilde{\mathsf{P}} \end{cases}$$

We assume that for every n the functions f_n and g_n have common support, thus we can define the likelihood-ratio process:

$$L_0 = 1 \quad , \quad L_n = \frac{g_n(X_1, \ldots, X_n)}{f_n(X_1, \ldots, X_n)}, \quad n \geq 1 \tag{1.6}$$

Then, Wald's likelihood ratio identity (or else the fundamental identity of sequential analysis) states that for an arbitrary $\{\mathcal{F}_n^X\}$-stopping time $\mathcal{T}$ we have:

$$\tilde{\mathsf{E}}[Y \quad {}_{\{\mathcal{T}<\infty\}}] = \mathsf{E}[Y\, L_{\mathcal{T}} \quad {}_{\{\mathcal{T}<\infty\}}] \tag{1.7}$$

where Y is any $\mathcal{F}_{\mathcal{T}}$-measurable random variable so that the right-hand side in (1.7) is finite.

Notice that it is not required that the random variables $\{X_n\}$ be independent and identically distributed under P or $\tilde{\mathsf{P}}$. Moreover, the stopping time $\mathcal{T}$ may be finite under only one of the two measures, i.e. we may have $\tilde{\mathsf{P}}(\mathcal{T} < \infty) < 1 = \mathsf{P}(\mathcal{T} < \infty)$. Finally, from (1.7) we can deduce that the likelihood ratio $\{L_n\}$ is a $(\mathsf{P}, \{\mathcal{F}_n^X\})$- martingale. Indeed, if we set $Y = 1$ we obtain $\mathsf{E}[L_{\mathcal{T}}] = 1 = \mathsf{E}[L_0]$ for every *bounded* $\{\mathcal{F}_n^X\}$- stopping time $\mathcal{T}$, which is a characterization of the martingale property.

1.6.2 Continuous-time techniques

1.6.2.1 Wald's identities

Let $(\Omega, \mathcal{F}, \mathsf{P}, \{\mathcal{F}_t\})$ be a filtered probability space which hosts a zero-mean local martingale $\{M_t\}_{t\geq 0}$ and a finite-variation process $\{A_t\}_{t\geq 0}$, which both have continuous paths. We also denote by $\{\langle M\rangle_t\}$ the quadratic variation process of $\{M_t\}$, i.e. the unique, $\{\mathcal{F}_t\}$-adapted, increasing, continuous process which makes $M^2 - \langle M\rangle$ is a continuous local martingale. We set $X_t = A_t + M_t$, $t \geq 0$.

If $\mathcal{T}$ is an $\{\mathcal{F}_t\}$-stopping time so that $\mathsf{E}[\langle M\rangle_{\mathcal{T}}] < \infty$, then we have:

$$\mathsf{E}[X_{\mathcal{T}}] = \mathsf{E}[A_{\mathcal{T}}] \quad , \quad \mathsf{E}[(X_{\mathcal{T}} - A_{\mathcal{T}})^2] = \mathsf{E}[\langle M\rangle_{\mathcal{T}}] \tag{1.8}$$

When $\{X_t\}$ is a Brownian motion with drift μ and diffusion coefficient σ, i.e. $X_t = \mu t + \sigma W_t$, then (1.8) implies that for any *integrable* $\{\mathcal{F}_t\}$-stopping time $\mathcal{T}$ we have:

$$\mathsf{E}[X_{\mathcal{T}}] = \mu\, \mathsf{E}[\mathcal{T}] \quad , \quad \mathsf{E}[(X_{\mathcal{T}} - \mu\, \mathcal{T})^2] = \sigma^2 \mathsf{E}[\mathcal{T}] \tag{1.9}$$

which is the direct generalization of (1.5).

1.6.2.2 Change of measure

We now restrict ourselves to the class of Itô processes, since we will focus on this class of stochastic processes later. Thus, let $(\Omega, \mathcal{F}, \mathsf{P}, \{\mathcal{F}_t\}_{t\geq 0})$ be a filtered probability space which hosts the standard Brownian motion $\{W_t\}$ and the processes $\{X_t\}$.

We say that $\{X_t\}$ is an *Itô process* (relative to the Brownian motion $\{W_t\}$), if there exist stochastic processes $\{b_t\}$ and $\{\sigma_t\}$, which are $\mathcal{B}[0,\infty)\times\mathcal{F}$-measurable, $\{\mathcal{F}_t\}$-adapted, satisfy

$$\mathsf{P}\Big(\int_0^t |b_s|\,ds < \infty,\ t \geq 0\Big) = \mathsf{P}\Big(\int_0^t |\sigma_s|^2\,ds < \infty,\ t \geq 0\Big) = 1 \tag{1.10}$$

and

$$\mathsf{P}\Big(X_t = \int_0^t b_s\,ds + \int_0^t \sigma_s\,dW_s, \quad t \geq 0\Big) = 1.$$

In order to avoid unnecessary technicalities we assume that for each $t \in [0,T]$ the random variables b_t and σ_t are functionals of the path of X up to time t, i.e. the processes $\{b_t\}$ and $\{\sigma_t\}$ are $\{\mathcal{F}_t^X\}$-adapted, where $\mathcal{F}_t^X = \sigma(X_s,\ 0 \leq s \leq t)$. Moreover, we assume that $(\Omega,\mathcal{F})$ is the space of continuous functions on $[0,\infty)$ accompanied by its Borel σ-algebra and $\{\mathcal{F}_t\}$ the filtration generated by the coordinate process $W_t(\omega) = \omega(t),\ \omega \in \Omega$. Therefore, P is the Wiener measure, the unique probability measure on the canonical path space under which the coordinate process $\{W_t\}$ is a Brownian motion.

Suppose now that there exist $\mathcal{B}[0,\infty)\times\mathcal{F}$-measurable, $\{\mathcal{F}_t\}$-adapted stochastic processes $\{\tilde{b}_t\}$ and $\{\theta_t\}$, so that $\sigma_t\theta_t = (\tilde{b}_t - b_t),\ t \geq 0$ and

$$\mathsf{P}\Big(\int_0^t |\tilde{b}_s|\,ds < \infty,\ t \geq 0\Big) = \mathsf{P}\Big(\int_0^t |\theta_s|^2\,ds < \infty,\ t \geq 0\Big) = 1 \tag{1.11}$$

and consider the positive $(\mathsf{P},\mathcal{F}_t)$-local martingale

$$L_t \equiv \exp\Big\{\int_0^t \theta_s\,dW_s - 0.5\int_0^t |\theta_s|^2\,ds\Big\}, \quad t \geq 0, \tag{1.12}$$

If $\{L_t\}$ is an $\{\mathcal{F}_t\}$-martingale, i.e. $\mathsf{E}[L_t] = 1, t \geq 0$, then we can define the probability measure $\tilde{\mathsf{P}}(A) = \mathsf{E}[L_t\mathcal{I}(A)],\ A \in \mathcal{F}_t$ and according to Girsanov's theorem there exists a $(\tilde{\mathsf{P}},\{\mathcal{F}_t\})$-Brownian motion $\{\tilde{W}_t\}$ so that

$$W_t = \int_0^t \theta_s\,ds + \tilde{W}_t \quad , \quad X_t = \int_0^t \tilde{b}_s\,ds + \int_0^t \sigma_s\,d\tilde{W}_s \tag{1.13}$$

The stochastic process $\{L_t\}$ is clearly a $(\mathsf{P},\mathcal{F}_t)$-martingale when the processes $\{b_t\}$, $\{\tilde{b}_t\}$ $\{\sigma_t\}$ reduce to constants and in general when they satisfy a Novikov or

Kazamaki condition, i.e.

$$\mathsf{E}[e^{0.5\int_0^t |\theta_s|^2\,ds}] < \infty \quad \text{or} \quad \mathsf{E}[e^{0.5\int_0^t \theta_s\,dW_s}] < \infty, \quad 0 \le t < \infty. \tag{1.14}$$

In order to obtain a version of Wald's likelihood ratio identity in this context, it is not necessary to assume (1.14), but it suffices to consider $\{\mathcal{F}_t\}$-stopping times $\mathcal{T}$, so that

$$\tilde{\mathsf{P}}\Big(\int_0^{\mathcal{T}} |\theta_s|^2\,ds < \infty\Big) = \mathsf{P}\Big(\int_0^{\mathcal{T}} |\theta_s|^2\,ds < \infty\Big) = 1, \quad t \ge 0, \tag{1.15}$$

Then:

$$\tilde{\mathsf{E}}[Y_{\mathcal{T}}] = \mathsf{E}[L_{\mathcal{T}}\,Y_{\mathcal{T}}] \tag{1.16}$$

where $\{Y_t\}$ is any $\{\mathcal{F}_t\}$-adapted process so that the above expectations are well-defined.

Notice that when the processes $\{b_t\}$, $\{\sigma_t\}$, $\{\tilde{b}_t\}$ reduce to constants, condition (1.15) implies that $\mathcal{T}$ must be finite under both P and $\tilde{\mathsf{P}}$.

Finally, if

$$\mathsf{P}\Big(\int_0^{\infty} \theta_s^2\,ds = \infty\Big) = \tilde{\mathsf{P}}\Big(\int_0^{\infty} \theta_s^2\,ds = \infty\Big) = 1 \tag{1.17}$$

then the localizing stopping times

$$\mathcal{T}_n = \inf\Big\{t \ge 0 : \int_0^t \theta_s^2\,ds \ge n\Big\} \tag{1.18}$$

clearly satisfy (1.15) and so does the stopping time

$$\mathcal{S} = \inf\{t \ge 0 : L_t \notin (-A, B)\}. \tag{1.19}$$

Again, when the processes $\{b_t\}$, $\{\sigma_t\}$, $\{\tilde{b}_t\}$ are constants, condition (1.17) is trivially satisfied.

Chapter 2

Decentralized sequential hypothesis testing

The structure of this chapter is as follows: we start with a review of centralized sequential testing with an emphasis on the Sequential Probability Ratio Test (SPRT). We then define and analyze the proposed decentralized sequential test; first in continuous time, when the observed process has continuous paths (Sec. 2.2) and when it has discontinuous paths (Sec. 2.3) and then in the discrete time case (Sec. 2.4).

2.1 Sequential testing under a centralized setup

Let $\{\xi_t = [\xi_t^1, \ldots, \xi_t^K]'\}_{t\geq 0}$ be a K-dimensional stochastic process, each component of which is observed *sequentially* at a different location or sensor. The flow of information locally at sensor i is described by the filtration $\{\mathcal{F}_t^i\}_{t\geq 0}$, whereas the flow of information at the whole sensor network is described by $\{\mathcal{F}_t\}_{t\geq 0}$, where

$$\mathcal{F}_t^i = \sigma(\xi_s^i,\, 0 \leq s \leq t) \quad , \quad \mathcal{F}_t = \sigma(\xi_s^i,\, 0 \leq s \leq t, i = 1, \ldots, K), \quad t \geq 0. \tag{2.1}$$

We assume that the sensors acquire their observations *sequentially* and that they communicate with a fusion center, as Fig. 2.1 suggests. The fusion center combines the information it receives from all sensors and is responsible for making the final decision.

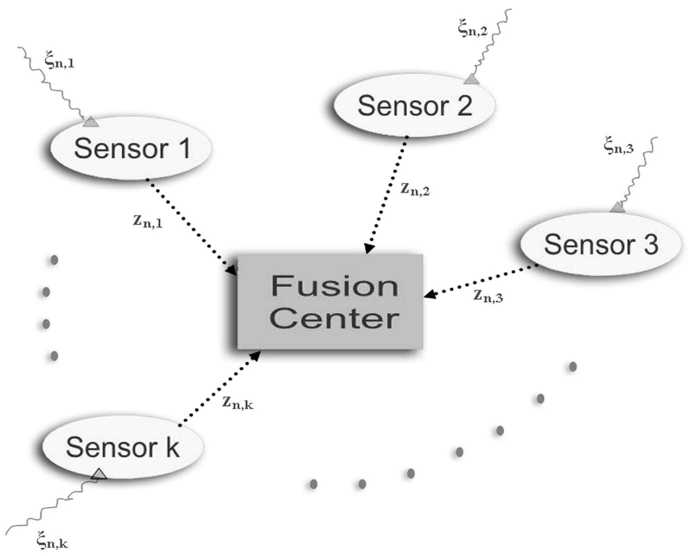


Figure 2.1: Sensor Network

We denote by P the distribution of the observed process $(\xi^1, \ldots, \xi^K)$ and we consider two *simple* hypotheses for P, i.e.

$$\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0 \quad , \quad \mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1 \, , \tag{2.2}$$

We assume that the probability measures P_0, P_1 are locally equivalent, but globally singular, in the sense that they are equivalent when they are restricted to the σ-algebra $\mathcal{F}_t$ for any $t \in [0, \infty)$, but singular when restricted to $\mathcal{F}_\infty$. We denote by $\{u_t\}$ the corresponding log-likelihood ratio process, i.e.

$$u_0 = 1 \quad , \quad u_t = \log \frac{d\mathsf{P}_1}{d\mathsf{P}_0}\Big|_{\mathcal{F}_t}, \quad 0 < t < \infty \tag{2.3}$$

The goal of the decision maker at the fusion center is to choose the correct hypothesis combining in an efficient way the information it receives from the sensors. Moreover, due to the sequential nature of the observations, the goal is to choose the correct hypothesis *as soon as possible*.

Thus, a *sequential test* consists of a *stopping rule*, which determines when the fusion center stops collecting data from the sensors, and a *decision rule* which determines which of the two hypotheses should be chosen. Since these rules cannot anticipate the future and must rely exclusively on the fusion center data, the stopping rule will be an $\{\tilde{\mathcal{F}}_t\}$-*stopping time* and the decision rule will be $\tilde{\mathcal{F}}_{\mathcal{T}}$-measurable *binary* random variable, where denote by $\{\tilde{\mathcal{F}}_t\}$ the fusion center filtration, which coincides with $\{\mathcal{F}_t\}$ under a centralized setup.

We will focus on the celebrated Sequential Probability Ratio Test (SPRT), which was proposed by Wald in his seminal work [63], and is defined as follows

$$\mathcal{S} = \inf\{t \geq 0 : u_t \notin (-A, B)\}$$
$$d_{\mathcal{S}} = \begin{cases} 1, & \text{if } u_{\mathcal{S}} \geq B \\ 0, & \text{if } u_{\mathcal{S}} \leq -A \end{cases} \tag{2.4}$$

where A, B are two positive constants. Thus, according to the SPRT, the decision maker observes the log-likelihood ratio process $\{u_t\}$ until it crosses either a positive or a negative threshold and chooses H_1 in the first case and H_0 in the second.

We review the properties of the SPRT first in continuous time when the sensors observe Itô processes and jump diffusions and then in discrete time when the sensors observe sequences of independent and identically distributed observations.

2.1.1 The case of Itô processes

2.1.1.1 Problem Formulation

Consider the hypothesis testing problem

$$\mathsf{H}_0 : \xi_t = \int_0^t \sigma_s \cdot dW_s \quad , \quad \mathsf{H}_1 : \xi_t = \int_0^t b_s \, ds + \int_0^t \sigma_s \cdot dW_s \tag{2.5}$$

where $\{W_t\}_{t\geq 0}$ is a K-dimensional Brownian motion, b_t $\{\mathcal{F}_t\}$-adapted K-dimensional vector and $\{\sigma_t\}$ an $\{\mathcal{F}_t\}$-adapted $K \times K$ matrix so that

$$\mathsf{P}_j\Big(\int_0^t \Big[|b_s| + |\sigma_s|^2 + \theta_s \cdot b_s\Big] \, ds < \infty\Big) = 1, \quad t \geq 0 \ , \ j = 0, 1 \tag{2.6}$$

where $\theta_t = b_t' \cdot (\sigma_t^{-1})' \sigma_t^{-1}$. Moreover, for the two measures to be globally singular, we need to assume that

$$\mathsf{P}_j\Big(\int_0^\infty \theta_s \cdot b_s \, ds = \infty\Big) = 1, \quad j = 0, 1. \tag{2.7}$$

The log-likelihood ratio process $\{u_t\}$ takes the form

$$u_t = \int_0^t \theta_s \cdot d\xi_s - 0.5 \int_0^t \theta_s \cdot b_s \, ds \tag{2.8}$$

and we can use it to implement the SPRT, which we defined in (2.4). We present a realization of the SPRT under (2.5) in Fig. 2.2.

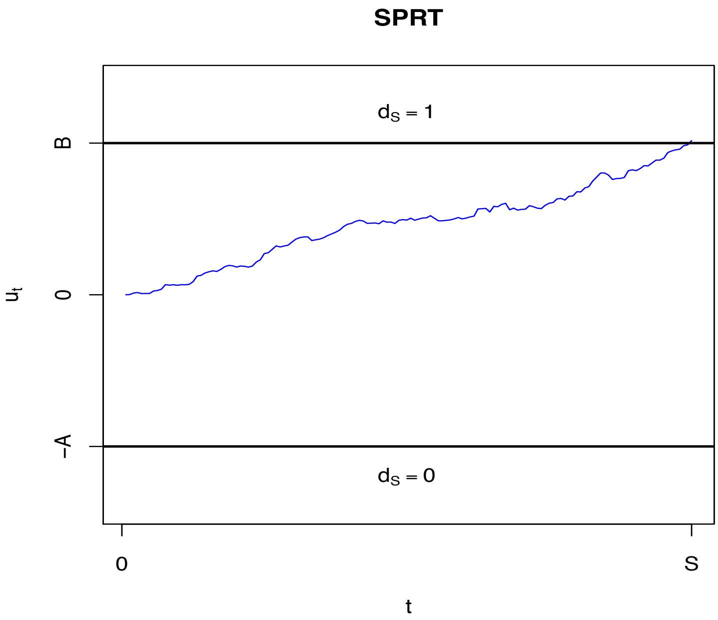


Figure 2.2: A realization of the SPRT

2.1.1.2 SPRT performance characteristics

We can compute explicitly the SPRT error probabilities in terms of the thresholds A, B. Indeed, with a a localization argument it can be shown that:

$$\mathsf{E}_j \Big[\int_0^{\mathcal{S}} \theta_s \cdot b_s \, ds \Big] < \infty, \quad j = 0, 1 \tag{2.9}$$

and that $\mathcal{S}$ is a finite stopping time under both hypotheses. Thus, we can perform change of measure and obtain

$$\mathsf{P}_0(d_{\mathcal{S}} = 1) = \mathsf{E}_1[e^{-u_{\mathcal{S}}} \quad _{\{d_{\mathcal{S}}=1\}}] \quad , \quad \mathsf{P}_1(d_{\mathcal{S}} = 0) = \mathsf{E}_0[e^{u_{\mathcal{S}}} \quad _{\{u_{\mathcal{S}}=-A\}}] \tag{2.10}$$

Moreover, since the process $\{u_t\}$ has continuous paths, we have $\{d_{\mathcal{S}} = 1\} = \{u_{\mathcal{S}} = B\}$ and $\{d_{\mathcal{S}} = 0\} = \{u_{\mathcal{S}} = -A\}$ and consequently

$$\begin{aligned} \mathsf{P}_0(d_{\mathcal{S}} = 1) &= \mathsf{E}_1[e^{-u_{\mathcal{S}}} \quad {}_{\{u_{\mathcal{S}}=B\}}] = e^{-B}\mathsf{P}_1(d_{\mathcal{S}} = 1) \\ \mathsf{P}_1(d_{\mathcal{S}} = 0) &= \mathsf{E}_0[e^{u_{\mathcal{S}}} \quad {}_{\{u_{\mathcal{S}}=-A\}}] = e^{-A}\mathsf{P}_0(d_{\mathcal{S}} = 0) \end{aligned} \tag{2.11}$$

Therefore, we obtain:

$$\mathsf{P}_0(d_{\mathcal{S}} = 1) = \frac{e^B - 1}{e^B - e^{-A}} \quad , \quad \mathsf{P}_1(d_{\mathcal{S}} = 0) = \frac{e^A - 1}{e^A - e^{-B}} \tag{2.12}$$

and

$$B = \log\Big(\frac{1 - \mathsf{P}_0(d_{\mathcal{S}} = 1)}{\mathsf{P}_1(d_{\mathcal{S}} = 0)}\Big) \ , \ A = \log\Big(\frac{1 - \mathsf{P}_1(d_{\mathcal{S}} = 0)}{\mathsf{P}_0(d_{\mathcal{S}} = 1)}\Big). \tag{2.13}$$

Thus, from (2.12) we have:

$$\begin{aligned} \mathsf{E}_1[u_{\mathcal{S}}] &= \mathsf{P}_1(d_{\mathcal{S}} = 1)B - \mathsf{P}_1(d_{\mathcal{S}} = 0)A = s(B, A) \\ -\mathsf{E}_0[u_{\mathcal{S}}] &= \mathsf{P}_0(d_{\mathcal{S}} = 0)A - \mathsf{P}_0(d_{\mathcal{S}} = 1)B = s(A, B) \end{aligned} \tag{2.14}$$

where

$$s(x, y) = \frac{(e^x - 1)x - (1 - e^{-y})y}{e^x - e^{-y}}, \quad x, y > 0 \tag{2.15}$$

2.1.1.3 SPRT Optimality

Consider the optimization problems

$$\inf_{(\mathcal{T}, d_{\mathcal{T}})} \mathsf{E}_1[u_{\mathcal{T}}] \quad \text{and} \quad \inf_{(\mathcal{T}, d_{\mathcal{T}})} \mathsf{E}_0[-u_{\mathcal{T}}] \tag{2.16}$$

where the infimum is taken over centralized sequential tests $(\mathcal{T}, d_{\mathcal{T}})$ which satisfy the following properties:

$$\mathsf{P}_0(d_{\mathcal{T}} = 1) \le \alpha \quad , \quad \mathsf{P}_1(d_{\mathcal{T}} = 0) \le \beta \tag{2.17}$$

where $\alpha, \beta > 0$ with $\alpha + \beta < 1$. Thus, our goal is to find the centralized sequential test that minimizes decision delay under both hypotheses while keeping both type-I and type-II error probabilities bounded by α and β, respectively.

Liptser and Shiryaev proved in [26] – using change of measure and Jensen's inequality– that the SPRT solves concurrently the two optimization problems in

(2.16) as long as the thresholds A, B are chosen so that the error probability constraints in (2.17) are satisfied with equality. This implies that the optimal SPRT thresholds A, B can be expressed in terms of α, β as follows

$$B = \log\Big(\frac{1-\alpha}{\beta}\Big) \ , \ A = \log\Big(\frac{1-\beta}{\alpha}\Big). \tag{2.18}$$

Moreover, the optimal decision delay is given by (2.14) and from (2.14) follows that

$$\mathsf{E}_1[u_{\mathcal{S}}] = |\log\alpha| + o(1) \quad , \quad -\mathsf{E}_0[u_{\mathcal{S}}] = |\log\beta| + o(1) \tag{2.19}$$

as $\alpha, \beta \to 0$ with $\alpha|\log\beta| + \beta|\log\alpha| = o(1)$.

Finally, we should note that both (2.18) and (2.14) are universal, in the sense that they do not depend on the particular form of the dynamics in (2.5), but only on the error probabilities α, β. Fig. 2.3 demonstrates the (universal) average run length (ARL) curve of the SPRT for problem (2.5), that is, the plot of $\mathsf{E}_1[u_{\mathcal{S}}]$ versus $\log\alpha$, where for simplicity we have set $\alpha = \beta$.

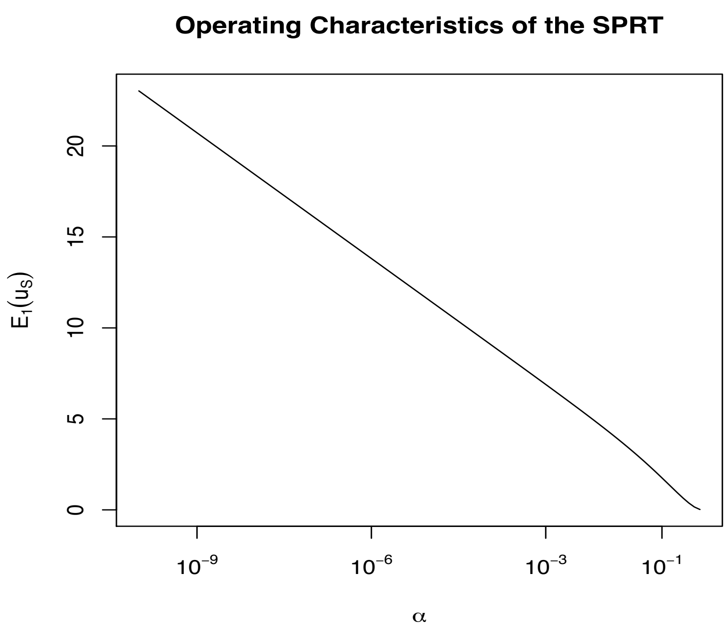


Figure 2.3: ARL curve of the SPRT for Itô processes

2.1.1.4 Connections to other optimality criteria

Problem (2.16) is a generalization of the original formulation of sequential hypothesis testing used in [63], where the goal is to find the sequential test that solves the following minimization problems

$$\inf_{(\mathcal{T},d_{\mathcal{T}})} \mathsf{E}_1[\mathcal{T}] \quad \text{and} \quad \inf_{(\mathcal{T},d_{\mathcal{T}})} \mathsf{E}_0[\mathcal{T}] \tag{2.20}$$

among centralized sequential tests $(\mathcal{T}, d_{\mathcal{T}})$ which satisfy (2.17).

Problems (2.20) and (2.16) are equivalent when $\{b_t\}$ and $\{\sigma_t\}$ reduce to a real vector $(b_1, \ldots, b_K)$ and a real matrix $[\sigma_{ij}]$, respectively, in which case the sensors observe correlated Brownian motions under both hypotheses. The optimality of the SPRT in the framework was established for the first by Shiryaev in [51] using the theory of optimal stopping.

Finally, we should mention that (2.16) is a special case of a more general optimality property of the SPRT, which was revealed by Irle [17] and is valid for even more general testing problems as long as $\{u_t\}$ has continuous paths. Thus, the SPRT solves the following optimization problems

$$\inf_{(\mathcal{T},d_{\mathcal{T}})} \mathsf{E}_1[g(e^{u_{\mathcal{T}}})] \quad \text{and} \quad \inf_{(\mathcal{T},d_{\mathcal{T}})} \mathsf{E}_0[g(e^{u_{\mathcal{T}}})] \tag{2.21}$$

where $g : [0, \infty) \to \mathbb{R}$ is an arbitrary convex function and $(\mathcal{T}, d_{\mathcal{T}})$ are centralized sequential tests that satisfy (2.17) and terminate almost surely under both hypotheses.

2.1.1.5 Sufficient statistics

The optimality properties of the SPRT imply that the log-likelihood ratio process $\{u_t\}$ is a *sufficient* statistic for the hypothesis testing problem (2.5), since the fusion center is able to implement the optimal sequential test without any performance loss *as long as it has full access to* $\{u_t\}$.

Moreover, from (2.8) follows that $u = \sum_{i=1}^{K} u^i$, where

$$u_t^i = \int_0^t \theta_s^i \, d\xi_s^i - 0.5 \int_0^t \theta_s^i \, b_s^i \, ds, \quad 0 \le t < \infty \tag{2.22}$$

where we denote by b^i, θ^i_t the i^{th} components of the vectors b_t, θ_t. Therefore, $(u^1, \ldots, u^K)$ is a vector of sufficient statistics, since for the implementation of the SPRT it suffices that the fusion center receives $\{u^1_t, \ldots, u^K_t\}$ instead of the raw sensor observations $\{\xi^1_t, \ldots, \xi^K_t\}$. Thus, if each process $\{u^i_t\}$ is $\{\mathcal{F}^i_t\}$-adapted, then sensor i could observe and transmit to the fusion center the path of $\{u^i_t\}$ instead of the path of $\{\xi^i_t\}$.

Of course, the process $\{u^i_t\}$ is not always $\{\mathcal{F}^i_t\}$-adapted in the general framework of the testing problem (2.5), but only in some (important) special cases. Indeed, this is true when the sensor processes $\xi^1, \ldots, \xi^K$ are independent, in which case $\{u^i_t\}$ corresponds to the marginal log-likelihood ratio of $\{\xi^i_t\}$, but also when the sensors observe *correlated* Brownian motions under each hypothesis, in which case each u^i_t is a deterministic function of ξ^i_t and t.

2.1.2 SPRT for jump-diffusions

2.1.2.1 Problem formulation

We now assume that the observed sensor at each sensor i admits the following decomposition:

$$\xi^i_t = \sigma^i Y^i_t + \sum_{j=1}^{N^i_t} X^i_j \ , \ Y^i_t = b^i t + W^i_t, \quad t \geq 0 \tag{2.23}$$

where $\{Y^i_t\}$ is a Brownian motion with drift $b^i\sigma^i$ and diffusion coefficient σ^i, $\{N^i_t\}$ a Poisson process with intensity λ^i and (X^i_n) a sequence of independent random variables with common density f^i. We assume that $\{W^i_t\}$, $\{N^i_t\}$ and $\{X^i_n\}$ are independent, therefore each process $\{\xi^i_t\}$ has stationary and independent increments. In particular, $\{\xi^i_t\}$ is a jump-diffusion process if $\sigma^i > 0$ and a compound Poisson process if $\sigma^i = 0$. We also assume that the triplets $(\{W^i_t\}, \{N^i_t\}, \{X^i_n\})$, $i = 1, \ldots, K$ are independent.

We consider the following testing problem

$$\mathsf{H}_0 : (b^i, \lambda^i f^i) = (b^i_0, \lambda^i_0, f^i_0), \ \forall i \quad , \quad \mathsf{H}_1 : (b^i, \lambda^i f^i) = (b^i_1, \lambda^i_1, f^i_1), \ \forall i \tag{2.24}$$

where λ_0^i, λ_1^i are known positive constants, b_0^i, b_1^i known real constants and $f_0^i(\cdot), f_1^i(\cdot)$ known densities with common support for each $i = 1, \ldots, K$.

Due to the assumption of independence across sensors, the log-likelihood ratio process $\{u_t\}$ admits the decomposition $u_t = \sum_{i=1}^K u_t^i$ for any $t \geq 0$, where

$$\begin{aligned} u_t^i &= [\mu_i Y_t^i - 0.5|\mu_i|^2 t] + [\rho_i N_t^i - (\lambda_1^i - \lambda_0^i)t] + \sum_{l=1}^{N_t^i} h_i(X_l^i) \\ &= \mu_i Y_t^i - \Big[0.5|\mu_i|^2 + (\lambda_1^i - \lambda_0^i)\Big] t + \sum_{l=1}^{N_t^i} \Big[\rho_i + h_i(X_l^i)\Big], \end{aligned}$$

where

$$\rho_i = \log \frac{\lambda_1^i}{\lambda_0^i} \quad , \quad h_i(\cdot) = \log \frac{f_1^i(\cdot)}{f_0^i(\cdot)} \tag{2.25}$$

and

$$\mu_i = \begin{cases} \frac{b_1^i - b_0^i}{\sigma^i}, & \text{if } \sigma_i > 0 \\ 0, & \text{if } \sigma_i = 0 \end{cases}$$

The optimality of the SPRT in this setup (and generally in the case of processes with stationary and independent increments) was conjectured initially by Kiefer et.al [5], but a formal proof of this statement was given by Irle and Schmitz in [18]. In particular, it was shown that the SPRT solves the optimization problems in (2.20) among sequential tests that satisfy (2.17). Notice that since $\mathsf{E}_1[u_t], \mathsf{E}_0[-u_t]$ are linear functions of t under the hypothesis testing problem (2.24), thus the optimization problems (2.16) and (2.20) are equivalent.

Since the log-likelihood ratio process $\{u_t\}$ for problem (2.24) does not have continuous paths, the random variable $u_{\mathcal{S}}$ is no longer binary, since it may exceed the thresholds $-A$ and B. Thus, the *overshoot* $\eta = (u_{\mathcal{S}} - B)^+ - (u_{\mathcal{S}} + A)^-$ is a non-trivial random variable and we cannot derive closed-form expressions for the thresholds A and B and the SPRT performance in terms of the error probabilities α, β.

Nevertheless, we can show that $B \leq |\log \alpha|$ and $A \leq |\log \beta|$ and we also have that

$$\mathsf{E}_1[u_{\mathcal{S}}] = (1-\beta)B - \beta A + \mathsf{E}_1[\eta] \quad , \quad \mathsf{E}_0[-u_{\mathcal{S}}] = (1-\alpha)A - \alpha B + \mathsf{E}_0[\eta] \tag{2.26}$$

Therefore, if the expected overshoots $\mathsf{E}_j[\eta]$, $j = 0, 1$ are asymptotically bounded, i.e. $\mathsf{E}_j[\eta] = \mathcal{O}(1)$, $j = 0, 1$ as $\alpha, \beta \to 0$, then we obtain

$$\mathsf{E}_1[u_{\mathcal{S}}] \leq |\log \alpha| + \mathcal{O}(1) \quad , \quad \mathsf{E}_0[-u_{\mathcal{S}}] \leq |\log \beta| + \mathcal{O}(1) \tag{2.27}$$

which is an asymptotic upper bound for the performance of the SPRT under (2.24) and will turn out to be sufficient for our purposes.

