RL-TR-91-96
Final Technical Report
June 1991

ON DISTRIBUTED SEQUENTIAL HYPOTHESIS TESTING

Syracuse University

Mohammad M. Al-Ibrahim and Pramod K. Varshney

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-91-96 has been reviewed and is approved for publication.

APPROVED:

VINCENT C. VANNICOLA
Project Engineer

APPROVED:

JAMES W. YOUNGBERG, LtCol, USAF
Deputy Director of Surveillance

FOR THE COMMANDER:

JAMES W. HYDE III
Directorate of Plans & Programs

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL(OCTS) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1991	Final May 89 - May 90

4. TITLE AND SUBTITLE
ON DISTRIBUTED SEQUENTIAL HYPOTHESIS TESTING

5. FUNDING NUMBERS
C - F30602-81-C-0169
PE - 61102F
PR - 2305
TA - J8
WU - PL

6. AUTHOR(S)
Mohammad M. Al-Ibrahim, Pramod K. Varshney

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Syracuse University
Department of Electrical & Computer Engineering
Syracuse NY 13210

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (OCTS)
Griffiss AFB NY 13441-5700

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RL-TR-91-96

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Vincent C. Vannicola/OCTS/(315) 330-4437

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Sequential hypothesis testing procedures are known to provide a significant advantage over fixed-sample-size (FSS) hypothesis testing in terms of the average test length. The problem of distributed FSS detection has been treated extensively in the literature recently. On the other hand, the problem of distributed sequential detection has received limited attention. This report deals with some distributed sequential detection procedures and their analysis. The performance of our testing schemes is determined in terms of average sample number (ASN) and operating characteristic (OC) functions. Most of the attention is given to parametric schemes in that a centralized testing scheme based on quantized multi-sensor data and two distributed schemes and without fusion and another with fusion are considered. The case of memoryless grouped data sequential (MLGDS) testing procedure is considered in detail. One distributed nonparametric sequential hypothesis testing procedure is also investigated. Numerical results are presented to demonstrate the superiority of distributed multisensor procedures over the classical single sensor schemes.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Sensor Fusion, Signal Processing, Detection Communication Systems, Electronic Engineering	150
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U/L

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1. Overview and Previous Work

Traditional surveillance and communication systems use centralized signal processing for the detection, identification and tracking of targets. In these systems, the received observations are transmitted to the central processor (detector) where classical hypothesis testing procedures are employed for signal processing [1,2]. The hypothesis testing procedures can be classified into two major categories depending on the test duration or the number of observations to be processed. In fixed-sample-size (FSS) hypothesis testing procedures a fixed number of observations are taken before the test can be executed. The sample size is determined by the performance level to be achieved. A variety of approaches such as the Bayesian approach and the Neyman-Pearson approach can be employed in FSS detection. Contrary to the FSS procedures, the sequential procedures have a random test duration depending on the actual observations and the test thresholds. These procedures provide a significant advantage over FSS test procedures in terms of test length. For specified values of the error probabilities α and β (where α and β are the error probabilities of the first and second kind, i.e., probabilities of false alarm and miss respectively), and known probability distributions, the sequential procedures, on the average, require significantly less number of observations than the FSS test procedures [3-5]. The sequential probability ratio test (SPRT) of Wald [1,5] is known to be optimal among all sequential test procedures.

Distributed detection systems have become an attractive alternative to the convential centralized detection systems for processing large quantities of data received from physically distributed sensors. The decomposition of processing is essential for controlling the complexity of data processing and to meet the needs for real-time results. Moreover, limited communication is often necessary due to physical bandwidth constraints combined with the desire to limit or reduce the effect of possible jamming, thereby increasing the system survivability. Distributed processing is the natural way to handle a situation in which various sensors with different sensing techniques, such as sonic, microwave and infra-red sensors are employed. These decentralized systems exhibit the advantages of higher reliability, survivability, and shorter decision time than their centralized counterparts [7,8].

Distributed FSS detection systems have received an increasing interest over the past several years. Different approaches have been used to analyze these systems. Tenney and Sandell [7] introduced distributed detection using a fixed fusion center and used the Bayesian approach to optimize the local detectors. Chair and Varshney [9] fixed the local detectors and used the Bayesian approach to optimize the fusion center. Hoballah and Varshney [10] using the Bayesian approach combined the results in [7,9] and optimized the entire distributed detection system. Reibman and Nolte [11] showed how to use numerical approaches to find globally optimum system for a particular detection problem. Srinivasan [12] considered the Neyman-Pearson criterion for system optimization. He fixed the fusion center and optimized the local detectors. Hoballah and Varshney [13] jointly optimized the entire distributed detection system

using the Neyman-Pearson criterion. Thomopoulos and Okello [14] considered a distributed system in which the primary detector communicates with the consultant detector only when the primary detector can not decide on one of the hypotheses. The majority of the distributed detection systems [7-13] allow the transmission of only one-bit decisions (hard decisions) from local detectors to the fusion center. However, some distributed detection systems [15-17] allow the transmission of multi-bit local decisions (soft decisions) to the fusion center. These soft decision distributed detection systems have the advantage of improved performance over their hard decision counterparts at the expense of increased communication (channel capacity) as well as increased analysis and processing complexity.

The decentralized sequential detection problem has been investigated in [18-20]. In [18], Teneketzis formulated and solved a decentralized version of the Wald problem with two decision makers. In his model, each detector was given the flexibility of either stopping and making a decision or continuing to the next stage. The coupling between the two local detectors was introduced through a common cost function and the objective was to minimize the cost. He showed that the person-by-person optimal policies of the local detectors are described by thresholds which are coupled. More specifically, the thresholds of detector 1 at any stage depend on the thresholds of detector 2 at all stages. For a two-detector N-stage detection system, he showed that the thresholds are determined by solving a set of 4N-2 nonlinear algebraic equations in 4N-2 unknowns. Hashemi and Rhodes [19] examined a two-step, two-detector, sequential hypothesis testing problem with data fusion center. They also discussed its straightforward extension to a decentralized multi-stage sequential

detection problem. Their model is different from the one examined in [18]. In [19], the sequential test is carried out at the fusion center and local detectors have no control over the termination of the test. Chair et al. [20] obtained a decentralized version of the SPRT for the same model as in [18]. However, they used the Neyman-Pearson approach for the solution of the problem. In [18-20], the computation of thresholds was quite complicated due to the fact that the thresholds are coupled at all stages. Moreover, the emphasis was solely on the computation of the optimal thresholds. The performance of the schemes was totally ignored. Therefore, it is desirable to explore alternate structures and schemes for distributed sequential detection where the thresholds are easier to obtain, and the performance of the schemes can be evaluated.

In this report, we consider alternate sequential detection schemes and evaluate their performance. The sequential procedures presented here are generalizations of the classical sequential procedures [1,21-27] to a distributed environment. Two important functions frequently used to evaluate the performance of an SPRT, namely the average sample number (ASN) function, and the operating characteristic (OC) function, are evaluated for the various schemes considered. These schemes have the desired advantages of distributed detection system over their centralized counterparts, and, in addition, their performance is easier to evaluate than the schemes in [18-20].

1.2. Report Organization

In this report, we consider some distributed sequential detection systems and evaluate their performance. Most of the distributed

systems considered here are parametric where an exact knowledge of the observation statistics is assumed. The only exception is the non-parametric distributed system considered in Chapter Seven, where the observation statistics are not known completely.

In Chapter Two, we describe the sequential probability ratio test (SPRT) of Wald [1]. Concepts such as average sample number (ASN) function and operating characteristic (OC) function, which are essential in evaluating the performance of sequential test are introduced in some detail. Alternate approaches to obtain the exact expressions [28,29] of the OC and ASN functions are also described.

In Chapter Three, an SPRT based on quantized multi-sensor observations is described and its performance is evaluated. It is shown that for the same level of performance, the ASN decreases monotonically by increasing the number of local sensors. The issue of optimal quantization for sequential detection is studied, and the optimal quantizer is shown to be a likelihood ratio quantizer. Moreover, we show that the quantizers obtained as a result of a joint optimization are the same as those obtained when each local quantizer is optimized independently. The specified performance level in terms of error probabilities is shown to be achievable despite channel errors at the expense of increased ASN's. An example is presented to illustrate the results obtained in this chapter.

In Chapter Four, we propose and analyze a simple sequential detection scheme using multiple sensors. The performance expressed in terms of the ASN is shown to improve monotonically with the number of local detectors (sensors) used. Moreover, we show that when truncated, the sequential test based on multiple sensors has a smaller increase in error proba-

bilities than its single detector counterpart. A numerical example is presented to illustrate the results of this chapter.

In Chapter Five, we generalize the SPRT of Wald to a distributed system consisting of two local detectors and a global decision maker (GDM). Each local detector performs an SPRT based on its own observations and communicates its local decision to the GDM. The GDM combines the local decisions according to a predetermined fusion rule, and decides either to stop and accept one of the two hypotheses, or to continue. The global error probabilities are shown to be functions of the local error probabilities and the fusion rule. The global test duration is obtained for various possible fusion rules and its average is derived. An example is presented for the case of two identical local detectors.

In Chapter Six, a modified sequential detection procedure is proposed and analyzed. The structure of the distributed system is the same as that considered in Chapter Five. Each local detector takes a group of N_o observations, performs a likelihood ratio test, and decides either in favor of one of the hypotheses or declares its inability to decide. The local decisions (including no decision) are transmitted to the GDM which combines the local decisions according to a predetermined fusion rule, and decides either to stop the local tests and accept one of the hypotheses or to continue. When the global decision is to continue, the detection process is repeated ignoring all previous stages. The binary and M-ary hypothesis testing problems are considered. A truncation scheme to limit the number of observations from being excessively large is proposed and analyzed. Numerical results are presented to illustrate the performance of the system.

In Chapter Seven, the nonparametric sequential conditional sign test

of Shin and Kassam [22] is first studied and modified. The modified test has fixed thresholds that are independent of the observations and, therefore, it does not require the table look-up operation. Both the sign and conditional sign nonparametric sequential tests [22] are generalized to a distributed system of two local sensors (or detectors). The resulting tests are shown to maintain the desirable nonparametric property. A numerical example is presented for illustration.

In Chapter Eight, a summary of results, conclusions, and suggestions for future research are presented.

CHAPTER TWO

PRELIMINARIES

In this chapter, we briefly review Wald's sequential probability ratio test (SPRT). Some terminology and notation are introduced. The modeling of SPRT as a finite state Markov chain is also discussed.

2.1. Wald's Sequential Probability Ratio Test

Consider the problem of testing a simple hypothesis H_0 versus a simple alternative H_1. Let $f(x,\theta)$ denote the probability density function of the random variable x representing the observation (sample) under consideration. Let H_0 be the hypothesis that $\theta = \theta_0$, and H_1 the hypothesis that $\theta = \theta_1$. Thus, the distribution of x is given by $f(x,\theta_j)$ when H_j is the true hypothesis , j=0,1. The successive observations on x are denoted by $x_1, x_2, \ldots x_n (n \geq 1)$ and they are assumed to be statistically independent and identically distributed (iid). For any positive integer n, the conditional probability density function of $\underline{x}_n$ given H_j is given by

$$f_{\underline{x}_n/H_j} = \prod_{i=1}^{n} f(x_i,\theta_j) \quad , \qquad j=0,1 \qquad (2.1)$$

where $\underline{x}_n \triangleq [x_1 x_2 \ldots x_n]$ is the observation vector.

Wald's SPRT for testing H_0 against H_1 is defined as follows: Two positive constants t_u and $t_\ell (t_u > t_\ell)$ are chosen. At each stage of testing (n, $n \geq 1$), the likelihood ratio function Λ_n defined below is computed.

$$\Lambda_n \triangleq f_{\underline{x}_n/H_1} / f_{\underline{x}_n/H_0} \tag{2.2}$$

The likelihood ratio function in (2.2) is compared to the thresholds t_u and t_ℓ as follows:

$$\Lambda_n \begin{cases} \geq t_u \text{ , stop and decide } H_1 \\ \leq t_\ell \text{ , stop and decide } H_0 \\ \text{otherwise, continue} \end{cases} \tag{2.3}$$

The choice of the test thresholds depends on the desired values of the error probabilities α and β, where α is the probability of deciding H_1 when H_0 is true, and β is the probability of deciding H_0 when H_1 is the true hypothesis. It has been shown [1] that the approximation of the thresholds

$$t_u \cong (1 - \beta)/\alpha \qquad t_\ell \cong \beta/(1 - \alpha) \tag{2.4}$$

is sufficiently accurate for all practical purposes and, consequently, it is widely used in the literature.

The monotonicity of the logarithm function enables us to derive a more convenient test by simply taking the logarithm of (2.3). The resulting test is given by

$$\log \Lambda_n \begin{cases} \geq \log t_u \text{ , stop and decide } H_1 \\ \leq \log t_\ell \text{ , stop and decide } H_0 \\ \text{otherwise, continue} \end{cases} \tag{2.5}$$

Taking the logarithm of the likelihood function in (2.2), we obtain

$$\log \Lambda_n = \sum_{i=1}^{n} \log \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} \tag{2.6}$$

Denoting the ith term of the sum in (2.6) by Z_i, i.e.,

$$Z_i \triangleq \log \frac{f(x_i,\theta_1)}{f(x_i,\theta_0)} \tag{2.7}$$

the test procedure can be described as follows: At each stage of the test (n, $n \geq 1$), the sum $Z_1 + Z_2 + \ldots + Z_n$ is computed, and the following test is performed

$$\sum_{i=1}^{n} Z_i \begin{cases} \geq \log t_u\ , & \text{stop and decide } H_1 \\ \leq \log t_\ell\ , & \text{stop and decide } H_0 \\ \text{otherwise,} & \text{continue} \end{cases} \tag{2.8}$$

The performance of an SPRT is characterized in terms of the operating characteristic (OC) and average sample number (ASN) functions. These are discussed next.

2.2. The OC Function of the SPRT

The operating characteristic function $L(\theta)$ of an SPRT is defined as the probability that the sequential testing procedure will terminate with the acceptance of the hypothesis H_0 when θ is the actual value of the parameter. This is necessary in situations where the exact knowledge of the parameter(s) of the distribution function is not available. An approximate formula for $L(\theta)$, neglecting the excess of the likelihood function Λ_n over the test thresholds t_u and t_ℓ at the termination stage is given by [1]

$$L(\theta) \cong \frac{t_u^{h(\theta)} - 1}{t_u^{h(\theta)} - t_\ell^{h(\theta)}} \tag{2.9}$$

where $h(\theta)$ is the unique nonzero solution of the equation

$$\int_{-\infty}^{\infty} \left[\frac{f(x,\theta_1)}{f(x,\theta_0)} \right]^{h(\theta)} f(x,\theta)\, dx = 1 \qquad (2.10a)$$

when x is a continuous random variable, and $h(\theta)$ is the unique nonzero solution of the equation

$$\sum_{x} \left[\frac{f(x,\theta_1)}{f(x,\theta_0)} \right]^{h(\theta)} f(x,\theta) = 1 \qquad (2.10b)$$

when x is a discrete random variable. The summation in (2.10b) is clearly over all possible values of x.

From (2.10a) and (2.10b), it is obvious that $h(\theta_0)=1$ and $h(\theta_1)=-1$. Therefore, it follows from (2.9) that for t_u and t_ℓ as given by (2.4), we have $L(\theta_0) = 1-\alpha$, and $L(\theta_1) = \beta$ as required.

