

Elements of Sequential Detection with Applications to Sensor Networks

Stefano Marano and Vincenzo Matta

Department of Information and Electrical Engineering (DIIIE), University of Salerno, Italy

1. Introduction

In this contribution the basic elements of sequential detection theory are presented and some applications to sensor networks are then addressed. Sequential detection was basically introduced by A. Wald in 1947 and deals with hypothesis testing problems assuming that the number of observations made available to the detector is virtually unbounded, hence not fixed in advance as it is the case of the more classical and well-known detection paradigms. From a conceptual view point, a sequential decision procedure is stopped when the specific realization of the data available to the detector is sufficiently informative to make a decision that satisfies prescribed error probability bounds. To compare with the Neyman-Pearson paradigm, note that for this latter the false alarm probability is prescribed, and the goal is to achieve the best detection probability, given a fixed number of samples available. In Wald's hypothesis test, instead, we are given both the false alarm and detection probabilities, and the attempt is to make a decision, compatible with those performance levels, using the minimum number of samples. Sequential tests outperform the classical decision procedures.

This notwithstanding, they are less known to non specialists, and perhaps less used in practical implementations. In part, this is due to the fact that sequential tests are less easy to be analyzed when the data are non independent and/or non identically distributed, although recent advances in that direction are available.

However, and this is relevant for the present work, there exist many applicative scenarios where the sequential paradigm naturally arises as a suitable framework. This is the case, for instance, of certain sensor network architectures where a mobile agent sequentially queries the nodes of the network, in order to retrieve data or local inferences stored at those sensors. Here the mobile nature of the agent implies that the data are made available to it in a sequential fashion. In addition, the typical dimensions of certain sensor networks make the assumption of an unbounded stream of data available for sampling, a reasonable mathematical model of the physical scenario.

This work is made of two parts. The first presents the theoretical background and tools of sequential detection, while the second addresses some practical applications. In presenting the basic elements of sequential analysis we are neither exhaustive nor mathematically advanced, beyond the typical tools of an electrical engineer. Rather, we hope to collect the main results for easy reference, and basic understanding of the applicative case studies. In discussing the applications to the sensor network, again, we have no pretence at all of

Source: Sensor and Data Fusion, Book edited by: Dr. ir. Nada Milisavljević,
ISBN 978-3-902613-52-3, pp. 490, February 2009, I-Tech, Vienna, Austria

exhaustiveness. Rather, we focus on some recent applications of sequential sampling to sensor networks, as investigated by our research group.

2. Basic elements of sequential detection

2.1 Martingales

The modern theory of martingales is due to Doob [2] that still remains the basic reference. Below, we only present the basic concepts and results. A random process $\{X_n\}_{n=1}^{\infty}$ is a martingale if, for $n = 1, 2, \dots$, we have

$$E[|X_n|] < \infty, \quad \text{and} \quad E[X_{n+1}|X_1, X_2, \dots, X_n] = X_n. \quad (1)$$

The term martingale was introduced in France to denote a gambling scheme in which the gambler doubles his bet at each step of the play, until he finally wins. More in general, a martingale is intended as a betting scheme designed to improve one's fortune. In effect, if we interpret X_n as the fortune of the gambler at step n , the above definition of martingale random process states that the gambler's fortune on the next play is, on the average, the same of his current fortune, irrespectively of the previous history; therefore, the martingale models a fair game [3-5].

The perhaps simplest examples of martingales are the sum $X_n = \sum_{i=1}^n W_i$ of independent random variables W_i 's satisfying the first of (1) and having zero mean, and the product $X_n = \prod_{i=1}^n W_i$ of independent random variables W_i 's with unit mean (again, if the first of (1) is satisfied).

A slightly general definition of martingale is as follows. The random process $\{X_n\}_{n=1}^{\infty}$ is a martingale with respect to the random process $\{Y_n\}_{n=1}^{\infty}$ if, for $n = 1, 2, \dots$,

$$E[|X_n|] < \infty, \quad \text{and} \quad E[X_{n+1}|Y_1, Y_2, \dots, Y_n] = X_n. \quad (2)$$

With the above definitions, the so-called Doob martingale can be introduced by considering an arbitrary random process $\{Y_n\}_{n=1}^{\infty}$, and a random variable X with $E[|X|] < \infty$. In fact, the process

$$X_n = E[X|Y_1, Y_2, \dots, Y_n] \quad (3)$$

is easily shown to be a martingale with respect to $\{Y_n\}_{n=1}^{\infty}$. Doob's martingale has important applications in various fields, including estimation theory since the sequence of optimal (in the mean square sense) estimators of a random variable X , given observations Y_1, \dots, Y_n , is the conditional mean $E[X|Y_1, \dots, Y_n]$ which, therefore, is nothing but a Doob martingale.

In the sequential sampling framework that we are interested in, the concept of stopping time is key. A random variable N taking values in $\{1, 2, \dots, \infty\}$ is a *random time* for the process $\{X_n\}_{n=1}^{\infty}$, if the event $\{N = n\}$ is determined by X_1, X_2, \dots, X_n . This means that we can decide if $N = n$ or $N \neq n$, by only observing the process X_i up to time n , while the samples X_{n+1}, X_{n+2}, \dots are irrelevant for that. A random time N is called *stopping time* for the process $\{X_n\}_{n=1}^{\infty}$ if $\Pr\{N < \infty\} = 1$. Therefore, N is a stopping time for $\{X_n\}$ if the event $\{N = n\}$ conditioned on knowing the past of the process $\{X_n\}_{n=1}^{\infty}$, does not depend upon what $\{X_n\}$ does in the future. Indeed, a more general formal definition of stopping time can be given just in terms conditional independence, which is relevant in cases where the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is only one actor of a more general probabilistic experiment [6].

Consider now a stopping time for the process $\{X_n\}_{n=1}^\infty$, and define the associated *stopped process* [5]: for $n = 1, 2, \dots$,

$$\bar{X}_n = \begin{cases} X_n & \text{if } n \leq N \\ X_N & \text{if } n > N \end{cases} \tag{4}$$

It can be shown that, if $\{X_n\}_{n=1}^\infty$ is a martingale, then the associated stopped process $\{\bar{X}_n\}_{n=1}^\infty$ is a martingale too.

By definition of martingale, taking the expectation to both the sides of the second equation in (1), we have $E[X_{n+1}] = E[X_n]$, $\forall n \geq 1$, and consequently,

$$E[X_n] = E[X_1]$$

for all n . Since $\{\bar{X}_n\}_{n=1}^\infty$ is a martingale too, we also have $E[\bar{X}_n] = E[\bar{X}_1] = E[X_1]$ (note that always $\bar{X}_1 = X_1$). Since the stopping time is finite with probability one, the original process $\{X_n\}_{n=1}^\infty$ will be eventually stopped, that is to say, there must exist a (sufficiently large) value of n such that $\bar{X}_n = X_N$. From that value of n on, the stopped process remains constant, implying that $\lim_n \bar{X}_n = X_N$. Taking the expectation

$$E[\lim_n \bar{X}_n] = E[X_N].$$

If we could exchange the limit with the statistical expectation operator, we would get the following important result

$$E[X_N] = E[\lim_n \bar{X}_n] = \lim_n E[\bar{X}_n] = E[X_1], \tag{5}$$

Indeed, under appropriate technical conditions the above exchange is legitimate and this result is known as the *martingale optional stopping theorem*, which states the following (see e.g., [4, 5]): Equation (5) holds true, provided that at least one of the following conditions is met

- the random variables $\{\bar{X}_n\}$ are uniformly bounded;
- the stopping time N is bounded, i.e., $\Pr\{N \leq k\} = 1$, for some $k \geq 1$;
- $E[N] < \infty$ and $E[|X_{n+1} - X_n| \mid X_1, \dots, X_n] < k < \infty$, for some $k \geq 1$.