2.1.3 SPRT in discrete-time

Suppose that each sensor i acquires sequentially the discrete-time observations ξ_t^i, $t = 0, 1, 2, \ldots$. We assume that $(\xi_t^1, \ldots, \xi_t^K)_{t \in \mathbb{N}}$ is a sequence of independent random vectors so that for each $t \in \mathbb{N}$:

$$\mathsf{H}_0 : (\xi_t^1, \ldots, \xi_t^K) \sim \mathsf{Q}_0 \quad , \quad \mathsf{H}_1 : (\xi_t^1, \ldots, \xi_t^K) \sim \mathsf{Q}_1 \tag{2.28}$$

where Q_0 and Q_1 are known Borel probability measures on $\mathbb{R}^K$. We assume that there is a probability measure Q that dominates both Q_0 and Q_1 and we denote by f_0 and f_1 the corresponding Radon-Nikodym derivatives. Then, the log-likelihood ratio process that corresponds to this hypothesis-testing problem takes the form:

$$u_t = \sum_{l=1}^{t} \log \frac{f_1(\xi_l^1, \ldots, \xi_l^K)}{f_0(\xi_l^1, \ldots, \xi_l^K)}, \quad t = 0, 1, 2, \ldots \tag{2.29}$$

and the SPRT $(\mathcal{S}, d_{\mathcal{S}})$ becomes:

$$\begin{aligned} \mathcal{S} &= \inf\{t \in \mathbb{N} : u_t \notin (-A, B)\} \\ d_{\mathcal{S}} &= \begin{cases} 1, & \text{if } u_{\mathcal{S}} \geq B \\ 0, & \text{if } u_{\mathcal{S}} \leq -A, \end{cases} \end{aligned} \tag{2.30}$$

Wald and Wolfowitz proved in [64] the optimality of the SPRT for the above testing problem in the sense of (2.20). Since the observations $\{(\xi_t^1, \ldots, \xi_t^K)\}_{t \in \mathbb{N}}$ are independent and identically distributed, it is straightforward to see that (2.16) and (2.20) are equivalent optimization criteria.

Moreover, the overshoot $\eta = (u_{\mathcal{S}} - B)^+ - (u_{\mathcal{S}} + A)^-$ is non-zero with probability 1 (see Fig. 2.4). Nevertheless, we can show that $B = \mathcal{O}(|\log \alpha|)$ and

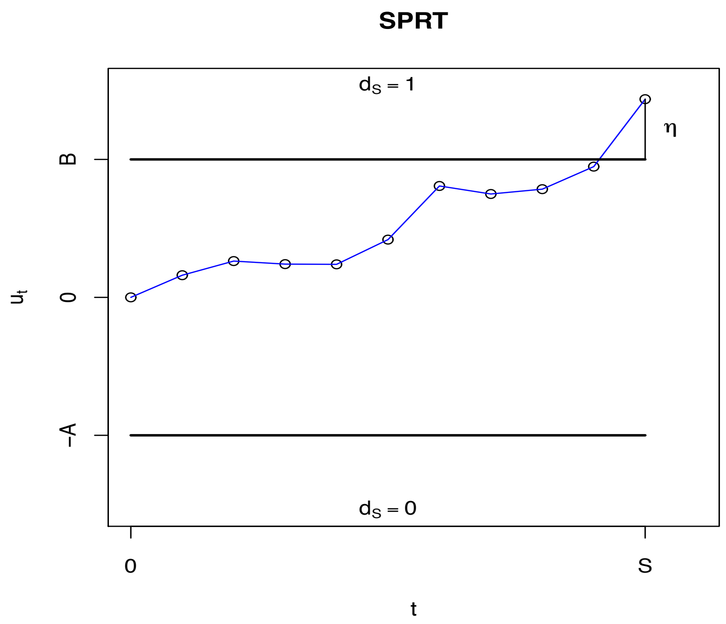


Figure 2.4: Discrete-time SPRT

$A = \mathcal{O}(|\log \beta|)$ as $\alpha, \beta \to 0$ and also

$$\mathsf{E}_1[u_{\mathcal{S}}] \leq |\log \alpha| + \mathcal{O}(1) \quad , \quad \mathsf{E}_0[-u_{\mathcal{S}}] \leq |\log \beta| + \mathcal{O}(1) \tag{2.31}$$

as long as u_1 has a finite second moment, in which case $\mathsf{E}_j[\eta], j = 0, 1$ are bounded *uniformly* in A and B [27]. Finally, notice that the optimality of the SPRT does not require the assumption of independence across sensors.

2.2 Decentralized sequential testing for Itô processes

In this section we propose and study a decentralized sequential test for the hypothesis testing problem (2.5), i.e. when each sensor observes the path of some Itô process under each hypothesis. We refer to the introduction for details on the decentralized setup and the relevant literature.

The fundamental assumption that we make is that each local sufficient statistic $\{u_t^i\}$ -defined in (2.22)- is $\{\mathcal{F}_t^i\}$-adapted. This means that the following will be valid as long as the sensors observe either independent Itô processes or correlated

Brownian motions under each hypothesis (see the 2.1.1.5). This assumption is crucial not only for the properties of the suggested decentralized scheme to hold, but also for the scheme to be implementable in the first place.

Finally, as we also discuss in the introduction, the scheme that we propose and analyze in this section is a continuous-time version of the decentralized sequential test proposed in [16] under a discrete-time setup. Despite the fact that these two versions have some important differences, we use the name D-SPRT (Decentralized SPRT) to describe the suggested scheme in both setups.

2.2.1 Communication scheme and quantization rule

We suggest that sensor i communicates with the fusion center at the $\{\mathcal{F}_t^i\}$-stopping times $(\tau_n^i)_{n\in\mathbb{N}}$ which are defined recursively as follows:

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i))\}, \quad n \in \mathbb{N}. \tag{2.32}$$

The thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ are positive constants, whose values are chosen by the designer of the scheme and are known at the fusion center. Thus, according to (2.32), sensor i monitors its local sufficient statistic u^i and communicates when its value has either increased by $\overline{\Delta}_i$ or decreased by $\underline{\Delta}_i$ in comparison to its value at the previous communication with the fusion center. We illustrate this communication scheme in Fig. 2.5.

Therefore, under (2.32), the number of signals that have been transmitted from sensor i up to time t is random and we will denote it by $m_t^i = \max\{n : \tau_n^i \leq t\}$. Moreover, the sensors do not transmit their signals concurrently to the fusion center, thus (2.32) induces *asynchronous* communication in the sensor network.

The communication scheme (2.32) is naturally coupled with the following quantization rule: *"at any time τ_n^i sensor i should inform the fusion center whether $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \overline{\Delta}_i$ or $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i$."* Sensor i can communicate this information by transmitting the signals:

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \overline{\Delta}_i \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i \end{cases} \tag{2.33}$$

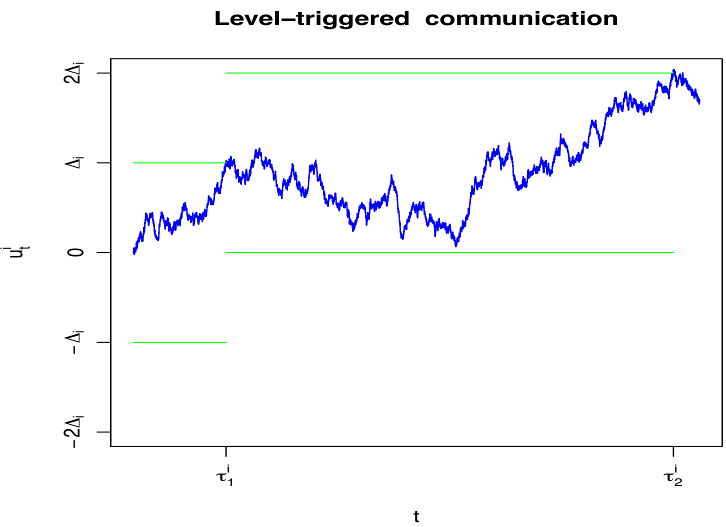


Figure 2.5: Level-triggered communication

and for that purpose it needs only a *binary* alphabet.

Moreover, since each process $\{u^i_t\}$ is continuously observed and has continuous paths, we actually have: $u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}} \in \{\overline{\Delta}_i, -\underline{\Delta}_i\}$ for every $n \in \mathbb{N}$. This means that we can replace the inequalities in (2.33) with equalities. Moreover, it implies that the fusion center can recover the *exact* value of u^i at any communication time τ^i_n, since for every $n \in \mathbb{N}$ we have:

$$u^i_{\tau^i_n} = \sum_{j=1}^{n} \Big[u^i_{\tau^i_j} - u^i_{\tau^i_{j-1}} \Big] = \sum_{j=1}^{n} \Big[\overline{\Delta}_i \, z^i_j - \underline{\Delta}_i \, (1 - z^i_j) \Big]. \tag{2.34}$$

2.2.2 Fusion center policy

Since the fusion center does not receive any information about u^i between communication times, we suggest that it approximates u^i at some arbitrary time t as follows

$$\tilde{u}^i_t = u^i_{\tau^i_n}, \quad t \in [\tau_n, \tau_{n+1}) \tag{2.35}$$

which –due to (2.34)– is equivalent to

$$\tilde{u}^i_t = \sum_{j=1}^{n} \Big[\overline{\Delta}_i \, z^i_j - \underline{\Delta}_i (1 - z^i_j) \Big], \quad t \in [\tau_n, \tau_{n+1}) \tag{2.36}$$

Thus, the process $\{\tilde{u}_t^i\}$ is defined to recover the exact value of u^i at any communication time τ_n^i and to stay flat in between. Consequently, it is a piecewise constant process which jumps at the communication times (τ_n^i) and its jumps are either upward of size $\overline{\Delta}_i$ or downward of size $\underline{\Delta}_i$. Notice that $\{\tilde{u}_t^i\}$ is a *model-free* approximation of $\{u_t^i\}$, since it does not rely on any distributional properties of $\{u_t^i\}$ but only on the continuity of its paths. We illustrate this approximation in Fig. 2.6. The policy that we suggest at the fusion center is simply to replace

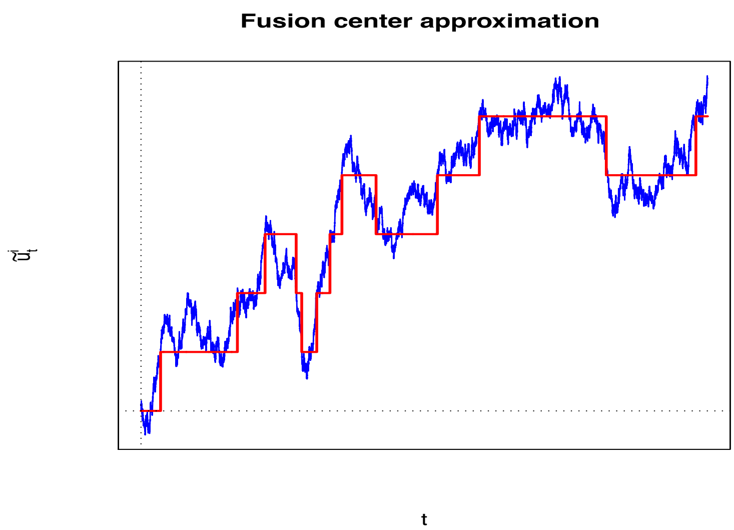


Figure 2.6: Fusion center local approximations

the global likelihood-ratio $u = \sum_{i=1}^K u^i$ in the the definition (2.4) of the SPRT by $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$. In other words, we suggest the following sequential test at the fusion center:

$$\tilde{\mathcal{S}} = \inf\{t \geq 0 : \tilde{u}_t \notin (-\tilde{A}, \tilde{B})\}$$
$$d_{\tilde{\mathcal{S}}} = \begin{cases} 1, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B} \\ 0, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \leq -\tilde{A} \end{cases} \tag{2.37}$$

where again the thresholds $\tilde{A}, \tilde{B}$ are chosen so that the error probability constraints in (2.16) are satisfied with equality. This is a valid decentralized sequential test, since its implementation at the fusion center requires only the transmission of the one-bit data $\{z_n^i\}$ from the sensors. Moreover, since $(\tilde{\mathcal{S}}, d_{\tilde{\mathcal{S}}})$ mimics the SPRT

$(\mathcal{S}, d_{\mathcal{S}})$, we call it *decentralized SPRT* (D-SPRT), adopting the term that was coined in [16] for the discrete-time analogue of this scheme.

2.2.3 Asymptotic optimality of order 2

The (asymptotic) performance of the D-SPRT is characterized by the following proposition:

Proposition 1. *For any fixed values of the sampling thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$, the D-SPRT is asymptotically optimal of order-2 in the sense of (2.16), i.e.*

$$\mathsf{E}_1[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_1[u_{\mathcal{S}}] = \mathcal{O}(1) \quad , \quad \mathsf{E}_0[-u_{\tilde{\mathcal{S}}}] - \mathsf{E}_0[-u_{\mathcal{S}}] = \mathcal{O}(1) \tag{2.38}$$

as $\alpha, \beta \to 0$ so that $\alpha|\log\beta| + \beta|\log\alpha| = \mathcal{O}(1)$. In particular,

$$\mathsf{E}_1[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_1[u_{\mathcal{S}}] \leq 3C + o(1) \quad , \quad \mathsf{E}_0[-u_{\tilde{\mathcal{S}}}] - \mathsf{E}_0[-u_{\mathcal{S}}] \leq 3C + o(1) \tag{2.39}$$

where $C = \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i)$.

It is an immediate corollary that in the Brownian case where the processes $\{b_t\}$ and $\{\theta_t\}$ reduce to real vectors $[b_1, \ldots, b_K]'$ and $[\theta_1, \ldots, \theta_K]'$, the D-SPRT is order-2 asymptotically optimal also in the sense of (2.20). In particular, we have:

$$\mathsf{E}_j[\tilde{\mathcal{S}}] - \mathsf{E}_j[\mathcal{S}] \leq \frac{3C}{0.5\sum_{i=1}^{K}\theta_i\, b_i} + o(1), \quad j = 0, 1. \tag{2.40}$$

Proof. We start by observing that for any time t we have:

$$|u_t - \tilde{u}_t| \leq C \quad , \quad |\tilde{u}_t - \tilde{u}_{t-}| \leq C. \tag{2.41}$$

The first inequality in (2.41) implies that $|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}| \leq C$. The second inequality provides a bound on the size of the jumps of the piecewise constant process $\tilde{u}$ and it implies that

$$\tilde{u}_{\tilde{\mathcal{S}}} - \tilde{B} \leq C \quad , \quad \tilde{u}_{\tilde{\mathcal{S}}} + \tilde{A} \geq -C, \tag{2.42}$$

since $\tilde{\mathcal{S}}$ corresponds to a jump time of $\tilde{u}$.

From the first inequality in (2.41) it follows that $\tilde{\mathcal{S}}$ is upper bounded pathwise by the following stopping time: $\mathcal{R} = \inf\{t \geq 0 : u_t \notin (-\tilde{A} - C, \tilde{B} + C)\}$ and consequently we have:

$$\mathsf{E}_j\Big[\int_0^{\tilde{\mathcal{S}}} \theta_s \cdot b_s \, ds\Big] \leq \mathsf{E}_j\Big[\int_0^{\mathcal{R}} \theta_s \cdot b_s \, ds\Big] < \infty, \quad j = 0, 1 \tag{2.43}$$

The first inequality in (2.43) follows from $\tilde{\mathcal{S}} \leq \mathcal{R}$ and the positivity of the quadratic form $\theta_t \cdot b_t = b_t' \cdot (\sigma_t^{-1})' \sigma_t^{-1} \cdot b_t$, whereas the second inequality from condition (2.7) and the fact that $\mathcal{R}$ is an SPRT test.

Therefore, with a change of measure we obtain

$$\begin{aligned} \alpha &= \mathsf{P}_0(d_{\tilde{\mathcal{S}}} = 1) = \mathsf{E}_1[e^{-u_{\tilde{\mathcal{S}}}} \ _{\{d_{\tilde{\mathcal{S}}}=1\}}] \\ &= \mathsf{E}_1\Big[e^{(-u+\tilde{u})_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}} \ _{\{\tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B}\}}\Big] \leq e^{C - \tilde{B}} \end{aligned}$$

Taking logarithms on both sides we have $\tilde{B} \leq |\log \alpha| + C$ and in a similar way we can show that $\tilde{A} \leq |\log \beta| + C$. Then, using these inequalities together with (2.41) and (2.42), we get:

$$\begin{aligned} \mathsf{E}_1[u_{\tilde{\mathcal{S}}}] &= \mathsf{E}_1[(u - \tilde{u})_{\tilde{\mathcal{S}}}] + \mathsf{E}_1[\tilde{u}_{\tilde{\mathcal{S}}}] \\ &\leq C + (\tilde{B} + C) \leq |\log \alpha| + 3C. \end{aligned} \tag{2.44}$$

The desired inequality (2.39) now follows from (2.19) and (2.44).

□

2.2.4 Comparison with the discrete-time centralized SPRT

As we discussed before, the D-SPRT is a valid decentralized sequential test, since its implementation requires the communication of only one-bit messages. However, since the fusion center is able to recover the exact sensor observations at the corresponding communication times, it is meaningful to compare the D-SPRT with the *discrete-time centralized SPRT*, which is based on the transmission of the exact sensor observations to the fusion center at the times $t = 0, h, 2h, \ldots$, where $h > 0$. Since we do not have closed-form expressions for the performance of neither of the two schemes, we need to resort to simulations.

We perform this comparison for the following hypothesis testing problem:

$$\mathsf{H}_0 : \xi_t^i = W_t^i \qquad \mathsf{H}_1 : \xi_t^i = W_t^i + b_i t, \quad t \geq 0 \tag{2.45}$$

where $(W^1, \ldots, W^K)$ is a K-dimensional Brownian motion and $b_1, \ldots, b_K$ known, non-zero constants. Thus, we assume that each sensor i observes either a standard Brownian (H_0) or a Brownian motion with constant drift b_i (H_1). Under this model for the observations, the increments $\{\xi_{nh}^i - \xi_{(n-1)h}^i\}$ are independent and identically distributed, and consequently the discrete-time centralized SPRT is also order-2 asymptotically optimal (with respect to the continuous-time centralized SPRT).

For the comparison to be fair, we need to equate the expected intersampling periods $\mathsf{E}_0[\tau_1^i]$ and $\mathsf{E}_1[\tau_1^i]$ with the constant period h, so that the two schemes require the same communication rate between sensors and fusion center *on average*. Indeed, using Wald's identity together with (2.14)-(2.15), we have:

$$\begin{aligned} \mathsf{E}_0[\tau_1^i] = \mathsf{E}_1[\tau_1^i] = h \Leftrightarrow \mathsf{E}_0[-u_{\tau_1^i}^i] &= \mathsf{E}_1[u_{\tau_1^i}^i] = 0.5|b_i|^2 h \\ \Leftrightarrow s(\overline{\Delta}_i, \underline{\Delta}_i) &= s(\underline{\Delta}_i, \overline{\Delta}_i) = 0.5|b_i|^2 h \end{aligned} \tag{2.46}$$

Thus, if we set $\overline{\Delta}_i = \underline{\Delta}_i = \Delta_i$, (2.46) becomes $s(\Delta_i, \Delta_i) = 0.5\,|b_i|^2\,h$ and for any given drift b_i we can compute the sampling period h that corresponds to the threshold Δ_i and vice-versa. For example, in the simulations in Fig. 2.7 we chose $K = 2$, $b_1 = b_2 = 1$ and $\overline{\Delta}_i = \underline{\Delta}_i = \Delta_i = 2$ for each sensor i, thus h had to be equal to 3.0462.

From Fig. 2.7 we can see that the distance between the D-SPRT and the optimal continuous-time SPRT remains bounded, which agrees with (2.39). Moreover, the D-SPRT exhibits a distinct performance improvement over the discrete time centralized SPRT which is applied after canonical deterministic sampling.

2.2.5 Rare communication / many sensors

The inequalities (2.39) imply that – with a *fixed* number of sensors and with a *fixed* communication rate between sensors and fusion center – the distance between the D-SPRT and the continuous-time centralized SPRT is bounded for any horizon

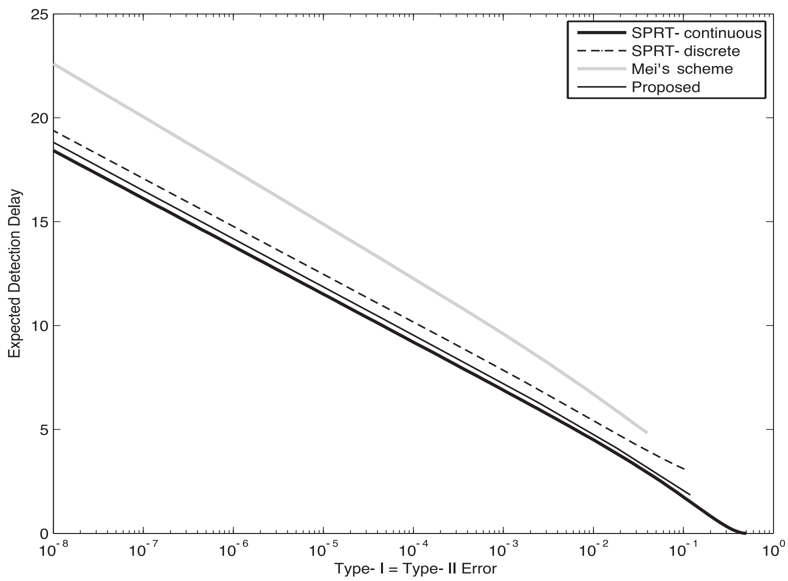


Figure 2.7: D-SPRT versus SPRT

of observations. The question we want to answer now is what happens if the communication between sensors and fusion center is very infrequent ($\overline{\Delta}_i, \underline{\Delta}_i \to \infty$) and/or the number of sensors very large ($K \to \infty$).

In this case, it is a direct consequence of (2.19) and (2.39) that the D-SPRT is asymptotically optimal of *order-1*, i.e.

$$\frac{\mathsf{E}_j[u_{\tilde{\mathcal{S}}}]}{\mathsf{E}_j[u_{\mathcal{S}}]} \to 1, \quad j = 0, 1, \tag{2.47}$$

as long as $C \equiv \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i) = o(|\log \alpha|) = o(|\log \beta|)$ as $\alpha, \beta \to 0$ with $\alpha|\log \beta| + \beta|\log \alpha| = \mathcal{O}(1)$. Thus, the D-SPRT remains asymptotically optimal, even with $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ or $K \to \infty$, as long as $\tilde{A}, \tilde{B}$ are much larger than $C = \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i)$, however in this case it loses its second-order optimality.

2.2.6 Level-triggered communication as repeated local SPRT's

Let us now restrict ourselves in the case that the sensor processes $\xi^1, \ldots, \xi^K$ are *independent*. Then, the local sufficient statistic $\{u_t^i\}$ -see (2.22)- corresponds to the log-likelihood ratio of $\{\xi_t^i\}$ and the level-triggered communication scheme (2.32)-(2.33) can be seen as a series of *repeated local SPRTs*. This observation leads to an interesting representation for the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ and an interesting interpretation for $\{\tilde{u}_t^i\}$.

Indeed, comparing (2.32)-(2.33) with (2.4), it becomes clear that the pair (τ_1^i, z_1^i) corresponds to an SPRT (2.4) with test statistic $\{u_t^i\}$ and thresholds $-\underline{\Delta}_i, \overline{\Delta}_i$. Thus, the message z_1^i that sensor i transmits to the fusion center at τ_1^i is the decision of this *local* SPRT; we can think of this message as a preliminary decision of sensor i for the testing problem (2.5) or more precisely for its local testing problem:

$$\begin{aligned} \mathsf{H}_0 &: \xi_t^i = \int_0^t \sigma_s^{ii}\, dW_s^i, \quad t \geq 0 \\ \mathsf{H}_1 &: \xi_t^i = \int_0^t b_s^i\, ds + \int_0^t \sigma_s^{ii}\, dW_s^i, \quad t \geq 0, \end{aligned} \tag{2.48}$$

At time τ_1^i sensor i repeats exactly the same procedure; thus, it starts a new SPRT with test-statistic $\{u_t^i - u_{\tau_1^i}^i\}_{t \geq \tau_1^i}$ and thresholds $-\underline{\Delta}_i$, $\overline{\Delta}_i$ and at time τ_2^i it transmits to the fusion center the decision z_2^i of this second local SPRT.

Similarly, at any time τ_{n-1}^i sensor i starts a new SPRT with test-statistic $\{u_t^i - u_{\tau_{n-1}^i}^i\}_{t \geq \tau_{n-1}^i}$ and thresholds $-\underline{\Delta}_i$, $\overline{\Delta}_i$ and at time τ_n^i it transmits to the fusion center the decision z_n^i of this n^{th} local SPRT.

Notice that for each sensor i the processes $\{u_t^i\}_{t \in [\tau_{n-1}^i, \tau_n^i]}$ and the pairs $(\tau_n^i - \tau_{n-1}^i, z_n^i)$ are *not* independent and identically distributed, even if the underlying process ξ^i has the Markov property. The only exception is the Brownian case $(b_t^i = b_i, \sigma_t^{ii} = \sigma_{ii})$, where each ξ^i not only has the strong Markov property but also *restarts probabilistically* at the stopping times (τ_n^i).

However, since $\tau_n^i - \tau_{n-1}^i$ is an SPRT stopping time (*for every* $n \in \mathbb{N}$) with thresholds $-\underline{\Delta}_i, \overline{\Delta}_i$, then $\mathsf{E}_j[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i]$ will be given by (2.14) with A, B replaced by $\underline{\Delta}_i, \overline{\Delta}_i$, that is:

$$\mathsf{E}_1[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] = s(\overline{\Delta}_i, \underline{\Delta}_i) \quad , \quad -\mathsf{E}_0[u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i] = s(\underline{\Delta}_i, \overline{\Delta}_i) \tag{2.49}$$

Similarly, although the error probabilities $\mathsf{P}_1(z_n^i = 0)$ and $\mathsf{P}_0(z_n^i = 1)$ vary with n, they are all described by (2.18) so that:

$$\overline{\Delta}_i = \log \frac{\mathsf{P}_1(z_n^i = 1)}{\mathsf{P}_0(z_n^i = 1)} \quad , \quad -\underline{\Delta}_i = \log \frac{\mathsf{P}_1(z_n^i = 0)}{\mathsf{P}_0(z_n^i = 0)}, \quad n \in \mathbb{N}. \tag{2.50}$$

Therefore, it becomes clear that $\overline{\Delta}_i$ corresponds to the log-likelihood ratio of the event $\{z_n^i = 1\}$ and $-\underline{\Delta}_i$ to the log-likelihood ratio of $\{z_n^i = 0\}$ *for every* $n \in \mathbb{N}$. This means that $\tilde{u}^i_{\tau^i_n}$ -defined in (2.34)- would be the log-likelihood ratio of the vector $(z_1^i, \dots, z_n^i)$ *if the signals* $z_1^i, \dots, z_n^i$ *were independent under each hypothesis.* However, this is true only in the Brownian case, where $z_1^i, \dots, z_n^i$ are not only independent but also identically distributed under each hypothesis. Therefore, we can think of $\tilde{u}^i_{\tau^i_n}$ as a *hypothetical* log-likelihood ratio for the vector $(z_1^i, \dots, z_n^i)$, which treats the signals $z_1^i, \dots, z_n^i$ as independent even if they are not.

2.2.7 Specification of the thresholds

The thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ determine the rate of communication between sensor i and the fusion center. From (2.42) it follows that smaller values for $\overline{\Delta}_i, \underline{\Delta}_i$ lead to more frequent communication and better D-SPRT performance. But in practice very frequent communication may be very expensive and the network may not be able to support it.

Thus, the choice of the thresholds $\overline{\Delta}_i$, $\underline{\Delta}_i$ at sensor i basically depends on how often sensor i is allowed to communicate with the fusion center. Consequently, we could specify the desired *expected* period of communication at sensor i under each hypothesis and choose $\overline{\Delta}_i$, $\underline{\Delta}_i$ in order to attain these targets. We actually performed such a specification with (2.46) in the Brownian case.

The problem with this approach is that –unless we are in the Brownian case– the times $(\tau_n^i - \tau_{n-1}^i)_n$ are not independent and identically distributed, thus their expectations $\mathsf{E}_j[\tau_n^i - \tau_{n-1}^i]$ will vary with n. Therefore, it is not possible to attain certain targets for $\mathsf{E}_j[\tau_n^i - \tau_{n-1}^i]$ for *every* $n \in \mathbb{N}$ with only the *two* free parameters $\overline{\Delta}_i, \underline{\Delta}_i$. In order to do that, we would need to adapt the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ at each communication time, but this choice has certain practical difficulties (see next section).

However, we can follow a different approach and specify target values for $\mathsf{E}_j[u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}}]$ instead of $\mathsf{E}_j[\tau_n^i - \tau_{n-1}^i]$ under each hypothesis H_j, $j = 0, 1$. Then, if the sensor processes are independent, from the previous section it follows that the quantities $\mathsf{E}_0[u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}}]$ and $\mathsf{E}_1[u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}}]$ have the same values for all $n \in \mathbb{N}$ and these values are given by (2.49). Therefore, it suffices to solve the *two* non-linear equations in (2.49) in order to compute the appropriate values for $\overline{\Delta}_i$ and $\underline{\Delta}_i$.

Apart from its practical advantages, this specification also has a very intuitive interpretation; we choose the thresholds $\overline{\Delta}_i$ and $\underline{\Delta}_i$ that guarantee that a certain amount of *information* has been accumulated between any two consecutive communications from sensor i.

2.2.8 Time-varying thresholds

We can generalize the D-SPRT by replacing $\overline{\Delta}_i$ and $\underline{\Delta}_i$ with two *sequences* of positive thresholds $\{\overline{\Delta}^i_n\}$ and $\{\underline{\Delta}^i_n\}$, so that sensor i communicates with the fusion center at the times

$$\tau^i_n = \inf\{t \geq \tau^i_{n-1} : u^i_t - u^i_{\tau^i_{n-1}} \notin (-\underline{\Delta}^i_n, \overline{\Delta}^i_n)\}, \quad n \in \mathbb{N} \tag{2.51}$$

and at time τ^i_n it transmits the binary signal:

$$z^i_n = \begin{cases} 1, & \text{if } u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}} \geq \overline{\Delta}^i_n \\ 0, & \text{if } u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}} \leq -\underline{\Delta}^i_n \end{cases} \tag{2.52}$$

The thresholds $\overline{\Delta}^i_n$ and $\underline{\Delta}^i_n$ can be random, as long as they are $\mathcal{F}^i_{\tau^i_{n-1}} \wedge \tilde{\mathcal{F}}^i_{\tau^i_n}$-measurable random variables. This means that each sensor i must specify the thresholds $\overline{\Delta}^i_n, \underline{\Delta}^i_n$ for its n^{th} transmission by time τ^i_{n-1} and must transmit these values (together with the message z^i_n) to the fusion center at time τ^i_n.

For example, if we wanted the expected time between any two consecutive communications to be equal to a fixed constant h under both hypotheses, then each pair $\overline{\Delta}^i_n, \underline{\Delta}^i_n$ should be chosen to satisfy $\mathsf{E}_j[\tau^i_n - \tau^i_{n-1} | \mathcal{F}^i_{\tau^i_{n-1}}] = h, \ j = 0, 1$. Notice however that there is no closed-form expression for the conditional expectation on the left-hand side, thus the specification of $\overline{\Delta}^i_n, \underline{\Delta}^i_n$ is not straightforward. Moreover, communicating the values of the thresholds to the fusion center requires a larger alphabet at each sensor and restricts the thresholds that can be chosen in practice.

Of course, the thresholds $\{\overline{\Delta}_n^i\}$ and $\{\underline{\Delta}_n^i\}$ do not have to depend on the sensor observations; instead, they can be deterministic sequences which become known at sensor i and at the fusion center from the beginning (at time $t = 0$). For example, the sequences $\{\overline{\Delta}_n^i\}$ and $\{\underline{\Delta}_n^i\}$ can be chosen in advance so that the same amount of information is accumulated at each sensor between any two communications. This can be achieved using (2.49), from which we obtain:

$$\mathsf{E}_1[u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}}] = s(\overline{\Delta}_n^i, \underline{\Delta}_n^i) \quad , \quad -\mathsf{E}_0[u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}}] = s(\underline{\Delta}_n^i, \overline{\Delta}_n^i) \tag{2.53}$$

In any case, assuming the communication scheme (2.51)-(2.52) can be implemented, we suggest that the fusion center approximates $\{u_t^i\}$ with

$$\tilde{u}_t^i = \sum_{j=1}^{n} \Big[\overline{\Delta}_j^i \, z_j^i - \underline{\Delta}_j^i (1 - z_j^i)\Big], \quad \tau_n^i \leq t < \tau_{n+1}^i, \tag{2.54}$$

and implements the D-SPRT (2.37) with test-statistic $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$.

It is clear that (2.54) is a generalization of (2.36) and it allows the fusion center to recover the values of each $\{u_t^i\}$ at the corresponding communication times (2.51). Moreover, $\{\tilde{u}_t\}$ satisfies the inequalities in (2.41) as long as $C \equiv \sup_n C_n < \infty$, where we define $C_n = \sum_{i=1}^K (\overline{\Delta}_n^i + \underline{\Delta}_n^i)$.

Consequently, it can be shown – in exactly the same way as in the case of constant thresholds– that the D-SPRT satisfies (2.38) and is order-2 asymptotically optimal as long as the thresholds $\{\overline{\Delta}_n^i\}$ and $\{\underline{\Delta}_n^i\}$ are uniformly bounded in n and i.