2.3. The ASN Function of the SPRT

Let n denote the number of observations required for the termination of the sequential test, and let $E_\theta(n)$ denote the expected value of n when θ is the true value of the parameter. $E_\theta(n)$ is a function of θ and is defined to be the average sample number (ASN). Following Wald [1] in neglecting the excess of Λ_n over the thresholds at the termination stage, it can be shown that $E_\theta(n)$ is given by

$$E_\theta(n) \cong \frac{L(\theta)\, \log t_\ell + [1- L(\theta)]\, \log t_u}{E_\theta(Z_i)}, \quad E_\theta(Z_i) \neq 0 \qquad (2.11)$$

where Z_i is as defined by equation (2.7), and $E_\theta(Z_i)$ is given by

$$E_\theta(Z_i) = \int_{-\infty}^{\infty} \log \left[\frac{f(x,\theta_1)}{f(x,\theta_0)} \right] \cdot f(x,\theta)\, dx \qquad (2.12)$$

when x is a continuous random variable. When x is a discrete random variable, equation (2.12) becomes

$$E_\theta(Z_i) = \sum_x \log \left[\frac{f(x,\theta_1)}{f(x,\theta_0)}\right] \cdot f(x,\theta) \qquad (2.13)$$

When $E_\theta(Z_i) = 0$, we denote θ' the value of θ such that $E_\theta(Z_i) = 0$. In this case [1], the expression for $E_{\theta'}(n)$ is given by

$$E_{\theta'}(n) \cong \frac{-\log t_\ell \cdot \log t_u}{E_{\theta'}(Z_i^2)} \qquad (2.14)$$

and the corresponding OC function is given by

$$L(\theta') \cong \frac{\log t_u}{\log t_u - \log t_\ell} \qquad (2.15)$$

It should be emphasized that the expressions for the OC and ASN functions as given by (2.9) and (2.11) respectively, are approximate in that they were derived under the assumption of no excess over the test thresholds. It is possible to derive the exact expressions for the OC and ASN functions only when the random variable Z_i defined by (2.7) can take a finite number of values which are integral multiples of a constant. Using the characteristic function of the test duration, Wald [1] derived expressions for the probability distribution of the test duration as well as exact expressions for the OC and ASN functions. The solution of Wald's equations has been simplified by Girshick [30] who reduced the solution to that of solving a set of simultaneous equations. In the remainder of this chapter, we present a totally different approach to the

computation of the exact values of the OC and ASN functions as well as the exact distribution of the test length (duration). The approach in the next section is based on a formulation of the sequential test as a random walk on a finite number of states using the transition probability matrix.

2.4. The Transition Probability Matrix Formulation of the Sequential Test

In this section, we describe an alternate approach for the evaluation of the exact distribution of the sequential test length as well as the corresponding OC and ASN functions. The approach requires that the step Z_i as defined by (2.7) be finite-valued and takes values that are integral multiples of a constant. This facilitates the formulation of the SPRT as a random walk (Markov chain) on a finite number of states and its complete description by the specification of its transition probability matrix. Proakis [28] described the random walk formulation approach and derived the exact distribution of the test length T and the ASN for quantized radar signals. The same approach is described in more detail in [29] which is the basis for the material presented here.

Consider a Markov chain with a state space made up of (N+1) states given by $0,1,2,\ldots,N$. Let S_n denote the state at stage n and $P_{ij}^{(n)}$ denote the probability of being in state j after n transitions starting in state i, i.e.,

$$P_{ij}^{(n)} = \Pr\{S_n = j / S_0 = i\} \tag{2.16}$$

States $0,1,\ldots,r-1$ are transient in that $\lim_{n\to\infty} P_{ij}^{(n)} = 0$ for $0\le i,j\le r-1$, while states $r,r+1,\ldots,N$ are absorbing, and $P_{ii} = 1$ for $r\le i\le N$. As n

approaches ∞, the process will ultimately be absorbed in one of the absorbing states.

The stochastic process is described by the transition probability matrix whose (ij)th element is $P_{ij}^{(1)} \triangleq P_{ij}$ and which can be represented in the partitioned form as follows

$$\underline{P} = \begin{bmatrix} \underline{Q} & \underline{R} \\ \underline{0} & \underline{I} \end{bmatrix} \tag{2.17}$$

where $\underline{0}$ is an (N-r+1) × r matrix all of whose components are zero, $\underline{I}$ is an (N-r+1) × (N-r+1) identity matrix and $Q_{ij} = P_{ij}$ for $0 \leq i,j \leq r-1$.

A straightforward matrix multiplication shows that the nth power of $\underline{P}$ is given by

$$\underline{P}^n = \begin{bmatrix} \underline{Q}^n & (\underline{I}+\underline{Q}+\ldots+\underline{Q}^{n-1})\underline{R} \\ \underline{0} & \underline{I} \end{bmatrix} \tag{2.18}$$

Let $\underline{W}^n \triangleq \underline{I} + \underline{Q} + \ldots + \underline{Q}^n$, upon rewriting $\underline{W}^n$, we obtain

$$\begin{aligned} \underline{W}^n &= \underline{I} + \underline{Q}(\underline{I} + \underline{Q} + \ldots + \underline{Q}^{n-1}) \\ &= \underline{I} + \underline{Q}\,\underline{W}^{n-1} \end{aligned} \tag{2.19}$$

In the limit, we have

$$\underline{W} = \lim_{n\to\infty} \underline{W}^n \tag{2.20}$$

Combining (2.19) and (2.20), we arrive at

$$\underline{W} = \underline{I} + \underline{Q}\,\underline{W} \tag{2.21}$$

which can be written as

$$(\underline{I} - \underline{Q})\underline{W} = \underline{I} \tag{2.22}$$

From (2.22), we observe that $\underline{W} = (\underline{I}-\underline{Q})^{-1}$, the inverse of $(\underline{I}-\underline{Q})$. The matrix $\underline{W}$ is called the fundamental matrix associated with $\underline{Q}$. The ASN is

obtained by counting the number of visits to all transient states knowing that the process started in the ith transient state. In other words, we have

$$ASN = \sum_{j=0}^{r-1} W_{ij} , \quad 0 \le i \le r-1 \tag{2.23}$$

Turning to the termination probabilities, i.e., the probabilities with which the process will be absorbed in one of the absorbing states. Recall that the states k=r,...,N are absorbing. Since such a state cannot be exited once entered, the probability of absorption in a particular absorbing state k up to time n and starting from initial state i, is simply

$$P_{ik}^{(n)} = Pr\{S_n = k/S_0 = i\} , \quad 0 \le i \le r-1 \text{ and } r \le k \le N \tag{2.24}$$

In terms of the test length T, we can write

$$P_{ik}^{(n)} = Pr\{T \le n \text{ and } S_T = k/S_0 = i\} , 0 \le i \le r-1 \text{ and } r \le k \le N \tag{2.25}$$

where $T = \min \{n \ge 0: r \le S_n \le N\}$ is the termination time. Let

$$U_{ik}^{(n)} = Pr\{T \le n \text{ and } S_T = k/S_0 = i\}, \quad 0 \le i \le r-1 \text{ and } r \le k \le N \tag{2.26}$$

Referring to (2.18), (2.24) and (2.25), we give the matrix $\underline{U}^n$ by

$$\underline{U}^n = (\underline{I} + \underline{Q} + \ldots + \underline{Q}^{n-1})\underline{R} \tag{2.27}$$

which can be simplified using (2.19) to obtain

$$\underline{U}^n = \underline{W}^{n-1}\underline{R} \tag{2.28}$$

Taking the limit as $n \to \infty$, we obtain the absorption probability matrix

$\underline{U}$ in terms of the fundamental matrix $\underline{W}$ as simply $\underline{U} = \underline{W}\,\underline{R}$, or

$$U_{ik} = \sum_{j=0}^{r-1} W_{ij} R_{jk} \quad , \; 0 \le i \le r-1 \text{ and } r \le k \le N \qquad (2.29)$$

It shculd be emphasized that equation (2.29) gives the exact value of the termination probability to state k starting from state i, and therefore, it can be used to ensure the specified error probabilities. On the other hand, (2.26) through (2.28) give the cumulative probability of termination in state k starting from state i. Therefore, summing over all k, we obtain

$$\Pr\{T \le n\} = \sum_{k=r}^{N} U_{ik}^{(n)} \qquad (2.30)$$

which gives the cumulative probability distribution of the test length given that the process started in state i, i.e., $S_0 = i$.

In the above analysis, we assume that the transition probability matrix is given and obtain the ASN and termination probabilities. However, while applying the theory to the sequential hypothesis testing problem, it is clear that we have to deal with two transition probability matrices each corresponding to one of the two hypotheses. Moreover, the set of absorbing states has to be partitioned into two groups each corresponding to a decision in favor of one of the hypotheses. The OC function can be easily found by summing the probabilities of termination in all the states assigned to H_0 decision provided that the transition probability matrix which represents the true parameters is used.

This formulation will be used in the later chapters to illustrate the concepts developed.

CHAPTER THREE

A SEQUENTIAL PROBABILITY RATIO TEST BASED ON MULTISENSOR DATA

3.1. Introduction

In binary hypothesis testing, it is well known that sequential test procedures provide a significant advantage over fixed-sample size (FSS) test procedures [1,3-5]. For prespecified values of the error probabilities α and β (where α is the probability of error of the first kind and β is the probability of error of the second kind), the sequential procedures, on the average, require a substantially less number of observations (samples) than FSS test procedures. The sequential probability ratio test (SPRT) of Wald [1] is known to be optimal among all possible sequential test procedures.

In this chapter, we generalize Wald's SPRT to a distributed system consisting of M local sensors and a global (central) decision maker as shown in Fig. 3.1. In Section 3.2, we define the observation model, formulate the centralized SPRT based on multisensor data, and derive an expression for the global average sample number ASN_j when the hypothesis H_j, $j=0,1$, is true. Moreover, we show that ASN_j is a monotonically decreasing function of the number of local sensors used. In most multisensor detection systems, a bandwidth constraint on the communication channels carrying data from local sensors to the global decision maker (GDM) is assumed. Therefore, it is often necessary to compress the data locally prior to transmission. In Section 3.3, we quantize each of the local observations into two levels and transmit the quantized value to the GDM for further processing. The issue of optimal

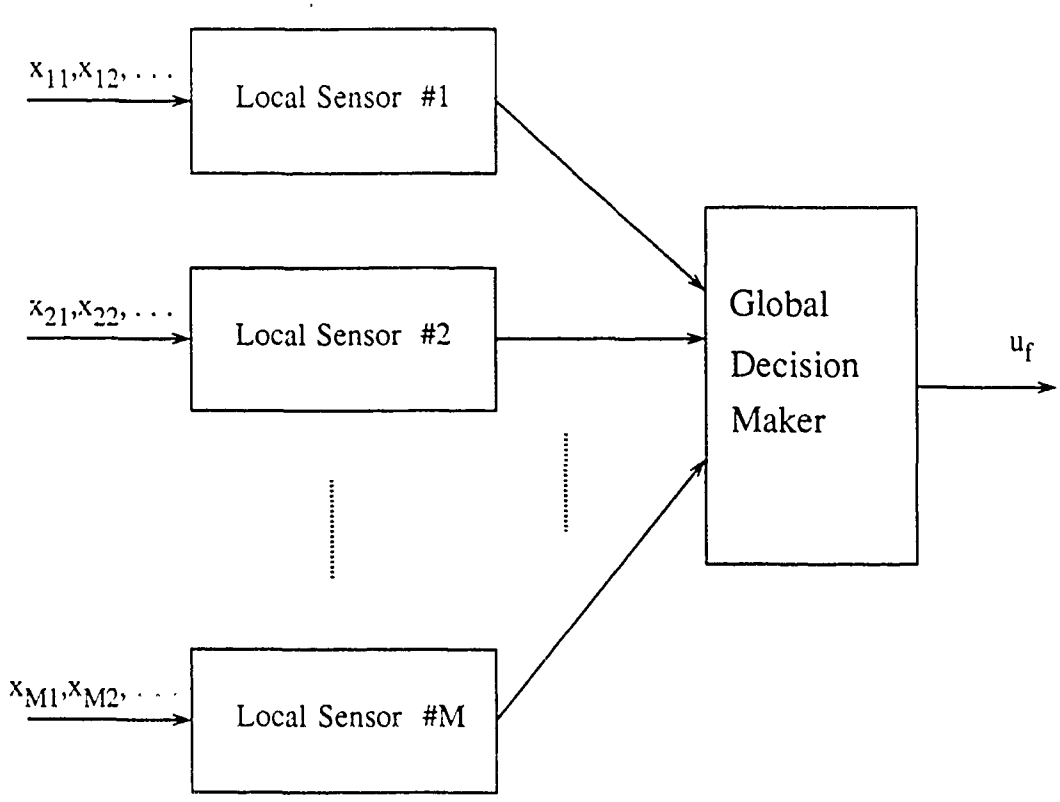

Fig. 3.1: A distributed system consisting of M local sensors and a global decision maker.

quantization is studied in some detail, and it is shown that a globally optimum design of the quantizers is obtained by optimizing the individual local quantizers independently. Also, it is shown that an optimal local quantizer is based upon the quantization of the local likelihood ratio (LR). In Section 3.4, we show that ASN_j is a monotonically decreasing function of the number of bits used for local quantization. In Section 3.5, we consider a distributed system consisting of two identical local sensors. The SPRT at the GDM is formulated as a random walk [28,29] on a finite state Markov chain and the effect of transmission errors is investigated. Two cases are studied. In the first case, it is shown that when the transmission errors are neglected, the result is an increase in the error probabilities over their specified values. It is shown in the second case that when the transmission errors are incorporated in the design, the specified error probabilities are satisfied at the expense of an increased ASN_j, j=0,1. In Section 3.6, we present some numerical results for the system described in Section 3.5. Finally, the results obtained in this chapter are discussed in Section 3.7.