2.2 Sequential probability ratio test

With the concepts introduced above, we can now elaborate on the likelihood ratio to derive the basic design formulas for the sequential test proposed by Wald. Let us consider the binary test between two simple hypotheses:

$$\begin{aligned} \mathcal{H}_0 &: Y_i \sim f_0(y), \\ \mathcal{H}_1 &: Y_i \sim f_1(y), \end{aligned} \tag{6}$$

where $f_{0,1}(y)$ are two known probability density functions of Y_i , under hypothesis 0, 1, and where the data Y_i , $i = 1, 2, \dots$ are iid (independent, identically distributed), for simplicity. As distinct feature, the number of such observations is not determined in advance but it can be virtually unbounded. Wald’s test, also known as SPRT (Sequential Probability Ratio Test) prescribes to proceed as follows. Let Λ_n be the likelihood ratio pertaining to the above

statistical test, using the first n samples Y_1, \dots, Y_n , available, that is: $\Lambda_n = \prod_{i=1}^n f_1(y_i)/f_0(y_i)$. We have [1]

$$\Lambda_n \begin{cases} \geq \gamma_1 & \text{decide } \mathcal{H}_1 \text{ and stop testing ,} \\ \leq \gamma_0 & \text{decide } \mathcal{H}_0 \text{ and stop testing ,} \\ \text{otherwise} & \text{take the } (n + 1)^{th} \text{ sample (i.e., increase } n\text{).} \end{cases} \quad (7)$$

As it is clear from this formulation, the actual number of samples processed in order to make a decision is not fixed and is instead a random quantity, whose actual value will depend on the specific realization of the observation process $\{Y_n\}_{n=1}^\infty$. The detection and false alarm probabilities of the test are defined as usual:

$$P_d = \Pr\{\text{the decision is for } \mathcal{H}_1 \mid \text{the actual state of the nature is } \mathcal{H}_1\}, \quad (8)$$

$$P_f = \Pr\{\text{the decision is for } \mathcal{H}_1 \mid \text{the actual state of the nature is } \mathcal{H}_0\}, \quad (9)$$

and there is an amazing simple relationship relating the pair (P_d, P_f) to the thresholds of the test (γ_0, γ_1) , as we shall promptly see.

Let us assume that hypothesis \mathcal{H}_0 is actually in force. Given the observations Y_1, \dots, Y_n , let us consider the random process build upon the likelihood ratio

$$\Lambda_n = \frac{\prod_{i=1}^n f_1(y_i)}{\prod_{i=1}^n f_0(y_i)}. \quad (10)$$

Under mild regularity conditions (i.e., assuming $E[|\Lambda_n| \mid \mathcal{H}_0] < \infty, \forall n$), the process $\{\Lambda_n\}_{n=1}^\infty$ is a martingale with respect to the observation process $\{Y_n\}_{n=1}^\infty$. Indeed, we have

$$\begin{aligned} E[\Lambda_{n+1} | Y_1, Y_2, \dots, Y_n] &= E\left[\frac{\prod_{i=1}^{n+1} f_1(y_i)}{\prod_{i=1}^{n+1} f_0(y_i)} \mid Y_1, Y_2, \dots, Y_n\right] \\ &= E\left[\frac{f_1(y_{n+1})}{f_0(y_{n+1})} \frac{\prod_{i=1}^n f_1(y_i)}{\prod_{i=1}^n f_0(y_i)} \mid Y_1, Y_2, \dots, Y_n\right] \\ &= E\left[\frac{f_1(y_{n+1})}{f_0(y_{n+1})}\right] \Lambda_n = \Lambda_n. \end{aligned} \quad (11)$$

Let us define the random time (with $\gamma_0 < 0$ and $\gamma_1 > 0$)

$$N = \min\{n \geq 1 : \Lambda_n \leq e^{\gamma_0} \text{ or } \Lambda_n \geq e^{\gamma_1}\} \quad (12)$$

and assume that $\Pr\{N < \infty\} = 1$, i.e., that N is actually a stopping time for the process $\{Y_n\}_{n=1}^\infty$. The associated stopped process can be defined as in (4), and we can invoke (under suitable regularity conditions, see above) the martingale optional stopping theorem, yielding

$$\begin{aligned} E[\Lambda_N] &= E[\Lambda_N \mid \Lambda_N \leq e^{\gamma_0}] \Pr\{\Lambda_N \leq e^{\gamma_0}\} + E[\Lambda_N \mid \Lambda_N \geq e^{\gamma_1}] \Pr\{\Lambda_N \geq e^{\gamma_1}\} \\ &= E[\Lambda_N \mid \Lambda_N \leq e^{\gamma_0}] (1 - P_f) + E[\Lambda_N \mid \Lambda_N \geq e^{\gamma_1}] P_f = E[\Lambda_1] = 1. \end{aligned} \quad (13)$$

Neglecting the *excess over the boundaries* (this is also known as Wald’s approximation), yields

$$E[\Lambda_N | \Lambda_N \leq e^{\gamma_0}] \approx e^{\gamma_0}, \tag{14}$$

$$E[\Lambda_N | \Lambda_N \geq e^{\gamma_1}] \approx e^{\gamma_1}, \tag{15}$$

that underly many of the approximate design formulas for sequential detectors. We therefore get

$$e^{\gamma_0} (1 - P_f) + e^{\gamma_1} P_f \approx 1. \tag{16}$$

Reasoning in the same way under hypothesis \mathcal{H}_1 , but using as martingale the the inverse of the likelihood ratio $1/\Lambda_n$, we get

$$e^{-\gamma_0} (1 - P_d) + e^{-\gamma_1} P_d \approx 1. \tag{17}$$

Putting together eqs. (16) and (17), immediately yields¹

$$\gamma_1 \approx \log \frac{P_d}{P_f}, \tag{18}$$

$$\gamma_0 \approx \log \frac{1 - P_d}{1 - P_f}. \tag{19}$$

We reiterate that the approximation involved follows from having neglected the excess over the boundaries².

Equations (18) and (19) relate the error probability of the SPRT to the thresholds and are therefore used to set the thresholds of the test, given prescribed performance level. Intriguingly, note that we can set the thresholds without knowing the statistics of the observations; clearly, the likelihood ratio does depend on these statistics.

Once that the desired error probabilities have been fixed, and that the thresholds are accordingly set, the main performance figure of the Wald’s test is the average sample number (ASN) $E[N]$. This is the expected number of samples needed to make a decision, and is also referred to as the (averaged) decision delay. To characterize $E[N]$ let us start from what is known in the literature as *Wald’s identity*. Let $W_i, i = 1, 2, \dots$ be a sequence of iid random variables, and let $X_n = \sum_{i=1}^n W_i$ their cumulative sum. Also, let

$$N = \min\{n \geq 1 : X_n \leq \gamma_0 \text{ or } X_n \geq \gamma_1\}. \tag{20}$$

Introducing the semi-invariant moment generating function of the random variable W_i , that is $\Gamma(r) = \log E[e^{rW_i}]$, and assuming that $\Gamma(r)$ is finite in an open interval Ω around the origin $r = 0$, we have that, for $r \in \Omega$,

¹ Logarithms are to natural base.