2.2.9 Non-parallel boundaries

Another way we can extend the communication scheme (2.32)-(2.33) is to use *linear* instead of parallel boundaries, so that sensor i communicates with the fusion center at the stopping times:

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : u_t^i - u^i_{\tau^i_{n-1}} \notin (-\underline{\Delta}_i + \underline{\delta}_i(t - \tau_{n-1}^i) \, , \, \overline{\Delta}_i - \overline{\delta}_i(t - \tau_{n-1}^i))\} \tag{2.55}$$

and at time τ_n^i it transmits the binary signal

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i} - u^i_{\tau^i_{n-1}} \geq \overline{\Delta}_i - \overline{\delta}_i(\tau_n^i - \tau_{n-1}^i) \\ 0, & \text{if } u_{\tau_n^i} - u^i_{\tau^i_{n-1}} \leq -\underline{\Delta}_i + \overline{\delta}_i(\tau_n^i - \tau_{n-1}^i) \end{cases} \tag{2.56}$$

where $\overline{\Delta}_i, \underline{\Delta}_i, \overline{\delta}_i, \underline{\delta}_i$ are *positive* constants, chosen by the designer of the scheme and known to the fusion center. Since the linear boundaries in (2.55) intersect (see Fig.2.8), the intercommunication times $(\tau_n^i - \tau_{n-1}^i)_n$ are bounded. This is an important difference with the case of parallel boundaries ($\overline{\delta}_i = \underline{\delta}_i = 0$) where the times between two consecutive communications are almost surely finite but not bounded.

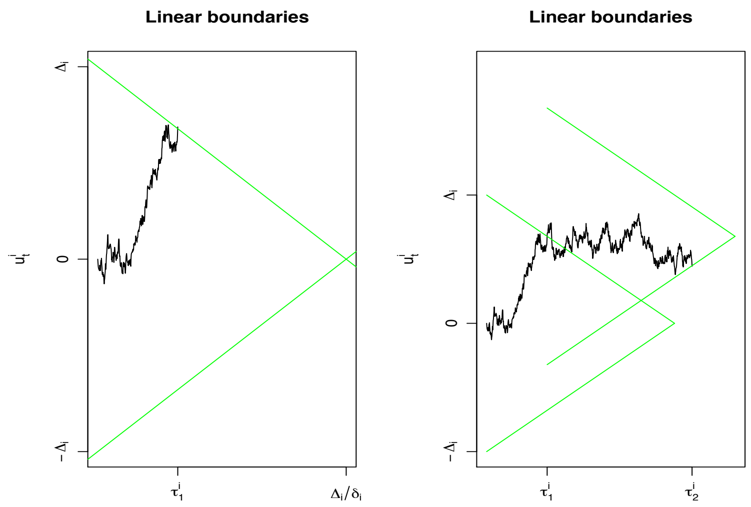


Figure 2.8: Event-triggered communication with linear boundaries.

We suggest that the fusion center approximates the process $\{u_t^i\}$ with

$$\tilde{u}_t^i = \sum_{j=1}^{m_t^i} \Big[[\overline{\Delta}_i - \overline{\delta}_i(\tau_j^i - \tau_{j-1}^i)]\, z_j^i + [-\underline{\Delta}_i + \underline{\delta}_i(\tau_j^i - \tau_{j-1}^i)](1 - z_j^i) \Big], \quad t \geq 0, \quad (2.57)$$

and the global log-likelihood ratio $u = \sum_{i=1}^K u^i$ with $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$ (see fig. 2.9). Then, $\{\tilde{u}_t\}$ satisfies the inequalities in (2.41), i.e. for all $t \geq 0$ we have: $|u_t - \tilde{u}_t| \leq C$ and $|\tilde{u}_t - \tilde{u}_{t-}| \leq C$, where $C = \sum_{i=1}^K (\overline{\Delta}_i + \underline{\Delta}_i)$.

Consequently, if we replace the global log-likelihood ratio $u = \sum_{i=1}^K u^i$ in the definition of the SPRT (2.4) with the test-statistic $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$, the resulting sequential test will satisfy (2.39) and consequently it will be asymptotically optimal of order-2; the proof for this result is exactly the same as in the case of parallel

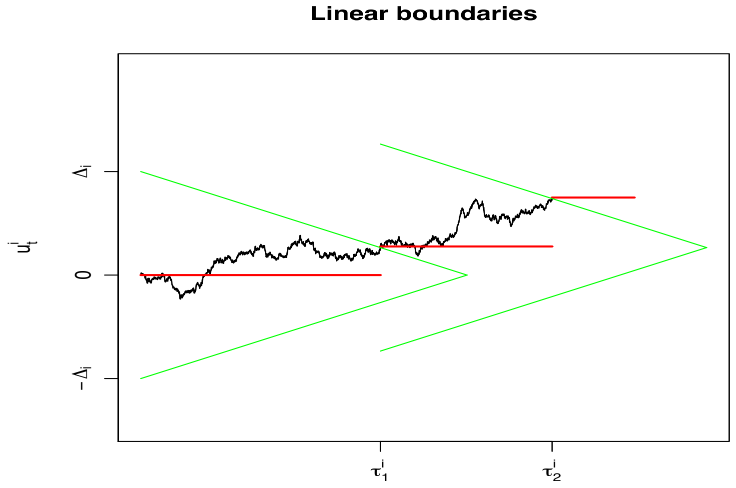


Figure 2.9: Local approximations at the fusion center

boundaries. However, we can no longer use (2.49) in order to specify the values of $\overline{\Delta}_i, \underline{\Delta}_i$.

There is nothing special about linear boundaries in the above argument. Thus, we can replace (2.55)-(2.56) with the following non-linear boundaries

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (g_i(t-\tau_{n-1}^i), h_i(t-\tau_{n-1}^i))\}$$
$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq g_i(t-\tau_{n-1}^i) \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq h_i(t-\tau_{n-1}^i) \end{cases} \tag{2.58}$$

where each $h_i(t)$ is a decreasing continuous function with $h_i(0) = \overline{\Delta}_i$ and each $g_i(t)$ an increasing continuous function with $g_i(0) = -\underline{\Delta}_i$.

2.3 Decentralized sequential testing for jump-diffusions

In the previous section we introduced and analyzed a decentralized sequential test for the hypothesis testing problem (2.5), i.e. when the sensors observe Itô processes under each hypothesis. This suggested decentralized scheme consisted of:

- the event-triggered communication scheme $(\tau_n^i) - (z_n^i)$, defined in (2.32)-

(2.33),

- the approximations $\{\tilde{u}_t^i\}$ to the local sufficient statistics $\{u_t^i\}$, defined in (2.36),
- and the fusion center policy $(\tilde{\mathcal{S}}, d_{\tilde{\mathcal{S}}})$, defined in (2.37).

We proved (2.41), which means that the performance loss of the D-SPRT is bounded for any fixed rate of communication and error probabilities. This property implies order-2 asymptotic optimality for small error probabilities and any fixed rate of communication and order-1 asymptotic optimality for small error probabilities and infrequent communication. Moreover, we saw that we can consider more general communication schemes, such as (2.51)-(2.52) or (2.55)-(2.56), without affecting the asymptotic optimality properties of the scheme.

In this section we consider the case where the observed sensor processes and consequently the log-likelihood ratio $\{u_t\}$ do not have continuous paths. In this case, the fusion center can no longer implement the model-free approximations (2.36) and we cannot have a bounded performance loss such as (2.41).

Our goal is to show that if we restrict ourselves to independent sensors which observe Lévy processes, then we can modify the fusion center policy so that the resulting scheme preserves the order-1 asymptotic optimality in a certain sense. This decentralized sequential scheme is the direct analogue of the scheme proposed in [16] in a discrete-time setup and their analysis are almost identical for that reason we do not prove here the results but refer to the discrete-time case.

2.3.1 Overshoot accumulation

As before, we assume that the global log-likelihood ratio $\{u_t\}$ admits the decomposition $u = \sum_{i=1}^K u^i$ so that each $\{u_t^i\}$ is $\{\mathcal{F}_t^i\}$-adapted and we consider the event-triggered communication scheme $(\tau_n^i) - (z_n^i)$ described by (2.32)-(2.33)

As soon as we remove the assumption that each $\{u_t^i\}$ has continuous paths, the fusion center is not able to recover the values of $\{u_t^i\}$ at the times $(\tau_n^i)_{n\in\mathbb{N}}$ and consequently it cannot use (2.35) to approximate $\{u_t^i\}$.

Indeed, (2.34) is no longer true and is replaced by:

$$u^i_{\tau^i_n} = \sum_{j=1}^{n} \Big[u^i_{\tau^i_j} - u^i_{\tau^i_{j-1}} \Big] = \sum_{j=1}^{n} \Big[\overline{\Delta}_i \, z^i_j - \underline{\Delta}_i \, (1 - z^i_j) \Big] + \sum_{j=1}^{n} \eta^i_j, \quad n \in \mathbb{N}, \tag{2.59}$$

where by $\{\eta^i_n\}$ we denote the *overshoots*

$$\eta^i_n = (u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}} - \overline{\Delta}_i)^+ - (u^i_{\tau^i_n} - u^i_{\tau^i_{n-1}} + \underline{\Delta}_i)^-, \quad n \in \mathbb{N}, \tag{2.60}$$

whose values do not become known to the fusion center (notice however that η^i_n is non-zero only if τ^i_n is a jump time of $\{u^i_t\}$).

Moreover, (2.60) implies that unlike the case of continuous-path observations, the approximation (2.35), i.e. $\tilde{u}^i_t = u^i_{\tau^i_n}$, $t \in [\tau_n, \tau_{n+1})$ is not equivalent to (2.36), i.e.

$$\tilde{u}^i_t = \sum_{j=1}^{m^i_t} \Big[\overline{\Delta}_i \, z^i_j - \underline{\Delta}_i (1 - z^i_j) \Big]. \tag{2.61}$$

However, the latter approximation is implementable, thus it would be interesting to see if it is possible to use this.

For simplicity let us assume that the jumps of $\{u^i_t\}$ are bounded, which implies that the overshoots $(\eta^i_n)_{n\in\mathbb{N}}$ are also bounded. Then, from (2.36) and (2.59) we have:

$$|\tilde{u}^i_t - u^i_t| \leq (\overline{\Delta}_i + \underline{\Delta}_i) + \sup_n |\eta^i_n| \, \tilde{m}^i_t, \quad t \geq 0, \tag{2.62}$$

where by $\tilde{m}^i_t$ we denote the number of transmissions from sensor i up to time t which were associated with a non-zero overshoot, i.e. $\tilde{m}^i_t = \max\{n : \tau^i_n \leq t \, , \, \eta^i_n \neq 0\}$.

Therefore, for the global approximation $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$ we have:

$$|\tilde{u}_t - u_t| \leq C + \theta \, \tilde{m}_t \quad , \quad |\tilde{u}_t - \tilde{u}_{t-}| \leq C, \quad t \geq 0, \tag{2.63}$$

where $\theta = \max_i \sup_n |\eta^i_n|$, $C = \sum_{i=1}^K (\overline{\Delta}_i + \underline{\Delta}_i)$, $\tilde{m}_t = \sum_{i=1}^K \tilde{m}^i_t$.

(2.63) is the analogue of (2.41) in the case of discontinuous processes (with bounded jumps) and shows that the distance between $\{u_t\}$ and $\{\tilde{u}_t\}$ is unbounded. This implies that the discontinuity of the sensor processes generates overshoots whose accumulation can deteriorate significantly the quality of the D-SPRT. Moreover, it implies that in this context we need a different analysis than in the case of sensor processes with continuous paths.

However, using the approximation (2.36) the fusion center ignores completely the overshoots $\{\eta_n^i\}$ and behaves as if the sensor processes had continuous paths. Therefore, it is reasonable to expect that if we replace (2.36) with an approximation that incorporates the underlying sensor dynamics and *approximates* the overshoots, we could mitigate the performance loss associated with the overshoot accumulation.

A natural way to do so is to work directly with the log-likelihood ratio that corresponds to the *fusion center filtration*, $\{\tilde{\mathcal{F}}_t\}$, instead of trying to approximate the log-likelihood ratio that corresponds to the global filtration in the sensor network, $\{\mathcal{F}_t\}$. This approach will lead to an approximation of the overshoots, however it will not enjoy the universality of (2.36), which is an approximation completely independent of the sensor dynamics. Indeed, in order to compute explicitly the log-likelihood ratio that corresponds to the fusion center filtration, we need to impose some structure on the sensors dynamics.

2.3.2 D-SPRT for jump-diffusions

2.3.2.1 Problem formulation

We assume that each sensor process $\{\xi_t^i\}$ has stationary and independent increments and its paths have finitely many jumps. Moreover, we assume that the sensor processes are independent under each hypothesis. Thus, we restrict ourselves to the testing problem described by (2.23) and (2.24).

In this context the log-likelihood ratio process $\{u_t\}$ can be written as the sum of the local log-likelihood ratios $\{u_t^i\}$. Moreover, we can apply the level-triggered communication scheme (2.32)-(2.33), which determines the flow of information at the fusion center.

It is important to emphasize that sensor i sends to the fusion center not only the bits $(z_n^i)_n$, but also -implicitly- the intercommunication periods $(\delta_n^i = \tau_n^i - \tau_{n-1}^i)_n$.

Thus, the local filtration $\{\mathcal{F}_t^i\}$ at sensor i is "approximated" at the fusion center by

$$\tilde{\mathcal{F}}_t^i = \sigma((z_j^i, \delta_j^i),\ j = 1, \ldots, m_t^i), \quad t \geq 0. \tag{2.64}$$

whereas the global filtration at the fusion center is:

$$\tilde{\mathcal{F}}_t = \sigma((z_j^i, \delta_j^i),\ j = 1, \ldots, m_t^i,\ i = 1, \ldots, K), \quad t \geq 0. \tag{2.65}$$

Due to the Lévy structure of the sensor processes and their independence under each hypothesis, the pairs $(z_n^i, \delta_n^i)_{n \in \mathbb{N}}$ are independent and identically distributed and we denote their joint pdf under H_j by $p_j^i(z, \delta)$, $j = 0, 1$.

Thus, the testing problem (2.24) at the sensors becomes at the fusion center

$$\mathsf{H}_0 : (z_n^i, \delta_n^i) \sim^{iid} p_0^i(z, \delta), \quad \forall\, i \quad , \quad \mathsf{H}_1 : (z_n^i, \delta_n^i) \sim^{iid} p_1^i(z, \delta), \quad \forall\, i \tag{2.66}$$

2.3.2.2 The likelihood ratio at the fusion center

First of all, we observe that we can write:

$$\begin{aligned} p_1^i(z, \delta) &= \pi_1^i\, g_1^i(\delta|z) \quad _{\{z=0\}} + (1 - \pi_1^i)\, g_1^i(\delta|z) \quad _{\{z=1\}}, \\ p_0^i(z, \delta) &= (1 - \pi_0^i)\, g_0^i(\delta|z) \quad _{\{z=0\}} + \pi_0^i\, g_0^i(\delta|z) \quad _{\{z=1\}}, \end{aligned} \tag{2.67}$$

where $\pi_1^i = \mathsf{P}_1(z_n^i = 0), \pi_0^i = \mathsf{P}_0(z_n^i = 1)$ and by $g_j^i(\delta|z)$ we denote the conditional pdf of δ_n^i *given* that $z_n^i = z$ under H_j, $j = 0, 1$.

Thus, π_1^i is the probability that sensor i exceeds the lower threshold $-\underline{\Delta}_i$ under H_1, whereas π_0^i is the probability that sensor i exceeds the upper threshold $\overline{\Delta}_i$ under H_0. In other words, we can think of π_0^i and π_1^i as the *local* type I and type II error probabilities, respectively.

Moreover, the marginal pdf's g_j^i of δ_n^i under H_j, $j = 0, 1$ take the form:

$$\begin{aligned} g_0^i(\delta) &= \pi_0^i\, g_0^i(\delta|z = 1) + (1 - \pi_0^i)\, g_1^i(\delta|z = 0) \\ g_1^i(\delta) &= \pi_1^i\, g_1^i(\delta|z = 0) + (1 - \pi_1^i)\, g_1^i(\delta|z = 1) \end{aligned} \tag{2.68}$$

Suppose now that we are at time t and that the fusion center has observed the $m_t^i = k$ pairs $(z_1^i, \delta_1^i), \ldots, (z_k^i, \delta_k^i)$ from sensor i. Since $\tau_k^i = \sum_{l=1}^k \delta_l^i$ we can write:

$$\{m_t^i = k\} = \{\delta_1^i + \cdots + \delta_k^i \leq t < \delta_1^i + \cdots + \delta_k^i + \delta_{k+1}^i\} = \{0 \leq t - \tau_k^i < \delta_{k+1}^i\} \tag{2.69}$$

and using the independence of the pairs we have:

$$\begin{aligned}
&\mathsf{P}_j(m_t^i = k; (z_1^i, \delta_1^i), \ldots, (z_k^i, \delta_k^i)) \\
&\quad = \mathsf{P}_j(0 \le t - \tau_k^i < \delta_{k+1}^i; (z_1^i, \delta_1^i), \ldots, (z_k^i, \delta_k^i)) \\
&\quad = [1 - G_j^i(t - \tau_k^i)] \left(\prod_{n=1}^{k} p_j^i(z_n^i, \delta_n^i) \right) \mathbb{1}_{\{\tau_k^i \le t\}},
\end{aligned} \tag{2.70}$$

where $G_j^i(\delta) = \int_0^\delta g_j^i(x)dx$ is the cdf of δ_n^i and $g_j^i(\delta)$ is the marginal pdf defined in (2.68).

Using (2.67) we have:

$$\begin{aligned}
\mathsf{P}_1(m_t^i = k; (z_1^i, \delta_1^i), \ldots, (z_k^i, \delta_k^i)) &= \left([1 - G_1^i(t - \tau_k^i)] \prod_{n=1}^{k} g_1^i(\delta_n^i | z_n^i) \mathbb{1}_{\{\tau_k^i \le t\}} \right) \\
&\times \prod_{n=1}^{k} (1 - \pi_1^i)^{z_n^i} (\pi_1^i)^{1 - z_n^i}
\end{aligned}$$

and similarly:

$$\begin{aligned}
\mathsf{P}_0(m_t^i = k; (z_1^i, \delta_1^i), \ldots, (z_k^i, \delta_k^i)) &= \left([1 - G_0^i(t - \tau_k^i)] \prod_{n=1}^{k} g_0^i(\delta_n^i | z_n^i) \mathbb{1}_{\{\tau_k^i \le t\}} \right) \\
&\times \prod_{n=1}^{k} (\pi_0^i)^{z_n^i} (1 - \pi_0^i)^{1 - z_n^i}
\end{aligned}$$

In both expressions, the second factor in the right-hand side is the likelihood of the 1-bit data $\{z_1^i, \ldots, z_k^i\}$ and the first factor is the likelihood of the intersampling periods $\{\delta_1^i, \ldots, \delta_k^i\}$ *conditioned* on the 1-bit data $\{z_1^i, \ldots, z_k^i\}$.

The corresponding likelihood ratio then is:

$$\left.\frac{d\mathsf{P}_1}{d\mathsf{P}_0}\right|_{\tilde{\mathcal{F}}_t^i} = e^{\tilde{u}_t^i} \times \frac{[1 - G_1^i(t - \tau_k^i)]}{[1 - G_0^i(t - \tau_k^i)]} \times \prod_{n=1}^{m_t^i} \frac{g_1^i(\delta_n^i | z_n^i)}{g_0^i(\delta_n^i | z_n^i)}, \quad t \ge 0 \tag{2.71}$$

where $\tilde{u}_t^i$ is the log-likelihood ratio of the 1-bit data that have been transmitted from sensor i up to time t, i.e.

$$\tilde{u}_t^i = \sum_{j=1}^{m_t^i} \lambda_j^i, \quad \lambda_j^i = \overline{\Lambda}_i z_j^i - \underline{\Lambda}_i (1 - z_j^i) \tag{2.72}$$

where

$$\overline{\Lambda}_i = \log \frac{1 - \pi_1^i}{\pi_0^i} \quad , \quad \underline{\Lambda}_i = \log \frac{1 - \pi_0^i}{\pi_1^i}. \tag{2.73}$$

Combining all sensors and using their independence, we obtain the likelihood ratio that refers to the total information $\tilde{\mathcal{F}}_t$ accumulated at the fusion center until time t:

$$\frac{d\mathsf{P}_1}{d\mathsf{P}_0}\bigg|_{\tilde{\mathcal{F}}_t} = e^{\tilde{u}_t} \times \mathscr{L}_t \tag{2.74}$$

where $\tilde{u}_t = \sum_{i=1}^{K} \tilde{u}_t^i$ is the log-likelihood ratio of the 1-bit signals that have been transmitted from *all* sensors up to time t, i.e.

$$\tilde{u}_t = \sum_{i=1}^{K} \sum_{j=1}^{m_t^i} [\overline{\Lambda}_i z_j^i - \underline{\Lambda}_i(1 - z_j^i)] \tag{2.75}$$

whereas $\mathscr{L}_t$ represents the *likelihood ratio* of the intersampling periods conditioned on the 1-bit data, i.e.

$$\mathscr{L}_t = \prod_{i=1}^{K} \left(\frac{[1 - G_1^i(t - \tau_k^i)]}{[1 - G_0^i(t - \tau_k^i)]} \times \prod_{n=1}^{m_t^i} \frac{g_1^i(\delta_n^i | z_n^i)}{g_0^i(\delta_n^i | z_n^i)} \right) \tag{2.76}$$

2.3.2.3 The partial log-likelihood ratio as test-statistic

The full log-likelihood ratio (2.74) is the ideal test-statistic at the fusion center. However, in order to use it, we need to be able to compute both $\{\tilde{u}_t\}$ and $\{\mathscr{L}_t\}$. The computation of $\{\tilde{u}_t\}$ requires the knowledge of the quantities $\{\overline{\Lambda}_i, \underline{\Lambda}_i\}$, which were defined in (2.73). These quantities are not known explicitly, however they can be pre-computed using simulations. The same is not true for $\{\mathscr{L}_t\}$ for which we neither have closed-form expressions nor we can use simulations.

Therefore, the full likelihood ratio at the fusion center is essentially decomposed into a tractable and an intractable part. We choose to use only the tractable part, thus *we approximate* $\{u_t\}$ *with the marginal log-likelihood ratio* $\{\tilde{u}_t\}$ *which was defined in (2.75).*

Notice that (2.75) has exactly the same form as (2.34) with the difference that the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ have been replaced by the log-likelihood ratios $\overline{\Lambda}_i$, $\underline{\Lambda}_i$. Therefore, the distance between $\{u_t\}$ and $\{\tilde{u}_t\}$ will remain unbounded.

We emphasize that the motivation for using the "partial" log-likelihood ratio $\{\tilde{u}_t\}$ is the simplicity of the resulting sequential test, which allows us to ignore the intractable term $\{\mathscr{L}_t\}$ and does not require the fusion center to record the interarrival times that correspond to each sensor.

Then, the sequential test at the fusion center is

$$\tilde{\mathcal{S}} = \inf\{t \geq 0 : \tilde{u}_t \notin (-\tilde{A}, \tilde{B})\}$$
$$d_{\tilde{\mathcal{S}}} = \begin{cases} 1, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B} \\ 0, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \leq -\tilde{A} \end{cases} \tag{2.77}$$

The sequential test (2.77) is the continuous-time analogue of the scheme proposed in [16], thus we will also call it D-SPRT.

2.3.3 Asymptotic optimality and design implications

We now state the performance of the D-SPRT in the above framework. As expected, the performance loss incurred by the D-SPRT is no longer bounded, however it turns out that their ratio of the D-SPRT over the SPRT performance will be close to 1 as long as $\overline{\Delta}_i, \underline{\Delta}_i$ are large but smaller than $\tilde{A}, \tilde{B}$. In particular, $\overline{\Delta}_i, \underline{\Delta}_i$ should be around the *square root* of $\tilde{A}$ and $\tilde{B}$.

First of all, in order to avoid unnecessary complications, we assume that there is some quantity Δ so that $\overline{\Delta}_i, \underline{\Delta}_i = \Theta(\Delta)$ as $\Delta \to \infty$ and that $\alpha = \Theta(\beta)$. Then we can prove the following result:

Proposition 2. *If we let $\alpha \to 0$ and $\Delta \to \infty$, then:*

$$|\mathsf{E}_j[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_j[u_{\mathcal{S}}]| \leq \frac{|\log \alpha|}{\Theta(\Delta)} + \Theta(\Delta), \quad j = 0, 1. \tag{2.78}$$

Consequently, the D-SPRT $(\tilde{\mathcal{S}}, d_{\tilde{\mathcal{S}}})$ is asymptotically optimal of order-1, in the sense that

$$\frac{\mathsf{E}_j[u_{\tilde{\mathcal{S}}}]}{\mathsf{E}_j[u_{\mathcal{S}}]} \longrightarrow 1, \quad j = 0, 1 \tag{2.79}$$

as long as $\Delta = o(|\log \alpha|)$.

Moreover, the optimal choice for the divergence rate of Δ is $\Delta = \mathcal{O}(\sqrt{|\log \alpha|})$ in which case then we have:

$$\frac{\mathsf{E}_j[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_j[u_{\mathcal{S}}]}{\mathsf{E}_j[u_{\mathcal{S}}]} = \mathcal{O}(|\log \alpha|^{-1/2}), \quad j = 0, 1, \tag{2.80}$$

We skip the proof of this result since it is essentially identical to the corresponding proof in the discrete-time context, which we present in the end of the next section.

It is interesting to note that despite the fact that the underlying processes are assumed to be continuously observed, the D-SPRT cannot be order-2 asymptotically optimal when the observed paths are not continuous.

On the other hand, under (2.24) the D-SPRT has exactly the same order-1 asymptotic optimality that it enjoys in the case of Itô processes (see (2.47)). Thus, the D-SPRT has the same behavior for both problems (2.5) and (2.24) when $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$. The difference of course is that under (2.5) smaller values for $\overline{\Delta}_i, \underline{\Delta}_i$ improve the performance of the D-SPRT, whereas this is no longer true under (2.24) due to the overshoot effect.

2.4 Decentralized sequential testing in discrete time

We now consider the discrete-time sequential testing problem (2.28), thus we assume that sensor i observes a sequence $(\xi_t^i)_{t\in\mathbb{N}}$ of independent and identically distributed random variables under each hypothesis. We additionally assume independence across sensors, thus the global log-likelihood ratio $\{u_t\}$ –defined in (2.29)– admits the decomposition $u_t = \sum_{i=1}^{K} u_t^i$, where $\{u_t^i\}$ is the marginal log-likelihood ratio that corresponds to the sequence $\{\xi_t^i\}$.

2.4.1 Level-triggered communication and overshoot effect

Before we define the suggested communication scheme in this framework, we should underline that in discrete time a sensor is able to communicate with the fusion center every time it takes an observation. In other words, every observation time can be a communication time, i.e. $\tau_n^i = n$, $n \in \mathbb{N}$. However, insisting on the idea that the sensors should communicate with the fusion center only when they have an important message to transmit, we propose the same communication scheme as in the continuous time case. Thus, we suggest that sensor i communicate with the fusion center at the times

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \quad n \in \mathbb{N}, \tag{2.81}$$

transmitting the messages

$$z_n^i = \begin{cases} 1, & \text{if} \quad u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \overline{\Delta}_i \\ 0, & \text{if} \quad u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i \end{cases} \tag{2.82}$$

which informs the fusion center whether u^i has increased *at least* by $\overline{\Delta}_i$ or decreased *at least* by $\underline{\Delta}_i$ in comparison to its value at the previous communication time, where $\overline{\Delta}_i, \underline{\Delta}_i$ are positive constants, fixed in advance and known to the fusion center. Notice that the number of times that sensor i has communicated with the fusion center up to time t is random, it takes values in $\{0, 1, \ldots, t\}$ and we denote it by $m_t^i = \max\{j : \tau_j^i \leq t\}$.

As in the case of Lévy processes, the fusion center cannot recover the value of u^i at the communication times $(\tau_n^i)_n$ using the signals $(z_n^i)_n$. Indeed, (2.34) is replaced by

$$u_{\tau_n^i}^i = \sum_{j=1}^{n} [\overline{\Delta}_i z_j^i - \underline{\Delta}_i (1 - z_j^i)] + \sum_{j=1}^{n} \eta_j^i, \quad n \in \mathbb{N} \tag{2.83}$$

where with η_n^i we denote the overshoot

$$\eta_n^i = (u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i - \overline{\Delta}_i)^+ + (u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i + \underline{\Delta}_i)^-, \quad n \in \mathbb{N} \tag{2.84}$$

to which the fusion center does not have access. Notice moreover that the overshoots (η_n^i) are non-zero with probability 1, thus every transmission is associated with an overshoot which deteriorates the D-SPRT performance (whereas in the case of jump diffusions the overshoots have positive probability to be 0). We illustrate the overshoot effect under a discrete-time setup in Fig. 2.10. Despite this difference, it is clear that the discrete-time framework shares a lot of similarities with the case of continuously observed Lévy processes, since the probabilistic structure in both problems is the same (independent sensors, iid observations/increments) and in both setups there is an overshoot effect. Therefore, we can argue in exactly the

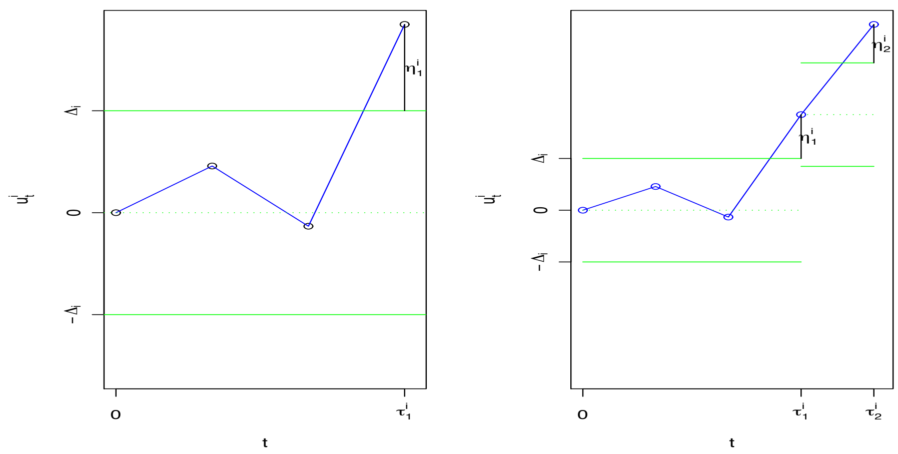


Figure 2.10: Overshoots in level-triggered communication

same way as in the jump-diffusion case and suggest the following approximation to the local log-likelihood ratio u_t^i

$$\tilde{u}_t^i = \sum_{j=1}^{n} [\overline{\Lambda}_i z_j^i - \underline{\Lambda}_i (1 - z_j^i)], \quad \tau_n^i \leq t < \tau_n^i \tag{2.85}$$

where

$$\overline{\Lambda}_i = \log \frac{\mathsf{P}_1(z_n^i = 1)}{\mathsf{P}_0(z_n^i = 1)} \quad , \quad -\underline{\Lambda}_i = \log \frac{\mathsf{P}_1(z_n^i = 0)}{\mathsf{P}_0(z_n^i = 0)}. \tag{2.86}$$

and the following sequential test at the fusion center

$$\tilde{\mathcal{S}} = \inf\{t \in \mathbb{N} : \tilde{u}_t \notin (-\tilde{A}, \tilde{B})\}$$
$$d_{\tilde{\mathcal{S}}} = \begin{cases} 1, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \geq \tilde{B} \\ 0, & \text{if } \tilde{u}_{\tilde{\mathcal{S}}} \leq -\tilde{A} \end{cases} \tag{2.87}$$

In Fig. 2.11 we show that using the statistic $\{\tilde{u}_t^i\}$ mitigates to some extent the overshoot effect.

2.4.2 Asymptotic optimality and design implications

Before we state the main result, we introduce some quantities that will be useful in the statement of the result and its proof. Thus, we denote by θ the maximal

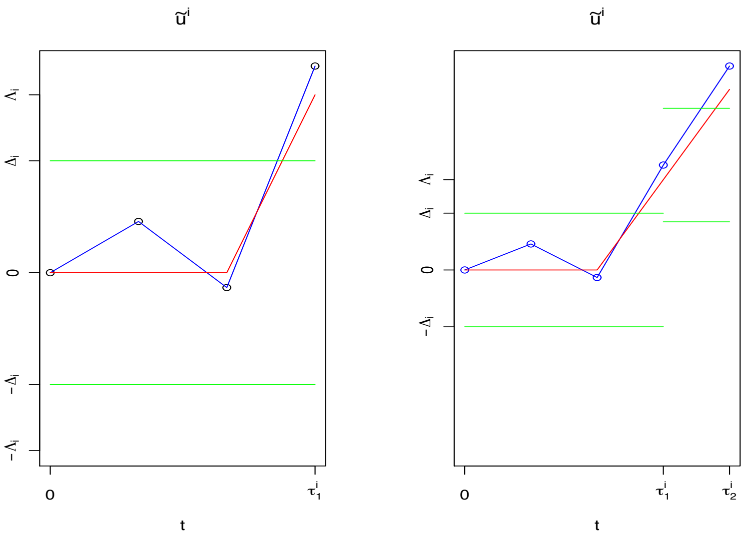


Figure 2.11: Fusion center approximation in discrete-time (continuation of 2.10)

expected overshoot:

$$\theta = \max_j \max_i \mathsf{E}_j[|\eta_1^i|]. \tag{2.88}$$

Moreover, we assume that there is some quantity Δ so that $\overline{\Delta}_i, \underline{\Delta}_i = \Theta(\Delta)$ as $\Delta \to \infty$ and that $\alpha, \beta \to 0$ so that $\alpha = \Theta(\beta)$. We then have the following result.