3.2. A Centralized SPRT for Multisensor Data

Consider the multisensor distributed system shown in Fig. 3.1 which consists of M local sensors and a global decision maker (GDM). The problem under consideration is that of testing a simple hypothesis H_0 versus a simple alternative H_1. Let $f(x_i,\theta_i)$ denote the probability density function (pdf) of the random variable X_i, i=1,2,...,M, which represents the observation at the ith local sensor. Let H_0 be the hypothesis that $\theta_i = \theta_{i0}$, and H_1 the hypothesis that $\theta_i = \theta_{i1}$. Therefore, the pdf of X_i is given by $f(x_i,\theta_{ij})$ wher H_j, j=0,1, is true. The successive

observations on X_i are denoted by $x_{i1}, x_{i2}, \ldots, x_{in}$ $(n \geq 1)$ and they are assumed to be iid. The observations are also assumed to be independent from one sensor to the other, i.e., x_{im} is independent of x_{ik} $(m \neq k)$, and $x_{\ell m}(\ell \neq i)$. Each local sensor is assumed to process its own observation(s) and transmit the analog value of the local test statistic to the GDM. At any observation time $n(n \geq 1)$, the GDM computes the central likelihood ratio (CLR) function Λ_{cn} as follows

$$\Lambda_{cn} \triangleq \prod_{i=1}^{M} \left[\prod_{k=1}^{n} \frac{f(x_{ik}, \theta_{i1})}{f(x_{ik}, \theta_{i0})} \right] \tag{3.1}$$

where $f(x_{ik}, \theta_{ij}) = f(x_i, \theta_{ij})$. Equation (3.1) can be written as

$$\Lambda_{cn} = \prod_{i=1}^{M} \left[\prod_{k=1}^{n-1} \frac{f(x_{ik}, \theta_{i1})}{f(x_{ik}, \theta_{i0})} \cdot \frac{f(x_{in}, \theta_{i1})}{f(x_{in}, \theta_{i0})} \right] \tag{3.2}$$

which is recognized as the product of the CLR at the (n-1)th stage $\Lambda_{c(n-1)}$ and the nth increment Λ_c^n, where Λ_c^n is given by

$$\Lambda_c^n = \prod_{i=1}^{M} \frac{f(x_{in}, \theta_{i1})}{f(x_{in}, \theta_{i0})} \tag{3.3}$$

Hence

$$\Lambda_{cn} = \Lambda_{c(n-1)} \cdot \Lambda_c^n$$

The centralized SPRT at stage n compares Λ_{cn} with two thresholds [1] A and B as follows

$$\Lambda_{cn} \begin{cases} \geq A \text{ , stop and decide } H_1 \\ \leq B \text{ , stop and decide } H_0 \\ \text{otherwise , continue} \end{cases} \tag{3.4}$$

where $A \cong \frac{1-\beta}{\alpha}$, and $B \cong \frac{\beta}{1-\alpha}$ are the test thresholds. From the monoto-

nicity of the logarithm function and the fact that Λ_{cn} as well as A and B are always positive, a more convenient test is derived by taking the logarithm of both sides of (3.4) to obtain

$$\log \Lambda_{cn} \begin{cases} \geq \log A \text{ , stop and decide } H_1 \\ \leq \log B \text{ , stop and decide } H_0 \\ \text{otherwise, continue} \end{cases} \tag{3.5}$$

Taking the logarithm of the CLR in (3.2), we obtain

$$\log \Lambda_{cn} = \sum_{i=1}^{M} \left[\sum_{k=1}^{n-1} \log \frac{f(x_{ik}, \theta_{i1})}{f(x_{ik}, \theta_{i0})} + \log \frac{f(x_{in}, \theta_{i1})}{f(x_{in}, \theta_{i0})} \right] \tag{3.6}$$

Let $Z_{ik} \triangleq \log \dfrac{f(x_{ik}, \theta_{i1})}{f(x_{ik}, \theta_{i0})}$, $i=1,2,\ldots,M$; $k \geq 1$, to obtain

$$\log \Lambda_{cn} = \sum_{i=1}^{M} \left[\sum_{k=1}^{n-1} Z_{ik} + Z_{in} \right] = \sum_{i=1}^{M} \sum_{k=1}^{n-1} Z_{ik} + \sum_{i=1}^{M} Z_{in} \tag{3.7}$$

Equation (3.7) shows that the local test statistic, Z_{ik}, $k \geq 1$, is simply the logarithm of the local LR function. In other words, at any stage of the test each local sensor computes the logarithm of its LR function based on the current observation and transmits the result to the GDM The GDM employs the received local test statistics to update its test statistic $\log \Lambda_{cn}$ according to (3.7). Because the successive observations at the ith sensor, $i=1,2,\ldots,M$, are iid random variables, it follows that Z_{ik}, $k \geq 1$, is also a sequence of iid random variables.

Assuming no excess over the test thresholds [1], it follows that when H_j is true and the test terminates, we have

$$\log \Lambda_{cn} = \begin{cases} \log A \text{ , with probability } p_j \\ \log B \text{ , with probability } (1-p_j) \end{cases} \tag{3.8}$$

where $p_0 \cong \alpha$ and $p_1 \cong (1-\beta)$. Equating the expected values of both sides of (3.8), we obtain

$$E[\log \Lambda_{cn}] \triangleq E\Big[\sum_{k=1}^{n} \sum_{i=1}^{M} Z_{ik} \Big] = p_j \log A + (1-p_j) \log B \tag{3.9}$$

However, $\sum_{k=1}^{n} \sum_{i=1}^{M} Z_{ik}$ is a random sum of iid random variables. Therefore, its expected value is known [29] to be

$$E\Big[\sum_{k=1}^{n} \sum_{i=1}^{M} Z_{ik} \Big] = E[n] \cdot E\Big[\sum_{i=1}^{M} Z_i \Big] = E[n] \cdot \sum_{i=1}^{M} E[Z_i] \tag{3.10}$$

Therefore, $ASN_j \triangleq E[n/H_j]$ is given by

$$ASN_j \cong \frac{p_j \log A + (1-p_j) \log B}{\sum_{i=1}^{M} E[Z_i/H_j]} \quad , \; j=0,1 \tag{3.11}$$

Since ASN_j, $j=0,1$, is a positive real number, it follows that the denominator and numerator in (3.11) must have the same sign. Moreover, it is clear that if the ith sensor is used alone, then the denominator in (3.11) is $E[Z_i/H_j]$, $i=1,2,\ldots,M$. Therefore, $E[Z_i/H_j]$ must have the same sign as the numerator of (3.11) provided that the ith sensor observations are sufficient for detection, i.e., the ith sensor is reliable for detection. The reciprocal of ASN_j in (3.11) is given by

$$(ASN_j)^{-1} = \frac{\sum_{i=1}^{M} E[Z_i/H_j]}{p_j \log A + (1-p_j) \log B} \quad , \quad j=0,1 \tag{3.12}$$

Let $g_{ij} = E[Z_i/H_j]/\{p_j \log A + (1-p_j) \log B\}$. Therefore

$$(ASN_j)^{-1} \cong \sum_{i=1}^{M} g_{ij} \tag{3.13}$$

For reliable local sensors, it is clear that $g_{ij} > 0$, i=1,2,...,M, and j=0,1. Therefore, it follows that

$$\sum_{i=1}^{M} g_{ij} > \sum_{i=1}^{M-1} g_{ij} > \ldots > \sum_{i=1}^{2} g_{ij} > g_{1j} \tag{3.14}$$

which shows that $(ASN_j)^{-1}$, the inverse of ASN_j is a monotonically increasing function of the number of local sensors. Consequently, it follows that ASN_j is a monotonically decreasing function of the number of local sensors used.

The centralized SPRT considered above is optimal in the sense that no quantization was employed. However, its implementation requires an error free transmission of analog observations or their analog sufficient statistics which is practically impossible to achieve. Therefore, we focus our attention in the next section on the case when each local sensor quantizes each of its individual observations into a binary-valued variable prior to transmission.

3.3. The Multisensor Centralized SPRT with Quantized Data

In the previous section, it was assumed that each local sensor is capable of computing the likelihood ratio function at its location and transmitting the exact computed value to the central processor exactly. In this section, we assume that each local sensor quantizes its individual observations into two distinct levels (one bit quantizer), namely 0 and 1. The quantized local observations are communicated to the GDM where a

centralized SPRT is performed. The binary quantization of the local observations greatly reduces the communication channel bandwidth required. In addition, it simplifies the implementation of the SPRT at the GDM.

Consider the system shown in Fig. 3.2, which consists of M local sensors followed by M local quantizers. The analog observations are as defined earlier in Section 3.2. The ith quantizer Q_i maps the successive observations of the ith local sensor into a sequence of 0's and 1's. Therefore, the hypothesis testing problem is given by

$$\begin{aligned} H_0 &: y_{ik} \equiv B(p_{i0}) \\ H_1 &: y_{ik} \equiv B(p_{i1}) \end{aligned} \qquad i=1,2,\ldots,, \text{ and } k \geq 1 \qquad (3.15)$$

where y_{ik} is the random variable representing the output of Q_i corresponding to the kth observation. $B(p_{ij})$ is a Bernoulli random variable with probability of success, i.e., $y_{ik} = 1$, equal to p_{ij} when H_j is true. The probability p_{ij} is given by

$$p_{ij} = \int_{R_{1i}} f(x_i, \theta_{ij})\, dx_i \qquad (3.16)$$

where R_{1i} is the region of the observation space in which y_i is assigned the value 1. Clearly, the design of the ith quantizer Q_i involves a unique determination of the region R_{1i}.

The centralized SPRT is performed at the central processor in a manner identical to that described in Section 3.2. The only difference is that $Z_{ik} = Z_i$ is now a discrete two-valued random variable whose distribution when H_j is true is given by

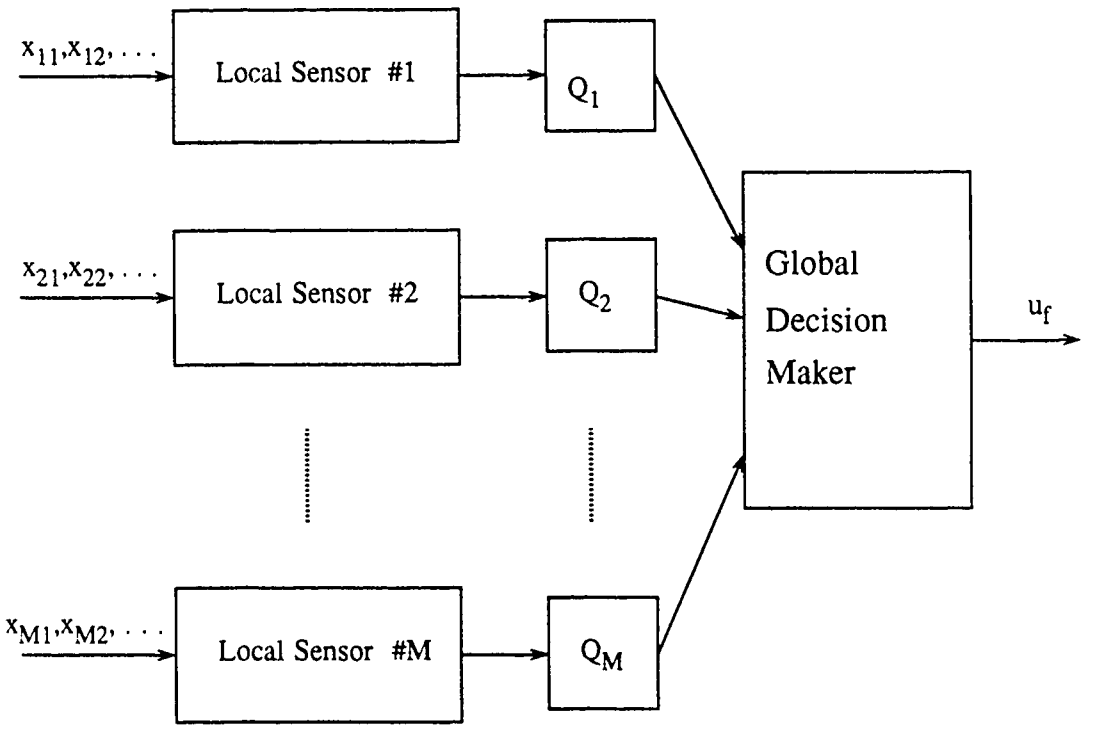

Fig. 3.2: A distributed system consisting of M local sensors followed by M local quantizers and a global decision maker.

$$Z_{ik} = \begin{cases} \log[(p_{i1})/(p_{i0})] & \text{, with probability } p_{ij} \\ \log[(1-p_{i1})/(1-p_{i0})] & \text{, with probability } 1-p_{ij} \end{cases} \tag{3.17}$$

From (3.17), we observe that the operation of the central processor can be described in terms of a finite number of possible steps ($\sum_{i=1}^{M} Z_i$), i.e., there are 2^M possible steps corresponding to the 2^M possible combinations of the local observations. The values of these possible steps are known and can be stored in a read-only-memory. To derive expressions for the ASN's we follow the same procedure as used in deriving equation (3.11). For the quantized observations case, (3.11) can be written as follows

$$ASN_j \cong \frac{p_j \log\frac{1-\beta}{\alpha} + (1-p_j) \log\frac{\beta}{1-\alpha}}{\sum_{i=1}^{M} \left[p_{ij} \log[(p_{i1})/(p_{i0})] + (1-p_{ij}) \log[(1-p_{i1})/(1-p_{i0})] \right]} \tag{3.18}$$

It should be emphasized that (3.18) is derived under the assumption of no excess over the test thresholds for given quantizers.

Next, we address the issue of optimal design of the quantizers Q_1, $Q_2, \ldots, Q_M$ and show that the optimal quantization is obtained by optimizing the local quantizers independently. To this end, let D_j denote the denominator of (3.18) when H_j is true, and let D_{ij} denote the contribution of the ith quantizer to D_j. Therefore, we can write

$$D_j = \sum_{i=1}^{M} D_{ij} \tag{3.19}$$

Note that the numerator of (3.18) is fixed. Therefore, in order to

minimize ASN_j, D_j must be maximized. However, maximization of the sum of any number of functions is equivalent to the maximization of the summands or individual terms. In other words, we can write

$$\max_{Q_1,Q_2,\ldots,Q_M} [D_j] = \sum_{i=1}^{M} \max_{Q_i} [D_{ij}] \tag{3.20}$$

Let R_{1i}^* be the region R_{1i} that corresponds to the optimal quantizer Q_i^*, and let p_{ij}^* be as defined in (3.16) with R_{1i} being replaced by R_{1i}^*. Suppose that H_1 is the true hypothesis, then with the optimal choice of Q_i^*'s the maximum value of D_1, D_1^*, is given by

$$D_1^* \triangleq \max_{Q_1,Q_2,\ldots,Q_M} [D_j] = \sum_{i=1}^{M} \left[p_{i1}^* \log \frac{p_{i1}^*}{p_{i0}^*} + (1 - p_{i1}^*) \log \frac{p_{i1}^*}{p_{i0}^*}\right] \tag{3.21}$$

The expression for the maximum value of D_0, D_0^*, is obtained from (3.18) and is given by

$$D_0^* = \sum_{i=1}^{M} \left[p_{i0}^* \log \frac{p_{i1}^*}{p_{i0}^*} + (1 - p_{i0}^*) \log \frac{p_{i1}^*}{p_{i0}^*}\right] \tag{3.22}$$

Observe that the optimization of D_1 does not, in general, lead to an optimal D_0 and vice versa. In other words, we can optimize either D_1 or D_0 or a weighted sum of them, i.e., we can optimize either ASN_1 or ASN_0 or the weighted sum given by

$$ASN = \gamma\ ASN_1 + (1-\gamma)\ ASN_0 \quad , \quad \gamma \in [0,1] \tag{3.23}$$

Next, we show that an optimal local quantizer is a likelihood ratio quantizer. We focus our attention on the ith quantizer Q_i whose contribution to the denominator of (3.18) is given by

$$D_{ij} = p_{ij} \log \frac{p_{i1}}{p_{i0}} + (1-p_{i1}) \log \frac{1-p_{i1}}{1-p_{i0}} \qquad (3.24)$$

When H_1 is the true hypothesis, we can write (3.24) as follows

$$D_{i1} = p_{i1} \log \frac{p_{i1}}{p_{i0}} + (1-p_{i1}) \log \frac{1-p_{i1}}{1-p_{i0}} \qquad (3.25)$$

We treat p_{i0} as a constant, and differentiate (3.25) with respect to p_{i1} to obtain

$$\frac{dD_{i1}}{dp_{i1}} = \left\{\log \frac{p_{i1}}{p_{i0}} + 1 - \log \frac{1-p_{i1}}{1-p_{i0}} - 1\right\} = \log \frac{p_{i1}}{p_{i0}} + \log \frac{1-p_{i0}}{1-p_{i1}} \qquad (3.26)$$

For $p_{i1} > p_{i0}$, p_{i1} and $p_{i0} \in (0,1)$, it is clear that $p_{i1}/p_{i0} > 1$ and $(1-p_{i0})/(1-p_{i1}) > 1$. Therefore, we conclude that

$$\frac{dD_{i1}}{dp_{i1}} = \log \frac{p_{i1}(1-p_{i0})}{p_{i0}(1-p_{i1})} > 0 \text{ , for } p_{i1} \text{ and } p_{i0} \in (0,1) \qquad (3.27)$$

Similarly, when H_0 is the true hypothesis, we can write (3.24) as follows

$$D_{i0} = p_{i0} \log \frac{p_{i1}}{p_{i0}} + (1-p_{i0}) \log \frac{1-p_{i1}}{1-p_{i0}} \qquad (2.28)$$

Differentiating (3.28) with respect to p_{i1}, we obtain

$$\frac{dD_{i0}}{dp_{i1}} = \left\{p_{i0} \cdot \frac{p_{i0}}{p_{i1}} \cdot \frac{1}{p_{i0}} + (1-p_{i0}) \frac{1-p_{i1}}{1-p_{i0}} \cdot \frac{(-1)}{1-p_{i0}}\right\} = \frac{p_{i0} - p_{i1}}{p_{i1}(1-p_{i1})} \qquad (3.29)$$

Equation (3.29) shows clearly that $\frac{dD_{i0}}{dp_{i1}}$ is negative for $p_{i1} > p_{i0}$ and nontrivial quantizer, i.e., p_{i1} and $p_{i0} \in (0,1)$. The numerator of the

right hand side in (3.18) is determined by the values of error probabilitie α and β. For small values of α and β, it is clear that when H_1 (H_0) is true, the numerator of (3.18) is positive (negative). Therefore, differentiating (3.18) with respect to p_{i1} and observing that the numerator which is denoted by $N(\alpha,\beta)$ is not a function of p_{ij}, we obtain

$$\frac{d(ASN_j)}{dp_{ij}} = \frac{-N(\alpha,\beta)}{D_{ij}^2} \cdot \frac{dD_{i0}}{dp_{i1}} < 0, \text{ for } j=0,1 \qquad (3.30)$$

From (3.30), we conclude that for any fixed p_{i0} and nontrivial quantizer ($p_{i1} > p_{i0}$ and p_{i1}, $p_{i0} \in (0,1)$), ASN_j is a monotonically decreasing function of p_{i1}. Consequently, the optimum quantizer is the one which maximizes p_{i1} corresponding to a fixed p_{i0}. This quantizer is obviously the one that employs a Neyman-Pearson criterion (likelihood ratio quantizer) with a probability of false alarm equal to p_{i0} ($0 < p_{i0} < 1$).