² The approximation involved in the above relationships is often accurate for practical purposes, but one can also resort to certain bounds. Specifically, let P_{da} and P_{fa} the *actual* detection and false alarm probabilities that one gets by setting the thresholds according to eqs. (18) and (19) in which the *nominal* values P_d and P_f of those probability are used. Then, see e.g. [7]:

$$1 - P_{da} \leq \frac{1 - P_d}{1 - P_f}, \quad P_{fa} \leq \frac{P_f}{P_d}.$$

$$\Gamma(r) = \log \mathbb{E}[e^{rW_i}] \quad (21)$$

Differentiating both sides of Wald's identity (21) and evaluating the result at $r = 0$, the so-called *Wald's equality* is obtained:

$$\mathbb{E}[X_N] = \mathbb{E}[N] \mathbb{E}[W]. \quad (22)$$

(Clearly, the above equality can be also proved more directly, without resorting to Wald's identity, see, e.g., [5].) On the other hand, the second derivative evaluated at $r = 0$ gives

$$\mathbb{E}[X_N] = \mathbb{E}[N] \text{VAR}[W], \quad (23)$$

which is useful in sequential analysis whenever $\mathbb{E}[W] = 0$, in which case Wald's equality is of limited use.

Suppose now that the true hypothesis is \mathcal{H}_0 , and let us replace the generic X_n with L_n , the *log-likelihood ratio* of hypothesis test (6):

$$L_n = \sum_{i=1}^n \log \frac{f_1(y_i)}{f_0(y_i)}. \quad (24)$$

With this assumption, it should be clear that the stopping time N in eq. (20) is exactly that defined in (12), and we are in fact faced with the earlier discussed sequential test. Clearly, L_n is the cumulative sum of the sequence of iid random variables $\log [f_1(y_i)/f_0(y_i)]$, whence application of Wald's equality yields

$$\mathbb{E}[N | \mathcal{H}_0] = \frac{\mathbb{E}[L_N | \mathcal{H}_0]}{\mathbb{E}\left[\log \frac{f_1(y_i)}{f_0(y_i)} | \mathcal{H}_0\right]}. \quad (25)$$

The numerator can be expanded by conditioning, and then approximated by Wald's approximations:

$$\begin{aligned} \mathbb{E}[L_N | \mathcal{H}_0] &= \mathbb{E}[L_N | L_N \leq \gamma_0, \mathcal{H}_0] \Pr\{L_N \leq \gamma_0 | \mathcal{H}_0\} \\ &+ \mathbb{E}[L_N | L_N \geq \gamma_1, \mathcal{H}_0] \Pr\{L_N \geq \gamma_1 | \mathcal{H}_0\} \\ &\approx \gamma_0 \Pr\{L_N \leq \gamma_0 | \mathcal{H}_0\} + \gamma_1 \Pr\{L_N \geq \gamma_1 | \mathcal{H}_0\} \\ &= \gamma_0 (1 - P_f) + \gamma_1 P_f \\ &\approx (1 - P_f) \log \frac{1 - P_d}{1 - P_f} + P_f \log \frac{P_d}{P_f}, \end{aligned} \quad (26)$$

where the last approximation follows from eqs. (18) and (19).

Defining the binary divergence (measured in nats) $D_b(\alpha | \beta)$ between two probability mass functions (pmf) $[\alpha, 1 - \alpha]$ and $[\beta, 1 - \beta]$ as [8]

$$D_b(\alpha | \beta) = \alpha \log \frac{\alpha}{\beta} + (1 - \alpha) \log \frac{1 - \alpha}{1 - \beta},$$

the numerator of (25) can be expressed (within the stated approximation) as follows

$$E[L_N | \mathcal{H}_0] \approx -D_b(P_f || P_d).$$

The divergence between two arbitrary probability density functions (pdfs) $f_0(y)$ and $f_1(y)$ is defined as [8]

$$D(f_0 || f_1) = \int f_0(y) \log \frac{f_0(y)}{f_1(y)} dy$$

implying that the denominator of (25) is $-D(f_0 | | f_1)$. We finally get $E[N | \mathcal{H}_0] = \frac{D_b(P_f || P_d)}{D(f_0 || f_1)}$. Elaborating exactly in the same way, assuming \mathcal{H}_1 true, we can compute $E[N | \mathcal{H}_0]$. Therefore, the final result is

$$E[N | \mathcal{H}_0] \approx \frac{D_b(P_f || P_d)}{D(f_0 || f_1)}, \tag{27}$$

$$E[N | \mathcal{H}_1] \approx \frac{D_b(P_d || P_f)}{D(f_1 || f_0)}. \tag{28}$$

Roughly speaking, the above approximations become tighter and tighter for small error probabilities $P_f, 1 - P_d \ll 1$, namely, when the no decision region between the thresholds is large: $\gamma_0 \ll 1$ and $\gamma_1 \gg 1$. Otherwise stated, the approximations are fair when the average number of samples collected is large enough.

It is worth noting how Wald popularized these formulas in 1947 in a form that did not involve the divergences, and in fact the divergence was defined by Kullback [9] in the context of information theory, a theory born just one year later, in 1948, with the work by Shannon.

2.3 Sequential detection with general test statistics

The fundamental results presented in the previous section trace the route for implementing the SPRT, as well as for computing simple approximations for performance evaluation. Of course, since the pioneering work by Wald, these results have been extended in many different directions, including, among many other, the case of dependent and non-identically distributed observations [10], sequential tests with arbitrary detection statistics [6], asymptotic results for vanishing signal-to-noise ratio [11], refined approximations for the excesses over boundaries [12], and so on. An exhaustive review of these concepts is clearly beyond the scope of the present work. In this section, we limit ourselves to consider a setting which slightly extends the classical SPRT framework. The mathematical results, which are perhaps less intuitive than the classical Wald’s formulas, turn out to be useful from an engineering perspective, as we next show in the sections devoted to sensor network applications.

Let $T_n = \sum_{i=1}^n t(Y_i)$, with Y_i iid random variables, and with $t(\cdot)$ being a certain transformation, in general different from the log-likelihood ratio. We consider the case that a sequential test is implemented, based upon the above T_n . More specifically, let N be the smallest n for which either $T_n \geq \gamma_1$ or $T_n \leq \gamma_0$. For concreteness, we assume that the $t(Y_i)$ has positive expectation under \mathcal{H}_1 and negative expectation under \mathcal{H}_0 , and that the two thresholds accordingly obey $\gamma_0 < 0 < \gamma_1$. This problem can be cast in the more general framework of random walks with two thresholds [6, 13, 14].

In order to assess the test performances, we need setting the thresholds in order to meet the prescribed false alarm and detection probabilities. To fix ideas, let us consider the case that $E[t(Y_i)] < 0$, that is, for our purposes, we are under \mathcal{H}_0 . Assuming that the semi-invariant moment generating function $\Gamma(r)$ of the random variable $t(Y_i)$ is finite in an open interval Ω around the origin, and that has a root $r^* > 0$, we can use Wald's identity with $r = r^*$, that is, $E[e^{r^*T_N}] = 1$, which is equivalent to

$$E[e^{r^*T_N} | T_N \leq \gamma_0] \Pr\{T_N \leq \gamma_0\} + E[e^{r^*T_N} | T_N \geq \gamma_1] \Pr\{T_N \geq \gamma_1\} = 1 \tag{29}$$

Note that, if T_n is the log-likelihood ratio, $r^* = 1$, and eq. (29) essentially translates into eq. (13), which, incidentally, was obtained by resorting to the theory of martingales. Now, different bounds and approximations can be derived from (29). One option is to neglect the overshoots, as previously done for the log-likelihood ratio, obtaining

$$e^{r^*\gamma_0} (1 - \Pr\{T_N \geq \gamma_1\}) + e^{r^*\gamma_1} \Pr\{T_N \geq \gamma_1\} \approx 1, \tag{30}$$

which is the counterpart of eq. (16).