Proposition 3. *If we let $\alpha \to 0$ and $\Delta \to \infty$, then:*

$$|\mathsf{E}_j[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_j[u_{\mathcal{S}}]| \leq \theta \, \frac{|\log \alpha|}{\Theta(\Delta)} + \Theta(\Delta), \quad j = 0, 1. \tag{2.89}$$

Moreover, if $\Delta = o(|\log \alpha|)$, the D-SPRT is asymptotically optimal of order-1, i.e.

$$\frac{\mathsf{E}_j[u_{\tilde{\mathcal{S}}}]}{\mathsf{E}_j[u_{\mathcal{S}}]} \longrightarrow 1, \quad j = 0, 1. \tag{2.90}$$

Finally, the optimal divergence rate for Δ is $\Delta = \mathcal{O}(\sqrt{|\log \alpha|})$, in which case we have:

$$\frac{\mathsf{E}_j[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_j[u_{\mathcal{S}}]}{\mathsf{E}_j[u_{\mathcal{S}}]} = \mathcal{O}(|\log \alpha|^{-1/2}), \quad j = 0, 1. \tag{2.91}$$

Therefore, in order to optimize the asymptotic performance of the D-SPRT, the local thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ should be large, but smaller than $\tilde{A}, \tilde{B}$; in particular, $\overline{\Delta}_i, \underline{\Delta}_i$ should be around the *square root* of $\tilde{A}, \tilde{B}$. This result reflects the trade-off in the heart of the discrete-time D-SPRT. On the one hand, small thresholds

$\{\overline{\Delta}_i, \underline{\Delta}_i\}$ allow the overshoot to be large and aggravate the inflicted performance loss. On the other hand, large thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ stabilize the overshoot but delay the communication between sensors and fusion center and the corresponding decision at the fusion center. We illustrate this trade-off in a "microscopic" level in Fig. 2.12 and in a "macroscopic" level in Fig. 2.13.

The proof of this proposition is quite involved and we present in the end of this chapter. Before we do that, in the next section we explore how the performance of the D-SPRT can be improved dramatically with *oversampling* at the sensors.

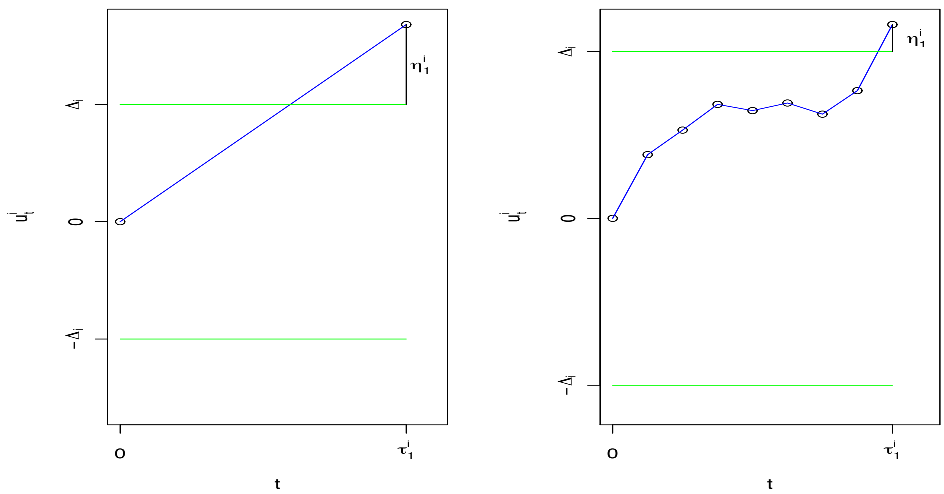


Figure 2.12: Small vs. Large Thresholds and overshoot effect

2.4.3 Oversampling and order-2 asymptotic optimality

The previous proposition implies that the performance loss of the discrete-time D-SPRT is unbounded and that asymptotic optimality can be achieved only if we let $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$. However, it is easy to see from (2.89) that for any fixed thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ the inflicted performance loss becomes asymptotically bounded as long as we let $\theta \to 0$ so that $\theta = \mathcal{O}(|\log \alpha|^{-1})$.

Therefore, if we fix the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$, the D-SPRT will have a bounded distance from the SPRT as long as the overshoot parameter θ vanishes at a certain

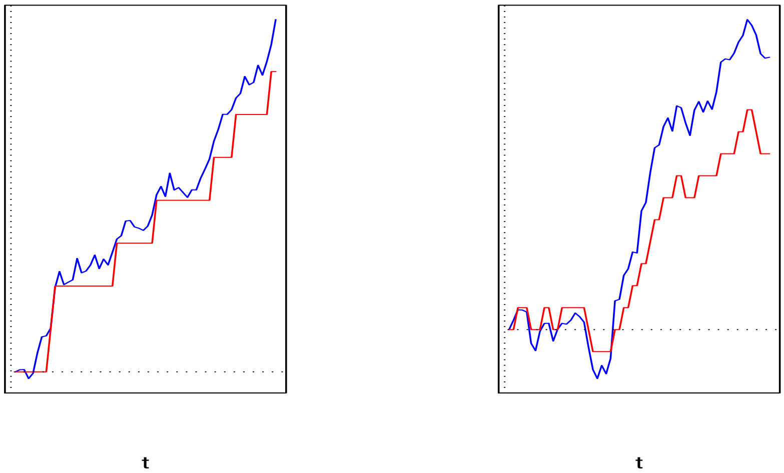


Figure 2.13: Small vs. large Thresholds and overshoot effect.

rate with the horizon of observations. Of course, for this result to be of practical interest, the designer of the scheme should be able to control the overshoot parameter θ and make it as small as possible, just like he can control the horizon of observations through $\tilde{A}, \tilde{B}$ and the rate of communication through $\overline{\Delta}_i, \underline{\Delta}_i$. It turns out that this is indeed possible in some cases, as the one that we now present.

Suppose that the underlying sensor processes are independent Brownian motions, so that the underlying hypothesis testing problem is:

$$\mathsf{H}_0 : \xi_t^i = W_t^i \qquad \mathsf{H}_1 : \xi_t^i = W_t^i + b_i t, \quad t \geq 0 \tag{2.92}$$

where $(W^1, \ldots, W^K)$ is a K-dimensional Brownian motion and $b_1, \ldots, b_K$ known constants. If each sensor i observes its underlying process $\{\xi_t^i\}$ continuously, then (2.92) is a special case of (2.5) and the resulting D-SPRT is *order-2* asymptotically optimal.

On the other hand, if all sensors observe their underlying process only at the discrete-times $t = 0, h, 2h, \ldots$, as it is the case in practice, then the hypothesis testing problem (2.92) becomes

$$\mathsf{H}_0 : \xi_{nh}^i - \xi_{(n-1)h} \sim^{iid} \mathcal{N}(0, h), \, \forall i \qquad \mathsf{H}_1 : \xi_{nh}^i - \xi_{(n-1)h} \sim^{iid} \mathcal{N}(b_i h, h), \, \forall i \tag{2.93}$$

which is a special case of (2.28), therefore (2.89) characterizes the asymptotic performance of the discrete-time D-SPRT.

We will now show that in the context of problem (2.93) we can control the overshoot parameter θ through the sampling period h at the sensors. In particular, we will prove that $\theta = \mathcal{O}(h^{1/4})$.

In order to establish this claim, we start by the log-likelihood ratio of the Brownian increments $\{\xi^i_{th} - \xi^i_{(t-1)h}\}_{t\in\mathbb{N}}$, which is

$$u^i_t = \sum_{n=1}^{t} \Big[-0.5|b_i|^2 h + b_i(\xi^i_{nh} - \xi^i_{(n-1)h}) \Big], \quad t \in \mathbb{N}. \tag{2.94}$$

Setting

$$\underline{\tau}^i_1 = \inf\{t > 0 : u^i_t \le -\underline{\Delta}_i\}; \quad \overline{\tau}^i_1 = \inf\{t > 0 : u^i_t \ge \overline{\Delta}_i\}. \tag{2.95}$$

we can write

$$\begin{aligned} \tau^i_1 &\equiv \inf\{t > 0 : u^i_t \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\} = \min\{\underline{\tau}^i_1, \overline{\tau}^i_1\} \\ \eta^i_1 &\equiv (u^i_{\tau^i_1} - \overline{\Delta}_i)^+ - (u^i_{\tau^i_1} + \underline{\Delta}_i)^- \\ &= (u^i_{\overline{\tau}^i_1} - \overline{\Delta}_i)_{\{u^i_{\tau^i_1} \ge \overline{\Delta}_i\}} - (u^i_{\underline{\tau}^i_1} + \underline{\Delta}_i)_{\{u^i_{\tau^i_1} \le -\underline{\Delta}_i\}} \end{aligned} \tag{2.96}$$

and we have the following upper bound on $\mathsf{E}_j[|\eta^i_1|]$:

$$\mathsf{E}_j[|\eta^i_1|] \le \mathsf{E}_j[u^i_{\overline{\tau}^i_1} - \overline{\Delta}_i] + \mathsf{E}_j[-(u^i_{\underline{\tau}^i_1} + \underline{\Delta}_i)]. \tag{2.97}$$

Then, from [27] we obtain the following upper bounds on the two terms of the right-hand side:

$$\begin{aligned} \sup_{\overline{\Delta}_i > 0} \mathsf{E}_j[u^i_{\overline{\tau}^i_1} - \overline{\Delta}_i] &\le \left[\frac{r+2}{r+1} \frac{\mathsf{E}_j[|\ell^i_1|^{r+1}]}{|\mathsf{E}_j[\ell^i_1]|} \right]^{1/r}, \\ \sup_{\underline{\Delta}_i > 0} \mathsf{E}_j[-(u^i_{\underline{\tau}^i_1} + \underline{\Delta}_i)] &\le \left[\frac{r+2}{r+1} \frac{\mathsf{E}_j[|\ell^i_1|^{r+1}]}{|\mathsf{E}_j[\ell^i_1]|} \right]^{1/r} \end{aligned} \tag{2.98}$$

where $r \ge 1$. Setting $r = 2$ we get:

$$\sup_{\overline{\Delta}_i > 0} \mathsf{E}_1[u^i_{\overline{\tau}^i_1} - \overline{\Delta}_i] = \mathcal{O}(h^{1/4}) \quad , \quad \sup_{\underline{\Delta}_i > 0} -\mathsf{E}_0[u^i_{\underline{\tau}^i_1} + \underline{\Delta}_i] = \mathcal{O}(h^{1/4}). \tag{2.99}$$

and consequently $\theta = \max_i \max_j \mathsf{E}_j[|\eta^i_1|] = \mathcal{O}(h^{1/4})$, which was the claim we wanted to prove. Therefore, the designer of the scheme can force θ to tend to 0 by letting $h \to 0$.

We can now state the following proposition.

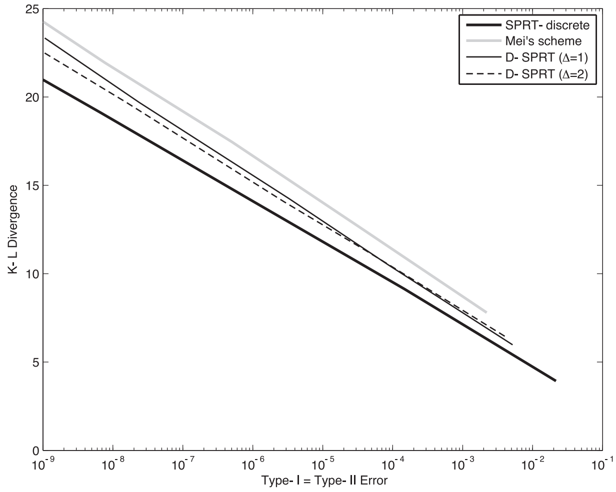
Proposition 4. *Consider the hypothesis testing problem (2.93) and the discrete-time D-SPRT where the thresholds* $\overline{\Delta}_i, \underline{\Delta}_i$ *are arbitrary and fixed. If we let* $h \to 0$ *and* $\alpha \to 0$ *so that* $h^{1/4} \cdot |\log \alpha| = \mathcal{O}(1)$*, the discrete-time D-SPRT becomes order-2 asymptotically optimal, i.e.* $\mathsf{E}_j[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_j[u_{\mathcal{S}}] = \mathcal{O}(1), \quad j = 0, 1.$

This proposition reconciles the behavior of the D-SPRT in discrete and continuous time. In particular, it specifies how frequent the sampling at the sensors must be for the assumption of "continuous-time" to be valid. Moreover, it implies that the sensors should sample their underlying continuous-time processes as frequently as possible without worrying very much about the choice of the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$. Indeed, if the sensor processes are sampled "sufficiently" fast, the resulting performance loss is bounded, whereas the D-SPRT performance becomes a decreasing function of the thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ and the latter will be determined exclusively by the cost of communication and the available budget.

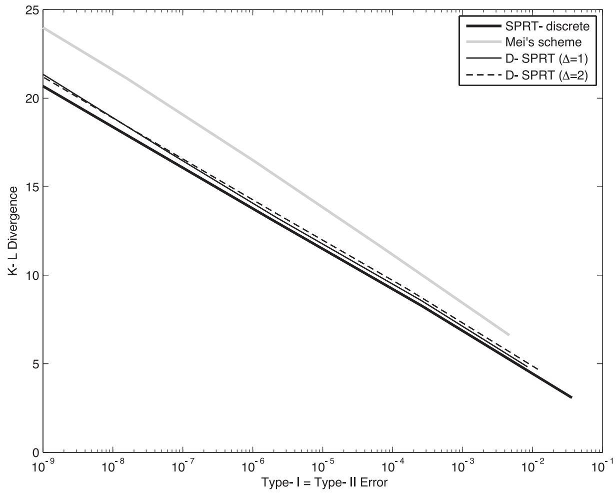
It is important to underline that this property of the D-SPRT is not at all trivial and is not necessarily shared by other decentralized schemes. We illustrate this point with some simulation experiments. We set $K = 2$, $b_1 = b_2 = 1$ and we consider two values for the sampling period h and the thresholds $\overline{\Delta}_i = \underline{\Delta}_i$, i.e. $h = 1, 0.1$ and $\overline{\Delta}_i = \underline{\Delta}_i = 1, 2$. We compare the discrete time D-SPRT with the optimal discrete time SPRT and also with the test suggested by Mei in [31], which is asymptotically optimal.

Fig. 2.14 depicts the K-L divergence of the competing schemes. We recall that in this case the K-L divergence is proportional to the expected detection delay. The reason that we decided to present the former measure instead of the latter is because the K-L divergence is independent of the size of the samples while the detection delay varies drastically with this quantity (smaller samples tend to need more time to reach the same threshold).

We observe that D-SPRT exhibits a notable performance improvement when we go from the value $h = 1$ to $h = 0.1$. This is in complete accordance with our previous analysis since $h = 0.1$ generates likelihood ratios and overshoots of



(a)



(b)

Figure 2.14: Performance of centralized and decentralized tests for Brownian Motions

smaller size than $h = 1$. The optimum SPRT on the other hand and Mei's scheme are relatively insensitive to this change of size in the samples. For D-SPRT, it is basically the error accumulation expressed though the difference $|u_t - \tilde{u}_t|$ that improves as we use smaller h, incurring an overall performance improvement. What is also worth emphasizing for the D-SPRT is that the communication frequency (expressed in continuous time) between the sensors and the fusion center *stays relatively unchanged* under both values of h while in the other two schemes it increases by a factor of 10.

Finally, in Fig. 2.14 we can also observe that the performance of the D-SPRT, as a function of the local threshold value $\underline{\Delta}_i = \overline{\Delta}_i = \Delta$, is not monotone. Indeed, case $\Delta = 2$ is better than $\Delta = 1$ for smaller values of α. Additionally, the error probability values where $\Delta = 2$ prevails are increasing with the size of the samples. This performance can be explained by our analysis. We recall that the optimum local threshold is $\Theta(\sqrt{\theta|\log\alpha|})$ suggesting that the error probability where any specific Δ is optimum is roughly $\alpha = \Theta(\exp(-\Delta^2/\theta))$. Consequently, a larger threshold delivers better performance at a smaller error probability and this value is an increasing function of the size θ of the samples.

2.4.4 Proof

In this section we prove inequality (2.89), which is the analogue of inequality (2.39) and guarantees the asymptotic optimality of the D-SPRT in discrete-time. We prove the result only under H_1 with the understanding that the proof is almost identical under H_0.

We start with the main idea of the proof and discuss the additional complications that are presented in the discrete-time setup. First of all, we observe that $|\tilde{u}_t - \tilde{u}_{t-1}| \leq C' = \sum_{i=1}^{K}(\overline{\Lambda}_i + \underline{\Lambda}_i)$, which is the analogue of the second inequality in (2.41). From this observation it is clear that $\tilde{u}_{\tilde{\mathcal{S}}} - \tilde{B} \leq C'$. Moreover, we recall that for the discrete-time centralized SPRT we have $\mathsf{E}_1[u_{\mathcal{S}}] \geq |\log\alpha| + o(1)$, therefore

–following the same steps as in the proof of (2.39)– we have:

$$\begin{aligned} \mathsf{E}_1[u_{\tilde{\mathcal{S}}} - u_{\mathcal{S}}] &\leq \mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] + \mathsf{E}_1[\tilde{u}_{\tilde{\mathcal{S}}}] - \mathsf{E}_1[u_{\mathcal{S}}] \\ &\leq \mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] + \tilde{B} + C' - |\log\alpha| + o(1) \\ &\leq \mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] + (\tilde{B} - |\log\alpha|) + C' + o(1) \end{aligned} \tag{2.100}$$

In the case of Itô processes, both $\mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|]$ and $\tilde{B} - |\log\alpha|$ were bounded by $C = \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i)$ due to the pathwise inequality $|u_t - \tilde{u}_t| \leq C$, $t \geq 0$. However, a similar inequality is no longer true in discrete time due to the emergence of the overshoots.

Therefore, in order to prove (2.89) we essentially have two establish appropriate upper bounds for $\tilde{B} - |\log\alpha|$ and $\mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|]$, i.e. to estimate the overshoot accumulation.

2.4.4.1 Connecting the thresholds $\tilde{A}, \tilde{B}$ with the error probabilities α, β

Lemma 1.

$$\tilde{B} \leq |\log\alpha|\ ,\ \tilde{A} \leq |\log\beta| \tag{2.101}$$

Proof. Let us assume for simplicity that at any given time the fusion center receives at most one message from the sensors. This will allow us to prove the result avoiding difficulties in the notation, but we remove this assumption in the end of the proof.

First of all, we denote by z_n the nth binary message that arrives at the fusion center *irrespectively of the sensor which sent it* and by k_n the identity of the sensor which transmitted the nth sample. The flow of information at the fusion center is then described by the filtration $\{\mathcal{C}_n\}$, where $\mathcal{C}_n = \sigma((z_1, k_1)\ldots,(z_n, k_n))$. The fusion center likelihood under H_j after the arrival of the first n messages is:

$$\begin{aligned} \mathsf{P}_j((z_1,k_1),\ldots,(z_n,k_n)) &= \mathsf{P}_j(k_1,\ldots,k_n)\prod_{l=1}^{n}\mathsf{P}_j(z_l|z_1,\ldots,z_{l-1},k_1,\ldots,k_n) \\ &= \mathsf{P}_j(k_1,\ldots,k_n)\cdot\prod_{l=1}^{n}\mathsf{P}_j(z_l|k_l) \end{aligned} \tag{2.102}$$

The first equality uses simply the definition of conditional probability. The second equality is based on the fact that z_l (the value of the lth transmitted message at the fusion center) is independent of all other messages $\{(z_j, k_j), j \neq l\}$ *conditionally on* k_l (the identity of the sensor from which the lth message was transmitted). For the third equality we simply used our notation that a sample coming from sensor i is denoted as z^i.

The likelihood ratio after the arrival of the nth message is

$$\frac{\mathsf{P}_1((z_1, k_1), \ldots, (z_n, k_n))}{\mathsf{P}_0((z_1, k_1) \ldots, (z_n, k_n))} = \phi_n \, e^{\tilde{v}_n} \tag{2.103}$$

where –recalling the definition of the log-likelihood ratios $\overline{\Lambda}_i, \underline{\Lambda}_i$– we define

$$\begin{aligned} \phi_n &= \frac{\mathsf{P}_1(k_1, \ldots, k_n)}{\mathsf{P}_0(k_1, \ldots, k_n)} \\ \tilde{v}_n &= \sum_{j=1}^{n} [\overline{\Lambda}_{k_j} \, z_j + \underline{\Lambda}_{k_j} \, (1 - z_j)]. \end{aligned} \tag{2.104}$$

The process $\tilde{v}_n$ is of course closely related to the process $\tilde{u}_t$. Note that $\tilde{u}_t$ is expressed in terms of the global time t, whereas $\tilde{v}_n$ in terms of the *number of messages* n received by the fusion center. To explicitly specify their dependence, let $\{\tau_n\}$ be the increasing sequence of communication times between *any* sensor and the fusion center, where τ_n is the time instant (in global time) that the fusion center receives its nth message. Then the two processes are related through the equality $\tilde{v}_n = \tilde{u}_{\tau_n}$.

The fusion center policy can also be expressed in terms of number of messages at the fusion center as

$$\begin{aligned} \tilde{\mathscr{N}} &= \inf\{n \in \mathbb{N} : \tilde{v}_n \notin (-\tilde{A}, \tilde{B})\} \\ d_{\tilde{\mathscr{N}}} &= \begin{cases} 1, & \text{if } \tilde{v}_{\tilde{\mathscr{N}}} \geq \tilde{B} \\ 0, & \text{if } \tilde{v}_{\tilde{\mathscr{N}}} \leq -\tilde{A}, \end{cases} \end{aligned} \tag{2.105}$$

and we clearly have $\tilde{\mathscr{T}} = \tau_{\tilde{\mathscr{N}}}$ and $d_{\tilde{\mathscr{T}}} = d_{\tilde{\mathscr{N}}}$. Now, $\tilde{\mathscr{N}}$ is a $\{\mathcal{C}_n\}$-stopping time which represents the number of messages that are collected by the fusion center until a decision is reached by D-SPRT, whereas $d_{\tilde{\mathscr{N}}}$ is $\mathcal{C}_{\tilde{\mathscr{N}}}$-measurable random variable which represents the D-SPRT decision rule.

Since $\{d_{\tilde{\mathscr{N}}} = 0\} = \{\tilde{v}_{\tilde{\mathscr{N}}} \leq -\tilde{A}\} \in \mathcal{C}_{\tilde{\mathscr{N}}}$, with a change of measure we have

$$\beta = \mathsf{P}_1(d_{\tilde{\mathscr{N}}} = 0) = \mathsf{E}_1[\ _{\{\tilde{v}_{\tilde{\mathscr{N}}} \leq -\tilde{A}\}}] = \mathsf{E}_0[e^{\tilde{v}_{\tilde{\mathscr{N}}}}\ \phi_{\tilde{\mathscr{N}}}\ \ _{\{\tilde{v}_{\tilde{\mathscr{N}}} \leq -\tilde{A}\}}] \leq e^{-\tilde{A}}\, \mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}], \tag{2.106}$$

and taking logarithms in both sides we obtain $\tilde{A} \leq |\log \beta| + \log \mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}]$, thus it suffices to show that $\mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = 1$. But $\{\phi_n\}$ is a likelihood ratio, thus it is a $(\mathsf{P}_0, \{\mathcal{C}_n\})$-martingale with P_0-expectation equal to 1. Therefore, it suffices to show that we can apply optional sampling theorem. This is possible due to the special form of the $\{\mathcal{C}_n\}$-stopping time $\tilde{\mathscr{N}}$.

Indeed, since $\tilde{\mathscr{N}}$ is a P_0-a.s. finite stopping time, it suffices to show that $\mathsf{E}_0[|\phi_{\tilde{\mathscr{N}}}|] < \infty$ and $\lim_{n\to\infty} \mathsf{E}_0[\phi_n\ \ _{\{n<\tilde{\mathscr{N}}\}}] = 0$. Since ϕ_n is a $\mathcal{C}_n$-measurable random variable and $\{n < \tilde{\mathscr{N}}\} \in \mathcal{C}_n$, from a change of measure we obtain

$$\begin{aligned} \mathsf{E}_0[\phi_n\ \ _{\{n<\tilde{\mathscr{N}}\}}] &= \mathsf{E}_1[e^{-\tilde{v}_n}\phi_n^{-1}\phi_n\ \ _{\{n<\tilde{\mathscr{N}}\}}] \\ &= \mathsf{E}_1[e^{-\tilde{v}_n}\ \ _{\{n<\tilde{\mathscr{N}}\}}] \leq e^{\max\{\tilde{A},\tilde{B}\}}\mathsf{P}_1(n < \tilde{\mathscr{N}}) \to 0 \end{aligned} \tag{2.107}$$

as $n \to \infty$. Notice that the inequality is due to the fact that $-\tilde{A} < \tilde{v}_n < \tilde{B}$ for $n < \tilde{\mathscr{N}}$, whereas for the limit we have used the fact that $\tilde{\mathscr{N}}$ is P_1-a.s. finite.

Similarly, we have

$$\mathsf{E}_0[|\phi_{\tilde{\mathscr{N}}}|] = \mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = \mathsf{E}_1[e^{-\tilde{v}_{\tilde{\mathscr{N}}}}\phi_{\tilde{\mathscr{N}}}^{-1}\phi_{\tilde{\mathscr{N}}}] = \mathsf{E}_1[e^{-\tilde{v}_{\tilde{\mathscr{N}}}}] \leq e^{\max\{\tilde{A},\tilde{B}\}+C'} < \infty \tag{2.108}$$

Therefore, we can apply the optional sampling theorem and obtain $\mathsf{E}_0[\phi_{\tilde{\mathscr{N}}}] = 1$, which proves the first inequality in (2.89). The second inequality can be shown in an analogous way.

Of course, two or more sensors can transmit a message at the same time, thus we need to modify the previous proof. However, this is straightforward; we can denote by z_n and k_n the *vector* of transmitted messages and labels, respectively, at the *nth* time the fusion center receives messages from the sensors. For example, if the first messages that the fusion center receives come concurrently from sensor 1 and sensor 3, then we have $k_1 = (1, 3)$. Moreover, if sensor 1 has transmitted an "upward" message and sensor 3 a "downward" one, then we have $z_1 = (1, 0)$ and the log-likelihood ratio becomes $\tilde{v}_1 = \overline{\Lambda}_1 - \underline{\Lambda}_3$. □

2.4.4.2 Asynchronous Wald's identities

Let us set $\zeta_n^i = f(\delta_n^i, z_n^i, \eta_n^i)$, where $\delta_n^i = \tau_n^i - \tau_{n-1}^i$ and $f(\delta, z, \eta)$ is an arbitrary Borel function. Since the triplets $(\delta_n^i, z_n^i, \eta_n^i)_{n\in\mathbb{N}}$ are independent and identically distributed, it follows that $\{\zeta_n^i\}$ is also a sequence of independent and identically distributed random variables under each hypothesis.

For each i we consider the following filtration $\mathcal{C}_n^i \equiv \mathcal{F}_{\tau_n^i},\ n \in \mathbb{N}$ and in the following lemma we connect $\{\mathcal{F}_t\}$-stopping times and $\{\mathcal{C}_n^i\}$-stopping times.

Lemma 2. *Let $\mathcal{T}$ be an arbitrary $\{\mathcal{F}_t\}$-stopping time. Then, for each i, the random variable $m_{\mathcal{T}}^i + 1$ is an $\{\mathcal{C}_n^i\}$-stopping time, where $m_t^i = \max\{n : \tau_n^i \leq t\}$.*

Proof. For any $n \in \mathbb{N}$ we have:

$$\begin{aligned}\{m_{\mathcal{T}}^i + 1 = n\} &= \{m_{\mathcal{T}}^i = n-1\} = \{\tau_{n-1}^i \leq \mathcal{T} < \tau_n^i\} \\ &= \{\mathcal{T} < \tau_{n-1}^i\}^c \cap \{\mathcal{T} < \tau_n^i\} \\ &= \{\mathcal{T} \leq \tau_{n-1}^i - 1\}^c \cap \{\mathcal{T} \leq \tau_n^i - 1\}\end{aligned} \tag{2.109}$$

Then, since $\mathcal{T}$ is an $\{\mathcal{F}_t\}$-stopping time, we have:

$$\{m_{\mathcal{T}}^i + 1 = n\} \in \mathcal{F}_{\tau_{n-1}^i - 1} \cap \mathcal{F}_{\tau_n^i - 1} \subset \mathcal{F}_{\tau_n^i} \equiv \mathcal{C}_n^i. \tag{2.110}$$

□

Lemma 3. *If $\mathcal{T}$ is a P_j-integrable $\{\mathcal{F}_t\}$-stopping time and $\mathsf{E}_j[|\zeta_1^i|] < \infty$, then:*

$$\mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i+1} \zeta_n^i\Big] = \mathsf{E}_j[m_{\mathcal{T}}^i + 1]\mathsf{E}[\zeta_1^i] \tag{2.111}$$

If moreover $\zeta_n^i \geq 0$, then:

$$\mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] \leq (\mathsf{E}_j[m_{\mathcal{T}}^i] + 1)\ \mathsf{E}[\zeta_1^i] \tag{2.112}$$

whereas if $|\zeta_n^i| \leq M$, then:

$$\Big|\mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] - \mathsf{E}_j[m_{\mathcal{T}}^i]\ \mathsf{E}_j[\zeta_1^i]\Big| \leq 2M \tag{2.113}$$

Proof. Since $\{\zeta_n\}$ is a $\{\mathcal{C}_n^i\}$-adapted sequence of independent and identically distributed random variables with finite mean and $m_{\mathcal{T}}^i+1$ an integrable $\{\mathcal{C}_n^i\}$-stopping time, (2.111) is a classical Wald's identity. The integrability of $m_{\mathcal{T}}^i+1$ follows from the integrability of $\mathcal{T}$, since $m_{\mathcal{T}} \leq \mathcal{T}$.

For (2.112) it suffices to observe that since $\zeta_n^i \geq 0$ we have:

$$\mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i+1} \zeta_n^i\Big] \leq \mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i+1} \zeta_n^i\Big] = \mathsf{E}_j[m_{\mathcal{T}}^i+1]\mathsf{E}[\zeta_1^i] \tag{2.114}$$

For (2.113), we have:

$$\begin{aligned} \mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i+1} \zeta_n^i\Big] - \mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] &= \mathsf{E}_j[\zeta_{m_{\mathcal{T}}^i+1}^i] \\ \mathsf{E}_j[m_{\mathcal{T}}^i+1]\,\mathsf{E}_j[\zeta_1^i] - \mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] &= \mathsf{E}_j[\zeta_{m_{\mathcal{T}}^i+1}^i] \\ \mathsf{E}_j[m_{\mathcal{T}}^i]\,\mathsf{E}_j[\zeta_1^i] - \mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] &= \mathsf{E}_j[\zeta_{m_{\mathcal{T}}^i+1}^i] - \mathsf{E}_j[\zeta_1^i] \end{aligned} \tag{2.115}$$

and consequently since $|\zeta_n^i| \leq M$ we have:

$$\Big|\mathsf{E}_j\Big[\sum_{n=1}^{m_{\mathcal{T}}^i} \zeta_n^i\Big] - \mathsf{E}_j[m_{\mathcal{T}}^i]\,\mathsf{E}_j[\zeta_1^i]\Big| \leq |\mathsf{E}_j[\zeta_{m_{\mathcal{T}}^i+1}^i]| + |\mathsf{E}_j[\zeta_1^i]| \leq 2M \tag{2.116}$$

□

Lemma 4. *Suppose that $\mathsf{E}_0[e^{\zeta_1^i}] = 1$. If $\mathcal{T}$ is a P_j-finite $\{\mathcal{F}_t\}$-stopping time, then*

$$\mathsf{E}_0\Big[\prod_{n=1}^{m_{\mathcal{T}}^i+1} e^{\zeta_n^i}\Big] = 1 \tag{2.117}$$

If moreover $\zeta_n^i \geq -M$ where M is some positive constant, then:

$$\mathsf{E}_0\Big[\prod_{n=1}^{m_{\mathcal{T}}^i} e^{\zeta_n^i}\Big] \leq e^M \tag{2.118}$$

Proof. (2.117) is a direct consequence of the so-called Wald's likelihood-ratio identity, since $\{\prod_{l=1}^n e^{\zeta_l}\}$ is a $\{\mathcal{C}_n^i\}$-martingale with P_0-expectation equal to 1 and $m_{\mathcal{T}}^i+1$ an $\{\mathcal{C}_n^i\}$-stopping time. For (2.118) we observe:

$$1 = \mathsf{E}_0\Big[\prod_{n=1}^{m_{\mathcal{T}}^i+1} e^{\zeta_n^i}\Big] = \mathsf{E}_0\Big[e^{\zeta_{m_{\mathcal{T}}^i+1}} \prod_{n=1}^{m_{\mathcal{T}}^i} e^{\zeta_n^i}\Big] \geq e^{-M}\,\mathsf{E}_0\Big[\prod_{n=1}^{m_{\mathcal{T}}^i} e^{\zeta_n^i}\Big] \tag{2.119}$$

which leads to the desired result.