3.4. Performance Enhancement by Means of Multi-level Quantization

The quantizers considered so far were binary valued. Since the quantizer does not discriminate between the observations in any single region of its two regions, the quantization leads to a performance degradation as reflected by an increase of both ASN_1 and ASN_0. However, this performance loss can be reduced by further partitioning of the observation space, i.e., by allowing more quantization levels. In this section, we show that ASN_j, $j=0,1$, is a monotonically decreasing function of the number of quantization levels when the quantization is based on the likelihood ratio function.

To prove our claim, it suffices to show that given any region R of the observation space, where R is specified in terms of the two thresholds t_1

and t_2 ($t_1 < t_2$) such that for any point $x \notin R$, we have $LR(t_1) < LR(x) < LR(t_2)$ as shown in Fig. 3.3. Then as a result of partitioning R into two regions, the denominator in the expression of ASN_j in equation (3.11) increases when H_1 is true (j=1) and decreases when H_0 is true (j=0). Let p_{rj}, j=0,1, be the probability of an observation falling in R when H_j is true with $0 < p_{rj} < 1$. Since the local quantizers were designed independently as shown in the previous section, it follows that the region R can be specified at any of the local observation spaces. Therefore, in what follows we simply drop the index i of the local quantizers for convenience. Let the region R be partitioned into two mutually exclusive subregions R^1 and R^2 such that

$$p_{r1} = p_{r1}^1 + p_{r1}^2$$
$$p_{r0} = p_{r0}^1 + p_{r0}^2 \tag{3.31}$$

where p_{rj}^i is the probability of an observation falling in the subregion R^i, i = 1,2, when H_j, j=0,1, is the true hypothesis. We assume without loss of generality that the likelihood ratio at any point in R^2 is higher or at least equal to its maximum in R^1. Therefore, it follows that

$$\frac{p_{r1}^2}{p_{r0}^2} > \frac{p_{r1}}{p_{r0}} > \frac{p_{r1}^1}{p_{r0}^1} \tag{3.32}$$

Upon writing the expression of ASN_j as in (3.18) after deleting the quantizer index i and looking at the contribution of the region R to the denominator, we observe that

$$D_j = \varepsilon + p_{rj} \log \frac{p_{r1}}{p_{r0}} \tag{3.33}$$

where ε is the contribution of all the regions in the observation space other than R. Similarly, the contribution of the subregions R^1 and R^2

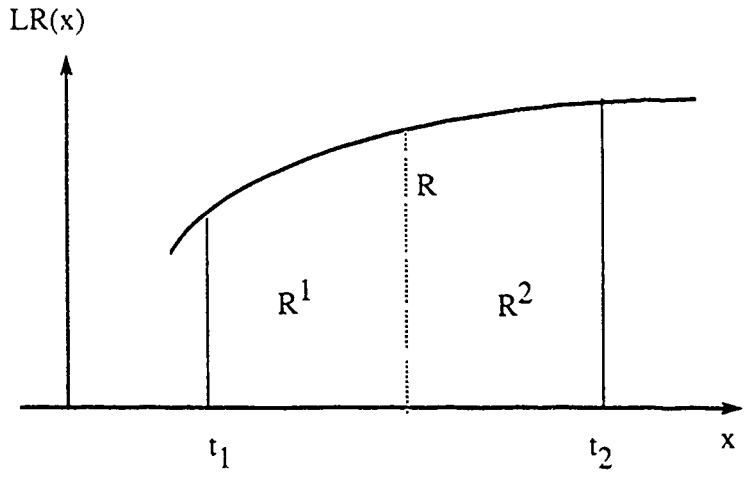

Fig. 3.3: A region R of the observation space such that $LR(t_1) < LR(x) < LR(t_2)$

can be incorporated in D_j to obtain

$$D_j = \varepsilon + p^1_{rj} \log \frac{p^1_{r1}}{p^1_{r0}} + p^2_{rj} \log \frac{p^2_{r1}}{p^2_{r0}} \tag{3.34}$$

Our objective is to show that ASN_j, j=0,1, is a monotonically decreasing function of the number of quantization levels. Therefore, comparing (3.33) with (3.34), we conclude that in order to prove the claimed monotonicity it suffices to show that

$$p^1_{r1} \log \frac{p^1_{r1}}{p^1_{r0}} + p^2_{r1} \log \frac{p^2_{r1}}{p^2_{r0}} > p_{r1} \log \frac{p_{r1}}{p_{r0}} \tag{3.35}$$

and

$$p^1_{r0} \log \frac{p^1_{r1}}{p^1_{r0}} + p^2_{r0} \log \frac{p^2_{r1}}{p^2_{r0}} < p_{r0} \log \frac{p_{r1}}{p_{r0}} \tag{3.36}$$

To prove the inequality (3.35) which corresponds to the hypothesis H_1 being true, we assume that $p^2_{r1} = c_1 p_{r1}$ and $p^2_{r0} = c_o p_{r0}$, where $c_1 > c_o$ as a consequence of (3.32). The left hand side of (3.35) can easily be written as

$$\text{L.H.S.} = (1-c_1)\, p_{r1} \log \frac{(1-c_1)p_{r1}}{(1-c_o)p_{r0}} + c_1 p_{r1} \log \frac{c_1 p_{r1}}{c_o p_{r0}} \tag{3.37}$$

Simplifying (3.37), we obtain

$$\text{L.H.S.} = (1-c_1)\, p_{r1}\left[\log \frac{p_{r1}}{p_{r0}} + \log \frac{1-c_1}{1-c_o}\right] + c_1 p_{r1}\left[\log \frac{p_{r1}}{p_{r0}} + \log \frac{c_1}{c_o}\right]$$

$$= p_{r1} \log \frac{p_{r1}}{p_{r0}} + p_{r1}\left[(1-c_1) \log \frac{1-c_1}{1-c_o} + c_1 \log \frac{c_1}{c_o}\right] \tag{3.38}$$

The first term on the R.H.S. of (3.38) is equal to the R.H.S. of (3.35). Therefore, the proof of (3.35) is complete once the second term on the R.H.S. of (3.38) is shown to be positive. We observe that $p_{r1} > 0$, and use the inequality $\log x < x-1$, $x \neq 1$, to obtain

$$(1-c_1)\log\frac{1-c_1}{1-c_o} = -(1-c_1)\log\frac{1-c_o}{1-c_1} > -(1-c_1)\left[\frac{1-c_o}{1-c_1}-1\right] = -(c_1-c_o) \quad (3.39)$$

and

$$c_1 \log\frac{c_1}{c_o} = -c_1 \log\frac{c_o}{c_1} > -c_1\left(\frac{c_o}{c_1}-1\right) = c_1-c_o \quad (3.40)$$

Combining (3.39) and (3.40), we obtain

$$(1-c_1)\log\frac{1-c_1}{1-c_o} + c_1 \log\frac{c_1}{c_o} > 0 \quad (3.41)$$

and, therefore, ASN_1 decreases monotonically by increasing the number of quantization levels. Similarly, when H_0 is the true hypothesis, the inequality (3.36) can be simplified yielding

$$\text{L.H.S.} = p_{r0} \log\frac{p_{r1}}{p_{r0}} + p_{r0}\left[(1-c_o) \log\frac{1-c_1}{1-c_o} + c_o \log\frac{c_1}{c_o}\right] \quad (3.42)$$

The first term on the R.H.S. of (3.42) is equal to the R.H.S. of (3.36) while the terms $(1-c_o) \log\frac{1-c_1}{1-c_o}$ and $c_o \log\frac{c_1}{c_o}$ can be simplified as follows

$$(1-c_o) \log\frac{1-c_1}{1-c_o} < (1-c_o)\left[\frac{1-c_1}{1-c_o}-1\right] = -(c_1-c_o) \quad (3.43a)$$

$$c_o \log\frac{c_1}{c_o} < c_o\left[\frac{c_1}{c_o}-1\right] = c_1-c_o \quad (3.43b)$$

From (3.43), it follows that

$$p_{r0}\left[(1-c_o)\ \log\frac{1-c_1}{1-c_o} + c_o\ \log\frac{c_1}{c_o}\right] < 0 \qquad (3.44)$$

and, therefore, ASN_0 decreases monotonically by increasing the number of quantization levels and the proof is complete.

3.5. The Effect of Transmission Errors

In Section 3.3, we have considered the quantization of each of the local observations into two levels, namely 1 and 0. There it was assumed that the quantized local observations are communicated without errors to the global decision maker where a centralized SPRT is performed. However, the assumption of error free transmission is not realistic, especially at high transmission rates. Therefore, we devote this section to a study of the effect of transmission errors on the error probabilities α and β. For the sake of clarity, we consider the case of two identical local detectors and the following observation model

$$\begin{aligned} H_0 &: y_{ik} \equiv B(\tilde{p}_0) \quad i = 1,2;\ k \geq 1 \\ H_1 &: y_{ik} \equiv B(\tilde{p}_1) \qquad \tilde{p}_1 = 1-\tilde{p}_0 > 0.5 \end{aligned} \qquad (3.45)$$

where $B(\bullet)$ is as defined in (3.15). The choice $\tilde{p}_1 = 1-\tilde{p}_0$ is intended for facilitating the formulation of the SPRT at the central detector as a random walk of equal steps [28,29]. The communication channel is assumed to be a binary symmetric channel in which the error rate is p_e as shown in Fig. 3.4.

The quantized observations encounter errors as they pass through the noisy channel. At the central detector, the observation model in (3.45) is still valid but with different values of parameters, namely $\tilde{p}_0$ and $\tilde{p}_1$

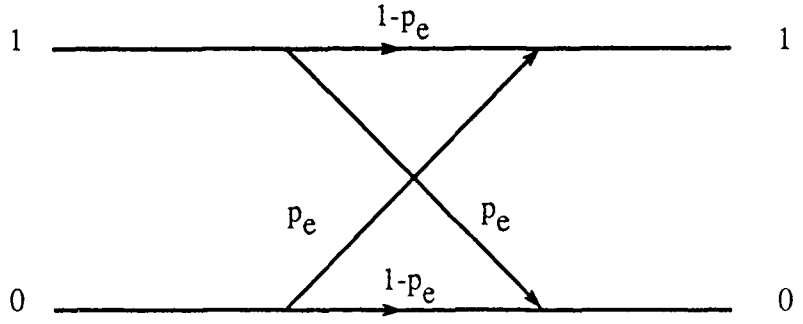

Fig. 3.4: A binary symmetric channel with error rate p_e.

are replaced by p_{0c} and p_{1c} respectively. When H_j, j=0,1, is the true hypothesis, we can solve for p_{jc} as follows

$$p_{jc} = \tilde{p}_j(1-p_e) + (1-\tilde{p}_j)\, p_e = \tilde{p}_j(1-2p_e) + p_e \qquad (3.46)$$

Suppose that H_1 is the true hypothesis, we obtain from (3.46) that $p_{1c} = \tilde{p}_1(1-p_e) + \tilde{p}_1 p_e$. Similarly, when H_0 is true, we obtain $p_{0c}=\tilde{p}_0(1-p_e) + \tilde{p}_1 p_e$. We note that

$$1 - p_{0c} = 1-\tilde{p}_0 + \tilde{p}_0 p_e - \tilde{p}_1 p_e = \tilde{p}_1(1-p_e) + \tilde{p}_0 p_e = p_{1c} \qquad (3.47)$$

From (3.47), we observe that the observation model at the central detector is symmetric and is given by

$$\begin{aligned} H_0 &: \omega_{ik} \equiv B(p_{0c}) \qquad i = 1,2,\ k \geq 1 \\ H_1 &: \omega_{ik} \equiv B(p_{1c}) \qquad p_{1c} = 1-p_{0c} > 0.5 \end{aligned} \qquad (3.48)$$

where ω_{ik} is the kth observation of the ith local detector as seen at the front end of the global processor.