The above technique gives also a direct way to derive upper bounds on the threshold crossing probabilities. Indeed, conditional on $T_N \geq \gamma_1$, $e^{r^*T_N} \geq e^{r^*\gamma_1}$, and being all terms on the right-hand-side of eq. (29) non-negative, we easily have

$$\Pr\{T_N \geq \gamma_1\} \leq e^{-r^*\gamma_1}. \tag{31}$$

A positive threshold crossing can be considered, under the assumption that $E[t(Y_i)] < 0$, as an event becoming rare as γ_1 grows, and this is consistent with the obtained *exponential* bound. All the reasoning can be applied to the random walk with $E[t(Y_i)] > 0$ (that is, under \mathcal{H}_1), obtaining the two formulas:

$$e^{-|r^{**}|\gamma_0} (1 - \Pr\{T_N \geq \gamma_1\}) + e^{-|r^{**}|\gamma_1} \Pr\{T_N \geq \gamma_1\} \approx 1, \tag{32}$$

and

$$\Pr\{T_N \leq \gamma_0\} \leq e^{-|r^{**}|\gamma_0},$$

where the non-zero root of $\Gamma(r)$ is now negative, and is denoted by r^{**} . We note explicitly that these exponential bounds do not work for the case that the random walk is zero-mean. Before concluding this section, we would like to report another useful tool, which extends the previous results to the characterization of the joint distribution of the stopping time N and the barriers. Assuming again that $E[t(Y_i)] < 0$ and $\Gamma(r^*) = 0$ for some $r^* > 0$, the following two bounds can be derived [6]:

$$\Pr\{N \geq n, T_N \geq \gamma_1\} \leq \begin{cases} e^{-r\gamma_1 + \Gamma(r)}; & \text{for } n > \gamma_1/\Gamma'(r^*), \quad r : \Gamma'(r) = \gamma_1/r, \\ e^{-r^*\gamma_1}; & \text{for } n \leq \gamma_1/\Gamma'(r^*), \end{cases}$$

and

$$\Pr\{N \leq n, T_N \geq \gamma_1\} \leq \begin{cases} e^{-r\gamma_1+n\Gamma(r)}; & \text{for } n < \gamma_1/\Gamma'(r^*), \quad r : \Gamma'(r) = \gamma_1/r, \\ e^{-r^*\gamma_1}; & \text{for } n \geq \gamma_1/\Gamma'(r^*) \quad . \end{cases}$$

In some sense, the above bounds furnish an interpretation of the value $n^* = \gamma_1/\Gamma'(r^*)$ as the typical value of the stopping time N conditional on the upper threshold crossing. This concludes our survey of the basic tools and results of sequential detection. In the following, we apply and extend these results by addressing several case studies.

3. Selected applications

We now present some applications of the sequential sampling theory, whose basic elements have been summarized in the previous sections; the applications are selected from the authors' recent works on the subject, which are somehow related to sensor networks. In the first example (Sect. 3.1) the decentralized architecture of the sensor network is key, and we take a genuine cross-layer perspective of the whole system that merges the detection layer with the (many-to-one) communication layer. The last two examples (Sects. 3.3 and 3.2) focus on the signal processing at the sensor level designed for improving the detection performances of the fusion center, and are therefore exploitable even in certain non decentralized systems. Whenever appropriate, for easy reference, we try to maintain the notation as close as possible to that of the original works to which we refer for more general discussion, in-depth description, and for many technicalities which are deliberately neglected in this presentation.

3.1 SENMA detection with censoring nodes [15]

Suppose that a WSN designed for solving a binary hypothesis test is made of many tiny remote units uniformly deployed over the surveyed area, and of a Mobile Agent (MA) having the role of fusion center. The remote units sense the environment and collect data relevant to the detection task, while the MA travels across the network domain and sequentially polls the sensors. Indeed, in the SENMA (Sensor Network with Mobile Agents) architecture proposed in [16], see also [17–19], at each successive MA's *snapshot* the nodes falling within its field of view are queried for delivering their data. Oppositely to the intrinsic nature of the remote units, the MA can be a very reliable device with large power capabilities and adequate communication/computational properties. In addition, its mobile nature greatly simplifies the sensors/MA communication tasks, thus making the SENMA architecture particularly suited for many practical applications being scalable, robust and simple to implement. In addition, as one might expect, the more important advantage of the SENMA over alternative network structures (*e.g.*, ad-hoc system) is in terms of energy saving for communications, a key parameter for sensor networks.

The MA collects the data delivered from the sensors and, as soon as a new observation is made available to it, this is included in the computed detection statistic. We assume that such a statistic is the cumulative sum of the log-likelihood ratios, resulting in the Wald's SPRT [1], discussed previously in this work.

The specific viewpoint taken in [15] is that the remote units do not necessarily deliver their data to the MA when they fall in its field of view. In order to further economize the energy burden, a censoring protocol is implemented [20–23]. Data are delivered only if they are

sufficiently informative: the sensor transmission is inhibited if the the locally computed log-likelihood of the measured data does not exceed (in modulus) a certain threshold level. In this way a communication session is activated, and the correspondent energy is spent by the sensor, only if the local observation is expected to contribute in a meaningful way to the final decision. Otherwise, such data will be never received by the MA and do not play any role in building the final statistics. A trade off clearly emerges between detection performances and energy consumption.

3.1.1 Network performances

Let us suppose that data collected by the remote units of the network are M -vectors of iid (independent and identically distributed) observations, and that different nodes observe iid data, as well. If we label with an index $n = 1, 2, \dots$, the (virtually, infinitely many) remote units, the basic hypothesis test under study is as follows

$$\begin{aligned} \mathcal{H}_0 : \mathbf{x}_n &= \mathbf{w}_n, \\ \mathcal{H}_1 : \mathbf{x}_n &= \mathbf{w}_n + \mu \mathbf{1}, \end{aligned} \tag{33}$$

where $\mathbf{1}$ represents a vector of all 1s. The vectors $\mathbf{w}_n = [w_{n1}, w_{n2}, \dots, w_{nM}]$ have iid components picked from a continuous random variable, whose probability density function (pdf) is $\phi(w)$, and is here assumed to be an even function with domain the whole real axis. The known parameter μ rules the amount of shift in mean that distinguishes the two alternative hypotheses.

Denoting by x_{nm} the m th observation taken at the n th sensor, the local log-likelihood is

$$L(\mathbf{x}_n) = \sum_{m=1}^M \log \frac{\phi(x_{nm} - \mu)}{\phi(x_{nm})},$$

and, in absence of censoring, the SPRT would be (see eq.(7))

$$\sum_{n=1}^j L(\mathbf{x}_n) \begin{cases} \geq \gamma_1 & \text{choose } \mathcal{H}_1, \\ \leq \gamma_0 & \text{choose } \mathcal{H}_0, \\ \text{otherwise} & \text{take another sample (increase } j). \end{cases} \tag{34}$$

Due to the censoring, when polled by the MA, the n th sensor of the network actually delivers data only if $L(\mathbf{x}_n) \notin [-\delta, \delta]$, where δ is the level of censoring. The delivering probability can be found to be

$$p_t(\delta) = \Pr[L(\mathbf{x}_n) \notin [-\delta, \delta]] = 1 - F_i(\delta) + F_i(-\delta) = F_1(-\delta) + F_0(-\delta), \tag{35}$$

where $F_i(y)$ is the CDF of the log-likelihood $L(\mathbf{x}_n)$, under hypothesis $i = 0, 1$.