□

2.4.4.3 The proof

In this section we present the final steps of the proof in a series of lemmas. Before we do so, we recall that we have set

$$C = \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i)\ ,\ C' = \sum_{i=1}^{K}(\overline{\Lambda}_i + \underline{\Lambda}_i) \tag{2.120}$$

and introduce some additional notation that will be important in the next lemmas. Thus, recalling that $\lambda_n^i = \overline{\Lambda}_i z_n^i - \underline{\Lambda}_i(1 - z_n^i)$ is the log-likelihood ratio of z_n^i, we set

$$\tilde{I}_0^i = \mathsf{E}_0[-\lambda_1^i]\ ,\ \tilde{I}_1^i = \mathsf{E}_1[\lambda_1^i]\ ,\ \tilde{I} = \min_{i,j} \tilde{I}_j^i. \tag{2.121}$$

and

$$R = \max_i\{\overline{\Lambda}_i - \overline{\Delta}_i\ ,\ \underline{\Lambda}_i - \underline{\Delta}_i\}\ ,\ \theta = \max_{i,j} \mathsf{E}_j[|\eta_1^i|] \tag{2.122}$$

The following lemma shows the log-likelihood ratios $\overline{\Lambda}_i, \underline{\Lambda}_i$ are always larger than the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$, but their distance remains bounded no matter how large the thresholds $\overline{\Delta}_i, \underline{\Delta}_i$ become. Moreover, it provides a lower bound for $\tilde{I}$ in terms of $\overline{\Delta}_i, \underline{\Delta}_i$.

Lemma 5. *We recall the definition of the function $s(x, y)$ in (2.15). Then:*

1.

$$0 \leq \overline{\Lambda}_i - \overline{\Delta}_i \leq \frac{\theta}{1 - e^{-\underline{\Delta}_i}} \quad , \quad 0 \leq \underline{\Lambda}_i - \underline{\Delta}_i \leq \frac{\theta}{1 - e^{-\overline{\Delta}_i}}, \tag{2.123}$$

2.

$$\tilde{I}_1^i = s(\overline{\Lambda}_i, \underline{\Lambda}_i) \geq s(\overline{\Delta}_i, \underline{\Delta}_i) \quad , \quad \tilde{I}_0^i = s(\underline{\Lambda}_i, \overline{\Lambda}_i) \geq s(\underline{\Delta}_i, \overline{\Delta}_i) \tag{2.124}$$

3. As $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$

$$C' = \Theta(\Delta) \quad , \quad \frac{1}{\tilde{I}} \leq \frac{1}{\Theta(\Delta)} \quad , \quad R = \frac{\theta}{1 + o(1)} \tag{2.125}$$

Proof. Given the inequalities (2.123), (2.124) follows from the definition of $\tilde{I}^i_0, \tilde{I}^i_1$ and the fact that the function $s(x, y)$ is increasing in each of its arguments. The inequality (2.125) follows also from (2.123), therefore it suffices to prove (2.123).

In order to lighten the notation we denote $z^i = z^i_1$, $\tau^i = \tau^i_1$. Then from a change of measure we have:

$$\begin{aligned} \mathsf{P}_0(z^i = 1) &= e^{-\overline{\Delta}_i}\, \mathsf{E}_1[e^{-(u^i_{\tau^i} - \overline{\Delta}_i)} \quad {}_{\{u^i_{\tau^i} - \overline{\Delta}_i\}}] \\ &= e^{-\overline{\Delta}_i}\, \mathsf{P}_1(z^i = 1)\, \mathsf{E}_1[e^{-(u^i_{\tau^i} - \overline{\Delta}_i)} | u^i_{\tau^i} \geq \overline{\Delta}_i] \end{aligned} \tag{2.126}$$

thus from the definition of $\overline{\Lambda}_i$ we have:

$$e^{-\overline{\Lambda}_i} = \frac{\mathsf{P}_0(z^i = 1)}{\mathsf{P}_1(z^i = 1)} = e^{-\overline{\Delta}_i}\, \mathsf{E}_1[e^{-(u^i_{\tau^i} - \overline{\Delta}_i)} | u^i_{\tau^i} \geq \overline{\Delta}_i] \tag{2.127}$$

and

$$\overline{\Lambda}_i - \overline{\Delta}_i = -\log \mathsf{E}_1[e^{-(u^i_{\tau^i} - \overline{\Delta}_i)} | u^i_{\tau^i} \geq \overline{\Delta}_i] \tag{2.128}$$

From this relationship it is obvious that $\overline{\Delta}_i \leq \overline{\Lambda}_i$. Moreover, from an application of conditional Jensen's inequality in the same relationship we have:

$$\overline{\Lambda}_i - \overline{\Delta}_i \leq \mathsf{E}_1[(u^i_{\tau^i} - \overline{\Delta}_i) |\, u^i_{\tau^i} \geq \overline{\Delta}_i] = \frac{\mathsf{E}_1[(u^i_{\tau^i} - \overline{\Delta}_i)^+]}{\mathsf{P}_1(u^i_{\tau^i} \geq \overline{\Delta}_i)} \leq \frac{\theta}{1 - e^{-\underline{\Delta}_i}} \tag{2.129}$$

which is what we wanted to prove. The final inequality is due to the definition of θ and the following change of measure:

$$\mathsf{P}_1(u^i_{\tau^i} \geq \overline{\Delta}_i) = 1 - \mathsf{P}_1(z^i = 0) = 1 - \mathsf{E}_0[e^{u^i_{\tau^i}} \quad {}_{\{u^i_\tau \leq -\underline{\Delta}_i\}}] \geq 1 - e^{-\underline{\Delta}_i}. \tag{2.130}$$

The other inequality in (2.123) can be proven in a similar way. □

The following lemma is the discrete-time analogue of the pathwise inequality (2.41).

Lemma 6. *Let m^i_t be the number of messages transmitted from sensor i to the fusion center up to time t, i.e. $m^i_t = \max\{n \in \mathbb{N} : \tau^i_n \leq t\}$. Then:*

$$|u_t - \tilde{u}_t| \leq C + \sum_{i=1}^{K} \sum_{n=1}^{m^i_t} \Big[|\eta^i_n| + \max\{\overline{\Lambda}_i - \overline{\Delta}_i \;,\; \underline{\Lambda}_i - \underline{\Delta}_i\} \Big], \quad t \in \mathbb{N} \tag{2.131}$$

Proof. Using the triangle inequality we have

$$\begin{aligned} |u_t - \tilde{u}_t| &\leq \sum_{i=1}^{K} |u_t^i - \tilde{u}_t^i| \\ &\leq \sum_{i=1}^{K} |u_t^i - u_{\tau_{m_t^i}}^i| + \sum_{i=1}^{K}\sum_{n=1}^{m_t^i} |u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i - \lambda_n^i| \\ &\leq C + \sum_{i=1}^{K}\sum_{n=1}^{m_t^i} |u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i - \lambda_n^i| \end{aligned} \tag{2.132}$$

Thus, it remains to show $|u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i - \lambda_n^i| \leq |\eta_n^i| + \max\{\overline{\Lambda}_i - \overline{\Delta}_i, \underline{\Lambda}_i - \underline{\Delta}_i\}$ and it suffices to prove this inequality for $n = 1$. Indeed, we have:

$$\begin{aligned} u_{\tau_1^i}^i - \lambda_1^i &= (u_{\tau_1^i}^i - \overline{\Delta}_i + \overline{\Delta}_i - \overline{\Lambda}_i)\, z_1^i + (u_{\tau_1^i}^i + \underline{\Delta}_i - \underline{\Delta}_i + \underline{\Lambda}_i)\,(1 - z_1^i) \\ &= (u_{\tau_1^i}^i - \overline{\Delta}_i)^+ + (\overline{\Delta}_i - \overline{\Lambda}_i)\, z_1^i - (u_{\tau_1^i}^i + \underline{\Delta}_i)^- - (\underline{\Delta}_i - \underline{\Lambda}_i)\,(1 - z_1^i) \end{aligned} \tag{2.133}$$

Then, from (2.123) we obtain:

$$|u_{\tau_1^i}^i - \lambda_1^i| \leq (u_{\tau_1^i}^i - \overline{\Delta}_i)^+ + (u_{\tau_1^i}^i + \underline{\Delta}_i)^- + \max\{\overline{\Lambda}_i - \overline{\Delta}_i\ ,\ \underline{\Lambda}_i - \underline{\Delta}_i\}, \tag{2.134}$$

which is what we wanted to prove.

□

Lemma 7.

$$\mathsf{E}_j[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] \leq (\theta + R)\left(\frac{|\log \alpha| + 3C'}{\tilde{I}} + K\right),\ j = 0, 1. \tag{2.135}$$

Proof. If we set $\zeta_n^i = |\eta_n^i| + \max\{\overline{\Lambda}_i - \overline{\Delta}_i\ ,\ \underline{\Lambda}_i - \underline{\Delta}_i\}$ in (2.112) and take expectations in (2.131), then – using the definition of θ and R– we obtain

$$\begin{aligned} \mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] &\leq \sum_{i=1}^{K}\Big(\mathsf{E}_1[m_{\tilde{\mathcal{S}}}^i] + 1\Big)\,\Big(\mathsf{E}_1[|\eta_1^i|] + \max\{\overline{\Lambda}_i - \overline{\Delta}_i\ ,\ \underline{\Lambda}_i - \underline{\Delta}_i\}\Big) \\ &\leq (\theta + R)\,(\mathsf{E}_1[m_{\tilde{\mathcal{S}}}] + K). \end{aligned} \tag{2.136}$$

where we denote by $m_t = \sum_{i=1}^{K} m_t^i$ the number of transmissions to the fusion center from *all* sensors up to time t.

Moreover, since λ_n^i is bounded by $\overline{\Lambda}_i + \underline{\Lambda}_i$, setting $\zeta_n^i = \lambda_n^i$ in (2.113) and recalling the definition of $\tilde{I}$ we have:

$$\mathsf{E}_1\Big[\sum_{n=1}^{m^i_{\tilde{\mathcal{S}}}} \lambda_n^i\Big] \geq \mathsf{E}_1[m^i_{\tilde{\mathcal{S}}}]\tilde{I} - 2(\overline{\Lambda}_i + \underline{\Lambda}_i) \tag{2.137}$$

and adding over i we obtain:

$$\mathsf{E}_1[\tilde{u}_{\tilde{\mathcal{S}}}] = \mathsf{E}_1\Big[\sum_{i=1}^{K}\sum_{n=1}^{m^i_{\tilde{\mathcal{S}}}} \lambda_n^i\Big] \geq \tilde{I}\mathsf{E}_1[m_{\tilde{\mathcal{S}}}] - 2C' \tag{2.138}$$

But since $\tilde{u}_{\tilde{\mathcal{S}}}$ cannot exceed $\tilde{B}$ more that C', we have $\tilde{u}_{\tilde{\mathcal{S}}} \leq \tilde{B} + C'$ and since $\tilde{B} \leq |\log\alpha|$ we obtain:

$$\mathsf{E}_1[m_{\tilde{\mathcal{S}}}] \leq \frac{|\log\alpha| + 3C'}{\tilde{I}} \tag{2.139}$$

and consequently (2.136) becomes

$$\mathsf{E}_1[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] \leq (\theta + R)\Big(\frac{|\log\alpha| + 3C'}{\tilde{I}} + K\Big), \tag{2.140}$$

which is what we wanted to prove.

□

The desired result now follows by substituting (2.101), (2.125) (2.135) into (2.100).

Chapter 3

Decentralized quickest detection

The structure of this chapter is the following: we start with a review of centralized sequential change-detection with an emphasis on the Cumulative Sums (CUSUM) test (Sec.3.1). We then define and analyze the proposed decentralized sequential detection rule, which we call D-CUSUM; first in continuous time, in the case of Itô processes (Sec.3.2.1) and then in discrete time, in the case of independent and identically distributed observations (Sec.3.2.2).

The D-CUSUM is the analogue of the decentralized sequential test (D-SPRT) that we analyzed in the previous chapter. This allows us to use some results from the previous chapter and obtain a more compact proof for the asymptotic optimality of D-CUSUM in discrete time.

3.1 Quickest detection under a centralized setup

Let $(\xi_t^1, \ldots, \xi_t^K)_{t \geq 0}$ be a K-dimensional stochastic process, each component of which is observed sequentially at a different location or sensor. We describe the flow of information locally at sensor i and globally at the sensor-network by the filtrations $\{\mathcal{F}_t^i\}_{t \geq 0}$ and $\{\mathcal{F}_t\}_{t \geq 0}$ respectively, where $\mathcal{F}_t^i = \sigma(\xi_s^i,\, 0 \leq s \leq t)$ and $\mathcal{F}_t = \sigma(\xi_s^i,\, 0 \leq s \leq t,\, i = 1, \ldots, K)$.

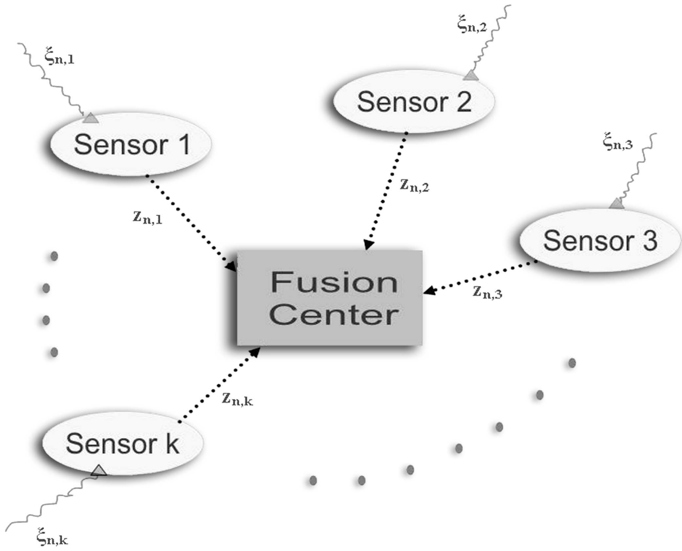


Figure 3.1: Sensor network

All sensors transmit their observations to a fusion center, as it is schematically shown in (3.1). In this section we consider a centralized setup, where the sensors transmit all their observations to the fusion center, therefore the filtration at the fusion center coincides with the global filtration $\{\mathcal{F}_t\}$ at the sensor network.

We assume that due to a disorder in the network or the emergence of a signal the dynamics of each component $\{\xi_t^i\}$ change abruptly and simultaneously at some *unknown* –but deterministic– time τ. Therefore, the distribution P of the stochastic process $\{\xi_t^1, \ldots, \xi_t^K\}_{t\geq 0}$ is parametrized by the time of the change τ and the measure P_τ represents the distribution of $\{\xi_t^1, \ldots, \xi_t^K\}$ *when the change occurs at time* τ. In particular, P_∞ corresponds to the pre-change distribution and P_0 to the post-change distribution.

We require that P_τ is equivalent to P_∞ when both measures are restricted to the σ-algebra $\mathcal{F}_t$ for all $t \in [\tau, \infty)$ and we denote the corresponding Radon-Nikodym derivative as follows

$$\frac{d\mathsf{P}_\tau}{d\mathsf{P}_\infty}\Big|_{\mathcal{F}_t} = e^{u_t - u_\tau}, \quad \tau \leq t < \infty. \tag{3.1}$$

where $u_0 = 1$. Therefore, e^{u_t} is the likelihood ratio of the post-change distribution P_0 versus the pre-change distribution P_∞ given $\mathcal{F}_t$.

The goal is to find a rule at the fusion center that detects the change as quickly as possible, while avoiding many false alarms in repeated applications of this rule. Due to the sequential nature of the observations, a (centralized) detection rule is a stopping time with respect to the fusion center filtration (which is $\{\mathcal{F}_t\}$ under a centralized setup).

Following Lorden [28] and Moustakides [33] we define as optimal centralized detection rule the $\{\mathcal{F}_t\}$-stopping time that minimizes the following criterion:

$$\mathcal{J}[\mathcal{T}] = \sup_{t\geq 0} \operatorname{esssup} \mathsf{E}_t\Big[(u_{\mathcal{T}} - u_t) \quad _{\{\mathcal{T}\geq t\}}|\mathcal{F}_t\Big] \tag{3.2}$$

among $\{\mathcal{F}_t\}$-stopping times $\mathcal{T}$ that satisfy the following false alarm constraint

$$\mathsf{E}_\infty[-u_{\mathcal{T}}] \geq \gamma \tag{3.3}$$

where γ is a fixed, positive constant. This constrained optimization problem was suggested by Moustakides in [33] and is a generalization of the approach proposed by Lorden in [28], according to which the optimal detection rule is the $\{\mathcal{F}_t\}$-stopping time that minimizes the following criterion:

$$\mathcal{J}_L(\mathcal{T}) = \sup_{t\geq 0} \operatorname{esssup} \mathsf{E}_t\Big[(\mathcal{T} - t)^+|\mathcal{F}_t\Big] \tag{3.4}$$

among $\{\mathcal{F}_t\}$-stopping times that satisfy the following false alarm constraint

$$\mathsf{E}_\infty[\mathcal{T}] \geq \gamma \tag{3.5}$$

In both problems, the optimal detection rule has the smallest worst-case conditional "*detection delay*" given the worst possible history up to the time of change among detection rules whose expected "*period of false alarms*" is at least equal to some constant γ. This constant is fixed in advance and expresses the tolerance to false alarms. Thus, both criteria take into account the worst case scenario not only with respect to the time of the change, but also with respect to the whole history up to the time of the change. However, they measure detection delay and penalize false alarms differently; in terms of the actual time in (3.4)-(3.5), in terms of the Kullback-Leibler divergence in (3.2)-(3.3).

Moreover, both criteria are closely associated with the Cumulative Sums (CUSUM) test, which is defined as follows:

$$\mathcal{S}_\nu = \inf\{t \geq 0 : u_t - \inf_{0 \leq s \leq t} u_s \geq \nu\}, \tag{3.6}$$

The CUSUM test was proposed by Page in [39] in '54, much earlier than the above criteria, which actually provided a strong theoretical support to the CUSUM rule. We now review the optimality properties of the CUSUM test with respect to (3.2)-(3.3) in the case of Itô processes and in the case of independent and identically distributed observations.

3.1.1 The case of Itô processes

Suppose that we have the following dynamics

$$\xi_t = {}_{\{t>\tau\}} \int_0^t b_s \, ds + \int_0^t \sigma_s \cdot dW_s, \quad t \geq 0 \tag{3.7}$$

in which case the log-likelihood ratio process $\{u_t\}$ takes the following form

$$\begin{aligned} u_t &= \int_0^t \theta_s \cdot d\xi_s - \frac{1}{2} \int_0^t \theta_s \cdot b_s \, ds, \quad 0 \leq t < \infty \\ \theta_t &= b_t' \cdot (\sigma_t^{-1})' \sigma_t^{-1} \end{aligned} \tag{3.8}$$

where $\{W_t\}$ is a K-dimensional Brownian Motion, $\{b_t\}$ a K-dimensional $\{\mathcal{F}_t\}$-adapted vector and $\{\sigma_t\}$ a $K \times K$ $\{\mathcal{F}_t\}$-adapted matrix.

Moreover, we assume that

$$\mathsf{P}_\infty\Big(\int_0^\infty \theta_s \cdot b_s \, ds = \infty\Big) = \mathsf{P}_0\Big(\int_0^\infty \theta_s \cdot b_s \, ds = \infty\Big) = 1 \tag{3.9}$$

Under condition (3.9), Moustakides showed in [34] that the CUSUM test minimizes criterion (3.2) as long as its threshold ν is chosen so that the the false alarm constraint (3.3) is satisfied with equality, i.e $\gamma = \mathsf{E}_\infty[-u_{\mathcal{S}_\nu}]$, which implies

$$\gamma = e^\nu - \nu - 1. \tag{3.10}$$

It should be mentioned that the CUSUM optimality with respect to (3.2)-(3.3) is among $\{\mathcal{F}_t\}$-stopping times $\mathcal{T}$ that satisfy the integrability conditions

$$\mathsf{E}_0\Big[\int_0^{\mathcal{T}} \theta_s \cdot \beta_s \, ds\Big] < \infty \quad , \quad \mathsf{E}_\infty\Big[\int_0^{\mathcal{T}} \theta_s \cdot \beta_s \, ds\Big] < \infty \tag{3.11}$$

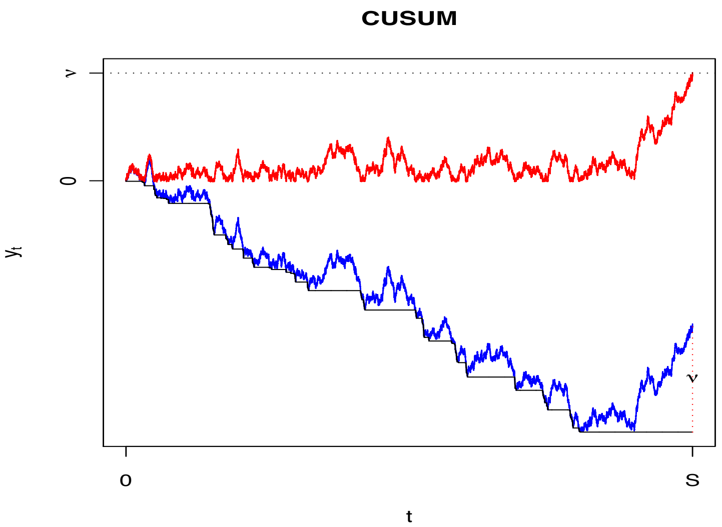


Figure 3.2: The CUSUM test for Itô processes

Note that these conditions are satisfied by the CUSUM stopping time $\mathcal{S}$ due to (3.9). (Actually, if the optimization problem (3.2)-(3.3) is slightly modified, the CUSUM becomes optimal among arbitrary stopping times that satisfy (3.3). We do not present this modification here and refer to [34] for details).

In the special case that the sensors observe drifted Brownian motions before and after the change, the CUSUM is also optimal with respect to Lorden's criterion, i.e. it minimizes (3.4) among stopping times that satisfy (3.5). This result had been earlier established by Shiryaev [52] and Beibel [?].

In [34] it was shown that the optimal performance $J[\mathcal{S}_\nu]$ is equal to the expected detection delay when the change occurs at time $\tau = 0$, i.e. $J[\mathcal{S}_\nu] = \mathsf{E}_0[u_{\mathcal{S}_\nu}]$, which leads to

$$J[\mathcal{S}_\nu] = e^{-\nu} + \nu - 1. \tag{3.12}$$

From (3.10) and (3.12) follows that the CUSUM threshold and the CUSUM performance are independent of the model dynamics (3.7) and completely determined by the design parameter γ. This is due to the fact that the problem (3.2)-(3.3) has incorporated the underlying dynamics (3.7). Moreover, we have $\log \gamma = \nu + o(1)$ and $J[\mathcal{S}_\nu] = \nu + o(1)$ as $\nu \to \infty$, which implies that for sufficiently large values

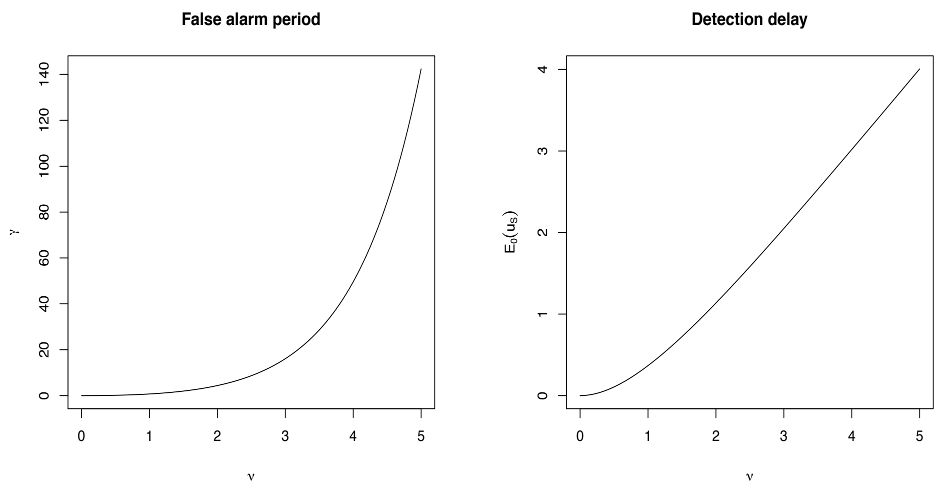


Figure 3.3: False alarm period γ and optimal performance $J(\mathcal{S}_\nu)$ as functions of the CUSUM threshold ν

of threshold ν the CUSUM "detection delay" is a *linear* function of ν, whereas "period of false alarms" is an *exponential* function of ν. This is shown in Fig. 3.3.

Finally, from (3.8) follows that the implementation of the CUSUM rule at the fusion center requires each sensor i to transmit to the fusion center the values of the process

$$u_t^i = \int_0^t \theta_s^i \, d\xi_s^i - 0.5 \int_0^t \theta_s^i b_s^i \, ds, \quad t \geq 0. \tag{3.13}$$

where θ_t^i, b_t^i are the i^{th} components of the vectors θ_t, b_t. It is natural to think of $\{u_t^i\}$ as the sufficient statistic that summarizes the observations from the i^{th} sensor. However, $\{u_t^i\}$ is not in general $\{\mathcal{F}_t^i\}$-adapted, thus locally observable at sensor i. This is always true when the sensor processes are independent, in which case $\{u_t^i\}$ corresponds to the local log-likelihood ratio process. It is also true when the sensor processes are correlated Brownian motions.

3.1.2 CUSUM in discrete-time

Suppose that each sensor i acquires sequentially the discrete-time observations $\{\xi_t^i\}_{t\in\mathbb{N}}$, where $\{\xi_t^i\}_{t\leq\tau}$ and $\{\xi_t^i\}_{t>\tau}$ are sequences of independent observations with common distribution P_∞ and P_0, respectively, where P_0 and P_∞ are known Borel probability measures on $\mathbb{R}^K$.

We assume that there is a probability measure that dominates both P_0 and P_∞ and we denote by f_0 and f_∞ the corresponding Radon-Nikodym derivatives. Thus, the log-likelihood ratio process $\{u_t\}$ takes the form

$$u_t = \sum_{l=1}^{t} \log \frac{f_0(\xi_l^1, \ldots, \xi_l^K)}{f_\infty(\xi_l^1, \ldots, \xi_l^K)}. \tag{3.14}$$

In this context, Moustakides [33] proved that the CUSUM test minimizes Lorden's criterion (3.4) among all stopping rules that satisfy the false alarm constraint (3.5). Since the acquired observations are independent and identically distributed, it is straightforward that the CUSUM test also minimizes (3.2) among stopping rules that satisfy (3.3). As in the continuous time case, the CUSUM threshold should be chosen so that the corresponding false alarm constraint be satisfied with equality, but now we no longer have closed-form expressions for the optimal threshold ν and the optimal performance in terms of γ. However, the asymptotic lower bound $\mathcal{J}[\mathcal{S}_\nu] \geq \log\gamma + o(1)$ as $\gamma \to \infty$ will be sufficient for our purposes.

3.2 Decentralized quickest detection

In this section we propose a novel decentralized detection rule, which combines level-triggered communication with a CUSUM test at the fusion center. The CUSUM test is applied on the sequentially transmitted one-bit messages which are sent asynchronously from the sensors, thus we call this scheme D-CUSUM (Decentralized CUSUM). We define and analyze it first in continuous and then in discrete time.

3.2.1 The case of Itô processes

We consider the change-detection problem (3.7), thus we assume that each sensor observes a standard Brownian motion up to the unknown time-change τ and adopts a random drift after τ. Moreover, we assume that each process $\{u_t^i\}$ –defined in (3.13)– is $\{\mathcal{F}_t^i\}$-adapted, so that sensor i is able to transmit its values to the fusion center. Consequently, the following analysis will hold only when the sensor processes are independent or when they are correlated Brownian motions.

3.2.1.1 Communication scheme and fusion center policy

We suggest that sensor i communicates with the fusion center at the $\{\mathcal{F}_t^i\}$-stopping times which are defined as follows:

$$\tau_n^i \quad = \inf\{t \geq \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \quad n \in \mathbb{N} \tag{3.15}$$

where $\overline{\Delta}_i, \underline{\Delta}_i$ are positive constants, fixed in advance and known to the fusion center. Therefore, a sensor communicates as soon as its locally observed sufficient statistic has either increased by $\overline{\Delta}_i$ or decreased by $\underline{\Delta}_i$ in comparison to its value at the previous communication time. At τ_n^i sensor i transmits the message

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i = \overline{\Delta}_i \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i = -\overline{\Delta}_i \end{cases} \tag{3.16}$$

where we have implicitly used the fact that $u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \in \{-\underline{\Delta}_i, \overline{\Delta}_i\}$, $\forall n \in \mathbb{N}$, since each process $\{u_t^i\}$ has continuous paths. Because of this fact, the fusion center is able to recover the exact value of $\{u_t^i\}$ at the corresponding communication times $\{\tau_n^i\}$ using only the transmitted messages $\{z_n^i\}$. In particular, we have:

$$\tilde{u}_{\tau_n^i}^i = \sum_{j=1}^{n} \Big[\overline{\Delta}_i \, z_j^i - \underline{\Delta}_i (1 - z_j^i)\Big], \quad n \in \mathbb{N}. \tag{3.17}$$

Since we have assumed independence across sensors, the fusion center does not receive any information for the process u^i between the communication times $\{\tau_n^i\}$, therefore it is reasonable to approximate the process u^i at some arbitrary time t as follows:

$$\tilde{u}_t^i = u_{\tau_n^i}^i, \quad t \in [\tau_n^i, \tau_n^i + 1). \tag{3.18}$$

or equivalently

$$\tilde{u}_t^i = \sum_{j=1}^{n} \Big[\overline{\Delta}_i \, z_j^i - \underline{\Delta}_i (1 - z_j^i)\Big], \quad t \in [\tau_n^i, \tau_n^i + 1). \tag{3.19}$$

In other words, we suggest that the fusion center approximate u_t^i with the most recently reproduced value of the process u^i; the resulting approximation $\{\tilde{u}_t^i\}$ is a piecewise constant process with upward jumps of size $\overline{\Delta}_i$ and downward jumps of size $-\underline{\Delta}_i$. Then, mimicking the CUSUM test (3.6), we propose the following detection rule at the fusion center

$$\tilde{\mathcal{S}} = \inf\{t \geq 0 : \tilde{u}_t - \inf_{0 \leq s \leq t} \tilde{u}_s \geq \tilde{\nu}\} \tag{3.20}$$

where $\tilde{u} = \sum_{i=1}^K \tilde{u}^i$ and the positive threshold $\tilde{\nu}$ is chosen so that the false alarm constraint in (3.3) be satisfied with equality. We call the stopping time $\tilde{\mathcal{S}}$ decentralized CUSUM (D-CUSUM), since it mimics the CUSUM test by replacing u with $\tilde{u}$ and can be implemented with the transmission of only one-bit messages from the sensors to the fusion center.

3.2.1.2 Asymptotic Optimality

The following theorem characterizes the performance of the D-CUSUM.

Proposition 5. *The D-CUSUM is order-2 asymptotically optimal for any fixed thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$. In particular, if we set $C = \sum_{i=1}^K (\overline{\Delta}_i + \underline{\Delta}_i)$, we have:*

$$J[\tilde{\mathcal{S}}] - J[\mathcal{S}_\nu] \leq 4C, \tag{3.21}$$

Moreover, if we let $\gamma \to \infty$ and $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ so that $\overline{\Delta}_i, \underline{\Delta}_i = o(\log \gamma)$, then the D-CUSUM is asymptotically optimal of order-1, i.e. $J[\tilde{\mathcal{S}}]/J[\mathcal{S}_\nu] \to 1$.

The proposition is based on the fact that the stopping time $\tilde{\mathcal{S}}$ is pathwise bounded from above and from below by two CUSUM stopping times. In order to emphasize this fact, we present it as a lemma and prove it before we prove the Proposition.