The central processor observes the incoming quantized observations and performs the SPRT. Let $\hat{p}_j$ be the probability of a local observation arriving as 1 at the central detector when H_j is the true hypothesis. Therefore, the observation model of the central detector is as follows:

$$H_j : y_{ck} = \begin{cases} 11\ , \text{ with prob. } \hat{p}_j^2 \\ 10 \text{ or } 01, \text{ with prob. } 2\hat{p}_j(1-\hat{p}_j)\ , \ k \geq 1 \\ 00, \text{ with prob. } (1-\hat{p}_j)^2 \end{cases} \qquad (3.49)$$

where y_{ck} is the kth observation of the central detector. From (3.49), it follows that the logarithm of the likelihood ratio function is a discrete random variable Z_{ck} whose probability distribution function under the hypothesis H_j is given by

$$H_j : Z_{ck} = \begin{cases} 2 \log (\hat{p}_1/\hat{p}_0), \text{ with prob. } \hat{p}_j^2 \triangleq P_j \\ 0 \quad , \text{ with prob. } 2\hat{p}_j(1-\hat{p}_j) \triangleq S_j \\ -2 \log (\hat{p}_1/\hat{p}_0) \ , \text{ with } (1-\hat{p}_j)^2 \triangleq Q_j \end{cases} \qquad j=0,1;\ k \geq 1 \qquad (3.50)$$

In (3.50), the positive step ($2 \log (\hat{p}_1/\hat{p}_0)$) is equal to the magnitude of the negative step ($-2 \log (\hat{p}_0/\hat{p}_1)$). Therefore, the central SPRT can be easily formulated as a random walk on a finite number of states (Markov chain) as shown in Fig. 3.5. The states labeled 0 and N are absorbing states with decisions in favor of H_0 and H_1 respectively. As long as the process is in one of the transient states labeled 1,2,...,N-1, the central test must be continued until for the first time the process reaches one of the absorbing states. For the Markov chain in Fig. 3.5, it is well known [29] that if the process starts in the transient state number Z, then the probability of absorption in the state 0 is given by

$$u_j = \frac{(Q_j/P_j)^Z}{1 - (Q_j/P_j)^N} \qquad (3.51)$$

and the average sample number ASN_j is given by

$$ASN_j = \frac{1}{P_j - Q_j} \left[(N-Z) - \frac{N(Q_j/P_j)^Z}{1 - (Q_j/P_j)^N} \right] \qquad (3.52)$$

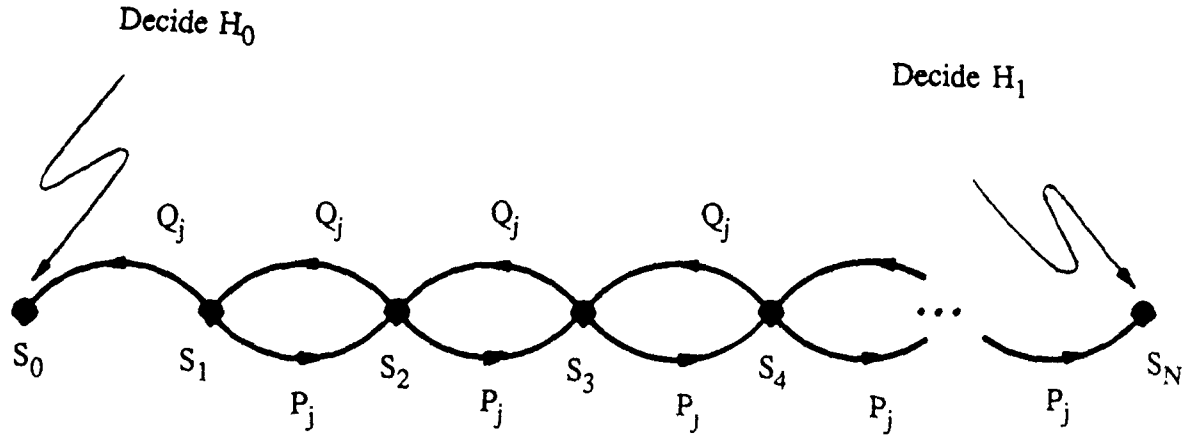

Fig. 3.5: A Markov chain of (N+1) states.

From (3.52), it follows that for a given α and β, we must have Z and N chosen such that $u_0 \geq 1-\alpha$ and $u_1 \leq \beta$. To this end let us distinguish between the following two cases:

Case 1: In this case, the channel is assumed to be error free while in fact it is not so, i.e., the design is based on the parameters $\hat{p}_0=\tilde{p}_0$ and $\hat{p}_1=\tilde{p}_1$ in (3.45) while the actual values of the parameters are $\tilde{p}_0(1-2p_e) + p_e$ and $\tilde{p}_1(1-2p_e) + p_e$ respectively. Therefore, the result will be higher error probabilities than specified. This fact can be shown by observing that

$$\frac{p_{0c}}{p_{1c}} = \frac{\tilde{p}_0(1-2p_e) + p_e}{\tilde{p}_1(1-2p_e) + p_e} = \frac{\tilde{p}_0}{\tilde{p}_1} \cdot \left[1 + \frac{p_e}{(1-2p_e)\tilde{p}_0}\right] / \left[1 + \frac{p_e}{(1-2p_e)\tilde{p}_1}\right] \quad (3.53)$$

Since $\tilde{p}_0 < \tilde{p}_1$, it follows that $\left\{\frac{p_e}{(1-2p_e)\tilde{p}_0} / \frac{p_e}{(1-2p_e)\tilde{p}_1}\right\} = \frac{\tilde{p}_1}{\tilde{p}_0} > 1$ and, therefore, we have

$$\frac{p_{0c}}{p_{1c}} > \frac{\tilde{p}_0}{\tilde{p}_1} \quad (3.54)$$

On the other hand, for sufficiently large N and Z, which is the case when the α and β are small, we can approximate the denominator in (3.51) by 1. Consequently, we can write

$$u_1 \triangleq \beta_c \cong \left(\frac{Q_1}{P_1}\right)^Z = \left(\frac{P_0}{P_1}\right)^Z = \left(\frac{p_{0c}}{p_{1c}}\right)^{2Z} \quad (3.55)$$

Similarly, it can be easily shown that

$$\alpha_c \triangleq 1-u_0 \cong \left(\frac{p_{0c}}{p_{1c}}\right)^{2(N-Z)} \quad (3.56)$$

where α_c and β_c are the actual error probabilities at t e central level when the channel is not error free as assumed. Equations (3.55) and (3.56) show clearly that $\alpha_c > \alpha$ and $\beta_c > \beta$ and hence the errors over

the channel will lead to an increase in the error probabilities. Similarly, it can be shown that ASN_j as given by (3.52) is dominated by $P_j - Q_j = (\tilde{p}_1 - \tilde{p}_0)(1-2p_e) < (\tilde{p}_1 - \tilde{p}_0)$, which will lead to an increase in ASN_j.

Case 2: In this case, the channel is assumed to be completely known, therefore, the central detector has a completely specified observation model under both hypotheses as given by (3.49) or (3.50) with $\hat{p}_j = p_{jc}$. The exact knowledge of the observation model at the front end of the central processor enables us to redesign the central detector taking the transmission errors into account. Consider Fig. 3.5 once again and, let Z_c and N_c be the initial and final states instead of Z and N respectively. Therefore, we can write

$$\alpha \geq \frac{(p_{0c}/p_{1c})^{2(N_c - Z_c)}}{1 - (p_{0c}/p_{1c})^{2N_c}} \triangleq \alpha_a \qquad (3.57)$$

and

$$\beta \geq \frac{(p_{0c}/p_{1c})^{2Z_c}}{1 - (p_{0c}/p_{1c})^{2N_c}} \triangleq \beta_a \qquad (3.58)$$

where the inequalities (3.57) and (3.58) are obtained by restricting the actual error probabilities α_a and β_a as shown on the right hand side of (3.57) and (3.58) to be less than or equal to their specified values α and β respectively. The fact that $p_{0c}/p_{1c} > \tilde{p}_0/\tilde{p}_1$ implies that Z_c, $N_c - Z_c$, and N_c are larger than their corresponding values in Case 1. The average sample number obviously increases with increasing Z and N and, therefore, ASN_j, j=0,1, tends to increase when the channel is noisy.

3.6. Numerical Results

The numerical results presented in this section are obtained assuming two identical local detectors as described in Section 3.5. The observation model is as described by (3.42) with $\tilde{p}_1$=0.54 and $\tilde{p}_0 = 1-\tilde{p}_1 = 0.46$ and the communication channel is as depicted in Fig. 3.4. The process of performing the central SPRT is formulated as a random walk as described in Section 3.5. The error probabilities are adjusted by suitable variation of the initial state Z or the number of states N+1 or both. In Table 3.1, we present some values of Z and N along with the resulting error probabilities and average sample numbers, for the case of one and two local sensors.

Table 3.1. The error probabilities and ASN's for an error free channel when one and two local sensors are used.

		ONE SENSOR				TWO SENSORS			
α	β	Z	N	ASN_1	ASN_0	Z	N	ASN_1	ASN_0
8.15×10^{-3}	8.15×10^{-3}	30	60	368.9	368.9	15	30	184.45	184.45
4.29×10^{-3}	4.29×10^{-3}	34	68	421.3	421.3	17	34	210.68	210.68
1.64×10^{-3}	1.64×10^{-3}	40	80	498.4	498.4	20	40	249.18	249.18
2.68×10^{-6}	2.68×10^{-6}	80	160	1000	1000	40	80	500	500
1.15×10^{-8}	1.03×10^{-6}	86	200	1425	1075	43	100	712.5	537.5
8.86×10^{-10}	1.09×10^{-7}	100	230	1625	1250	50	115	812.5	625
5.4×10^{-7}	4.4×10^{-9}	120	210	1125	1500	60	105	562.5	750
7.45×10^{-7}	8.86×10^{-10}	130	218	1100	1625	65	109	550	812.5

The numerical results presented in Table 3.1 are obtained under the assumption that $p_e = 0$. However, the effect of channel errors on the performance of the central detector is demonstrated in Table 3.2 for the two cases described in Section 3.5. When $p_e = 0$, the choice $Z = N/2 = 43$ yields $\alpha = \beta = 1.03\times10^{-6}$ and $ASN_1 = ASN_0 = 537.5$. As p_e varies while Z and N are fixed, the error probabilities α and β as well as the average sample numbers ASN_1 and ASN_0 varies as shown under Case 1. On the other hand, when p_e varies while Z and N are adjusted accordingly as described in Case 2 of Section 3.5, the error probabilities are maintained while ASN_1 and ASN_0 are higher than their values in Case 1.

Table 3.2. The effect of transmission errors on the performance of the central SPRT.

	CASE 1		CASE 2		
p_e	$\alpha_c=\beta_c$	$ASN_1=ASN_0$	Z=N/2	$\alpha_a=\beta_a$	$ASN_1=ASN_0$
0	1.03×10^{-6}	537.5	43	1.03×10^{-6}	537.5
0.02	1.79×10^{-6}	559.9	45	0.065×10^{-6}	585.9
0.04	3.11×10^{-6}	584.2	47	0.955×10^{-6}	638.6
0.06	5.4×10^{-6}	610.8	49	0.994×10^{-6}	696.0
0.08	9.39×10^{-6}	639.9	52	0.833×10^{-6}	773.8
0.1	1.63×10^{-5}	671.9	54	0.973×10^{-6}	843.7
0.12	2.84×10^{-5}	707.2	57	0.938×10^{-6}	937.5
0.15	6.49×10^{-5}	767.8	62	0.917×10^{-6}	1107.1

3.7. Discussion

In this chapter, we studied a centralized SPRT based on multisensor data, and showed that the average sample numbers are monotonically decreasing functions of the number of local sensors. We also studied the issue of quantization of the local observations and showed that an optimal quantizer is obtained by using a set of independent local likelihood ratio quantizers, when the local observations are independent both spatially and temporally. Furthermore, we showed that by increasing the number of quantization levels, the average sample numbers decrease monotonically. The formulation of the central SPRT as a random walk on a finite state machine of fixed structure (Markov chain) is considered for the case of two identical local sensors. The numerical results in Table 3.1 show clearly that when both Z and N are even, the average sample numbers are exactly half of their counterparts in the case of single local sensor. When the transmission errors are ignored, the result is shown to be an increase of the error probabilities and average sample numbers over their specified values as demonstrated in Case 1 of Table 3.2. In addition, we showed that when the transmission errors are known, they can be incorporated in the design of the central SPRT to maintain the specified error probabilities. In this case, the only effect of transmission errors is to increase the average sample numbers as shown in Case 2 of Table 3.2.

CHAPTER FOUR

A SIMPLE MULTI-SENSOR SEQUENTIAL DETECTION PROCEDURE

4.1. Introduction

In this chapter, a simple multi-sensor decentralized sequential test is proposed and analyzed. In Section 4.2, we describe the proposed multi-sensor sequential test and show that it satisfies the prespecified error probabilities α and β. The ASN of the proposed test is shown to be a monotonically decreasing function of the number of sensors used. Moreover, we show that the increase in the error probabilities as a result of truncation of the multi-sensor sequential test at any stage n_T is less than its counterpart in the single sensor case, i.e. a single sensor sequential test also truncated at stage n_T. In Section 4.3, we employ the memoryless grouped data sequential detection (MLGDS) test procedure [21] at the local sensors in the multi-sensor decentralized scheme. The ASN is shown to be a monotonically decreasing function of the number of sensors used. This result shows the feasibility of using our testing scheme based on multiple sensors that employ the MLGDS test procedure at the local sensors to outperform Wald's SPRT based on a single detector. In other words, the decision time of the multi-sensor system using MLGDS local detectors decreases when more sensors are added. At some point, the performance loss due to the grouping of samples and memoryless nature of the MLGDS test is overcome and from then on the performance in terms of average decision time is better than Wald's SPRT. Our scheme compensates for the performance loss due to the grouping of samples and the memoryless

nature of the MLGDS procedure by adding more detectors. In Section 4.4, we propose and analyze a truncation scheme for the case when the sensors employ the MLGDS test procedure. Numerical results are presented in Section 4.5 for illustration. Finally, the results of this chapter are discussed in Section 4.6.

4.2. The Proposed Multi-Sensor Sequential Test

Consider the problem of testing a simple hypothesis H_0 versus a simple alternative H_1. The multi-sensor observation system is as shown in Fig. 4.1 which consists of M local detectors and a supervisor (global) detector. Let $f(x_i, \theta)$ denote the pdf of the random variable x_i, $i = 1,2,\ldots,M$, representing the observation at the ith local detector.

Let H_0 be the hypothesis that $\theta = \theta_0$, and H_1 the hypothesis that $\theta = \theta_1$. Therefore, the pdf of x_i is given by $f(x_i, \theta_j)$ when H_j is true, $j=0,1$. The successive observations on x_i are denoted by $x_{i1}, x_{i2}, \ldots, x_{in}$ $(n \geq 1)$ and they are assumed to be iid. The observations are assumed to be independent from one local detector to the other, i.e., x_{im} is independent of $x_{ik} (m \neq k)$, and $x_{\ell m} (\ell \neq i)$. The local sequential detectors SD_i, $i = 1,2,\ldots, M$, are designed to satisfy the prespecified values of the system error probabilities α and β. Each local detector performs an SPRT based on its own sequence of observations and transmits its decision to the supervisor detector as soon as a decision on one of the two hypotheses is reached. Nothing is transmitted if the decision at the local detector is to continue taking observations. The supervisor detector accepts the first incoming local decision and declares it to be the global decision. Also, it informs all the local detectors to terminate their tests when the global decision is reached. It should be mentioned that when two or more

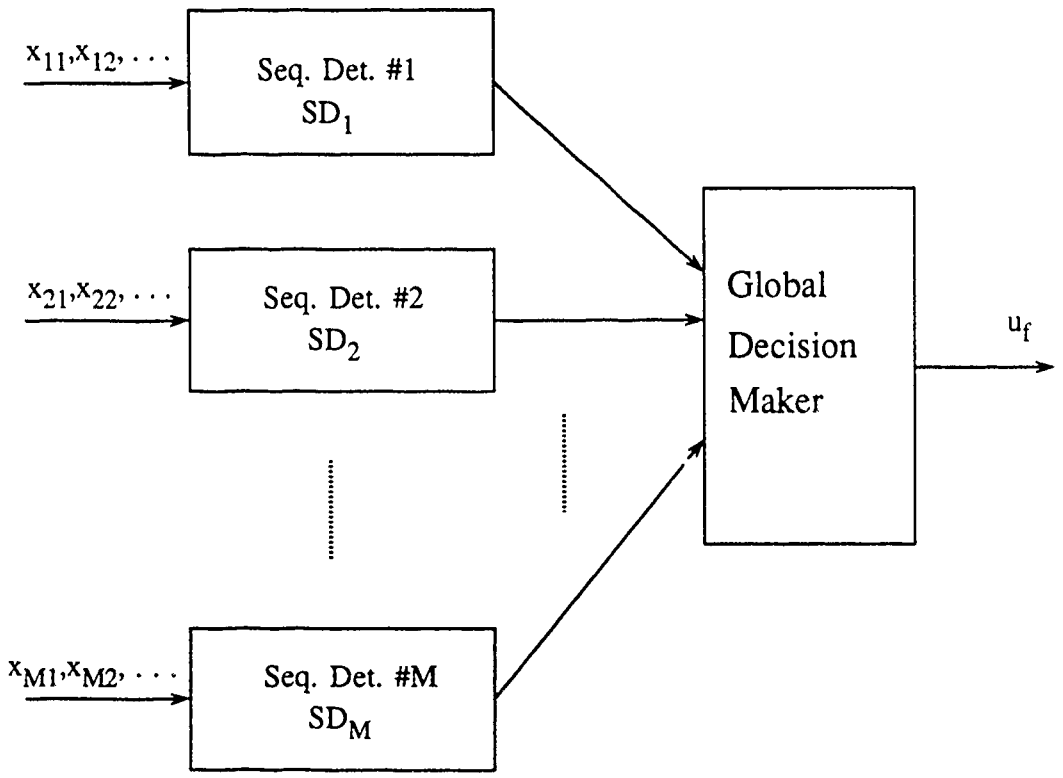

Fig. 4.1: Block diagram of the simple multi-sensor sequential detection scheme..

local decisions are available to the supervisor detector at the same time, only one of them will be accepted by the supervisor detector and this choice is arbitrary.