Let N_t be the random number of sensors that actually deliver data to the MA, as opposed to N_v , the number nodes encountered by the MA in its travel across the surveyed area. Denoting by $\mathcal{I}(\cdot)$ the indicator function, we have $N_t = \sum_{n=1}^{N_v} \mathcal{I}\{L(\mathbf{x}_n) \notin [-\delta, \delta]\}$, where N_v is a valid stopping time [6], so that Wald's equality (22) yields

$$E[N_t] = E[N_v]E[\mathcal{I}\{L(\mathbf{x}_n) \notin [-\delta, \delta]\}] = E[N_v]p_t(\delta). \tag{36}$$

It makes sense to adopt N_t as a proxy of the energy consumption, and N_v as a proxy of detection performance in terms of detection delay for achieving a desired level of error probabilities. Therefore, the above relationship emphasize the trade off between detection performances (more data yield better performances) and energy saving, tuned by the censoring level δ .

We need now to introduce the following quantity:

$$s(\delta) = \frac{1}{p_t(\delta)} \int_{|\xi|>\delta} f_1(\xi) \log \frac{f_1(\xi)}{f_0(\xi)} d\xi = \frac{1}{p_t(\delta)} \int_{|\xi|>\delta} \xi f_1(\xi) d\xi. \tag{37}$$

Denoting by P_d and P_f the desired detection and false alarm probabilities set at the design stage, see eqs. (8) and (9), in [15] the framework of sequential analysis earlier discussed is exploited, to show that (via proper modification of the the techniques leading to eqs. (27) and (28)):

$$E[N_t|\mathcal{H}_0] = \frac{D_b(P_f||P_d)}{s(\delta)}, \quad E[N_t|\mathcal{H}_1] = \frac{D_b(P_d||P_f)}{s(\delta)},$$

and

$$E[N_v|\mathcal{H}_0] = \frac{D_b(P_f||P_d)}{p_t(\delta)s(\delta)}, \quad E[N_v|\mathcal{H}_1] = \frac{D_b(P_d||P_f)}{p_t(\delta)s(\delta)}.$$

It can be also shown that $s(\delta)$ is monotonically increasing in δ , while the product $p_t(\delta)s(\delta)$ monotonically decreases with δ . This implies that the larger is δ , the more energy the network saves but the larger is the detection delay.

3.1.2 Optimization

To compare our censored system with respect to the absence of censoring, let us define $E[N]$ as the average number of sensors resulting from assuming $\delta = 0$ (no censoring). The following quantities can be introduced

$$\eta_t = \frac{E[N_t|\mathcal{H}_0]}{E[N|\mathcal{H}_0]} = \frac{E[N_t|\mathcal{H}_1]}{E[N|\mathcal{H}_1]} = \frac{s(0)}{s(\delta)} \tag{38}$$

and

$$\eta_v = \frac{E[N_v|\mathcal{H}_0]}{E[N|\mathcal{H}_0]} = \frac{E[N_v|\mathcal{H}_1]}{E[N|\mathcal{H}_1]} = \frac{\eta_t}{p_t(\delta)}. \tag{39}$$

Now, let us consider the single-sample local log-likelihood $L(x_n)$ and let us denote by a and b^2 its mean and variance:

$$a = E[L(x_n)|\mathcal{H}_1] = -E[L(x_n)|\mathcal{H}_0] \tag{40}$$

$$b^2 = \text{VAR}[L(x_n)|\mathcal{H}_1] = \text{VAR}[L(x_n)|\mathcal{H}_0]. \tag{41}$$

Also let

$$\Delta = \frac{\delta}{b\sqrt{M}}, \quad \rho = \frac{a\sqrt{M}}{b}. \tag{42}$$

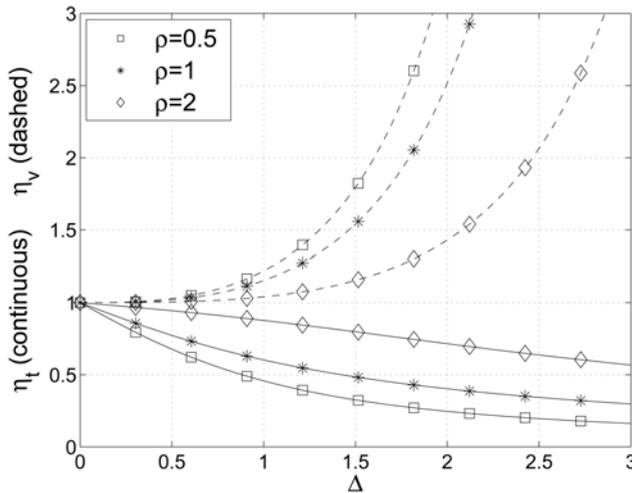


Fig. 1. Performance of the designed censored system, η_t and η_v , as function of the parameter Δ for different values of ρ .

An approximate analysis providing amenable formulas for system performances can be made by using the Central Limit Theorem, see [15], yielding:

$$\eta_t \approx \left(1 - \frac{1}{\sqrt{2\pi\rho^2}} \frac{e^{-\frac{(\Delta + \rho)^2}{2}} - e^{-\frac{(\Delta - \rho)^2}{2}}}{Q(\Delta + \rho) + Q(\Delta - \rho)} \right)^{-1}, \tag{43}$$

$$\eta_v \approx \frac{\eta_t}{Q(\Delta + \rho) + Q(\Delta - \rho)}, \tag{44}$$

where $Q(\cdot)$ is the standard Gaussian complementary CDF.

By drawing η_t and η_v as functions of Δ for different values of ρ , we get the curves for the system optimization, as depicted in Fig. 1, from which the desired system operative point in terms of detection delay and sensors energy consumption can be decided. Figure 2 provides the same information and insight of Fig. 1.

An interesting behavior is also observed when the sensors, provided that their observations are informative enough, can only send to the MA the *hard* decisions (*i.e.*, a binary value) taken at a local level. In this case, different from the previous case that the censored log-likelihood are transmitted, it is possible to prove that an optimal censoring level exists, minimizing the detection delay. Examples of applications, as well as detailed discussions of the above aspects, are addressed in [15], to which the reader is referred for details.