Lemma 8.

$$\mathcal{S}_{\tilde{\nu}-2C} \leq \tilde{\mathcal{S}} \leq \mathcal{S}_{\tilde{\nu}+2C}. \tag{3.22}$$

Proof of Lemma 8. We start by introducing the following notation: $m_t = \inf_{0\le s\le t} u_s$, $\tilde{m}_t = \inf_{0\le s\le t} \tilde{u}_s$ and $y_t = u_t - m_t$, $\tilde{y}_t = \tilde{u}_t - \tilde{m}_t$. Then, the detection rules $\mathcal{S}_\nu$ and $\tilde{\mathcal{S}}$ take the form:

$$\mathcal{S}_\nu = \inf\{t \ge 0 : y_t \ge \nu\} \quad , \quad \tilde{\mathcal{S}} = \inf\{t \ge 0 : \tilde{y}_t \ge \tilde{\nu}\} \tag{3.23}$$

From the definition of the approximations $\{\tilde{u}_t^i\}$ and $\{\tilde{u}_t\}$ and the fact that each process $\{u_t^i\}$ has continuous paths we have:

$$|u_t - \tilde{u}_t| \le \sum_{i=1}^K |u_t^i - \tilde{u}_t^i| \le \sum_{i=1}^K (\overline{\Delta}_i + \underline{\Delta}_i) = C, \quad t \ge 0. \tag{3.24}$$

Using this relationship we also obtain:

$$m_t - C = \inf_{0\le s\le t}(u_t - C) \le \inf_{0\le s\le t} \tilde{u}_t \le \inf_{0\le s\le t}(u_t + C) = m_t + C, \tag{3.25}$$

thus $|m_t - \tilde{m}_t| \le C, \quad t \ge 0$. Consequently, we have:

$$|y_t - \tilde{y}_t| \le |u_t - \tilde{u}_t| + |m_t - \tilde{m}_t| \le 2C, \quad t \ge 0. \tag{3.26}$$

which leads to the desired result. □

Proof of Proposition 5. We start with the observation that for any stopping time $\mathcal{T}$ that satisfies the integrability conditions (3.11) we can write

$$\mathsf{E}_\infty[-u_{\mathcal{T}}] = \mathsf{E}_\infty\Big[\frac{1}{2}\int_0^{\mathcal{T}} \theta_s \cdot b_s \, ds\Big] = \mathsf{E}_\infty\Big[\frac{1}{2}\int_0^{\mathcal{T}} [b_s' \cdot (\sigma_s^{-1})'\sigma_s^{-1} b_s] \, ds\Big] \tag{3.27}$$

and

$$\mathcal{J}[\mathcal{T}] = \sup_{\tau\ge 0} \operatorname{esssup} \mathsf{E}_\tau\Big[\frac{1}{2}\Big(\int_\tau^{\mathcal{T}} [b_s' \cdot (\sigma_s^{-1})'\sigma_s^{-1} b_s] \, ds\Big)^+ |\mathcal{F}_\tau\Big] \tag{3.28}$$

The conditions (3.11) are satisfied by the CUSUM-stopping times $\mathcal{S}_{\tilde{\nu}-2C}, \mathcal{S}_{\tilde{\nu}+2C}$ and consequently by the D-CUSUM stopping time $\tilde{\mathcal{S}}$, due to (3.22).

Therefore, from (3.22) and (3.27) it becomes clear that

$$\mathsf{E}_\infty[-u_{\mathcal{S}_{\tilde{\nu}-2C}}] \le \mathsf{E}_\infty[-u_{\tilde{\mathcal{S}}}] \le \mathsf{E}_\infty[-u_{\mathcal{S}_{\tilde{\nu}+2C}}] \tag{3.29}$$

Moreover, since the thresholds ν and $\tilde{\nu}$ are chosen so that $\mathcal{S}_\nu$ and $\tilde{\mathcal{S}}$ satisfy (3.3) with equality, i.e. $\mathsf{E}_\infty[-u_{\tilde{\mathcal{S}}}] = \mathsf{E}_\infty[-u_{\mathcal{S}_\nu}] = \gamma$, we obtain:

$$\mathsf{E}_\infty[-u_{\mathcal{S}_{\tilde{\nu}-2C}}] \le \mathsf{E}_\infty[-u_{\mathcal{S}_\nu}] \le \mathsf{E}_\infty[-u_{\mathcal{S}_{\tilde{\nu}+2C}}] \tag{3.30}$$

If we now introduce the functions

$$\psi(x) = \mathsf{E}_\infty[-u_{\mathcal{S}_x}] \quad \text{and} \quad \phi(x) = \mathcal{J}(\mathcal{S}_x), \quad x \geq 0\,, \tag{3.31}$$

the inequalities (3.30) translate to

$$\psi(\tilde{\nu} - 2C) \leq \psi(\nu) \leq \psi(\tilde{\nu} + 2C). \tag{3.32}$$

From (3.12) we have that

$$\psi(x) = e^x - x - 1 \quad , \quad \phi(x) = e^{-x} + x - 1, \quad x \geq 0\,, \tag{3.33}$$

which implies that both ψ and ϕ are *strictly increasing* real functions and consequently from (3.32) we obtain $|\nu - \tilde{\nu}| \leq 2C$.

From (3.22) and (3.28) we have $\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}_\nu] \leq \mathcal{J}[\mathcal{S}_{\tilde{\nu}+2C}] - \mathcal{J}[\mathcal{S}_\nu]$. We can now obtain (3.21) if we use (3.33) as follows:

$$\begin{aligned} \mathcal{J}[\mathcal{S}_{\tilde{\nu}+2C}] - \mathcal{J}[\mathcal{S}_\nu] &= \phi(\tilde{\nu} + 2C) - \phi(\nu) \\ &= [e^{-\tilde{\nu}-2C} + (\tilde{\nu} + 2C) - 1] - [e^{-\nu} + \nu - 1] \\ &= (e^{-\tilde{\nu}-2C} - e^{-\nu}) + (\tilde{\nu} - \nu) + 2C \leq 4C \end{aligned} \tag{3.34}$$

where the inequality follows from the fact that $|\nu - \tilde{\nu}| \leq 2C$ and consequently $-\tilde{\nu} - 2C \leq -\nu$.

Finally, from (3.21) we have

$$\frac{\mathcal{J}[\tilde{\mathcal{S}}]}{\mathcal{J}[\mathcal{S}_\nu]} = 1 + \frac{\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}_\nu]}{\mathcal{J}[\mathcal{S}_\nu]} \leq 1 + \frac{4C}{\phi(\nu)} = 1 + \frac{4C}{\log\gamma}\,\frac{\log\gamma}{\phi(\nu)} \tag{3.35}$$

Thus, if we let $\overline{\Delta}_i, \underline{\Delta}_i, \gamma \to \infty$ so that $\overline{\Delta}_i, \underline{\Delta}_i = o(\log\gamma)$, from (3.12) we obtain $\limsup \frac{\mathcal{J}[\tilde{\mathcal{S}}]}{\mathcal{J}[\mathcal{S}_\nu]} \leq 1$. But this implies the order-1 asymptotic optimality of $\tilde{\mathcal{S}}$, since we also have $\mathcal{J}[\tilde{\mathcal{S}}] \geq \mathcal{J}[\mathcal{S}_\nu]$ due to the optimality of the CUSUM test.

□

3.2.1.3 Extensions

We can modify the D-CUSUM in a straightforward way and incorporate more general communication schemes. Thus, we can have time-varying thresholds as in (2.51)- (2.52) or linear (or non- linear) intersecting boundaries as in (2.55)-(2.56). The previous analysis will remain valid as long as inequality (3.24) remains valid.

3.2.1.4 Comparison with the discrete-time centralized CUSUM in the Brownian case

As we did for the testing problem, it is meaningful to compare the D-CUSUM with the *discrete-time centralized CUSUM*, which is based on the transmission to the fusion center of the exact sensor observations at the times $t = 0, h, 2h, \ldots$, where $h > 0$. We perform this comparison for the following change-detection problem:

$$\xi_t^i = W_t^i + \ _{\{t \geq \tau\}}\, b_i t, \quad t \geq 0, \quad i = 1, \ldots, K \tag{3.36}$$

where $(W^1, \ldots, W^K)$ is a K-dimensional Brownian motion and $b_1, \ldots, b_K$ known constants. Thus, we assume that each sensor i observes a standard Brownian up to time τ, which adopts a constant drift b_i after τ.

Under this model, the increments $\{\xi_{nh}^i - \xi_{(n-1)h}^i\}$ are independent and identically distributed, and consequently the discrete-time centralized CUSUM is also order-2 asymptotically optimal (with respect to the continuous-time centralized CUSUM). Therefore, since the asymptotic performance of the two schemes is similar, we need to resort to simulations. Moreover, for the comparison to be fair, we need to equate the expected intersampling periods before and after the change $\mathsf{E}_\infty[\tau_1^i]$ and $\mathsf{E}_0[\tau_1^i]$ with the constant period h, so that the two schemes require the same communication rate between sensors and fusion center *on average.*

Using Wald's identity together with (2.14)-(2.15), we have:

$$\begin{aligned} \mathsf{E}_0[\tau_1^i] = \mathsf{E}_\infty[\tau_1^i] = h \Leftrightarrow \mathsf{E}_0[-u_{\tau_1^i}^i] = \mathsf{E}_\infty[u_{\tau_1^i}^i] = 0.5\,|b_i|^2\,, h \\ \Leftrightarrow s(\overline{\Delta}_i, \underline{\Delta}_i) = s(\underline{\Delta}_i, \overline{\Delta}_i) = 0.5\,|b_i|^2\, h \end{aligned} \tag{3.37}$$

Then, if we set $\overline{\Delta}_i = \underline{\Delta}_i = \Delta_i$, (3.37) becomes $s(\Delta_i, \Delta_i) = 0.5\,|b_i|^2\, h$ and for any given drift b_i we can compute the sampling period h that corresponds to the threshold Δ_i and vice-versa.

For the simulations in Fig.3.4 we chose $K = 2$, $b_1 = b_2 = 1$ and $\overline{\Delta}_i = \underline{\Delta}_i = \Delta_i = 2$ for each sensor i, thus h had to be equal to 3.0462.

In Fig.3.4 we can see that the distance between the D-CUSUM and the optimal continuous-time CUSUM remains bounded. Moreover, the D-CUSUM exhibits

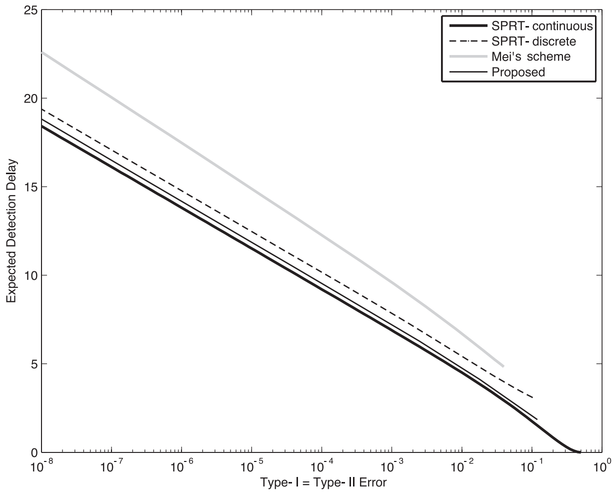


Figure 3.4: D-CUSUM versus CUSUM

a distinct performance improvement over the discrete time centralized CUSUM which is applied after canonical deterministic sampling.

3.2.2 D-CUSUM in discrete time

We now consider the discrete-time case and in particular the change-detection problem (3.40). We assume that the observations are independent across sensors, so that the log-likelihood ratio process $\{u_t\}$ can be written as the sum of the local log-likelihood ratios, i.e. $u_t = \sum_{i=1}^{K} u_t^i$, where

$$u_t^i = \sum_{j=1}^{t} \log \frac{f_0^i(\xi_j^i, \ldots, \xi_j^i)}{f_\infty^i(\xi_j^i, \ldots, \xi_j^i)}. \tag{3.38}$$

3.2.2.1 Communication scheme and fusion center policy

As in the continuous-time setup, we suggest that sensor i communicate with the fusion center at the times

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : u_t^i - u_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \quad n \in \mathbb{N} \tag{3.39}$$

and transmit the messages

$$z_n^i = \begin{cases} 1, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \geq \overline{\Delta}_i \\ 0, & \text{if } u_{\tau_n^i}^i - u_{\tau_{n-1}^i}^i \leq -\underline{\Delta}_i \end{cases} \tag{3.40}$$

Unlike the case of continuous-time observations, the fusion center can no longer recover the values of the process u^i at the corresponding communication times $(\tau_n^i)_{n\in\mathbb{N}}$, because it does not have access to the *overshoots*:

$$\eta_n^i = (u^i_{\tau_n^i} - u^i_{\tau_{n-1}^i} - \overline{\Delta}_i)^+ - (u^i_{\tau_n^i} - u^i_{\tau_{n-1}^i} + \underline{\Delta}_i)^-, \quad n \in \mathbb{N}. \tag{3.41}$$

Thus, the approximations (3.18) and (3.19) are no longer equivalent and approximation (3.18) is no longer implementable. Of course, it is still possible to use approximation (3.18), however, by doing so we ignore the overshoots and this is likely to deteriorate the performance of the resulting decentralized scheme. Therefore, working as in the testing problem, we choose to approximate the process $\{u_t^i\}$ with the log-likelihood ratio of the messages $\{z_n^i\}$ that have been transmitted from sensor i up to time t, i.e.

$$\tilde{u}_t^i = \sum_{j=1}^{n}\Big[\overline{\Lambda}_i z_j^i - \underline{\Lambda}_i(1-z_j^i)\Big], \quad \tau_n^i \le t < \tau_{n+1}^i \tag{3.42}$$

where

$$\overline{\Lambda}_i = \log\frac{\mathsf{P}_0(z_n^i = 1)}{\mathsf{P}_\infty(z_n^i = 1)} \quad , \quad -\underline{\Lambda}_i = \log\frac{\mathsf{P}_0(z_n^i = 0)}{\mathsf{P}_\infty(z_n^i = 0)} \tag{3.43}$$

The computation of $\tilde{u}_t$ depends on the knowledge of the quantities $\{\overline{\Lambda}_i, \underline{\Lambda}_i\}$; these are not known explicitly, however they can be pre-computed using simulations.

We emphasize that this is a partial-likelihood approach, since we ignore the contribution to the total log-likelihood ratio of the inter-communication times $\tau_n^i - \tau_{n-1}^i$. The reason is that the form of this log-likelihood is intractable. Otherwise, it would make perfect sense to use their contribution as well.

After choosing this approximation, we mimic the CUSUM test and suggest the following detection rule at the fusion center

$$\tilde{\mathcal{S}} = \inf\{t \in \mathbb{N} : \tilde{u}_t - \inf_{0\le s\le t} \tilde{u}_s \ge \tilde{\nu}\} \tag{3.44}$$

where $\tilde{u}_t = \sum_{i=1}^K \tilde{u}_t^i$. We call $\tilde{\mathcal{S}}$ decentralized CUSUM (D-CUSUM), as we did in the continuous-path case, although there is a considerable difference in the construction of the fusion center policy.

3.2.2.2 Asymptotic optimality

The following proposition characterizes the (asymptotic) performance of the discrete-time D-CUSUM and implies its asymptotic optimality. For simplicity, we assume that there is a quantity Δ so that $\overline{\Delta}_i, \underline{\Delta}_i = \Theta(\Delta)$ for all $i = 1, \ldots, K$ as $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$. This assumption implies that the frequency of communication is similar in all sensors. Moreover, in order to simplify the notation, we drop the threshold from the CUSUM stopping time $\mathcal{S}_\nu$, which we now denote as $\mathcal{S}$.

Proposition 6. *If we let $\gamma \to \infty$ and $\Delta \to \infty$, then*

$$\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] \leq \frac{\log \gamma}{\Theta(\Delta)} + \Theta(\Delta) \tag{3.45}$$

Therefore, the D-CUSUM is asymptotically optimal of order-1, i.e. $\mathcal{J}[\tilde{\mathcal{S}}]/\mathcal{J}[\mathcal{S}] \to 1$, as $\Delta = o(\log \gamma)$.

Proof. We start by showing that inequality (3.45) implies the (order-1) asymptotic optimality of the D-CUSUM stopping time $\tilde{\mathcal{S}}$. Indeed, recalling the asymptotic lower bound $\mathcal{J}[\mathcal{S}] \geq \log \gamma + o(1)$ on the performance of the discrete-time centralized CUSUM, we obtain:

$$\frac{\mathcal{J}[\tilde{\mathcal{S}}]}{\mathcal{J}[\mathcal{S}]} = 1 + \frac{\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}]}{\mathcal{J}[\mathcal{S}]} \leq 1 + \frac{1}{\Theta(\Delta)} \frac{\Theta(\Delta)}{\log \gamma} = 1 + o(1) \tag{3.46}$$

where the last equality follows if we let $\Delta \to \infty$ and $\gamma \to \infty$ so that $\Delta = o(\log \gamma)$.

We now turn to the proof of (3.45). The Lorden-performance for the CUSUM and the D-CUSUM corresponds to the expected delay when the change occurs at $\tau = 0$, therefore:

$$\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = \mathsf{E}_0[u_{\tilde{\mathcal{S}}}] - \mathsf{E}_0[u_{\mathcal{S}}] \leq \mathsf{E}_0[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] + \mathsf{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] - \mathsf{E}_0[u_{\mathcal{S}}] \tag{3.47}$$

Since $\tilde{u}_0 = 0$, it is clear that $\tilde{u}_t \leq \tilde{y}_t$ for every $t \geq 0$. Moreover, since the overshoot $\tilde{y}_{\tilde{\mathcal{S}}} - \tilde{\nu}$ cannot be larger than $C' = \sum_{i=1}^K (\overline{\Lambda}_i + \underline{\Lambda}_i)$, we obtain $\tilde{u}_{\tilde{\mathcal{S}}} \leq \tilde{\nu} + C'$. Thus, from (3.47) and the asymptotic lower bound $\mathcal{J}[\mathcal{S}_\nu] \geq \log \gamma + o(1)$ we obtain:

$$\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}_\nu] \leq \mathsf{E}_0[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] + (\tilde{\nu} + C') - (\log \gamma + o(1)) \tag{3.48}$$

Moreover, working in exactly the same way as in the testing problem, we can show that $C', C = \Theta(\Delta)$, $\tilde{I} \geq \Theta(\Delta)$, $R = \frac{\theta}{1+o(1)}$ and we can obtain the following upper bound

$$\mathsf{E}_0[|u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}|] \leq (\theta + R)\Big[\frac{\tilde{\nu} + 3C'}{\tilde{I}} + K\Big] \tag{3.49}$$

where θ, R and $\tilde{I}$ are defined as in (2.122) and (2.121), respectively.

Thus, the crucial task, which cannot be shown as in the testing problem, is to connect the threshold $\tilde{\nu}$ with the parameter γ. In order to do so, we denote by z_n the nth binary message that arrives at the fusion center *irrespectively of the sensor which sent it* and by k_n the identity of the sensor which transmitted the nth sample. The flow of information at the fusion center is then described by the filtration $\{\mathcal{C}_n\}$, where $\mathcal{C}_n = \sigma((z_1, k_1) \ldots, (z_n, k_n))$.

For the simplicity of the notation we assume that at any given time the fusion center receives at most one message from the sensors. In the more general case where two or more sensors transmit a message at the same time, the following proof will remain valid as long as we denote by z_n and k_n the *vector* of transmitted messages and labels, respectively, at the *nth* time the fusion center receives messages from the sensors.

For example, if the first messages that the fusion center receives come concurrently from sensor 1 and sensor 3, then we have $k_1 = (1, 3)$. Moreover, if sensor 1 has transmitted an "upward" message and sensor 3 a "downward" one, then we have $z_1 = (1, 0)$.

Then, working in the same way as in the testing problem, we obtain the following likelihood ratio before and after the change

$$\mathscr{L}_n = \frac{\mathsf{P}_0((z_1, k_1), \ldots, (z_n, k_n))}{\mathsf{P}_\infty((z_1, k_1) \ldots, (z_n, k_n))} = e^{\phi_n + \tilde{v}_n} \tag{3.50}$$

where –recalling the definition of the log-likelihood ratios $\overline{\Lambda}_i, \underline{\Lambda}_i$– we define

$$e^{\phi_n} = \frac{\mathsf{P}_0(k_1, \ldots, k_n)}{\mathsf{P}_\infty(k_1, \ldots, k_n)} \quad , \quad \tilde{v}_n = \sum_{j=1}^{n} [\overline{\Lambda}_{k_j}\, z_j + \underline{\Lambda}_{k_j}\, (1 - z_j)]. \tag{3.51}$$

The process $\tilde{v}_n$ is closely related to the process $\tilde{u}_t$; their difference is that $\tilde{u}_t$ is expressed in terms of global time , whereas $\tilde{v}_n$ in terms of times the fusion center has received messages from the sensors. To explicitly specify their dependence, let $\{\tau_n\}$ be the increasing sequence of communication times between *any* sensor and the fusion center, where τ_n is the time instant that the fusion center receives a message for the *nth* time. Then, the two processes are related through the equality $\tilde{v}_n = \tilde{u}_{\tau_n}$.

The fusion center policy can now be expressed as follows

$$\tilde{\mathscr{N}} = \inf\{n \in \mathbb{N} : \tilde{v}_n - \min_{j=1,\ldots,n} \tilde{v}_j \geq \tilde{\nu}\} \tag{3.52}$$

and we clearly have $\tilde{\mathscr{T}} = \tau_{\tilde{\mathscr{N}}}$. Thus, $\tilde{\mathscr{N}}$ is a $\{\mathcal{C}_n\}$-stopping time which represents the number of times the fusion center received messages from the sensors until an alarm is raised by the D-CUSUM.

We can also obtain an alternative representation of the fusion center policy, which will be very useful for the proof of the desired result. In particular, we can write the D-CUSUM stopping time $\tilde{\mathscr{N}}$ as a sum of repeated D-SPRTs. Thus, if we define the following stopping times

$$\mathscr{T}_j = \inf\{n \geq \mathscr{T}_{j-1} : v_n - v_{\mathscr{T}_{j-1}} \notin (0, \tilde{\nu})\} \tag{3.53}$$

then $\tilde{\mathscr{N}} = \mathscr{T}_{\mathcal{R}}$, where

$$\mathcal{R} = \inf\{j \in \mathbb{N} : v_{\mathscr{T}_j} - v_{\mathscr{T}_{j-1}} \geq \tilde{\nu}\} \tag{3.54}$$

The notation we introduced will help us obtain the desired connection between the threshold $\tilde{\nu}$ and γ. We recall that $\tilde{\nu}$ is chosen so that $\mathsf{E}_\infty[-u_{\tilde{\mathcal{S}}}] = \gamma$. From Wald's identity we have $\gamma = \mathsf{E}_\infty[-u_{\tilde{\mathcal{S}}}] = I_\infty \, \mathsf{E}_\infty[\tilde{\mathcal{S}}]$, where $I_\infty = \mathsf{E}_\infty[-u_1]$. Moreover, we have $\mathcal{R} \leq \mathscr{N} \leq K\tilde{\mathcal{S}}$. Notice that both inequalities are very crude; the first one becomes equality when the process $\{\tilde{v}_n\}$ exceeds $\tilde{\nu}$ before 0, whereas the second becomes equality when all sensors communicate with the fusion center at every time t till $\tilde{\mathcal{S}}$.

However, the inequality $\mathcal{R} \leq K\tilde{\mathcal{S}}$ is going to be sufficient for our purposes. Indeed, combining this inequality with $\gamma = I_\infty \, \mathsf{E}_\infty[\tilde{\mathcal{S}}]$ we obtain

$$K\,\gamma \geq I_\infty \, \mathsf{E}_\infty[\mathcal{R}], \tag{3.55}$$

Thus, it suffices to find a lower bound for $\mathsf{E}_\infty[\mathcal{R}]$ in terms of $e^{\tilde{\nu}}$. In order to do that, we start with the layered representation of the expectation

$$\mathsf{E}_\infty[\mathcal{R}] = \sum_{n=0}^{\infty} \mathsf{P}_\infty(\mathcal{R} > n) = \sum_{n=0}^{\infty} \mathsf{E}_\infty[A_n] \tag{3.56}$$

where $A_n = \; {}_{\{\tilde{v}_{\mathscr{T}_1}<0,\ldots,\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}<0\}}$, $n \in \mathbb{N}$.

Then:

$$\begin{aligned}
\mathsf{E}_\infty[A_n] &= \mathsf{E}_\infty[A_{n-1} \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}<0\}}] \\
&= \mathsf{E}_\infty[A_{n-1}\,(1 - \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}\geq\tilde{\nu}\}})] \\
&= \mathsf{E}_\infty[A_{n-1}] - \mathsf{E}_\infty[A_{n-1} \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}\geq\tilde{\nu}\}}]
\end{aligned} \tag{3.57}$$

From a change of measure and the law of iterated expectation we obtain:

$$\begin{aligned}
&\mathsf{E}_\infty[A_{n-1} \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}\geq\tilde{\nu}\}}] \\
&= \mathsf{E}_0[\mathscr{L}^{-1}_{\mathscr{T}_{n-1}} \, A_{n-1} \, e^{-(\phi_{\mathscr{T}_n}-\phi_{\mathscr{T}_{n-1}})-(\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}})} \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}\geq\tilde{\nu}\}}] \\
&= \mathsf{E}_0[\mathscr{L}^{-1}_{\mathscr{T}_{n-1}} \, A_{n-1} \, \mathsf{E}_0[e^{-(\phi_{\mathscr{T}_n}-\phi_{\mathscr{T}_{n-1}})-(\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}})} \; {}_{\{\tilde{v}_{\mathscr{T}_n}-\tilde{v}_{\mathscr{T}_{n-1}}\geq\tilde{\nu}\}}|\mathcal{F}_{\mathscr{T}_{n-1}}]] \\
&\leq e^{-\tilde{\nu}}\mathsf{E}_0[\mathscr{L}^{-1}_{\mathscr{T}_{n-1}} \, A_{n-1} \, \mathsf{E}_0[e^{-(\phi_{\mathscr{T}_n}-\phi_{\mathscr{T}_{n-1}})}|\mathcal{F}_{\mathscr{T}_{n-1}}]] \\
&= e^{-\tilde{\nu}}\mathsf{E}_0[\mathscr{L}^{-1}_{\mathscr{T}_{n-1}} \, A_{n-1}] = e^{-\tilde{\nu}}\mathsf{E}_\infty[A_{n-1}]
\end{aligned} \tag{3.58}$$

Therefore, for every $n \in \mathbb{N}$ we have $\mathsf{E}_\infty[A_n] \geq (1 - e^{-\tilde{\nu}})\mathsf{E}_\infty[A_{n-1}]$, thus $\mathsf{E}_\infty[A_n] \geq (1 - e^{-\tilde{\nu}})^n$ and consequently

$$\mathsf{E}_\infty[\mathcal{R}] = \sum_{n=0}^{\infty} \mathsf{E}_\infty[A_n] \geq \sum_{n=0}^{\infty} (1 - e^{-\tilde{\nu}})^n = e^{\tilde{\nu}} \tag{3.59}$$

Then, recalling (3.55) we obtain

$$\tilde{\nu} \leq \log \frac{K}{I_\infty} + \log \gamma \tag{3.60}$$

which –combined with (3.48) and (3.49)– gives the desired result.

□

Chapter 4

Decentralized parameter estimation

The structure of this chapter is the following; we introcude centralized parameter estimation for a class of Itô processes and argue in favor of a sequential formulation of the problem in Section 4.1. We define and analyze the suggested decentralized estimator in Section 4.2, whereas we consider the case of correlated sensors in Section 4.3.

4.1 Centralized parameter estimation

Let $(\Omega, \mathcal{G}, \mathsf{P}, \{\mathcal{G}_t\})$ be a filtered probability space which hosts the K-dimensional Brownian motion $\{W_t\}_{t\geq 0}$ and the K-dimensional stochastic process $\{\xi_t\}_{t\geq 0}$. In contrast to the Brownian motion $\{W_t\}$ which is non-observable, the process $\{\xi_t\}$ is observed and, in particular, each component $\{\xi_t^i\}$ is observed at a different location (or sensor). Thus, the local history at sensor i up to time t is $\mathcal{F}_t^i = \sigma(\xi_s^i,\, 0 \leq s \leq t)$, whereas the global history up to time t is $\mathcal{F}_t = \sigma(\xi_s,\, 0 \leq s \leq t)$. As it is shown schematically in Fig. 4.1, there is a global decision maker (fusion center), which receives observations from all sensors and is responsible for combining them in order to make the final decision.

We assume that the distribution of $\{\xi_t\}$ is known up to a parameter λ and we

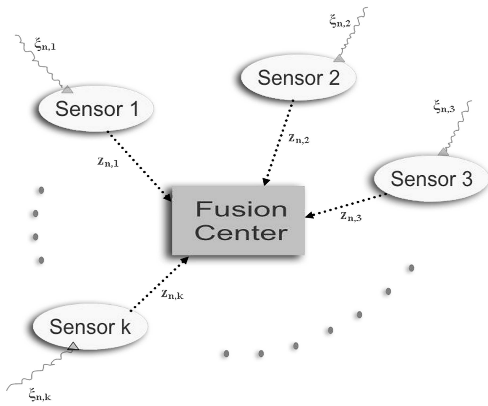


Figure 4.1: Sensor Network

denote it by μ_λ. Thus, the underlying probability measure P is also parametrized by λ and we denote it by P_λ. The goal is to estimate λ using the data acquired by the fusion center, not the sensor observations. When these two coincide, we say that we are in a *centralized* setup and an estimator of λ is simply an $\{\mathcal{F}_t\}$-adapted process. This is going to be our focus on this section.

4.1.1 The Brownian case

Suppose that the sensors observe independent Brownian motions, so that:

$$\xi_t^i = \lambda\, b_i\, t + W_t^i, \quad t \geq 0, \quad i = 1, \ldots, K, \tag{4.1}$$

where $b_1, \ldots, b_K$ are *known* constants. Thus, we assume that the absolute drift in each Brownian motion is unknown, but the *relative* drifts are known. For example, Fig. 4.2 shows the paths of two independent drifted Brownian motions, whose drifts have different signs and sizes.

The *local* likelihood for λ at location i and at time t is the Radon-Nikodym derivative of μ_λ^i with respect to μ_0^i when the two measures are restricted to the σ-algebra $\mathcal{F}_t^i$, i.e.

$$\mathscr{L}_t^i(\lambda) \equiv \frac{d\mu_\lambda^i}{d\mu_0^i}\Big|_{\mathcal{F}_t^i} = \exp\{\lambda\, b_i\, \xi_t^i - 0.5\, \lambda^2 |b_i|^2 t\,\} \tag{4.2}$$

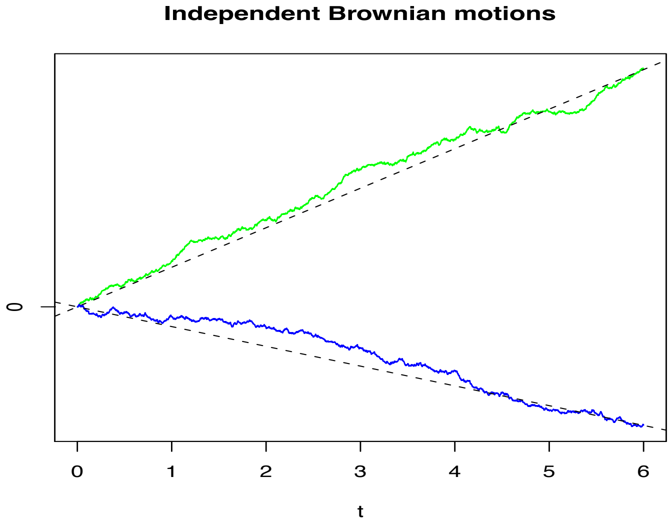


Figure 4.2: Two Brownian paths

Then, the *local* MLE at sensor i for estimating λ is:

$$\lambda_t^i = \frac{\xi_t^i}{b_i t}, \quad t \geq 0. \tag{4.3}$$

Due to the independence between sensors, the global likelihood at time t is the product of the local likelihoods, i.e. $\mathscr{L}_t(\lambda) = \prod_{i=1}^K \mathscr{L}_t^i(\lambda)$, and the global MLE at the fusion center is a weighted average of the local MLEs, i.e.

$$\lambda_t = \sum_{i=1}^K w_i \lambda_t^i, \quad t \geq 0 \tag{4.4}$$

where

$$w_i = \frac{|b_i|^2}{\sum_{i=1}^K |b^i|^2}, \quad i = 1, \dots, K. \tag{4.5}$$

Thus, the fusion center trusts more the local MLEs which correspond to sensors with stronger signals. We illustrate this behavior in fig. 4.3, which is a continuation of fig. 4.2.

The Fisher information at time t is:

$$I_t(\lambda) \quad = \quad \mathsf{E}_\lambda\Big[\Big(\frac{\partial}{\partial\lambda} \log \mathscr{L}_t(\lambda)\Big)^2\Big] = \sum_{i=1}^K |b^i|^2\, t \equiv A_t \tag{4.6}$$

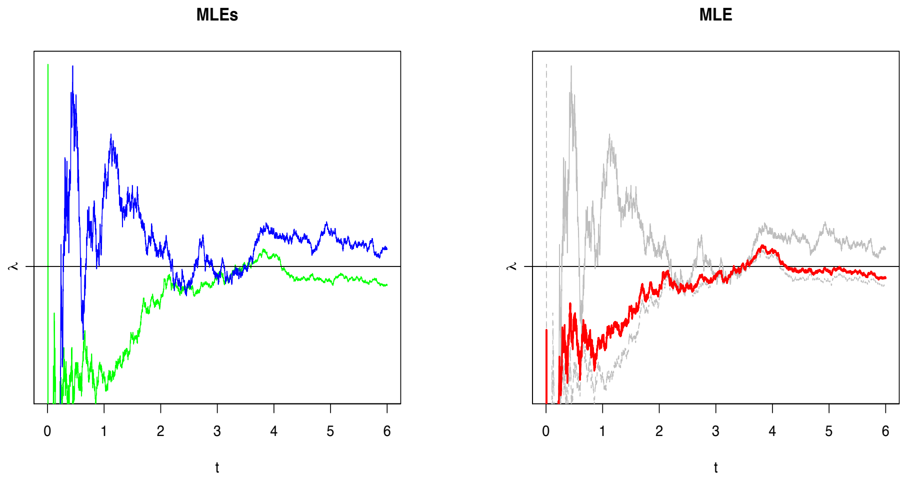


Figure 4.3: The global MLE trusts more the local MLE which corresponds to the drifted Brownian motion (see fig. 4.2) with the largest slope.

and it is straightforward to see that under P_λ we have:

$$\sqrt{A_t}(\lambda_t - \lambda) \sim \mathcal{N}(0,1), \quad t \geq 0 \tag{4.7}$$

Thus, λ_t is an unbiased and normally distributed estimator of λ *for every* $t \geq 0$. Moreover, at any given time t, λ_t is an *optimal* estimator of λ, in the sense that its variance is equal to the Cramer-Rao lower bound A_t^{-1} [20].