Next we show that the overall system satisfies the required error probabilities α and β. Let γ_{ij} denote the probability that the global decision is based on the decision of the ith local detector when H_j is the true hypothesis. Then, we have

$$\sum_{i=1}^{M} \gamma_{ij} = 1 \qquad j = 0,1 \tag{4.1}$$

Let e_{ij} denote the probability of error at the ith local detector when H_j is the true hypothesis. Expressing the global error probabilities α_g and β_g in terms of e_{ij}'s and γ_{ij}'s, we obtain

$$\alpha_g = \sum_{i=1}^{M} e_{i0}\, \gamma_{i0} \tag{4.2}$$

and

$$\beta_g = \sum_{i=1}^{M} e_{i1}\, \gamma_{i1} \tag{4.3}$$

However, the design of the local detectors to satisfy α and β implies that $e_{i0} = \alpha$ and $e_{i1} = \beta$ for $i = 1,2,\ldots,M$. Therefore, it follows from (4.2), (4.3) and (4.1) that $\alpha_g = \alpha$ and $\beta_g = \beta$ as required.

The SPRT at SD_i is based on its local observations and is performed in a manner identical to Wald [1], i.e., at stage n ($n \geq 1$), the likelihood ratio function Λ_{in} is computed:

$$\Lambda_{in} \triangleq \prod_{k=1}^{n} \frac{f(x_{ik},\, \theta_1)}{f(x_{ik},\, \theta_0)} \tag{4.4}$$

This likelihood ratio is compared with two thresholds as follows

$$\Lambda_{in} \begin{cases} \geq t_{iu}, \text{ stop and decide } H_1 \\ \leq t_{i\ell}, \text{ stop and decide } H_0 \\ \text{otherwise, continue.} \end{cases} \tag{4.5}$$

where t_{iu} and $t_{i\ell}$ are the test thresholds of SD_i, which can be approximated [1] by

$$t_{iu} \cong \frac{1 - e_{i1}}{e_{i0}} = \frac{1 - \beta}{\alpha}$$

and (4.6)

$$t_{i\ell} \cong \frac{e_{i1}}{1 - e_{i0}} = \frac{\beta}{1 - \alpha}$$

Note that the number of samples necessary for the termination of the sequential test at SD_i is a discrete random variable T_i. We denote by $f_{Ti}(n)$ and $F_{Ti}(n)$ the pdf and the cumulative distribution function (cdf) of the test duration T_i, where

$$f_{T_i}(n) = Pr\{T_i = n\}$$

$$n \geq 1 \tag{4.7}$$

$$F_{T_i}(n) = Pr\{T_i \leq n\}$$

The global test length is also a discrete random variable denoted by T_g and its pdf and cdf are denoted by $f_{Tg}(n)$ and $F_{Tg}(n)$ respectively.

Now we show that the global ASN decreases monotonically with the increase in the number of local detectors. We observe that for the M detector system in Fig. 4.1 the event that the global test terminates after the nth stage is the same as the event that none of the (SD_i)'s has terminated at stage n or before. Therefore, we can write

$$\Pr\{T_g > n\} = \Pr\{T_1 > n,\ T_2 > n, \ldots,\ T_M > n\}\ ,\quad n \geq 1 \tag{4.8}$$

Using the independence of local detectors, we may write

$$\Pr\{T_g > n\} \triangleq [1 - F_{T_g}(n)] = [1 - F_{T_1}(n)] \times [1 - F_{T_2}(n)] \times \ldots \times [1 - F_{T_M}(n)] \tag{4.9}$$

Since $[1 - F_{T_i}(n)] < 1$ (n finite), $i = 1,2,\ldots,M$, it follows from (4.9) that $\Pr\{T_g > n\}$ decreases as the number of local detectors increases. The global ASN is given by

$$ASN = \sum_{n=0}^{\infty} \Pr\{T_g > n\} \tag{4.10}$$

Since for each value of n, $\Pr\{T_g > n\}$ decreases with increasing M, it follows from (4.10) and (4.9) that ASN decreases monotonically with M. Also,

$$ASN_i = \sum_{n=0}^{\infty} \Pr\{T_i > n\} \quad i = 1,2,\ldots,M \tag{4.11}$$

Since $\Pr\{T_g > n\} < \Pr\{T_i > n\}$ for $M > 1$, the ASN is strictly less than any ASN_i, $i = 1,2,\ldots,M$. It should be emphasized that the above analysis is valid under both hypotheses.

From equation (4.9), it can be observed that the probability of termination of the test after the nth stage is small since it is the product of the corresponding probabilities of the local tests. However, it is well known [1] that for the SPRT there does not exist any definite upper bound for the number n of samples to be observed before reaching a decision. Large values of n are possible with a nonzero probability. Therefore, it is necessary from an implementation point of view to set a

definite upper bound n_T on n. This can be achieved by truncating the sequential procedure at $n=n_T$, i.e., by specifying a new decision rule for the acceptance or rejection of H_0 at the n_Tth stage if the sequential test did not terminate prior to stage n_T.

Next we show that the increase in error probabilities due to truncation of the global test at stage n_T is a decreasing function of M. Let $\alpha_T(n_T)$ and $\beta_T(n_T)$ be the global error probabilities at the truncation stage n_T. Recall that all local detectors are identical, therefore, if we consider only one of the local sequential detectors, say SD_i, $i = 1,2,\ldots,M$, we can write

$$\alpha_T(n_T) = Pr\{\Lambda_{in_T} \geq v/H_0 \text{ and } t_{i\ell} < \Lambda_{in} < t_{iu} \text{ for all } 1 \leq n \leq n_T\} \quad (4.12)$$

where v is the threshold at the truncation stage. Similarly, we can write

$$\beta_T(n_T) = Pr\{\Lambda_{in_T} < v/H_1 \text{ and } t_{i\ell} < \Lambda_{in} < t_{iu} \text{ for all } 1 \leq n \leq n_T\} \quad (4.13)$$

Let $\alpha_s(n_T)$ and $\beta_s(n_T)$ be the error probabilities satisfied by the sequential portion of the test. The optimality of Wald's SPRT [1] implies that no other sequential procedure can satisfy the same error probabilities with a smaller ASN. However for the truncated test the ASN is obviously smaller. Therefore, at least one of the error probabilities must be higher than required. In our analysis, we assume $\alpha_T(n_T) > \alpha_s(n_T)$ and $\beta_T(n_T) > \beta_s(n_T)$ following the convention in the literature [1,31]. For the truncated test, the overall error probabilities $\alpha_o(M)$ and $\beta_o(M)$ for the M detector case are given by:

$$\alpha_o(M) = \alpha_s(n_T)\ Pr\{T_g \leq n_T/H_0\} + \alpha_T(n_T)\ Pr\{T_g > n_T/H_0\} \quad (4.14)$$

$$\beta_o(M) = \beta_S(n_T)\ Pr\{T_g \le n_T/H_1\} + \beta_T(n_T)\ Pr\{T_g > n_T/H_1\} \qquad (4.15)$$

and for the single detector case, say for the ith detector, we have

$$\alpha_o(1) = \alpha_s(n_T)\ Pr\{T_i \le n_T/H_0\} + \alpha_T(n_T)\ Pr\{T_i > n_T/H_0\} \qquad (4.16)$$

$$\beta_o(1) = \beta_s(n_T)\ Pr\{T_i \le n_T/H_1\} + \beta_T(n_T)\ Pr\{T_i > n_T/H_1\} \qquad (4.17)$$

Simplifying (4.14) and (4.16), we obtain

$$\alpha_o(M) = \alpha_o(n_T) - [\alpha_T(n_T) - \alpha_s(n_T)]\ Pr\{T_g \le n_T/H_1\} \qquad (4.18)$$

and

$$\alpha_o(1) = \alpha_T(n_T) - [\alpha_T(n_T) - \alpha_s(n_T)]\ Pr\{T_g \le n_T/H_1\} \qquad (4.19)$$

respectively. Taking the difference $\alpha_o(1) - \alpha_o(M)$, we arrive at

$$\alpha_o(1) - \alpha_o(M) = [\alpha_T(n_T) - \alpha_s(n_T)] \times [Pr\{T_g \le n_T/H_0\} - Pr\{T_i \le n_T/H_0\}]$$

As shown before, the second term on the right hand side of (4.20) increases with increasing M. Therefore, $\alpha_o(M)$ is a decreasing function of M. In the same manner, we can show that $\beta_o(M)$ is also a monotonically decreasing function of M.

In the next section, we focus our attention on the analysis of the multi-sensor decentralized scheme when the local detectors use the MLGDS procedure of Lee and Thomas [21]. In this case, simple expressions can be obtained for the test length probability distribution at both the local and global levels. Also the expression for $\alpha_s(n_T)$, $\beta_s(n_T)$, $\alpha_T(n_T)$, and $\beta_T(n_T)$ are simpler to find unlike the case considered above where each local detector performs Wald's SPRT.

4.3. The Multi-Sensor Decentralized Scheme using the MLGDS Procedure

Consider the problem of testing a simple hypothesis H_0 versus a simple location alternative H_1. The observation model is as defined in

Section 4.2 except for the fact that $f(x_i, \theta_j)$ is replaced by $f(x_i-\theta_j)$. In other words, the observation model is given by

$$H_0 : x_i \sim f(x - \theta_0)$$
$$, \; i \geq 1 \qquad (4.21)$$
$$H_1 : x_i \sim f(x - \theta_1).$$

where $\theta_1 > \theta_0$ and $x_i \sim f(x-\theta_j)$ means that the ith observation x_i of any local detector has a pdf $f(x-\theta_j)$ when $H_j, j = 0,1$, is the true hypothesis. The pdf $f(\cdot)$ is assumed to be continuous and symmetric, i.e., $f(x)=f(-x)$. The multi-sensor decentralized scheme using identical MLGDS detectors can be described as follows. At the nth stage ($n \geq 1$), and at each local sensor, using the current group of N_o samples, a sufficient test statistic $T_{N_o}(\underline{x}_{N_o})$ is formed, where

$$\underline{x}_{N_o} = \left[x_{(n-1)N_o+1}, \; x_{(n-2)N_o+1}, \ldots, \; x_{nN_o}\right] \qquad (4.22)$$

Since the successive observations are assumed to be iid, it follows that the successive test statistics are iid. Once the test statistic is formed, the following test is performed at each local detector

$$T_{N_o}(\underline{x}_{N_o}) \begin{cases} \geq A \text{ , stop and decide } H_1 \\ \leq B \text{ , stop and decide } H_0 \\ \text{otherwise, discard all previous stages} \\ \text{and proceed to the next stage.} \end{cases} \qquad (4.23)$$

where A and B are the test thresholds with $A > B$. The values of A, B, and N_o are predetermined such that the prespecified values of α and β are satisfied and the global ASN of the test is minimized.

Let $p_0 = Pr\{T_{N_o}(\underline{x}_{N_o}) \geq A/H_0\}$, $q_0 = Pr\{T_{N_o}(\underline{x}_{N_o}) \leq B/H_0\}$, and $r_0 =$

$1 - (p_0+q_0)$. Similarly let $p_1 = Pr\{T_{N_o}(\underline{x}_{N_o}) \geq A/H_1\}, q_1 = Pr\{T_{N_o}(\underline{x}_{N_o}) \leq B/H_1\}$, and $r_1 = 1 - (p_1+q_1)$. Since the local detectors as well as their observations are identical at all stages, it follows that the global decision will be based on any one of the local decisions with equal probability, i.e.,

$$\gamma_{1j} = \gamma_{2j} = \ldots = \gamma_{Mj} \quad , \quad j = 0,1 \tag{4.24}$$

Since the local sequential test can terminate at any stage, the error probabilities can be expressed as

$$\alpha = p_0 + r_0 p_0 + r_0^2 p_0 + \ldots = \frac{p_0}{1 - r_0} \tag{4.25}$$

and

$$\beta = q_1 + r_1 q_1 + r_1^2 q_1 + \ldots = \frac{q_1}{1 - r_1} \tag{4.26}$$

Equations (4.25) and (4.26) give the necessary conditions imposed on A and B for a particular choice of the package size N_o to satisfy α and β. We observe that the event of test termination is independent from stage to stage, and from one detector to another. Therefore, the global test length is geometrically distributed random variable. When H_j is true, its pdf is given

$$f_{T_g/H_j}(t_g/H_j) = (1 - r_j^M)(r_j^M)^{n-1} \bullet \delta(t_g - nN_o) \quad , \quad n \geq 1 \tag{4.27}$$

where $\delta(\cdot)$ is the Kronecker delta function and (r_j^M) is the probability that all of the local tests fail to decide on a hypothesis and must continue at any single stage. From (4.27), we can express the ASN under the hypothesis H_j to be

$$ASN_j = E[T_g/H_j] = \frac{N_o}{1-(r_j^M)} \tag{4.28}$$

As was pointed out in [21], we must select the package size N_o and the test thresholds A, and B before the test can be executed. The thresholds A and B must obviously satisfy (4.25) and (4.26), and, in general, minimize a weighted sum of ASN_1 and ASN_0 as given by

$$ASN = \eta\ ASN_0 + (1-\eta)\ ASN_1 \quad , \quad \eta \in [0,1] \tag{4.29}$$

An analytical solution of this optimization problem does not appear to be feasible and, therefore, a numerical solution is necessary. In the following we propose an algorithm for the determination of the optimal package size N_o and the test thresholds.

Step 1: Pick a package of size $N_o = n_o \in [1, N_{FSS}]$ (where N_{FSS} is the number of observations such that $r_j = 0$), and choose $\tilde{A}(n_o)$ with

$$P_0(n_o) \triangleq \int_{\tilde{A}(n_o)}^{\infty} f_{T_{n_o}}(\underline{x}_{n_o}/H_0)\ d\ T_{n_o} < \alpha \tag{4.30}$$

and

$$P_1(n_o) \triangleq \int_{\tilde{A}(n_o)}^{\infty} f_{T_{n_o}}(\underline{x}_{n_o}/H_1)\ d\ T_{n_o} < 1-\beta \tag{4.31}$$

Step 2: For $P_0(n_o)$ in (4.30) there exists a single threshold $\tilde{B}_0(n_o)$ such that α is satisfied. Let

$$q_0(n_o) \triangleq \int_{-\infty}^{\tilde{B}_0(n_o)} f_{T_{n_o}}(\underline{x}_{n_o}/H_0)\, d\, T_{n_o} ,$$

it follows that α is given by

$$\alpha = \frac{P_0(n_o)}{P_0(n_o) + q_0(n_o)} \tag{4.32}$$

Similarly, let $q_1(n_o) \triangleq \int_{-\infty}^{\tilde{B}_1(n_o)} f_{T_{n_o}}(\underline{x}_{n_o}/H_1)\, d\, T_{n_o}$,

and obtain $\tilde{B}_1(n_o)$ such that β as given by

$$\beta = \frac{q_1(n_o)}{q_1(n_o) + P_1(n_o)} \tag{4.33}$$

is satisfied.