3.2 Pre-processing at sensor level for detection after transmission over noisy channels

Suppose that the sensors of a network are connected by dedicate channels (parallel architecture) to a fusion center, *i.e.*, some unit devoted to the task of data fusion, and assume

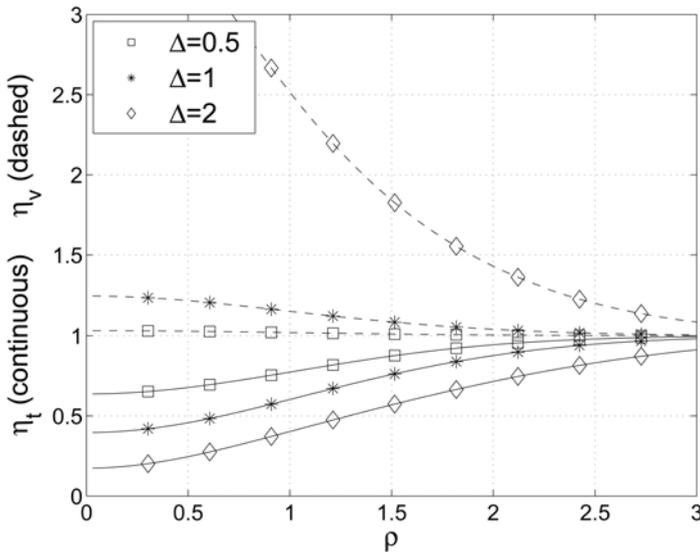


Fig. 2. Performance of the designed censored system, η_t and η_v , as function of the parameter ρ for different values of Δ .

also that such channels are noisy. The issue is to understand if some processing of the data measured at the sensor would increase the detection capabilities of the fusion center. Specifically, we are faced with a detection problem in which remotely observed data are delivered to a fusion center through a certain channel. The fusion center is designed to decide between two mutually exclusive statistical hypotheses, basing its decision upon the received data whose statistical distribution is determined by the underlying hypothesis. Should the observations made at the remote sensor be somehow processed before delivering them over the channel?

We assume that the fusion center implements a sequential test. Motivated by eqs. (27) and (28), it makes sense to choose as a measure of the detection performances the divergence between the distributions under the two hypotheses since this directly impacts the average sample number, *i.e.*, the detection delay. Therefore, the above question can be rephrased in terms of divergence: can we increase the divergence at the output of a noisy channel, by elaborating on its input? It is obvious that, if the channel were ideal (noiseless) the answer is certainly negative in view of the data processing inequality. On the other hand, for noisy channels, the answer is in some case affirmative. Let us limit the following discussion to the case where the noisy channel has *binary* input and output alphabets, and let us model the sought sensor processing as a further channel having as input the original measured data and whose output are the transformed data, to be sent over the physical noisy channel. Formally, we consider the statistical test

$$\begin{aligned} \mathcal{H}_0 &: I \sim p, \\ \mathcal{H}_1 &: I \sim q, \end{aligned} \tag{45}$$

where p and q are two arbitrary pmfs (column vectors) with alphabet $\{1, 2\}$, that rule the random variable I modeling the (iid) sensor observations. The sensor delivers data to the

fusion center by a discrete memoryless channel, $J \rightarrow K$, whose input may be different from I . Indeed a possible sensor processing may take place, which is also modeled as a channel $I \rightarrow J$. The physical channel is

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \tag{46}$$

where $C_{kj} = \Pr\{K = k \mid J = j\}$, $k, j = 1, 2$, while the processing channel is

$$\mathbf{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \tag{47}$$

where $H_{ji} = \Pr\{J = j \mid I = i\}$, $j, i = 1, 2$. A convenient, self-explaining, notation is as follows

$$I \xrightarrow{\mathbf{H}} J \xrightarrow{\mathbf{C}} K, \tag{48}$$

where \mathbf{C} is given, while the task is to find a matrix \mathbf{H} that maximizes the detection performance at the remote site that observes K .

Let x and y be the pmfs of J under the two hypotheses, and, similarly, let w and z be the pmfs of K . We have:

$$x = \mathbf{H}p, \quad y = \mathbf{H}q, \quad w = \mathbf{C}\mathbf{H}p, \quad z = \mathbf{C}\mathbf{H}q. \tag{49}$$

The described problem can be cast in the form of an optimization

$$\max_{\mathbf{H}} D(\mathbf{C}\mathbf{H}p \parallel \mathbf{C}\mathbf{H}q), \tag{50}$$

and the following claim can be proved by elementary convex analysis tools: only the following four matrices are candidates for solving the posed optimization problem

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}.$$

Clearly, the last two matrices should be ignored since they both lead to zero divergence in terms of the variable K . Given that the first matrix is the identity, the only possibility for improving the detection performance based on K is to try with the upper-right matrix, that is to say, to try with a symbol flipping: if $I = 1$ is observed, then $J = 2$ is presented at the input of the physical channel \mathbf{C} , and vice versa.

To elaborate, let us assume that the original detection problem is "difficult", in the sense that the hypotheses p and q are very close each other. For instance:

$$q_1 = p_1 + \epsilon, \quad q_2 = p_2 - \epsilon, \tag{51}$$

where $|\epsilon|$ is small enough. Expanding in series around $\epsilon = 0$, we can find

$$\begin{aligned}
 D(\mathbf{C}p \parallel \mathbf{C}q) &\approx \epsilon^2/2 (1 - C_{11} - C_{22})^2 \\
 &\quad [(1 - C_{11} - C_{22})p_1 + C_{22}]^{-1} \\
 &\quad [1 - C_{22} - (1 - C_{11} - C_{22})p_1]^{-1}, \\
 D(\mathbf{C}\mathbf{F}p \parallel \mathbf{C}\mathbf{F}q) &\approx \epsilon^2/2 (1 - C_{11} - C_{22})^2 \\
 &\quad [(1 - C_{11} - C_{22})p_1 + C_{11}]^{-1} \\
 &\quad [1 - C_{11} - (1 - C_{11} - C_{22})p_1]^{-1}.
 \end{aligned}$$

If we want that $D(\mathbf{C}\mathbf{F}p \parallel \mathbf{C}\mathbf{F}q) > D(\mathbf{C}p \parallel \mathbf{C}q)$ we must have that

$$C_{11}(1 - C_{11})(2p_1 - 1) > C_{22}(1 - C_{22})(2p_1 - 1).$$

The conclusion is that the sensor pre-processing improves the final detection performances if and only if

$$\begin{aligned}
 \text{either} \quad & C_{11}(1 - C_{11}) > C_{22}(1 - C_{22}) \text{ and } p_1 > 1/2, \\
 \text{or} \quad & C_{11}(1 - C_{11}) < C_{22}(1 - C_{22}) \text{ and } p_1 < 1/2.
 \end{aligned} \tag{52}$$

We are currently working on extensions of the topic here briefly described, with a more general formulation, including non binary observations and channels. For the time being, it is important to emphasize that for *binary* and *symmetric* physical channels (i.e., $C_{11} = C_{22}$) there is no way of improving the performances. We have evidences, however, that for non binary cases the problem exhibits much more structure and provides more useful insights from a practical perspective. What remains true in more general settings, however, is that the optimal pre-processing is deterministic, in the sense that given the input I the output J can be determined with probability one, a circumstance with a precise physical meaning.

3.3 Noise enhanced sequential detectors [24]

Let us consider a *fully decentralized* sensor network without fusion center, designed for an inference task. In this typical architecture, each node senses the environment and collects data about a phenomenon to be monitored, think for instance of a binary detection problem where the challenge is to decide which of two possible statistical distributions actually rules the observations. The lack of fusion center is remediated by suitable inter-node communication protocols that allow the system to exchange data up to make the final decision. These data, due to often unavoidable physical constraints, are here assumed to be some nonlinear transformation $t(\cdot)$ of the original observations.

Specifically, the i^{th} node of the network computes $t(X_i)$, where X_i is the sensed sample, and delivers such a value to one of its neighbors, say node j . This latter computes $t(X_i) + t(X_j)$ and delivers that to node k , and so on. The decision process is sequential: as soon as the value computed at some node exceeds given thresholds the decision is taken, and the task is terminated.