Overall, we conclude that for the estimation of λ at some fixed time t under the Brownian model (4.1), it suffices that each sensor i transmits to the fusion center the value of its local MLE at time t, λ_t^i, or equivalently its observed value at time t, ξ_t^i.

4.1.2 Ornstein-Uhlenbeck type processes

We now assume that each ξ_t^i is governed by the following SDE

$$\xi_t^i = \lambda \int_0^t b_s^i ds + W_t^i, \quad t \geq 0, \quad i = 1, \ldots, K \tag{4.8}$$

where each $\{b_t^i\}$ is an $\{\mathcal{F}_t^i\}$-adapted process. Then, the local likelihood at sensor i is

$$\mathscr{L}_t^i(\lambda) \equiv \frac{d\mu_\lambda^i}{d\mu_0^i}\Big|_{\mathcal{F}_t^i} = \exp\{\lambda\, B_t^i - 0.5\, \lambda^2 A_t^i\} \tag{4.9}$$

where

$$A_t^i = \int_0^t |b_s^i|^2 ds \quad , \quad B_t^i = \int_0^t b_s^i \, d\xi_s^i, \quad t \geq 0. \tag{4.10}$$

Again, the global likelihood at time t, $\mathscr{L}_t(\lambda)$, is the product of the local likelihoods and the global MLE takes the form

$$\lambda_t = \frac{B_t}{A_t} = \frac{\sum_{i=1}^K B_t^i}{\sum_{i=1}^K A_t^i}, \quad t \geq 0. \tag{4.11}$$

In order to compute the Fisher information, we start by defining the process

$$M_t = \sum_{i=1}^K M_t^i = \sum_{i=1}^K \int_0^t b_s^i \, dW_s^i, \quad t \geq 0, \tag{4.12}$$

which is a square-integrable martingale with quadratic variation process $\{A_t\}$.

Then, we can obtain the following decompositions for B_t and λ_t in terms of A_t and M_t:

$$B_t = \lambda A_t + M_t \quad , \quad \lambda_t = \lambda + \frac{M_t}{A_t}, \quad t \geq 0 \tag{4.13}$$

and compute the Fisher information at time t as follows:

$$I_t(\lambda) \equiv \mathsf{E}_\lambda\Big[\Big(\frac{\partial}{\partial\lambda}\log \mathscr{L}_t(\lambda)\Big)^2\Big] = \mathsf{E}_\lambda[(B_t - \lambda A_t)^2] = \mathsf{E}_\lambda[M_t^2] = \mathsf{E}_\lambda[A_t], \quad t \geq 0. \tag{4.14}$$

Thus, the process $\{A_t\}$ is the (global) *observed* Fisher information up to time t, whereas $\{A_t^i\}$ is the local *observed* Fisher information at sensor i.

There are many differences in the properties of the MLE in the general case (4.8) in comparison to the Brownian case (4.1). First of all, Fisher information $I_t(\lambda)$ is no longer a linear function of t and is no longer independent of the true value of λ. Moreover, the global MLE remains a weighted average of the local MLEs, but now the weights $\{w_t^i\}$ are no longer constant over time but stochastic processes themselves. More specifically, we have

$$\lambda_t = \sum_{i=1}^K w_t^i \lambda_t^i, \quad \lambda_t^i = \frac{B_t^i}{A_t^i} \ , \ w_t^i = \frac{A_t^i}{A_t} \tag{4.15}$$

Notice that the weights $\{w_t^i\}$ are the normalized observed Fisher informations that correspond to the different sensors. Thus, at any time t the global MLE gives different weights to the local MLEs depending on the information that they have

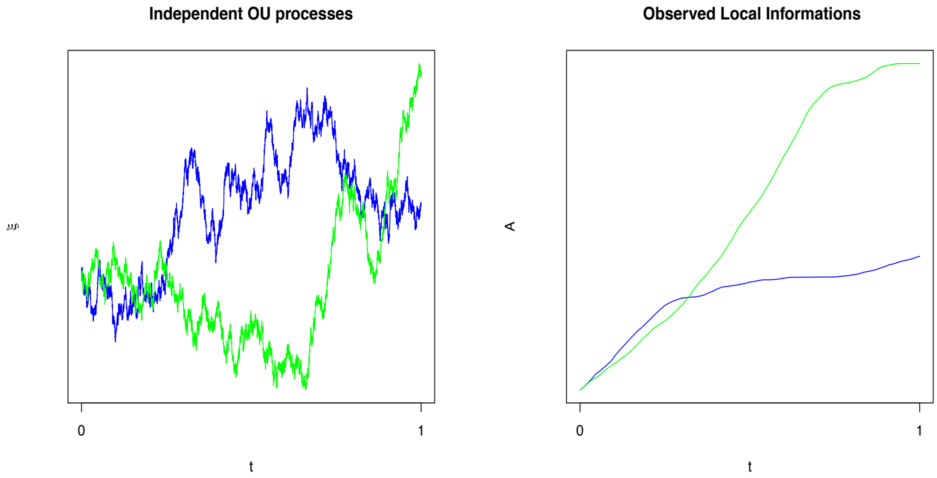


Figure 4.4: Two Ornstein-Uhlenbeck paths and their corresponding observed Fisher information processes

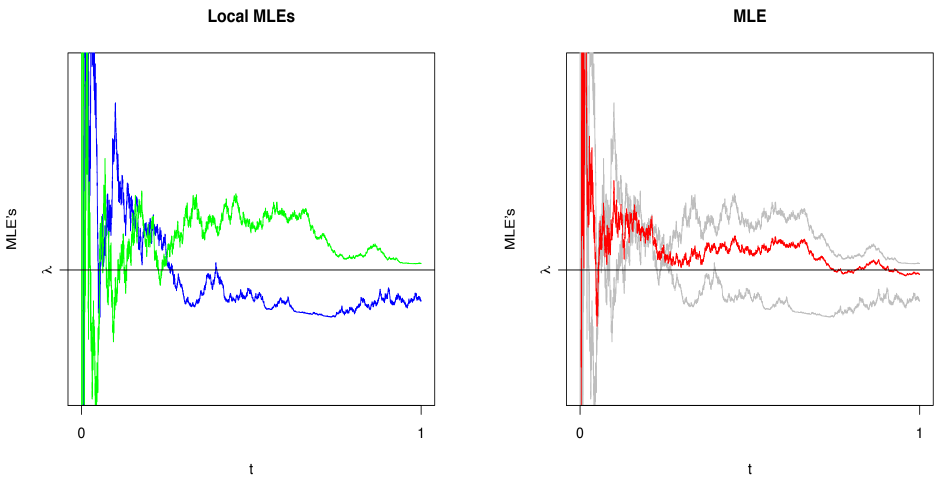


Figure 4.5: The corresponding local MLEs in a continuation of Fig. 4.4

.

accumulated up to time t; the difference with the Brownian case is that these weights are random and time-varying (see Fig. 4.4 and Fig. 4.5 for an illustration of this phenomenon in the case of two independent Ornstein-Uhlenbeck processes with the same dynamics).

Furthermore, unlike the Brownian case, the MLE at time t, λ_t, cannot be computed at the fusion center if each sensor transmits only its final observed value ξ_t^i or the corresponding value of its local MLE λ_t^i. Finally, the MLE is no longer unbiased, Gaussian and optimal in a mean square sense, although it is possible to recover these properties asymptotically (see [25] for an analysis in the Ornstein-Uhlenbeck case and [20] for more general ergodic diffusion processes).

It turns out that a *sequential* version of the MLE can recover all these properties in a non-asymptotic sense. The need for a sequential estimator can arise, if we follow the approach in [25] and we fix – not the horizon of observations– but the (Fisher) *information* that is available for decision.

We can define a centralized sequential estimator as a pair $(\mathcal{T}, \delta_{\mathcal{T}})$, where $\mathcal{T}$ is an $\{\mathcal{F}_t\}$-stopping time and $\delta_{\mathcal{T}}$ an $\mathcal{F}_{\mathcal{T}}$-measurable random variable; $\mathcal{T}$ is the stopping rule at which the decision maker at the fusion center stops collecting observations from the sensors and $\delta_{\mathcal{T}}$ the estimator of λ it uses at time $\mathcal{T}$.

Following [25], we define as optimal centralized sequential estimator the solution to the following constrained optimization problem:

$$\inf_{(\mathcal{T},\delta_{\mathcal{T}})} \mathsf{E}_\lambda[(\delta_{\mathcal{T}} - \lambda)^2] \quad \text{with} \quad I_{\mathcal{T}}(\lambda) = \mathsf{E}_\lambda[A_{\mathcal{T}}] \leq \gamma, \tag{4.16}$$

where γ is a fixed, positive constant. Notice that in the Brownian case (4.1), problem (4.16) reduces to finding the $\mathcal{F}_T$-measurable estimator with the minimum mean square error, where $T = \frac{\gamma}{\sum_{i=1}^K |b_i|^2}$, and the solution is the MLE λ_T, which we defined in (4.4).

Liptser and Shiryayev [25] proved that a solution to (4.16) is given by the sequential version of the MLE:

$$\mathcal{S} = \inf\{t \geq 0 : A_t = \gamma\} \quad , \quad \lambda_{\mathcal{S}} = \left(\frac{B}{A}\right)_{\mathcal{S}} = \frac{B_{\mathcal{S}}}{\gamma} \tag{4.17}$$

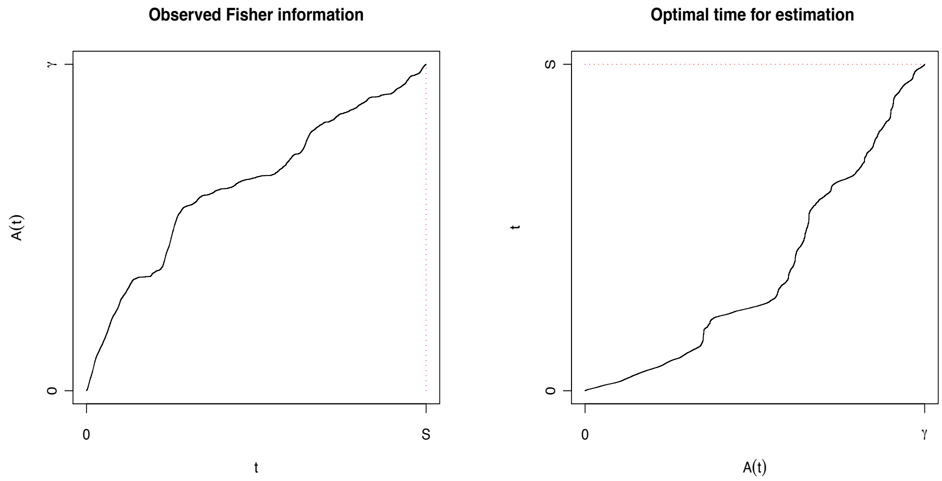


Figure 4.6: Sequential MLE

according to which the fusion center stops collecting observations when the *observed* Fisher information $\{A_t\}$ becomes equal (due to path-continuity) to γ and uses the MLE to estimate λ at this time.

Notice that the stopping rule $\mathcal{S}$ is determined exclusively by the constraint in (4.16), which it satisfies with equality. In particular, the sequential (4.17) uses all the available *information* for the estimation of λ. This is a direct generalization of the strategy that is followed in the fixed-horizon problem, where the estimation takes place at the end of the available observation horizon. We illustrate the stopping time $\mathcal{S}$ in Fig. 4.6.

As in the non-sequential case, the proof of the optimality of the sequential MLE relies on a lower bound on the mean square error of sequential estimators. More specifically, it is shown in [25] that if $(\mathcal{T}, \delta_{\mathcal{T}})$ is an *unbiased* sequential estimator of λ, then $\text{MSE}_\lambda(\delta_{\mathcal{T}}) \geq (\mathsf{E}_\lambda[A_{\mathcal{T}}])^{-1}$. This is the sequential analogue of the Cramer-Rao lower bound and it implies that for all sequential estimators $(\mathcal{T}, \delta_{\mathcal{T}})$ that satisfy the constraint in (4.16) we have $\text{MSE}_\lambda(\delta_{\mathcal{T}}) \geq \gamma^{-1}$.

Thus, for the sequential MLE $(\mathcal{S}, \lambda_{\mathcal{S}})$ to solve problem (4.16), it suffices to show that $\lambda_{\mathcal{S}}$ is an unbiased estimator of λ and that its variance attains the lower bound $\frac{1}{\gamma}$. Indeed, in [25] it is shown that if

$$\mathsf{P}_\lambda(A_\infty = \infty) = 1 \tag{4.18}$$

then

$$\sqrt{\gamma}(\lambda_{\mathcal{S}} - \lambda) \sim \mathcal{N}(0, 1). \tag{4.19}$$

Thus, the estimator $\lambda_{\mathcal{S}}$ is unbiased, Gaussian and has the minimum possible mean square error in the sense of (4.16). The sequential MLE $(\mathcal{S}, \lambda_{\mathcal{S}})$ enjoys these properties in such generality, because it takes into account the *natural clock* $t \to A_t$ which is embedded in the dynamics of (4.8). This is also reflected in the proof of (4.19), which is based on a time-change argument. Overall, this result suggests that it is the *sequential* MLE $(\mathcal{S}, \lambda_{\mathcal{S}})$ that should be used for the estimation of λ in the general setup (4.8).

Finally, we notice that condition (4.18) guarantees that the stopping rule $\mathcal{S}$ is finite P_λ-a.s. We can also think of (4.18) as a regularity condition, since it implies that $It(\lambda)$ is an increasing and *unbounded* function of time t. This condition is satisfied trivially in the Brownian case and in [25] it is shown that (4.18) is also satisfied in the Ornstein-Uhlenbeck case, i.e. when $b_t^i = \xi_t^i$ in (4.8).

4.2 Decentralized parameter estimation

In this section we assume that each sensor can use for its communication with the fusion center only an alphabet of finite length, thus it can only transmit *quantized* versions of its observations. Under this *decentralized* setup, the implementation of the optimal centralized estimators that we previously discussed becomes impossible, thus there is a need for alternative schemes that incorporate the reality of small-length alphabets at the sensors. In this chapter, we develop an efficient decentralized sequential estimator based on the idea that the sensors should communicate whenever they have to transmit an "important" message. This is the same idea that we used in the problems of sequential testing and change-detection, i.e. we suggest that the sensors communicate whenever certain locally observed statistics cross some thresholds.

4.2.1 The Brownian case

We first consider decentralized parameter estimation in the Brownian case, thus we assume that the sensor dynamics are described by (4.1) and the goal is to estimate λ at some fixed time t.

In this setup, the implementation of the centralized MLE, λ_t, requires from sensor i the transmission of its value only at time t, ξ_t^i. However, if each sensor has only a *binary* alphabet $\{b_0, b_1\}$ in its disposal, an accurate transmission of one *Gaussian* observation may require multiple transmissions, depending of course on the desired precision. Therefore, even in this very simple case, we need to take into account the fact that each sensor can use only two letters –b_0 and b_1– in order to communicate its messages.

The main idea in the suggested decentralized scheme is that the times at which the sensors communicate with the fusion center should *not* be fixed in advance. Instead, the communication times should be triggered by the local observations at the sensors. In particular, we suggest that each sensor i communicates with the fusion center at the times

$$\tau_n^i = \inf\{t \geq \tau_{n-1}^i : \xi_t^i - \xi_{\tau_{n-1}^i}^i \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \quad n \in \mathbb{N} \tag{4.20}$$

transmitting the letter b_1 when $z_n^i = 1$ and the letter b_0 when $z_n^i = 0$, where

$$z_n^i = \begin{cases} 1, & \text{if } \xi_{\tau_n^i}^i - \xi_{\tau_{n-1}^i}^i = \overline{\Delta}_i \\ 0, & \text{if } \xi_{\tau_n^i}^i - \xi_{\tau_{n-1}^i}^i = -\overline{\Delta}_i \end{cases} \tag{4.21}$$

and $\overline{\Delta}_i, \underline{\Delta}_i$ are fixed positive constants, known to the fusion center.

Under the communication scheme (4.20), the number of messages transmitted by sensor i up to time t is random and we will denote it by m_t^i, i.e. $m_t^i = \max\{n : \tau_n^i \leq t\}$. Moreover, every time it receives letter b_1 (b_0) from sensor i, the fusion center knows that the process ξ^i has increased by $\overline{\Delta}_i$ (decreased by $\underline{\Delta}_i$) since the last communication from sensor i.

Therefore, the fusion center is able to recover the exact value of ξ^i at the times

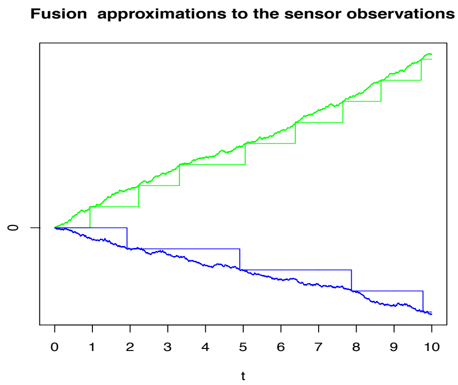


Figure 4.7: Communication times and fusion center approximations in the case of two independent Brownian motions with drifts of different sign and size and equal local thresholds in both sensors.

$\{\tau_n^i\}$, since

$$\xi^i_{\tau^i_n} = \sum_{j=1}^{n} [\overline{\Delta}_i z^i_j - \underline{\Delta}_i (1 - z^i_j)], \quad n \in \mathbb{N}. \tag{4.22}$$

Between communication times the fusion center does not receive any information, thus we suggest that it approximates the value of the process ξ^i at some arbitrary time t with the most recently reproduced value, i.e.

$$\tilde{\xi}^i_t = \xi^i_{\tau^i_n}, \quad \tau^i_n \leq t < \tau^i_{n+1} \tag{4.23}$$

or equivalently

$$\tilde{\xi}^i_t = \sum_{j=1}^{n} [\overline{\Delta}_i z^i_j - \underline{\Delta}_i (1 - z^i_j)] = \xi^i_{\tau^i_{m^i_t}}, \quad \tau^i_n \leq t < \tau^i_{n+1} \tag{4.24}$$

We illustrate these approximations in Fig. 4.7.

Using these approximations, we can mimic the centralized MLE and estimate λ as follows:

$$\tilde{\lambda}_t = \sum_{i=1}^{K} w_i \tilde{\lambda}^i_t , \quad \tilde{\lambda}^i_t = \frac{\tilde{\xi}^i_t}{b_i t}, \tag{4.25}$$

where the weights $\{w_i\}$ are given by (4.5). We call $\{\tilde{\lambda}_t\}$ *decentralized MLE* (D-MLE), since it mimics the centralized MLE and can be implemented with the

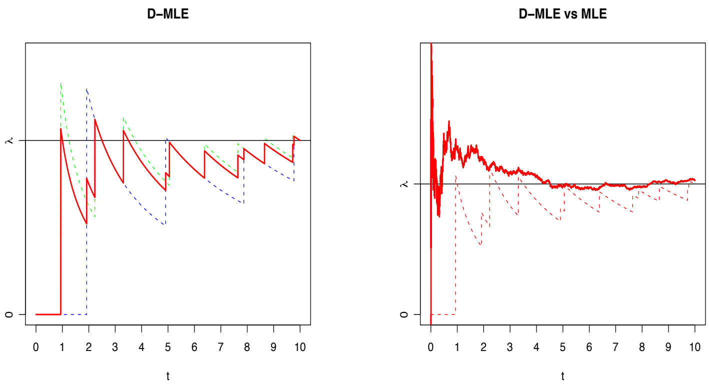


Figure 4.8: The left-hand side graph illustrates that the D-MLE is a weighted average of the local D-MLEs (similarly to the centralized MLE). The right-hand side graph presents the D-MLE against the centralized global MLE.

transmission of one-bit messages from the sensors to the fusion center. We illustrate the D-MLE in comparison to the local D-MLEs and the centralized MLE in Fig. 4.8.

Contrary to the MLE, the D-MLE $\tilde{\lambda}_t$ spreads the communication load throughout $[0, t]$, instead of forcing *all* required messages from all sensors to be transmitted at time t. This property is especially desirable when one is interested in estimating λ repeatedly, that is, not only at t but also at previous times. In this case, the implementation of the centralized MLE clearly requires much heavier communication load, whereas for the D-MLE estimation at these intermediate points comes for free.

The D-MLE is a very flexible decentralized estimator, since the parameters at each sensor, $\overline{\Delta}_i$ and $\underline{\Delta}_i$, are determined by the designer of the scheme. The choice of these parameters is characterized by the following trade-off; small values for $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ lead to frequent communication and good statistical properties for the D-MLE $\tilde{\lambda}_t$, since $\tilde{\lambda}_t$ approaches the centralized MLE λ_T as $\overline{\Delta}_i, \underline{\Delta}_i \to 0$ for any i and t. However, if the communication between sensors and fusion center is expensive, the sensors may not have the luxury to transmit messages to the fusion center very often. This is the reason for the introduction of the decentralized setup in the first place and this implies a practical desire for large values for $\{\overline{\Delta}_i, \underline{\Delta}_i\}$.

Therefore, the D-MLE is a very appealing decentralized estimator, because it recovers the statistical properties of the MLE in an asymptotic sense (as $t \to \infty$) even with *rare* communication between sensors and fusion center ($\overline{\Delta}_i, \underline{\Delta}_i \to \infty$), as the following proposition suggests.

Proposition 7. *1. If $t \to \infty$ and $\overline{\Delta}_i, \underline{\Delta}_i \to \infty$ so that $\overline{\Delta}_i, \underline{\Delta}_i = o(t)$, then $\tilde{\lambda}_t$ is a consistent estimator of λ in probability and in mean square.*

2. If additionally $\overline{\Delta}_i, \underline{\Delta}_i = o(\sqrt{t})$, $\tilde{\lambda}_t$ is asymptotically normal and optimal, i.e.

$$\sqrt{t}(\tilde{\lambda}_t - \lambda) \to \mathcal{N}(0,1)\ ,\ \frac{MSE(\tilde{\lambda}_t)}{MSE(\lambda_t)} \to 1.$$

The proposition remains valid for any fixed values of the thresholds $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ and is a special case of a more general proposition that we state and prove in the next subsection.

4.2.2 Ornstein-Uhlenbeck type processes

We now assume that the sensor dynamics are given by (4.8) and our goal is –as in the Brownian case– to apply level-triggered communication scheme and mimic the optimal centralized estimator. However, in order to do so, we now need a *3-letter* alphabet $\{a, b_0, b_1\}$ at each sensor and each sensor should transmit messages at *two* distinct sequences of communication times.

Indeed, from the optimality of the sequential MLE $(\mathcal{S}, \lambda_{\mathcal{S}})$ with respect to problem (4.16), it follows that the processes $\{A_t^i, B_t^i\}$ –defined in (4.10)– are *sufficient* statistics for the estimation of λ. Thus, if the fusion center is able to reconstruct their values from the received sensor messages, then it can implement the optimal stopping rule $\mathcal{S}$ and compute the corresponding MLE $\lambda_{\mathcal{S}}$. Since sensor i observes the path of $\{\xi_t^i\}$ continuously and $\{b_t^i\}$ is $\{\mathcal{F}_t^i\}$-adapted, the pair $\{A_t^i, B_t^i\}$ is observable at sensor i.

Based on this observation, we suggest that sensor i communicate with the fusion center at the following sequences of stopping times:

$$\begin{aligned} \tau_n^{i,A} &= \inf\{t \geq \tau_{n-1}^{i,A} : A_t^i - A^i_{\tau_{n-1}^{i,A}} \geq c_i\}, \quad n \in \mathbb{N} \\ \tau_n^{i,B} &= \inf\{t \geq \tau_{n-1}^{i,B} : B_t^i - B^i_{\tau_{n-1}^{i,B}} \notin (-\underline{\Delta}_i, \overline{\Delta}_i)\}, \quad n \in \mathbb{N} \end{aligned} \tag{4.26}$$

where $\overline{\Delta}_i, \underline{\Delta}_i, c_i$ are positive constants, fixed in advance and known to the fusion center.

At the times $\{\tau_n^{i,A}\}$ sensor i transmits the letter a, whereas at the times $\{\tau_n^{i,B}\}$ it transmits the letter b_1 when $z_n^{i,B} = 1$ and the letter b_0 when $z_n^{i,B} = 0$, where

$$z_n^i = \begin{cases} 1, & \text{if } B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} = \overline{\Delta}_i \\ 0, & \text{if } B^i_{\tau_n^{i,B}} - B^i_{\tau_{n-1}^{i,B}} = -\overline{\Delta}_i \end{cases} \tag{4.27}$$

Thus, when it receives the letter a from sensor i, the fusion center understands that the process A^i has increased *exactly by* c_i since the last time sensor i transmitted an a. Similarly, when it receives the letter $b_1(b_0)$ from sensor i, the fusion center understands that the process B^i has increased (decreased) *exactly by* $\overline{\Delta}_i$ ($\underline{\Delta}_i$) since the last time sensor i transmitted either b_1 or b_0.

Due to the path-continuity of the processes A^i, B^i, the fusion center can recover the exact values of A^i at the times $\{\tau_n^{i,A}\}$ and B^i at the times $\{\tau_n^{i,B}\}$ using the information (4.26)-(4.27) it receives from the sensors, since

$$\begin{aligned} A^i_{\tau_n^{i,A}} &= \sum_{j=1}^{n} [A^i_{\tau_j^{i,A}} - A^i_{\tau_{j-1}^{i,A}}] = n\, c_i, \quad n \in \mathbb{N} \\ \tilde{B}^i_{\tau_n^{i,B}} &= \sum_{j=1}^{n} [B^i_{\tau_j^{i,B}} - B^i_{\tau_{j-1}^{i,B}}] = \sum_{j=1}^{n} \Big[\overline{\Delta}_i z_j^i - \underline{\Delta}_i (1 - z_j^i)\Big], \quad n \in \mathbb{N} \end{aligned} \tag{4.28}$$

Then, since the fusion center does not receive any information between communication times, we suggest that it approximates A_t^i and B_t^i at some arbitrary time t with the corresponding most recently recovered values, i.e.

$$\begin{aligned} \tilde{A}_t^i &= A^i_{\tau_n^{i,A}}, \quad \tau_n^{i,A} \leq t < \tau_{n+1}^{i,A} \\ \tilde{B}_t^i &= B^i_{\tau_n^{i,B}}, \quad \tau_n^{i,B} \leq t < \tau_{n+1}^{i,B} \end{aligned} \tag{4.29}$$

and we propose the following sequential estimator at the fusion center

$$\tilde{\mathcal{S}} = \inf\{t \geq 0 : \tilde{A}_t \geq \gamma - 2c\} \quad , \quad \tilde{\lambda}_{\tilde{\mathcal{S}}} = \Big(\frac{\tilde{B}}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}} \tag{4.30}$$

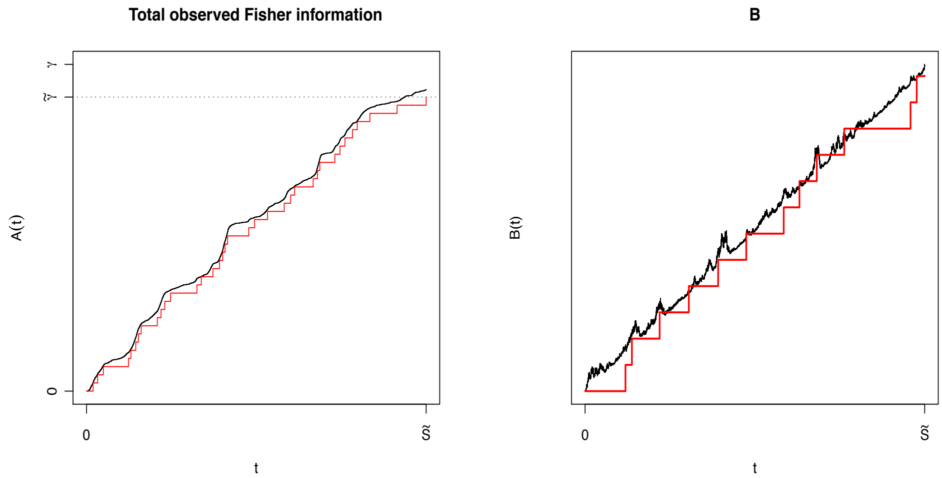


Figure 4.9: The fusion center approximations of the processes A, B

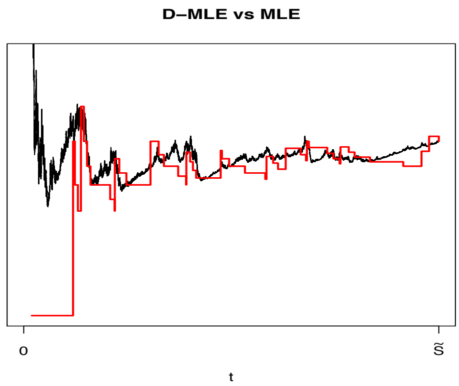


Figure 4.10: D-MLE versus MLE

where

$$\tilde{A} = \sum_{i=1}^{K} \tilde{A}^i \ , \ \tilde{B} = \sum_{i=1}^{K} \tilde{B}^i \ , \ c = \sum_{i=1}^{K} c_i. \tag{4.31}$$

Thus, $(\tilde{\mathcal{S}}, \tilde{\lambda}_{\tilde{\mathcal{S}}})$ mimics the sequential MLE $(\mathcal{S}, d_{\mathcal{S}})$ by replacing A^i, B^i with $\tilde{A}^i, \tilde{B}^i$ and γ by $\gamma - 2c$. We illustrate the global approximations $\tilde{A}$ and $\tilde{B}$ and the D-MLE stopping time $\tilde{\mathcal{S}}$ in Fig. 4.9, whereas we plot the corresponding D-MLE versus the centralized MLE in Fig. 4.10. Moreover, as in the Brownian case, we use the term D-MLE for the suggested decentralized sequential estimator $(\tilde{\mathcal{S}}, \tilde{\lambda}_{\tilde{\mathcal{S}}})$.

The following lemma describes the main properties of the suggested decentralized scheme. Before we state it, we recall (4.31) and introduce the following notation $C \equiv \sum_{i=1}^{K}(\overline{\Delta}_i + \underline{\Delta}_i)$.

Lemma 9. *For any choice of* γ, $\{\overline{\Delta}_i, \underline{\Delta}_i\}$ *and* $\{c_i\}$ *we have:*

1. $|\tilde{A}_t - A_t| \leq c \quad , \quad |\tilde{B}_t - B_t| \leq C, \quad t \geq 0$
2. $\tilde{\mathcal{S}} \leq \mathcal{S}$
3. *The D-MLE satisfies the constraint of problem (4.16), i.e.* $\mathsf{E}[A_{\tilde{\mathcal{S}}}] \leq \gamma$. *In particular,*
$$\gamma - 2c \leq \tilde{A}_{\tilde{\mathcal{S}}} \leq A_{\tilde{\mathcal{S}}} \leq \gamma \tag{4.32}$$

Proof. From the definition of $\tilde{A}^i, \tilde{B}^i$ and the continuity of the paths of A^i, B^i, we have: $\tilde{A}_t^i \leq A_t^i \leq \tilde{A}_t^i + c_i$ and $|\tilde{B}_t^i - B_t^i| \leq \overline{\Delta}_i + \underline{\Delta}_i$ for every $t \geq 0$. The first claim then follows by adding these inequalities over i. The second claim now follows easily:

$$\mathcal{S} = \inf\{t \geq 0 : A_t \geq \gamma\} \geq \inf\{t \geq 0 : \tilde{A}_t + c \geq \gamma\} \geq \tilde{\mathcal{S}}, \tag{4.33}$$

For the third claim let us first observe that $\{A_t\}$ is a piecewise constant, increasing process with jumps bounded by c and the stopping time $\tilde{\mathcal{S}}$ corresponds to a jump time of $\{A_t\}$. Then, it becomes clear that the overshoot $\tilde{A}_{\tilde{\mathcal{S}}} - (\gamma - 2c)$ is upper bounded by c, so that $\tilde{A}_{\tilde{\mathcal{S}}} \leq \gamma - c$. From this observation and the first part of the lemma we obtain (4.32), since

$$\gamma - 2c \leq \tilde{A}_{\tilde{\mathcal{S}}} \leq A_{\tilde{\mathcal{S}}} \leq \tilde{A}_{\tilde{\mathcal{S}}} + c \leq \gamma. \tag{4.34}$$

□

Based on this lemma, we are able to establish the following asymptotic properties of the D-MLE.

Proposition 8. *As* $\gamma, c_i, \overline{\Delta}_i, \underline{\Delta}_i \to \infty$

1. $\tilde{\lambda}_{\tilde{\mathcal{S}}}$ *converges to* λ *in probability and in mean square, if* $\overline{\Delta}_i, \underline{\Delta}_i, c_i = o(\gamma)$.

2. $\tilde{\lambda}_{\tilde{\mathcal{S}}}$ *is asymptotically normal and optimal, i.e.*

$$\sqrt{\gamma}(\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda) \to \mathcal{N}(0,1)\ ,\ MSE(\tilde{\lambda}_{\tilde{\mathcal{S}}})/MSE(\lambda_{\mathcal{S}}) \to 1,$$

if $\overline{\Delta}_i, \underline{\Delta}_i, c_i = o(\sqrt{\gamma})$.