Step 3: We have one of the following three cases:

a) $\tilde{B}_0(n_o) < \tilde{B}_1(n_o)$. This case represents the situation where $\bar{A}(n_o)$ is higher than the optimal $A(n_o)$, because any value of $B(n_o)$, $\tilde{B}_0(n_o) < B_1(n_o) < \tilde{B}_1(n_o)$, will yield smaller values of α and β than specified. Therefore, choose a smaller value of $\bar{A}(n_o)$ and repeat.

b) $\tilde{B}_0(n_o) > \tilde{B}_1(n_o)$. In this case, there is no value of $B(n_o)$ that will satisfy the required α and β. Therefore, choose a higher value of $\bar{A}(n_o)$ and repeat.

c) $\tilde{B}_0(n_o) = \tilde{B}_1(n_o)$. This case represents the solution which we are seeking. Therefore, let $A(n_o)=\bar{A}(n_o)$ and $\tilde{B}_1(n_o) = B(n_o)=\tilde{B}_0(n_o)$.

Step 4: Once the thresholds for the selected n_o are determined, we repeat the process to determine the thresholds for (n_o-1). Com-

pute the corresponding ASN given by (4.29). If the ASN for n_o is less than the ASN for (n_o-1), we increase n_o. Otherwise, we choose a smaller value of n_o.

Step 5: The algorithm must be continued until for the first time at n_o^*, the process of increasing or decreasing n_o is reversed. The optimal package size is obviously $N_o^* = n_o^*$.

The above algorithm is based on the fact that r_j is monotonically decreasing function of the package size N_o. This fact has been proven for equal error probabilities and normally distributed observations in [21]. To prove the above fact in the general case, we proceed as follows. Let N_o and $(1 - r_j(N_o))$ be any package size and the corresponding probability of decision on one of the hypotheses when H_j is true. Let $(N_o+\Delta N_o)$ and $[1-r_j(N_o+\Delta N_o)]$ be another package size and the corresponding probability of a decision on one of the hypotheses. If we view the increment ΔN_o as an independent package and apply the MLGDS procedure to the individual packages of size N_o and ΔN_o, we may write

$$r_j(N_o + \Delta N_o) = r_j(N_o) \cdot r_j(\Delta N_o) \tag{4.34}$$

since $r_j(\Delta N_o)$ is strictly less than one for $\Delta N_o \geq 1$, it follows that $r_j(N_o + \Delta N_o) < r_j(N_o)$ and, therefore, the proof is complete.

Returning to the multi-sensor decentralized scheme, we observe from (4.28) that ASN_j is a monotonically increasing function of the number of local detectors M. In the remainder of this section, we consider the truncation of the sequential test at the n_Tth stage. A package of observations of the same size N_o is taken at one of the local detectors and the following single threshold test is performed.

$$T_{N_o}(\underline{x}_{N_o}) \begin{cases} \geq v, \text{ decide } H_1 \\ < v, \text{ decide } H_0 \end{cases} \tag{4.35}$$

where v is the test threshold at the (n_T+1)th stage. Let $\alpha_T(N_o)$ and $\beta_T(N_o)$ be the error probabilities at the truncation stage with

$$\alpha_T(N_o) = Pr\{T_{N_o}(\underline{x}_{N_o}) \geq v/H_0\}$$

$$\beta_T(N_o) = Pr\{T_{N_o}(\underline{x}_{N_o}) < v/H_1\} \tag{4.36}$$

The overall global error probabilities $\alpha_o(M)$ and $\beta_o(M)$ are as given in (4.14) and (4.15) respectively. Upon writing (4.14) and (4.15) explicitly and observing that $\alpha_s(n_T) = \alpha$ and $\beta_s(n_T) = \beta$, we obtain

$$\alpha_o(M) = \alpha + (\alpha_T - \alpha)(r_0^M)^{n_T} \tag{4.37}$$

and

$$\beta_o(M) = \beta + (\beta_T - \beta)(r_1^M)^{n_T} \tag{4.38}$$

Since $r^j < 1$, $j = 0,1$, it follows from (4.37) and (4.38) that the increase in error probabilities $\{\alpha_o(M) - \alpha\}$ and $\{\beta_o(M) - \beta\}$ is a monotonically decreasing function of M, the number of sensors, as well as the truncation stage n_T. Consequently, the M detector system is less sensitive to truncation.

4.4. The Proposed Truncation Scheme

We have shown that the probability of the global test length T_g exceeding a specified truncation stage n_T is less for the multi-sensor scheme than for the single detector case. However, T_g is still a random variable which can assume excessively large values. To avoid this

problem which will limit the practical use of sequential procedures, truncation is necessary. The truncation problem is not simple to analyze [31-33] except for the MLGDS procedure in [21].

The proposed truncation scheme for the system under consideration is as follows: choose A* and B* (the test thresholds) such that

$$\alpha^* \triangleq \frac{P_0^*}{1 - r_0^*} < \alpha \tag{4.39}$$

and

$$\beta^* \triangleq \frac{q_1^*}{1 - r_1^*} < \beta \tag{4.40}$$

where P_0^*, r_0^*, q_1^*, and r_1^* are as defined earlier when the thresholds A*, and B* are used instead of A, and B respectively. The local and global sequential tests are performed as usual. If no decision is reached up to the n_Tth stage, only one local detector is allowed to take one more package of observations of the same size and test against a single threshold v as described in (4.35). Writing the expressions for the overall error probabilities, we obtain

$$\alpha_o(M) = \alpha^* + (\alpha_T(N_o) - \alpha^*)(r_0^{*M})^{n_T} \tag{4.41}$$

and

$$\beta_o(M) = \beta^* + (\beta_T(N_o) - \beta^*)(r_1^{*M})^{n_T} \tag{4.42}$$

where $\alpha_T(N_o)$ and $\beta_T(N_o)$ are as defined by (4.36). Since $\alpha^* < \alpha$ and $\beta^* < \beta$ and the term $(r_j^{*M})^{n_T}$ is a decreasing function of n_T; there exists a value of n_T at which $\alpha_o(M) \leq \alpha$ and $\beta_o(M) \leq \beta$. The smallest value of n_T at which both $\alpha_o(M)$ and $\beta_o(M)$ become smaller than α and β will be used for truncation. It should be emphasized that for different values of α^* and

β^*, the solution for n_T is different. However, the determination of α^* and β^* for some value of the package size is not simple because of the continuous nature of α^* and β^*. In other words, all values $\alpha^* \in (0,\alpha)$ and $\beta^* \in (0,\beta)$ are possible. On the other hand, since the ASN of the untruncated test is less than that of the truncated test for the same α and β, we can say that a reasonable choice of α^* and β^* is that which leads to a small increase in the ASN's over their untruncated counterparts while maintaining the constraints on α and β. The average sample number of the truncated test is given by

$$ASN_j^T = N_o \frac{1 - (r_j^{*M})^{n_T+1}}{1 - r_j^{*M}} \qquad (4.43)$$

when the hypothesis H_j is true, $j=0,1$.

4.5. Numerical Results

In this section, we present some numerical results for the multi-sensor decentralized scheme using MLGDS procedure. The observation model is given by

$$H_0: \quad x_i \sim N(0, \sigma^2)$$
$$H_1: \quad x_i \sim N(\theta, \sigma^2)$$

where $N(\theta, \sigma^2)$ is the normal density function with mean θ, and variance σ^2. We assume that $\theta = 0.2$ and $\sigma^2 = 1$. It is well known that for $\alpha = \beta$, the number of observations necessary for the optimal FSS linear detector (Neyman-Pearson) is the solution of the equation

$$N_{FSS} = [\frac{2\sigma}{\theta} \Phi^{-1}(1 - \alpha)]^2$$

where $\Phi(\bullet)$ is the cdf of the normal density function of zero mean and unit variance. The ASN of Wald's SPRT is approximated by the following equation [1]

$$ASN_W \cong \frac{(1-\beta)\ \log \frac{1-\beta}{\alpha} + \beta\ \log \frac{\beta}{1-\alpha}}{\theta^2/2\sigma^2}$$

In Table 4.1, the optimal package size N_M^* is determined numerically for the case of M local detectors, M = 1,2,3. This optimal value depends on the specified error probabilities α and β as well as the number of local detectors. The values of N_{FSS}, ASN_W, ASN(M), and N_M^* are given for some values of α and β. Note that for smaller values of α and β, it takes more sensors to compensate for the loss due to grouping of data.

Table 4.1: Optimal package sizes N_M^* and their corresponding ASN(M)

$\alpha=\beta$	N_{FSS}	ASN_W	N_1^*	ASN(1)	N_2^*	ASN(2)	N_3^*	ASN(3)
10^{-2}	541	225.2	208	312.0	163	219.1	143	181.8
10^{-3}	954	344.6	359	492.1	289	359.9	246	305.9
10^{-4}	1383	460.4	511	664.2	406	499.4	364	431.1
10^{-5}	1818	575.6	659	830.7	534	638.9	485	557.2
10^{-6}	2259	690.8	808	993.5	663	776.3	600	683.4

The increase in error probabilities as a result of truncation is demonstrated by the numerical results in Table 42. The package size is set at N_1^* (given in Table 4.1) irrespective of the number of local detectors. The central test is truncated at the third stage. The

resulting error probabilities α_o and β_o, and the ASN of the truncated test (ASN^T) are given for some values of α and β.

Table 4.2: Overall error probabilities α_o and β_o, and the ASN for the truncated test

	ONE DETECTOR		TWO DETECTORS		THREE DETECTORS	
$\alpha=\beta$	$\alpha_o=\beta_o$	ASN^T	$\alpha_o=\beta_o$	ASN^T	$\alpha_o=\beta_o$	ASN^T
10^{-2}	1.24×10^{-2}	308.1	1.009×10^{-2}	234	1.003×10^{-2}	216
10^{-3}	1.55×10^{-3}	489.5	1.011×10^{-3}	387.3	1.0002×10^{-3}	366.2
10^{-4}	2.45×10^{-4}	662.3	1.018×10^{-4}	539.7	1.0002×10^{-4}	527.3
10^{-5}	5.52×10^{-4}	829.2	1.04×10^{-5}	688.4	1.0004×10^{-5}	664.9
10^{-6}	15.56×10^{-6}	992.3	1.095×10^{-6}	837.2	1.0006×10^{-6}	813.3

The truncation scheme proposed in Section 4.4 is applied to the distributed system of two and three local detectors. The package sizes are the optimal for the untruncated test as given in Table 3.1. However, the sequential portion of the test is designed to satisfy α_s and β_s ($\alpha_s < \alpha$, $\beta_s < \beta$) and the truncation stage M is obtained by requiring the overall error probabilities α_o and β_o to be less than or equal to α and β respectively. In Table 4.3, we present values of M, ASN, ASN^T, α_o, β_o, α_s, and β_s as obtained for different values of α and β.

Table 4.3: Overall error probabilities and ASN's of the proposed truncation scheme.

	TWO SENSORS					THREE SENSORS				
$\alpha=\beta$	M	ASN	ASN^T	$\alpha_s=\beta_s$	$\alpha_o=\beta_o$	M	ASN	ASN^T	$\alpha_s=\beta_s$	$\alpha_o=\beta_o$
10^{-2}	5	219.1	220.2	9.79×10^{-3}	9.9×10^{-3}	5	181.8	182.8	9.81×10^{-2}	9.86×10^{-3}
10^{-3}	5	359.9	361.5	9.74×10^{-4}	9.88×10^{-4}	5	305.9	307.2	9.72×10^{-4}	9.72×10^{-5}
10^{-4}	6	499.4	501.5	9.65×10^{-5}	9.76×10^{-5}	5	431.1	432.6	9.65×10^{-5}	9.94×10^{-3}
10^{-5}	6	638.9	640.9	9.58×10^{-6}	9.81×10^{-6}	6	557.2	559.2	9.62×10^{-6}	9.69×10^{-6}
10^{-6}	7	776.3	779	9.56×10^{-7}	9.64×10^{-7}	6	683.4	685.7	9.58×10^{-7}	9.86×10^{-7}

4.6. Discussion

In this chapter, we proposed a simple multi-sensor decentralized sequential detection procedure and derived the expressions for the ASN's. The case when each local detector employs the MLGDS procedure of Lee and Thomas [21] is considered in detail and a truncation scheme is analyzed. We showed that the performance of the multi-sensor scheme is a monotonically increasing function of the number of local sensors used. The increase in error probabilities due to truncation of the sequential procedure at any stage is shown to be a decreasing function of the number of local detectors. The numerical results presented show clearly the improved performance and support the analysis. In the simple multi-sensor scheme, each local detector was designed independently to satisfy the specified error probabilities, i.e., no coupling between local thresholds

was allowed. It is expected that if coupling between local thresholds is allowed and local decisions are all taken into account in the process of arriving at the global decision, the resulting system performance will be further improved. In the next two chapters, we will study distributed sequential detection systems in which the local thresholds are coupled and all local decisions are fused at the global decision maker to arrive at the global decision.

CHAPTER FIVE

A DECENTRALIZED SEQUENTIAL TEST WITH DATA FUSION

5.1. Introduction

In Chapter Four, we studied a simple multi-sensor sequential detection procedure and showed that its performance improves monotonically with the increase in the number of sensors used. The sequential test in Chapter Four is carried out locally and only local decisions are communicated to the global decision maker. The local thresholds are not coupled because no explicit data fusion rule is employed. All the thresholds are designed to satisfy the global error probabilities α and β.

In this chapter, we allow the local thresholds to be coupled. Each local detector performs a SPRT based on its own observations, and communicates its decision to the global decision maker (fusion center) whenever it decides on a hypothesis. Nothing is communicated if the local decision is to continue. The global decision maker combines the incoming local decisions according to a predetermined fusion rule to come up with the global (final) decision. In Section 5.2, we describe the sequential detection system, and define the observation model. The global error probabilities are functions of the local error probabilities and the fusion rule. An analysis for different possible fusion rules is presented in Section 5.3, with emphasis on the relationship between the local and global error probabilities. Moreover, the expected global test length is derived in terms of the local error probabilities and the pdf's of the local test length. The case of three local detectors is considered

briefly in Section 5.4. An example is presented in Section 5.5 to illustrate the design of the local tests for prespecified values of α and β, and for different fusion rules. In Section 5.6, we discuss the results obtained in this chapter.