With this model in mind, motivated by recent advances in noise enhanced and stochastic resonance detection, in [24] the question is posed if adding a “noise” sample, say W_i , to the measurement made at node i before computing the nonlinearity $t(\cdot)$, could provide any benefit in terms of the final detection performances. This at first glance counterintuitive question may have (surprisingly?) a positive answer: there exist cases in which adding noise is beneficial!

3.3.1 Problem formalization

Let us consider first the original shift-in-mean binary hypothesis test:

$$\begin{aligned} \mathcal{H}_0 &: X_i \sim f_X(x + A), \\ \mathcal{H}_1 &: X_i \sim f_X(x - A). \end{aligned}$$

Here $i = 1, 2, \dots, \infty$, represents the sensor number, X_i is the observation made at node i (observations are iid), and the pdf $f_X(x)$ is an even function. According to formulation (7), the SPRT for this problem would be

$$\sum_{i=1}^n \log \frac{f_X(x_i - A)}{f_X(x_i + A)} \begin{cases} \geq \gamma_1 & \Rightarrow \text{decide } \mathcal{H}_1, \\ \leq -\gamma_0 & \Rightarrow \text{decide } \mathcal{H}_0, \\ \text{otherwise} & \Rightarrow \text{go ahead,} \end{cases} \tag{53}$$

but, in many cases of interest, implementing such an optimal SPRT is unfeasible [25]. Therefore, we consider sub-optimal sequential detectors and, as explained before, we also contaminate the original observations with iid noise: this latter effect amounts to consider the noise contaminated observables $Y_i = X_i + W_i$, in place of the original X_i . The noise density is also assumed even-symmetric: $f_W(w) = f_W(-w)$, and that of the contaminated samples becomes the convolution of the two: $f_Y(y) = f_X(x) * f_W(w)$.

The said sub-optimality of the detector amounts to work with a decision statistic in the form:

$$T_n = \sum_{i=1}^n t(y_i), \tag{54}$$

where $t(y)$ is a bounded and non-decreasing odd function. To simplify the analysis, let the error probabilities $1 - P_d$ and P_f of the sequential test, see eqs. (8) and (9), be equal, and denote by P_e such value. We have

$$T_n \begin{cases} \geq \gamma & \text{decide } \mathcal{H}_1 \text{ and stop testing,} \\ \leq -\gamma & \text{decide } \mathcal{H}_0 \text{ and stop testing,} \\ \text{otherwise} & \text{take the } (n + 1)^{th} \text{ sample (i.e., increase } n), \end{cases}$$

where the thresholds γ and $-\gamma$ are symmetric as consequence of the assumed problem symmetries.

The above test is not a standard SPRT, since T_n is not the log-likelihood. Therefore, the system performances can be obtained as discussed in Sect. 2.3. First, Wald’s equality (25), under hypothesis \mathcal{H}_1 , yields, in the regime of small P_e ,

$$E[N] \approx \frac{\gamma}{E[t(X + W) | \mathcal{H}_1]}. \tag{55}$$

Furthermore, the exponential bound (31) is used:

$$P_e \leq e^{-\gamma r^*}, \tag{56}$$

where r^* solves

$$E \left[e^{r^* t(X+W)} | \mathcal{H}_0 \right] = 1. \tag{57}$$

The denominator of (55) can be written

$$E[t(X + W) | \mathcal{H}_1] = \iint f_X(x - A) f_W(w) t(x + w) dx dw,$$

and, introducing the function

$$h_1(w) = \int f_X(x - A) t(x + w) dx, \tag{58}$$

yields

$$E[t(X + W) | \mathcal{H}_1] = E_W[h_1(W)].$$

Similarly, we have

$$E \left[e^{r t(X+W)} | \mathcal{H}_0 \right] = \iint f_X(x + A) f_W(w) e^{t(x+w)r} dx dw,$$

which, introducing the function

$$h_2(w, r) = \int f_X(x + A) e^{t(x+w)r} dx, \tag{59}$$

yields

$$E \left[e^{r t(X+W)} | \mathcal{H}_0 \right] = E_W[h_2(W, r)].$$

The two above functions allow us to write compact formulas for the system analysis. Indeed, in [24] it is found that

$$\begin{aligned} E[N] &\approx \frac{\gamma}{E_W[h_1(W)]}, \\ P_e &\approx e^{-\gamma r^*}, \quad r^* : E_W[h_2(W, r^*)] = 1. \end{aligned} \tag{60}$$

The optimal performance-enhancing noise density $f_W(w)$ must minimize the expected sample number, without increasing the error probability. Using our approximations and bounds, this amounts to

$$\max_{f_W: r^* \geq r_0^*} E_W[h_1(W)], \tag{61}$$

where r_0^* corresponds to the noise-free case.

3.3.2 Example: sign detector

Assume now that the nonlinearity is the sign function, which amounts to one-bit quantized observations

$$t(x) = \text{sign}(x).$$

The performance functions $h_1(w)$ and $h_2(w, r)$ can be found to be

$$h_1(w) = 2F_X(A + w) - 1 \quad (62)$$

$$h_2(w, r) = e^{-r} F_X(A - w) + e^r [1 - F_X(A - w)], \quad (63)$$

where $F_X(\cdot)$ is the cumulative distribution function of the random variable X . By introducing the quantity

$$p = \Pr[X + W > 0; \mathcal{H}_1] = E_W[F_X(A + W)], \quad (64)$$

we finally get

$$E_W[h_1(W)] = 2p - 1, \quad (65)$$

$$\begin{aligned} E_W[h_2(W, r)] &= e^{-r} E_W[F_X(A + W)] + e^r E_W[1 - F_X(A + W)] \\ &= e^r - 2p \sinh(r). \end{aligned} \quad (66)$$

Enforcing $E_W[h_2(W, r)] = 1$ yields as solution

$$r^* = \log \frac{p}{1-p}. \quad (67)$$

Equation (61) requires a constrained maximization of $E_W[h_1(W)] = 2p - 1$, where the constraint is $r^* = \log \frac{p}{1-p} \geq r_0^*$, with r_0^* being the value of r^* corresponding to the absence of injected noise. However this can be shown to be equivalent to its unconstrained counterpart, which amounts to simply find the maximum achievable value of p . Remarkably, it is also possible to prove that the optimizing noise density can be chosen in the class of the *coin flipping* distributions:

$$f_W(w) = \frac{1}{2} [\delta(w - w_0) + \delta(w + w_0)].$$

We report now the evidences of some numerical simulations aimed at checking the goodness of the found formulas, as well as the potential benefits of adding noise to the observations for detection purposes. We choose for the observation density a mixture of Gaussians:

$$f_X(x) = \frac{1}{2\sqrt{2\pi}} \left[e^{-\frac{(x-\mu/2)^2}{2}} + e^{-\frac{(x+\mu/2)^2}{2}} \right].$$

Moreover, we assume that the detection threshold γ is fixed in such a way to yield, in absence of noise, an error probability $P_{e0} \approx 10^{-2}$.

In Fig. 3, top plot, the ASN is displayed as function of different injected noise depth w_0 . The theoretical formulas reasonably match the simulation points, and an optimal value of w_0 minimizing the sample number is clearly present. To get the complete perspective, the actual error probability P_e is also displayed in the bottom plot of the same figure. It can be seen that, for any w_0 , while the ASN decreases, the error probability is kept below the value P_{e0} . Remarkably, in correspondence of the optimal w_0 , the actual P_e is in effect orders of magnitude smaller than the design value 10^{-2} .