Thus, $\tilde{\lambda}_{\tilde{\mathcal{S}}}$ recovers the properties of its centralized counterpart $\lambda_{\mathcal{S}}$ for large values of γ as long as the design parameters $c_i, \overline{\Delta}_i, \underline{\Delta}_i$ are "large" but smaller than γ, ideally around the square root of γ and smaller.

Proof. 1. *Consistency.*

We start with the following representation of the D-MLE

$$\tilde{\lambda}_{\tilde{\mathcal{S}}} = \Big(\frac{\tilde{B}-B}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}} + \Big(\frac{A}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\lambda_{\tilde{\mathcal{S}}} = \Big(\frac{\tilde{B}-B}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}} + \Big(\frac{A}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\Big[\lambda+\Big(\frac{M}{A}\Big)_{\tilde{\mathcal{S}}}\Big], \quad (4.35)$$

where the first equality can be derived with simple algebra and the second using (4.13). Then we have:

$$\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda = \Big(\frac{\tilde{B}-B}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}} + \Big(\frac{A-\tilde{A}}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\lambda + \Big(\frac{A}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\Big(\frac{M}{A}\Big)_{\tilde{\mathcal{S}}} \quad (4.36)$$

which –using the lemma and the triangle inequality– gives:

$$|\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda| \leq \frac{C}{\gamma-2c} + \frac{c}{\gamma-2c}|\lambda| + \frac{\gamma-c}{\gamma-2c}\,\frac{|M_{\tilde{\mathcal{S}}}|}{\gamma-2c} \quad (4.37)$$

Using the inequality $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and taking expectations, we obtain:

$$\frac{1}{3}\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda)^2] \leq \Big(\frac{C}{\gamma-2c}\Big)^2 + \Big(\frac{c}{\gamma-2c}\Big)^2\lambda^2 + \Big(\frac{\gamma-c}{\gamma-2c}\Big)^2\,\frac{\mathsf{E}_\lambda[M_{\tilde{\mathcal{S}}}]^2}{(\gamma-2c)^2} \quad (4.38)$$

But from the Cauchy-Schwartz inequality and Itô's isometry for the square integrable martingale $\{M_t\}$ we have:

$$\mathsf{E}_\lambda[|M_{\tilde{\mathcal{S}}}|] \leq \sqrt{\mathsf{E}_\lambda[M_{\tilde{\mathcal{S}}}^2]} = \sqrt{\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}}]} \leq \sqrt{\gamma}, \quad (4.39)$$

therefore:

$$\frac{1}{3}\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda)^2] \leq \Big(\frac{C}{\gamma-2c}\Big)^2 + \Big(\frac{c}{\gamma-2c}\Big)^2\lambda^2 + \Big(\frac{\gamma-c}{\gamma-2c}\Big)^2\,\frac{\gamma}{(\gamma-2c)^2}. \quad (4.40)$$

From this expression it is clear that $\mathsf{E}_\lambda[(\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda)^2] \to 0$ even if $\overline{\Delta}_i, \underline{\Delta}_i, c_i \to \infty$ as long as $\overline{\Delta}_i, \underline{\Delta}_i, c_i = o(\gamma)$.

2. *Asympotic Normality*

We start with the following representation, which is easily obtained with some simple algebraic manipulations:

$$\begin{aligned}\sqrt{\gamma}\,(\tilde{\lambda}_{\tilde{\mathcal{S}}}-\lambda) &= \sqrt{\gamma}\,\Big(\frac{\tilde{B}-B}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}+\sqrt{\gamma}\,\Big(\frac{A}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}(\lambda_{\tilde{\mathcal{S}}}-\lambda_{\mathcal{S}})\\ &\quad+\sqrt{\gamma}\,\Big(\frac{A-\tilde{A}}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\lambda_{\mathcal{S}}+\sqrt{\gamma}\,(\lambda_{\mathcal{S}}-\lambda).\end{aligned} \tag{4.41}$$

Thus, if we show that the first three terms vanish as $\gamma \to \infty$, the asymptotic normality of $\tilde{\lambda}_{\tilde{\mathcal{S}}}$ will follow from Slutsky's theorem and the exact normality of $\lambda_{\mathcal{S}}$, i.e. $\sqrt{\gamma}\,(\lambda_{\mathcal{S}}-\lambda) \sim \mathcal{N}(0,1)$.

Using the lemma we can see that

$$\sqrt{\gamma}\,\Big(\frac{|\tilde{B}-B|}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\leq \frac{C\sqrt{\gamma}}{\gamma-2c} \quad , \quad \sqrt{\gamma}\,\Big(\frac{A-\tilde{A}}{\tilde{A}}\Big)_{\tilde{\mathcal{S}}}\leq \frac{c\sqrt{\gamma}}{\gamma-2c}, \tag{4.42}$$

thus the first and the third term in (4.41) will vanish as $\gamma, \overline{\Delta}_i, \underline{\Delta}_i, c_i \to \infty$ as long as $\overline{\Delta}_i, \underline{\Delta}_i, c_i = o(\sqrt{\gamma})$ (it can be easily shown with a time-change argument that the centralized MLE $\lambda_{\mathcal{S}}$ is a consistent estimator of λ as $\gamma \to \infty$).

It remains to show that the second term in (4.41) converges to 0 in probability or –a fortiori– in mean square. Since $A_{\tilde{\mathcal{S}}}/\tilde{A}_{\tilde{\mathcal{S}}} \to 1$ as $c_i = o(\gamma)$, it suffices to show that $\mathsf{E}_\lambda[\gamma(\lambda_{\tilde{\mathcal{S}}}-\lambda_{\mathcal{S}})^2] \to 0$. Indeed, using the representation (4.13), Itô's isometry and the lemma, we obtain:

$$\begin{aligned}\gamma\,\mathsf{E}_\lambda\Big[\Big(\lambda_{\tilde{\mathcal{S}}}-\lambda_{\mathcal{S}}\Big)^2\Big] &= \gamma\,\mathsf{E}_\lambda\Big[\Big(\frac{M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}-\frac{M_{\mathcal{S}}}{A_{\mathcal{S}}}\Big)^2\Big]\\ &= \gamma\Big\{\mathsf{E}_\lambda\Big[\Big(\frac{M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big)^2\Big]+\mathsf{E}_\lambda\Big[\Big(\frac{M_{\mathcal{S}}}{A_{\mathcal{S}}}\Big)^2\Big]-2\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}}{A_{\mathcal{S}}}\,\frac{M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\Big\}\\ &\leq \gamma\Big\{\frac{\mathsf{E}_\lambda[M_{\tilde{\mathcal{S}}}^2]}{(\gamma-2c)^2}+\frac{\mathsf{E}_\lambda[M_{\mathcal{S}}^2]}{\gamma^2}-2\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}\,M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}\gamma}\Big]\Big\}\\ &= \frac{\gamma\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}}]}{(\gamma-2c)^2}+\frac{\mathsf{E}_\lambda[A_{\mathcal{S}}]}{\gamma}-2\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}\,M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\\ &\leq \frac{\gamma^2}{(\gamma-2c)^2}+1-2\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}\,M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\\ &= \frac{\gamma^2}{(\gamma-2c)^2}+1-2\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}\,M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\end{aligned} \tag{4.43}$$

Thus, it it suffices to show that $\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}}\, M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \to 1$. More specifically, since

$$\mathsf{E}_\lambda\Big[\frac{M_{\mathcal{S}} M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] = \mathsf{E}_\lambda\Big[\frac{(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}) M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] + \mathsf{E}_\lambda\Big[\frac{M^2_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}] \tag{4.44}$$

we need to prove that $\mathsf{E}_\lambda\Big[\frac{M^2_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \to 1$ and $\mathsf{E}_\lambda\Big[\frac{(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}) M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \to 0$.

Indeed, using the lemma and (4.39) we have:

$$\frac{\gamma - 2c}{\gamma} \le \frac{\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}}]}{\gamma} \le \mathsf{E}_\lambda\Big[\frac{M^2_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \le \frac{\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}}]}{\gamma - 2c} \le \frac{\gamma}{\gamma - 2c} \tag{4.45}$$

thus $\mathsf{E}_\lambda\Big[\frac{M^2_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \to 1$ as long as $c_i = o(\gamma)$.

Finally, using Cauchy-Schwartz inequality, Itô's isometry and the lemma, we have:

$$\begin{aligned}\Big|\mathsf{E}_\lambda\Big[\frac{(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}) M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\Big| &\le \frac{1}{\gamma - 2c}\mathsf{E}_\lambda[|M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}|\;|M_{\tilde{\mathcal{S}}}|] \\ &\le \frac{1}{\gamma - 2c}\sqrt{\mathsf{E}_\lambda[(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}})^2]\;\mathsf{E}_\lambda[M^2_{\tilde{\mathcal{S}}}]} \\ &= \frac{1}{\gamma - 2c}\sqrt{\mathsf{E}_\lambda[A_{\mathcal{S}} - A_{\tilde{\mathcal{S}}}]\;\mathsf{E}_\lambda[A_{\tilde{\mathcal{S}}}]}\end{aligned} \tag{4.46}$$

But since $A_{\mathcal{S}} = \gamma$ and $A_{\tilde{\mathcal{S}}} \ge \gamma - 2c$ we have $A_{\mathcal{S}} - A_{\tilde{\mathcal{S}}} \le 2c$, which implies

$$\Big|\mathsf{E}_\lambda\Big[\frac{(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}) M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big]\Big| \le \frac{\sqrt{2c\gamma}}{\gamma - 2c} \tag{4.47}$$

Thus, $\mathsf{E}_\lambda\Big[\frac{(M_{\mathcal{S}} - M_{\tilde{\mathcal{S}}}) M_{\tilde{\mathcal{S}}}}{A_{\tilde{\mathcal{S}}}}\Big] \to 0$ as long as $c_i = o(\gamma)$, which finishes the proof.

3. *Asymptotic optimality*

We start by setting – for notational convenience– $f(c) = \frac{\gamma - c}{\gamma - 2c}$ in (4.37), which gives:

$$|\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda| \le \frac{C}{\gamma - 2c} + \frac{|\lambda|\,c}{\gamma - 2c} + \frac{f(c)\,|M_{\tilde{\mathcal{S}}}|}{\gamma - 2c} \tag{4.48}$$

Taking squares in both sides we have:

$$|\tilde{\lambda}_{\tilde{\mathcal{S}}} - \lambda|^2 \le \frac{C^2 + c^2\lambda^2 + f^2(c)\;M^2_{\tilde{\mathcal{S}}} + 2\,C\,c\,|\lambda| + 2|\lambda|\,f(c)\,c\,|M_{\tilde{\mathcal{S}}}| + 2f(c)\,C\,|M_{\tilde{\mathcal{S}}}|}{(\gamma - 2c)^2} \tag{4.49}$$

Taking expectations and using (4.39) we obtain:

$$\mathrm{MSE}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}}) \le \frac{C^2 + c^2\lambda^2 + f^2(c)\,\gamma + 2\,C\,c\,|\lambda| + 2|\lambda|\,f(c)\,c\,\sqrt{\gamma} + 2f(c)\,C\,\sqrt{\gamma}}{(\gamma - 2c)^2} \tag{4.50}$$

Since $\mathrm{MSE}_\lambda(\lambda_{\mathcal{S}}) = \gamma^{-1}$, we have:

$$\begin{aligned}\frac{\mathrm{MSE}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}})}{\mathrm{MSE}_\lambda(\lambda_{\mathcal{S}})} \leq \frac{C^2\gamma}{(\gamma-2c)^2} + \lambda^2\frac{c^2\gamma}{(\gamma-2c)^2} + f^2(c)\,\frac{\gamma^2}{(\gamma-2c)^2} \\ + 2\,|\lambda|\frac{C\,c\,\gamma}{(\gamma-2c)^2} + 2|\lambda|\,f(c)\frac{c\,\gamma^{3/2}}{(\gamma-2c)^2} + 2f(c)\,\frac{C\,\gamma^{3/2}}{(\gamma-2c)^2}\end{aligned} \tag{4.51}$$

At this point it is clear that if we let $c \to \infty$ and $\gamma \to \infty$ so that $c = o(\gamma)$, then $f(c) \to 1$ and the previous inequality becomes:

$$\frac{\mathrm{MSE}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}})}{\mathrm{MSE}_\lambda(\lambda_{\mathcal{S}})} \leq \mathcal{O}\Big(\frac{C^2}{\gamma}\Big) + \mathcal{O}\Big(\frac{c^2}{\gamma}\Big) + 1 + \mathcal{O}\Big(\frac{C\,c}{\gamma}\Big) + \mathcal{O}\Big(\frac{c}{\sqrt{\gamma}}\Big) + \mathcal{O}\Big(\frac{C}{\sqrt{\gamma}}\Big) \tag{4.52}$$

Therefore, we obtain $\mathrm{MSE}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}})/\mathrm{MSE}_\lambda(\lambda_{\mathcal{S}}) \leq 1$ as $c = o(\sqrt{\gamma})$ and $C = o(\sqrt{\gamma})$. But this finishes the proof since the exact optimality of $\lambda_{\mathcal{S}}$ implies $\mathrm{MSE}_\lambda(\tilde{\lambda}_{\tilde{\mathcal{S}}}) \geq \mathrm{MSE}_\lambda(\lambda_{\mathcal{S}})$.

□

4.3 Correlated sensors

In this section we propose a modification of the D-MLE in the general case where the sensors are correlated.

4.3.1 The Brownian case

Suppose that the sensors observe drifted correlated Brownian motions, i.e.

$$\xi_t^i = \lambda b_i t + \sum_{j=1}^{K} \sigma_{ij} W_t^j, \quad t \geq 0, \quad i = 1, \ldots, K. \tag{4.53}$$

We set $\theta_i = \sum_{j=1}^{K} \alpha_{ij} b_j$, where α_{ij} is the (i,j)-element of the matrix $(\sigma^{-1})'\sigma^{-1}$ and $\sigma = [\sigma_{ij}]$. Then, the likelihood function of λ at time t has the form:

$$\mathscr{L}_t(\lambda) = \exp\Big\{\lambda \sum_{i=1}^{K} \theta_i\,\xi_t^i - 0.5\lambda^2 \sum_{i=1}^{K} \theta_i\,b_i\,t\Big\}, \tag{4.54}$$

and the centralized MLE of λ is:

$$\lambda_t = \sum_{i=1}^{K} r_i \frac{\xi_t^i}{t} \quad , \quad r_i = \frac{\theta_i}{\sum_{i=1}^{K} \theta_i\,b_i}, \tag{4.55}$$

Thus, $\xi_t^1, \ldots, \xi_t^K$ remain sufficient statistics for the computation of the MLE at time t, λ_t, even if the underlying Brownian motions at the sensors are correlated; consequently, the sensors can implement the communication scheme (4.20)-(4.21) without any modification and the fusion center can approximate $\{\xi_t^i\}$ with the process $\{\tilde{\xi}_t^i\}$ which was defined in (4.23). The only difference is that the overall estimator of λ at the fusion center now mimics (4.55) instead of (4.25), thus the D-MLE becomes:

$$\tilde{\lambda}_t = \sum_{i=1}^{K} r_i \frac{\tilde{\xi}_t^i}{t}, \quad t \geq 0 \tag{4.56}$$

Notice that in the case of independent sensors, θ_i reduces to b_i, r_i to w_i – defined in (4.4)– and the estimator (4.55) to (4.25).

4.3.2 Correlated Itô processes

We now assume that the evolution of $\{\xi_t^i\}$ is governed by the following stochastic differential equation

$$\xi_t^i = \lambda \int_0^t b_s^i \, ds + \sum_{j=1}^{K} \int_0^t \sigma_s^{ij} \, dW_s^j, \quad t \geq 0, \quad i = 1, \ldots, K \tag{4.57}$$

where

$$b_t^i = f_i(\xi_t^1, \ldots, \xi_t^K) \quad , \quad \sigma_t^{ij} = g_{ij}(\xi_t^1, \ldots, \xi_t^K), \quad t \geq 0 \tag{4.58}$$

with $f_i : \mathbb{R}^K \to \mathbb{R}$, $g_{ij} : \mathbb{R}^K \to \mathbb{R}$ being known Borel functions for each i, j (so that (4.57) has a unique strong solution).

Then, the likelihood function and the MLE of λ at time t have the following form:

$$\mathscr{L}_t(\lambda) = \exp\{\lambda B_t - 0.5\lambda^2 A_t\} \quad , \quad \lambda_t = \frac{B_t}{A_t} = \frac{\sum_{i=1}^K B_t^i}{\sum_{i=1}^K A_t^i} \tag{4.59}$$

where

$$B_t^i = \int_0^t \theta_s^i \, d\xi_s^i \quad , \quad A_t^i = \int_0^t \theta_s^i \, b_s^i \, ds \quad , \quad \theta_t^i = \sum_{j=1}^{K} \alpha_t^{ij} \, b_t^j, \quad i = 1, \ldots, K \tag{4.60}$$

and by α_t^{ij} we denote the (i, j)-element of the matrix $(\sigma_t^{-1})'\sigma_t^{-1}$, where $\sigma_t = [\sigma_t^{ij}]$.

The main difficulty in this framework stems from the fact that sensor i cannot compute the processes $\{A_t^i\}$,$\{B_t^i\}$ using only its local observations; it can at best approximate them and in order to do so efficiently it needs information about the observed processes at the other sensors.

For that reason we suggest that each sensor i communicate with all other sensors at the times:

$$\sigma_n^i = \inf\{t \geq \sigma_{n-1}^i : \xi_t^i - \xi_{\sigma_{n-1}^i}^i \notin (-\underline{\Gamma}_i, \overline{\Gamma}_i)\}, \quad n \in \mathbb{N} \tag{4.61}$$

transmitting the messages

$$w_n^i = \begin{cases} 1, & \text{if } \xi_{\sigma_n^i}^i - \xi_{\sigma_{n-1}^i}^i = \overline{\Gamma}_i \\ 0, & \text{if } \xi_{\sigma_n^i}^i - \xi_{\sigma_{n-1}^i}^i = -\underline{\Gamma}_i \end{cases} \tag{4.62}$$

Using $\{\sigma_n^i, w_n^i\}_{n\in\mathbb{N}}$, all other sensors can approximate the observed process at sensor i as follows:

$$\hat{\xi}_t^i = \sum_{l=1}^{n} [\overline{\Gamma}_i w_l^i - \underline{\Gamma}_i(1 - w_l^i)], \quad \sigma_n^i \leq t < \sigma_{n+1}^i \tag{4.63}$$

This *between-sensor* communication scheme allows each sensor i to implement the

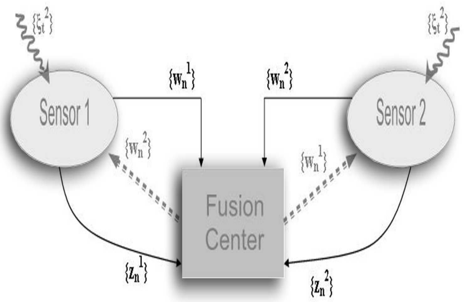


Figure 4.11: Feedback from the fusion center

communication scheme (4.26)-(4.27) and compute the D-MLE (4.30) replacing A_t^i, B_t^i with

$$\hat{A}_t^i = \int_0^t \hat{\theta}_s^i \, \hat{b}_s^i \, ds \quad , \quad \hat{B}_t^i = \int_0^t \hat{\theta}_s^i \, d\xi_s^i, \quad t \geq 0. \tag{4.64}$$

where

$$\hat{b}_t^i = f_i(\hat{\xi}_t^1, \ldots, \hat{\xi}_t^{i-1}, \xi_t^i, \hat{\xi}_t^{i+1} \ldots, \hat{\xi}_t^K) \quad , \quad \hat{\theta}_t^i = g_i(\hat{\xi}_t^1, \ldots, \hat{\xi}_t^{i-1}, \xi_t^i, \hat{\xi}_t^{i+1} \ldots, \hat{\xi}_t^K) \tag{4.65}$$

However, the resulting D-MLE does not estimate the corresponding continuous-time centralized MLE λ_t , but its approximation

$$\hat{\lambda}_t = \frac{\sum_{i=1}^{K} \hat{B}_t^i}{\sum_{i=1}^{K} \hat{A}_t^i}. \tag{4.66}$$

Thus, it is clear that the between-sensor communication should be very frequent – or equivalently the values of $\overline{\Gamma}_i, \underline{\Gamma}_i$ should be as small as possible– for the resulting decentralized estimator to be reliable. However, this can increase significantly the communication load in the sensor network. Therefore, choosing the thresholds $\{\overline{\Gamma}_i, \underline{\Gamma}_i\}$ optimally is not a trivial task and requires the introduction of criteria that penalize properly excessive communication between sensors.

Finally, we should note that if the sensors can communicate directly with each other, they can implement the suggested communication scheme using the 3-letter alphabet $\{a, b_0, b_1\}$. Indeed, sensor i will transmit to all other sensors at time σ_n^i the letter $b_1(b_0)$ if $w_n^i = 1$ $(w_n^i = 0)$. On the other hand, if the sensors communicate only with the fusion center, then the messages $\{w_n^i\}$ will be in the form of feedback from the fusion center, in which case each sensor will need two additional letters (since the letters b_0, b_1 will be reserved for the transmission of the messages $\{z_n^i\}$.)

Chapter 5

Conclusions

We consider three statistical problems – hypothesis testing, change detection and parameter estimation– under a sequential, decentralized setup. Thus, the relevant information is acquired sequentially by remote sensors, these transmit *quantized* versions of their observations to a central processor (fusion center) and the latter is responsible for making the final decision. The problem is to choose optimally a quantization rule at the sensors and a fusion center policy that will rely only on the transmitted quantized messages.

We suggest that the sensors transmit messages at stopping times of their observed filtrations and we propose fusion center policies that mimic the corresponding optimal centralized schemes.

In decentralized sequential testing and change-detection, we prove that when the sensors observe independent Itô processes or correlated Brownian motions, the resulting decentralized schemes inflict bounded performance loss for any fixed communication rate, thus they are order-2 asymptotically optimal. When the sensors take discrete-time, independent and identically distributed observations, we prove that the resulting decentralized schemes are order-1 asymptotically optimal. In decentralized parameter estimation, we prove that when the sensors observe independent Itô processes whose drift is observable up to an unknown, common parameter, then the resulting estimator is consistent, asymptotically normal and efficient (in a mean-square-error sense) even with rare communication.

All the above decentralized schemes that we suggest induce asynchronous communication between sensors and fusion center. This complicates considerably the their analysis and requires the introduction of some new tools, such as the asynchronous Wald's identities that we prove in Sec. 2.4.4.2. However, all these schemes can be implemented easily, they demand limited local memory and do not require any communication between sensors.

The decentralized schemes that we suggest rely on the existence of sufficient statistics which are observable locally at the sensors. We can always find such statistics when we assume independence across sensors, but also when the sensors observe correlated Brownian motions. However, apart from this special case, when we remove the assumption of independence across sensors, it is no longer possible to apply the same techniques.

In Sec.4.3.2, we describe a decentralized scheme that uses communication between sensors – or equivalently feedback from the fusion center– when the sensors observe correlated diffusions (we can obtain analogous schemes for the testing or the detection problem). Unlike the case of independent sensors, we no longer have bounded performance loss and a detailed, rigorous analysis becomes much more challenging. Therefore, it remains an open problem to find easily implementable, asymptotically optimal and efficient decentralized schemes when the sensor processes are correlated.

The ideas in this thesis could be applied with small modifications to more complicated statistical problems, such as multiple hypothesis testing, change-detection where the distribution before and after the change is not fully specified, estimation of many parameters. Another direction of research would be to assume that the communication between sensors and fusion center is noisy, i.e. the fusion center may sometimes receive wrong messages. It would be interesting to propose appropriate modifications of the suggested decentralized schemes and examine the additional inflicted performance loss in this case.

Finally, whereas it is an advantage that the suggested schemes can be typically implemented with a small alphabet, it would be useful to consider more general schemes that can exploit larger alphabets, reducing more in this way the corresponding performance loss.

Bibliography

[1] M.M. Al-Ibrahim and P.K. Varshney,"A simple multi-sensor sequential detection procedure," *in Proc. 27th IEEE Conf. Decision Contr, Austin, TX*, pp. 2479-2483, 1988.

[2] K.J. Astrom and B. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," *in Proc. 41st IEEE Conf. on Decision and Control, Las Vegas NV,2002*, IEEE Control Systems Society, 2002, pp. 2011-2016

[3] I.V. Basawa and B.L.S. Prakasa Rao, *Statistical inference for stochastic processes,* Academic Press Ins. (London), 1980.

[4] Basseville, M. and Nikiforov, I.V.,*Detection of abrupt changes: Theory and Applications*, Engelwood Cliffs, NJ: Prentice-Hall, 1993.

[5] A. Dvoretzky, J. Kiefer and J. Wolfowitz, " Sequential decision problems for processes with continuous time parameter. Testing hypotheses." *Ann. Math Statist.*, vol. 24, pp. 254-264, 1953.

[6] G. Fellouris and G.V. Moustakides, "Asymptotically optimum tests for decentralized sequential testing in continuous time", *Proceedings 11th International Conference on Information Fusion, Fusion'2008, Cologne, Germany*, 2008.

[7] G. Fellouris and G.V. Moustakides, "Asymptotically optimum tests for decentralized change detection",*Proceedings 2008 International Workshop on Applied Probability, IWAP'2008, Compiegne, France*, 2008.

[8] G. Fellouris and G.V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication", submitted to the IEEE Transactions on Information Theory. Sept. 2009.

[9] G. Fellouris and G.V. Moustakides, "Decentralized sequential hypothesis testing in discrete time," *Proceedings of the 2nd International Workshop on Sequential Methodologies, IWSM'2009, Troyes, France*, 2009.

[10] N. Ghasemi, N. and S. Dey, "A constrained mDP approach to dynamic design for HMM state estimation," *IEEE Trasnsactions on Signal Processing*, vol. 57, no. 3, 1203-1209, 2009.

[11] B.K. Ghosh and P.K. Sen, *Handbook of sequential analysis.* Marcel Dekker, New York, 1991.

[12] H.R Hashemi and I.B. Rhodes,"Decentralized Sequential Detection," *IEEE Trans. Inf. Th.*, vol. 35, pp. 509-520, 1989.

[13] Hawkins, D. M. and Olwell, D.H., *Cumulative Sum Control Charts and Charting for Quality Improvement*, New York: Springer-Verlag, 1998.

[14] Huang, M. and Dey, S., "Dynamic Quantization for Multisensor Estimation Over Bandlimited Fading Channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp.4696-4702, 2006.

[15] Huang, M. and Dey, S. , "Dynamic Quantizer Design for Hidden Markov State Estimation Via Multiple Sensors With Fusion Center Feedback," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp.2887-2896, 2006.

[16] A.M. Hussain, "Multisensor distributed sequential detection," *IEEE Trans. Aer. Elect. Syst.*, vol. 30, no. 3, pp. 698-708, 1994.

[17] A. Irle, "Extended optimality of sequential ratio tests," *The Annals of Statistics*, vol. 12, no. 1, pp. 380-386, 1984.

[18] A. Irle and N. Schmitz, " On the Optimality of the SPRT for processes with continuous time parameter," *Math. Operationsforsch. Statist.*, 1981.

[19] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer, New York, 1991.

[20] Y.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Proesses,* Springer, 2002.

[21] T.L. Lai, "Likelihood ratio identities and their applications to sequential analysis," *Sequential Analysis*, vol. 23, no. 4, pp. 467-497, 2004.

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

[23] T.L. Lai, Sequential change-point detection in quality control and dynamical systems, J. Roy. Statist. Soc. Ser. B, vol. 57, pp. 613658, 1995.

[24] T.L. Lai, Information bounds and quick detection of parameter changes in stochastic systems, IEEE Trans. Inform. Theory, vol. 44, pp.29172929, Nov. 1998.

[25] R.L Liptser and A.N. Shiryaev, *Statistics of Random Processes, I General Theory*, 1st edition, Springer, New York, 1977.

[26] R.L Liptser and A.N. Shiryaev, *Statistics of Random Processes, II Applications*, 2nd edition, Springer, New York, 2001.

[27] G. Lorden, "On excess over the boundary," *Ann. Math. Stat.*, vol. 41, no. 2, pp.520-527, 1970.

[28] G. Lorden, "Procedures for reacting to a change in distribution,". *Ann. Math. Stat.*, vol 42, pp.1897-1908, 1971.

[29] Luo, A.Q., An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network, *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp.735-744, 2005.

[30] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, pp. 206227, 1985.

[31] Y.Mei, "Asymptotic optimality theory for sequential hypothesis testing in sensor networks," *IEEE Trans. Inf. Th.*, vol. 54, pp. 2072-2089 , 2008.

[32] Y. Mei, "Information bounds and quickest change detection in decentralized decision systems, " *IEEE Tr. Inf. Th.*, vol. 51, no. 7, pp.2669-2681, 2005.

[33] G.V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Annals of Statistics*, vol. 14, no. 4, pp. 1379-1387, 1986.

[34] G.V. Moustakides, "Optimality of the CUSUM procedure in continuous time," *Annals of Statistics*, vol 32, no. 1, pp. 302-315, 2004.

[35] G.V. Moustakides, " Decentralized CUSUM change Detection," *Proc. 9th Int. Conf. Inf. Fusion*, Fusion 2006, Florence, Italy.

[36] G.V. Moustakides, "Sequential Change Detection Revisited," *The Annals of Statistics*, vol. 36, no. 2, pp. 787-807, 2008.

[37] G.V. Moustakides, G.V., "Quickest Detection of Abrupt Changes for a Class of Random Processes," *IEEE Transactions on Information Theory*, vol. 44, no. 5, 1965-1968, 1998.

[38] G.V. Moustakides, "Sequential rate-change detection in homogeneous Poisson processes", to be submitted to the Annals of Statistics.

[39] E.S. Page, " Continuous inspection schemes," *Biometrika*, vol 41. pp.100-115, 1954.

[40] G. Peskir and A.N. Shiryaev, *Optimal Stopping and Free- Boundary Problems*, Birkhäuser Lectures in Mathematics, ETH Zürich, 1996.

[41] G. Peskir and A.N. Shiryaev, " Sequential Testing Problems for Poisson Processes", Ann. Statist., vol. 28, no. 3, pp. 837-859, 2000.

[42] H.V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press UK, 2009.

[43] B.L.S. Prakasa Rao, *Statistical Inference for Diffusion Type Processes,* London: Arnold, 1985.

[44] M. Rabi, "Packet based inference and control," *Ph.D. thesis*, University of Maryland, College Park, 2006

[45] M. Rabi, G.V. Moustakides and J.S. Baras, "Adaptive sampling for linear state estimation," *Submitted to the SIAM journal on control and optimization*, 2009

[46] A. Ribeiro and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part II: unknown probability density function," *IEEE Transactions on Signal Processing*, vol. 54, no. 7, pp.2784-2796, 2006.

[47] Y. Ritov, "Decision Theoretic Optimality of the CUSUM Procedure," *The Annals of Statisics*, **18**, 1464-1469, 1990.

[48] S.M. Ross, *Applied Probability Models with Optimization Applications.* Dover Publications, Inc., New York, 1969.

[49] V.N.S. Samarasooriya and P.K. Varshney, "Sequential approach to asynchronous decision fusion," *Opt. Eng.*, vol. 35, no. 3, pp. 625-633, 1996.

[50] W.A. Shewhart, *Economic Control of Quality of Manufactured Product*, New York: Van Nostrand, 1931.

[51] A.N. Shiryaev, *Optimal Stopping Rules*, Springer, New York, 1978.

[52] A.N. Shiryayev, "Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time," *Russ. Math. Surv.*, vol 51, pp. 750-751 , 1996.

[53] D. Siegmund, *Sequential analysis, tests and confidence intervals.* Springer, New York, 1985.

[54] Z.G. Stoumbos, M.R. Reynolds, T.P. Ryan and W.H. Woodall , "The State of Statistical Process Control as We Proceed into the 21st Century," *Journal of the American Statistical Association*, vol. 95, no. 451, pp.992-998, 2000.

[55] A.G. Tartakovsky and H. Kim, "Performance of Certain Decentralized Distributed Change Detection Procedures," *Proc. 9th Int. Conf. Inf. Fusion*, Fusion 2006, Florence, Italy.

[56] A.G. Tartakovsky and V.V. Veeravalli, "An Efficient Sequential Procedure for Detecting Changes in Multichannel and Distributed Systems," *Proc. 5th Int. Conf. Inf. Fusion*, Fusion 2002, Annapolis, USA.

[57] R.R. Tenney and N.R. Sandell Jr., " Detection with distributed sensors," *IEEE Trans. Aerospace Elect. Syst.*, vol. AES-17, pp. 501-510, 1981.

[58] J.N. Tsitsiklis, "On threshold rules in decentralized detection," *in Proc. 15th IEEE Conf. Decision Contr., Athens, Greece*, pp. 232-236, 1986.

[59] J.N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, Greenwich, CT: JAI Press, 1990.

[60] V.V. Veeravalli, "Comments on 'Decentralized Sequential Detection', " *IEEE Trans. Inf. Th.*,vol. 38, pp. 1428-1429, 1992.

[61] V.V. Veeravalli, T. Basar and H.V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Tran. Inf. Th.*, vol. 39, no. 3, pp. 433-442, 1993.

[62] V.V. Veeravalli, "Sequential Decision Fusion: Theory and Applications," *J. Franklin Inst.*, vol. 336, pp. 301-322, 1999.

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

[64] A. Wald and J. Wolfowitz, " Optimum character of the sequential probability ratio test," *Ann. Math. Statist*, vol. 19, pp.326-339, 1948.