4. Conclusions

In many instances of Wireless Sensor Networks (WSNs) designed for detection purposes, the fusion center is a mobile device that sequentially queries the nodes of the network. Such

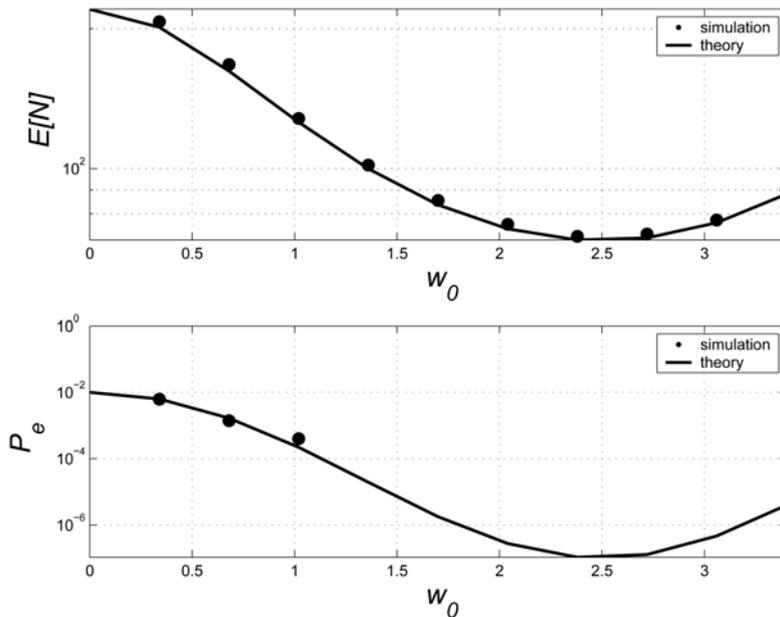


Fig. 3. Sequential detector with fixed threshold γ , for the sign nonlinearity. The expected sample number (top panel) and the error probability (bottom panel) of the proposed sequential detector are displayed as function of the injected noise depth w_0 . The simulation parameters are $A = 1$, $\mu = 5$, and $P_{e0} = 10^{-2}$. Simulation points are obtained by 10^4 Monte Carlo trials and, in bottom panel, we draw only the simulation points that do not fall under the accuracy of the numerical procedure.

a sequential architecture, known as the SENMA structure, fits well the sequential detection paradigms, *i.e.*, the SPRT and its variants. Aside from the SENMA scenarios, the typical tools of sequential detection are exploited much more in general, in various guises, in a variety of WSN applications. This paper provides a succinct introduction to sequential analysis and presents several examples of applications to detection problems.

The main aim of this paper is to introduce the reader to the very powerful tool of sequential analysis, providing the basic insights and useful entry points to some topical literature. The specific issues presented in the first part of this work are selected in order to provide the necessary theoretical background for the applicative examples discussed in the second part of the paper. These latter examples, on the other hand, reflect the authors' recent research on the subject. As a consequence, neither the theory nor the applications are exhaustive or complete in any sense. However, the paper is rather self-consistent and can help in gain a first understanding of the topic.

5. References

[1] A. Wald, *Sequential Analysis*. New York: Dover, 1947.
 [2] J. Doob, *Stochastic Processes*. New York: John Wiley & Sons, 1953.
 [3] G. Shafer and V. Vovk, *Probability and finance. It's only a game*. New York: John Wiley & Sons, 2001.

- [4] S. Karlin and H. M. Taylor, *A first Course in Stochastic Processes*, 2nd ed. San Diego, CA: Academic Press, 1975.
- [5] S. Ross, *Stochastic Processes*, 2nd ed. New York: John Wiley & Sons, Inc., 1996.
- [6] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer Academic Publisher, 1996.
- [7] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [8] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.
- [9] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.
- [10] A. G. Tartakovsky, "Sequential composite hypothesis testing with dependent nonstationary observations," *Problems of Information Transmission*, vol. 17, pp. 18–28, 1981.
- [11] T. L. Lai, "Pitman efficiencies of sequential tests and uniform limit theorems in nonparametric statistics," *The Annals of Statistics*, vol. 6, no. 5, pp. 1027–1047, 1978.
- [12] A. G. Tartakovsky and I. A. Ivanova, "Approximations in sequential rules for testing composite hypotheses and their accuracy in the problem of signal detection from postdetector data," *Prob. Inform. Transm.*, vol. 28, no. 1, pp. 55–64, Jan.-Mar. 1992.
- [13] W. Feller, *An Introduction to Probability and Its Applications*, Volume 2. New York: John Wiley & Sons, 1971.
- [14] D. R. Cox and H. D. Miller, *The theory of stochastic processes*. Chapman & Hall, 1964.
- [15] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *IEEE Trans. Signal Processing*, vol. 55, no. 11, pp. 5497–5505, Nov. 2007.
- [16] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proceedings of MILCOM 2003*, Boston MA, Oct. 2003.
- [17] B. M. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE Aerosp. Electron. Syst. Mag*, vol. 20, no. 8, Aug. 2005.
- [18] S. Appadwedula, V. Veeravalli, and D. Jones, "Energy-efficient detection in sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 693–702, Apr. 2005.
- [19] —, "Energy-efficient detection in sensor networks," in *Proceedings of Intern. Conf. on Information Fusion (FUSION 2003)*, Cairns, Queensland, Australia, July 2003, pp. 764–771.
- [20] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.
- [21] S. Appadwedula, V. Veeravalli, and D. Jones, "Robust and locally-optimum decentralized detection with censoring sensors," in *Proceedings of Intern. Conf. on Information Fusion (FUSION 2002)*, Annapolis, MD, USA, July 2002, pp. 56–63.
- [22] N. Patwari, I. Hero, A.O., and B. Sadler, "Hierarchical censoring sensors for change detection," in *IEEE Workshop on Statistical Signal Processing*, Sept. 27-Oct. 29 2003, pp. 21–24.
- [23] P. Willett and L. Tong, "One aspect to cross-layer design in sensor networks," in *Proceedings of MILCOM 2004*, Monterey, CA, USA, Oct. 2004, pp. 688–693.
- [24] M. Guerriero, S. Marano, V. Matta, and P. Willett, "Stochastic resonance in sequential detectors," *IEEE Trans. Signal Processing*, in print.
- [25] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I-Fixed detectors," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3172–3184, July 2007.



Sensor and Data Fusion

Edited by Nada Milisavljevic

ISBN 978-3-902613-52-3

Hard cover, 436 pages

Publisher I-Tech Education and Publishing

Published online 01, February, 2009

Published in print edition February, 2009

Data fusion is a research area that is growing rapidly due to the fact that it provides means for combining pieces of information coming from different sources/sensors, resulting in ameliorated overall system performance (improved decision making, increased detection capabilities, diminished number of false alarms, improved reliability in various situations at hand) with respect to separate sensors/sources. Different data fusion methods have been developed in order to optimize the overall system output in a variety of applications for which data fusion might be useful: security (humanitarian, military), medical diagnosis, environmental monitoring, remote sensing, robotics, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Stefano Marano and Vincenzo Matta (2009). Elements of Sequential Detection with Applications to Sensor Networks, Sensor and Data Fusion, Nada Milisavljevic (Ed.), ISBN: 978-3-902613-52-3, InTech, Available from:

http://www.intechopen.com/books/sensor_and_data_fusion/elements_of_sequential_detection_with_applications_to_sensor_networks

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821