5.2. A Description of the Distributed System

Consider the problem of testing a simple hypothesis H_0 versus a simple alternative H_1. The system consists of two sequential detectors and a global decision maker as shown in Fig. 5.1. Let $f(x,\theta_x)$ and $f(y, \theta_y)$ denote the pdf's of the random variables x and y, which represent the observations of the first and second sequential detectors respectively. Let H_0 be the hypothesis that $\theta_x = \theta_{x0}$, $\theta_y = \theta_{y0}$, and H_1 the hypothesis that $\theta_x = \theta_{x1}$ and $\theta_y = \theta_{y1}$. Thus, the pdf's of x and y are given by $f(x, \theta_{xi})$ and $f(y, \theta_{yi})$ when H_i is true, i=0,1. The successive observations on x are denoted by $x_1, x_2, \ldots, x_n \triangleq \underline{x}_n (n \geq 1)$ and they are assumed to be samples of iid random variables. Similarly, the successive observations on y are denoted by $y_1, y_2, \ldots, y_n \triangleq \underline{y}_n$ $(n \geq 1)$ and they are assumed to be iid random variables. For any two positive integers n,m, the joint pdf of $\underline{x}_n$ and $\underline{y}_m$ conditioned on the hypothesis H_i is given by

$$f_{\underline{x}_n, \underline{y}_m / H_i} = \prod_{j=1}^{n} f(x_j , \theta_{xi}) \cdot \prod_{j=1}^{m} f(y_j , \theta_{yi}) \quad , i = 0,1 \qquad (5.1)$$

Based on its own observations, each local detector performs an SPRT. The SPRT at the first detector is defined as follows: Two positive constants A_1 and $B_1 (A_1 > B_1)$ are chosen and at each stage n, $n \geq 1$, of testing, the likelihood ratio function $\Lambda_{\underline{x}_n}$ is computed:

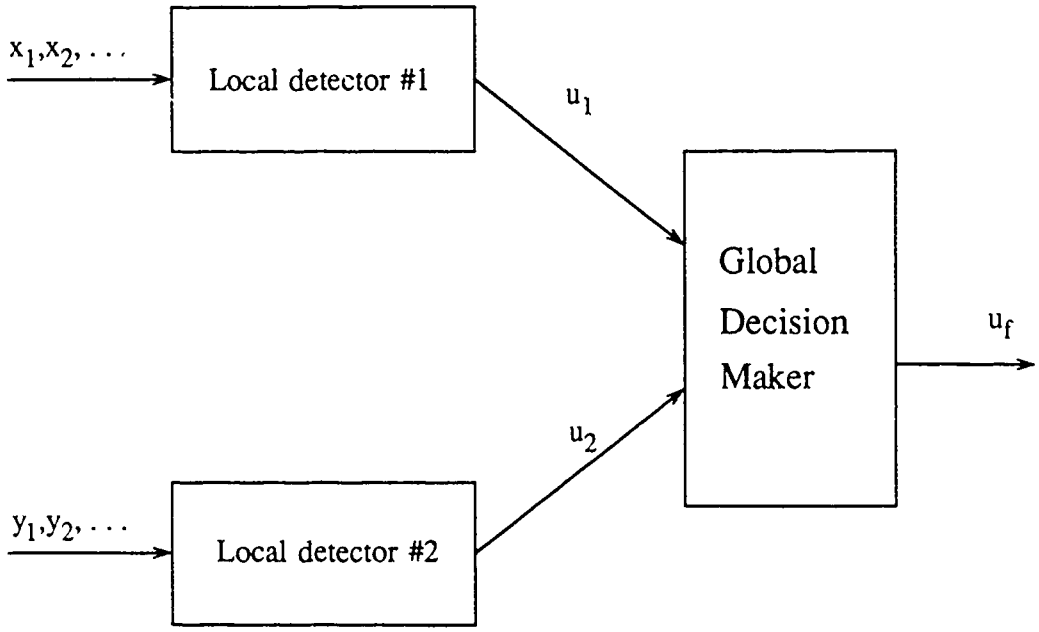

Fig. 5.1: A distributed system consisting of two local detectors and a global decision maker.

$$\Lambda_{\underline{x}_n} \triangleq \prod_{j=1}^{n} \frac{f(x_j, \theta_{x1})}{f(x_j, \theta_{x0})} \tag{5.2}$$

The likelihood ratio in (5.2) is compared with the constants (thresholds) A_1 and B_1 as follows

$$\Lambda_{\underline{x}_n} \begin{cases} \geq A_1, & \text{decide } H_1 \\ \leq B_1, & \text{decide } H_0 \\ \text{otherwise,} & \text{continue} \end{cases} \tag{5.3}$$

A similar SPRT is implemented at the second detector with A_2 and B_2 as the thresholds. The choice of the thresholds A_k and B_k, $k = 1,2$, depends on the values of the local error probabilities α_k and β_k. It has been shown [1] that the thresholds are approximated in terms of α_k and β_k as follows:

$$A_k \cong (1 - \beta_k)/\alpha$$

$$B_k \cong \beta_k/(1 - \alpha_k) \qquad , k = 1,2 \tag{5.4}$$

A more convenient local test can be obtained by taking the logarithm of both sides of (5.3). The resulting tests at the two detectors are given by

$$\sum_{j=1}^{n} v_{kj} \begin{cases} \geq \log A_k, & \text{decide } H_1 \\ \leq \log B_k, & \text{decide } H_0 \\ \text{otherwise,} & \text{continue.} \end{cases} \qquad k = 1,2 \tag{5.5}$$

where

$$v_{1j} \triangleq \log [f(x_j, \theta_{x1})/f(x_j, \theta_{x0})]$$

and

$$v_{2j} \triangleq \log [f(y_j, \theta_{y_1})/f(y_j, \theta_{y_0})]$$

As mentioned above, the local error probabilities α_k and β_k are functions of the local thresholds A_k and B_k, $k = 1,2$. By adjusting the

local thresholds we can vary the corresponding local error probabilities. Since the observations are independent from one local detector to the other, the local decisions are also independent. As in the distributed fixed-sample-size systems using Neyman-Pearson criterion, we can not obtain the global error probabilities without the knowledge of the fusion rule. In other words, the global error probabilities are functions of the local error probabilities and the fusion rule. Conversely, given the global error probabilities α and β, we can choose the fusion rule first, and then solve for α_k and β_k, $k = 1,2$, or equivalently A_k and B_k such that α and β are satisfied and the global average sample number is minimum. The global average sample number ASN_g is defined as the average number of observations necessary for reaching a global decision. Clearly, an optimal solution will require the choice of the fusion rule among all possible fusion rules and solving for the local thresholds such that ASN_g is minimized. However, for a distributed system consisting of two local detectors and binary local decisions (assuming that local detectors have reached a hypothesis decision), we observe that the number of possible binary fusion rules is

$$2^4 = \sum_{i=0}^{4} \binom{4}{i} = \sum_{i=0}^{4} \frac{4!}{i!(4-i)!} .$$

For sequential hypothesis testing, the fusion rule can be ternary to allow the occurrence of the event continue. In this case, the number of possible fusion rules is obviously more than 2^4 and is given by

$$\sum_{i,j=0}^{4} \frac{4!}{i!j!(4-i-j)!} = 3^4 .$$

For an optimal solution, all possible fusion rules must be enumerated and the one yielding the lowest ASN_g must be chosen. In the following section, we study only three different possible fusion rules and evaluate their performance.

5.3. Analysis of Some Possible Fusion Rules

The fusion rules of interest are those which take into account the decisions of both local detectors, and satisfy the monotone property of the likelihood ratio function at the global level. Therefore, we consider only three fusion rules and analyze them.

i) AND Fusion Rule: According to this fusion rule, the global decision is $H_1 (u_g = H_1)$ if and only if both local decisions are $H_1 (u_1 = H_1$ and $u_2 = H_1)$, i.e.,

$$u_g \overset{\Delta}{=} f(u_1, u_2) \begin{cases} H_1, \text{ if } u_1 = H_1 \text{ and } u_2 = H_1 \\ H_0, \text{ if } u_1 = H_0 \text{ or } u_2 = H_0 \\ \text{continue, elsewhere .} \end{cases} \tag{5.6}$$

From (5.6), it follows that the global error probabilities are given by

$$\begin{aligned} \alpha &= \alpha_1 \cdot \alpha_2 \\ (1-\beta) &= (1-\beta_1) \cdot (1-\beta_2) \end{aligned} \tag{5.7}$$

It is clear that depending on the first incoming local decision, the global decision will be reached no later than the decision time of the second incoming local decision. Since an SPRT terminates in a finite number of steps (stages) with probability one [1], it follows that a global decision will be reached in a finite number of steps with probability one.

From (5.7), we observe that $\alpha_k > \alpha$ and $\beta_k < \beta$, $k = 1,2$. Also, since α_k, $\beta_k << 1$ we can further approximate (5.4) to obtain

$$\begin{aligned} A_k &\cong \frac{1}{\alpha_k} < \frac{1}{\alpha} \cong A \\ B_k &\cong \beta_k < \beta \cong B \end{aligned} \tag{5.8}$$

From (5.8), we observe that the use of AND fusion rule requires lower local thresholds than the centralized test. This will lead to a shorter local average decision time (ASN) when H_1 is true and a longer local average decision time when H_0 is the true hypothesis.

The two main functions usually used in evaluating the performance of an SPRT are the power function and the ASN function. The power function $P_{Dg}(\theta_x,\theta_y)$ is defined as the probability of deciding H_1 when the actual parameters of the distributions are θ_x and θ_y instead of θ_{x1} and θ_{y_1} respectively. From (5.7), it follows that $P_{Dg}(\theta_x,\theta_y)$ is simply the product of the local power functions $P_{Dx}(\theta_x)$ and $P_{Dy}(\theta_y)$ at the first and second local sequential detectors respectively. Therefore, we can write

$$P_{Dg}(\theta_x,\theta_y) = P_{Dx}(\theta_x) \cdot P_{Dy}(\theta_y) \qquad (5.9)$$

The computation of $P_{Dg}(\theta_x,\theta_y)$ is thus achieved by computing $P_{Dx}(\theta_x)$ and $P_{Dy}(\theta_y)$ at the local detectors exactly as described in [1].

The second function of interest is the expected duration of the global test denoted by $E[T_g/H_i]$ when H_i is the true hypothesis. Let T_k denote the test length of the kth, k=1,2 sequential detector. Let A_n^i, B_n^i, C_n^i, D_n^i, and F_n^i be the events that $\{T_1=n,\ T_2>n/H_i\}$, $\{T_1>n,\ T_2=n/H_i\}$, $\{T_1=n,\ T_2<n/H_i\}$, $\{T_1<n,\ T_2=n/H_i\}$, and $\{T_1=T_2=n/H_i\}$ respectively. Observe that if a global decision is reached at the nth stage, then one of the above mutually exclusive events must have occurred. When the first incoming decision is H_0 the global test will terminate without the need to wait for the other local detector to decide. This corresponds to the events A_n^i and B_n^i. On the other hand, if the first incoming local decision is H_1, then the global decision maker must wait for the other local detector to

decide. This corresponds to the events C_n^i and D_n^i. When both decisions arrive at the same time, the event F_n^i occurs. To compute the probabilities of the events A_n^i, B_n^i, C_n^i, D_n^i and F_n^i the exact probability distribution functions of T_1 and T_2 when H_i, i=0,1, is true are required. The distribution function of the test length T_k is known [1,28,29] only when the increments v_{ki} as defined in equation (5.5) are integral multiples of a constant. In this case, the exact probability distribution function of T_g can be derived. Let $\alpha_k^i(n)$ and $\beta_k^i(n)$ be the probabilities that the kth sequential detector terminates at the nth stage when H_i is true accepting H_1 and H_0 respectively, i.e.

$$\alpha_k^i(n) \triangleq \Pr\{T_k=n,\ u_k= H_1/H_i\}$$

$$\beta_k^i(n) \triangleq \Pr\{T_k=n,\ u_k=H_0/H_i\}$$

$$k=1,2 \text{ and } i=0,1$$

$$n \geq 1 \qquad (5.10)$$

Let $\tilde{\alpha}_k^i(n) = \sum_{j=1}^{n} \alpha_k^i(j)$ and $\tilde{\beta}_k^i(n) = \sum_{j=1}^{n} \beta_k^i(j)$. From (5.6) and (5.10), it follows that the probability of reaching a global decision at the nth stage when H_i is true is given by

$$\begin{aligned}\Pr\{T_g=n/H_i\} &= \beta_1^i(n)[1 - \tilde{\alpha}_2^i(n) - \tilde{\beta}_2^i(n)] + \beta_2^i(n)\ [1 - \tilde{\alpha}_1^i(n) - \tilde{\beta}_1^i(n)] \\ &\quad + [\beta_1^i(n) + \alpha_1^i(n)]\ \tilde{\alpha}_2^i(n-1) + \tilde{\alpha}_1^i(n-1)[\tilde{\beta}_2^i(n) + \tilde{\alpha}_2^i(n)] \\ &\quad + [\alpha_1^i(n) + \beta_1^i(n)]\ [\alpha_2^i(n) + \beta_2^i(n)] \qquad (5.11)\end{aligned}$$

Simplifying (5.11), we obtain

$$\begin{aligned}\Pr\{T_g=n/H_i\} &= \beta_1^i(n)[1 - \tilde{\beta}_2^i(n-1)] + \beta_2^i(n)[1-\tilde{\beta}_1^i(n)] \\ &\quad + \alpha_1^i(n)\ \tilde{\alpha}_2^i(n) + \alpha_2^i(n)\ \tilde{\alpha}_1^i(n-1) \qquad (5.12)\end{aligned}$$

To simplify the analysis, we restrict the local detectors to be identical. In this case, T_1 and T_2 are identically distributed under both hypotheses. Therefore, equation (5.12) can be simplified further to obtain

$$\Pr\{T_g=n/H_i\} = \beta^i(n)[2-\tilde{\beta}^i(n) - \tilde{\beta}^i(n-1)] + \alpha^i(n)\ [\tilde{\alpha}^i(n) - \tilde{\alpha}^i(n-1)] \quad (5.13)$$

where $\alpha^i(n) = \alpha_k^i(n), \beta^i(n) = \beta_k^i(n)$, $\tilde{\alpha}^i(n) = \tilde{\alpha}_k^i(n)$, and $\tilde{\beta}^i(n) = \tilde{\beta}_k(n)$, $k=0,1$. Equation (5.13) can be used to numerically compute the average global test duration. However, it does not provide enough insight to the problem. We observe that if H_i is the true hypothesis, then the obtained local decision will satisfy the local error probabilities. Therefore, we can write

$$\Pr\{T_g=n/H_i\} = \Pr\{u_1=H_0/H_i\} \cdot \Pr(A_n^i) + \Pr\{u_2=H_0/H_i\} \cdot \Pr(B_n^i)$$
$$+ \Pr\{u_2=H_1/H_i\} \cdot \Pr(C_n^i) + \Pr\{u_1=H_1/H_i\} \cdot \Pr(D_n^i) + \Pr(F_n^i) \quad (5.14)$$

From the definitions of the minimum and maximum of two random variables [29], we observe that

$$\Pr\{\min(T_1,T_2) = n/H_i\} = \Pr(A_n^i) + \Pr(B_n^i) + \Pr(F_n^i)$$

and

$$\Pr\{\max(T_1,T_2) = n/H_i\} = \Pr(C_n^i) + \Pr(D_n^i) + \Pr(F_n^i)$$

Therefore, it follows from (5.14)-(5.16) that T_g is given by

$$T_g = \begin{cases} \min(T_1,T_2) \text{ with probability } P_{\min}(H_i) \\ \\ \max(T_1,T_2) \text{ with probability } P_{\max}(H_i) \end{cases} \quad (5.17)$$

when H_i is the true hypothesis, and where $P_{\min}(H_i) = \Pr(u_1 = H_0/H_i) =$

$Pr(u_2 = H_0/H_i) \triangleq Pr(u_\ell = H_0/H_i)$ and $P_{max}(H_i) = Pr(u_1 = H_1/H_i) = Pr(u_2 = H_1/H_i) \triangleq Pr(u_\ell = H_1/H_i)$. Equation (5.17) shows that the global test duration is either the maximum or the minimum of the local test durations depending on whether the first incoming decision is H_1 or H_0 respectively. From (5.17), it follows that the expected global test duration conditioned on the hypothesis $H_i, i=0,1$, is as follows:

$$E[T_g/H_i] = Pr(u_\ell=H_0/H_i) \cdot E[\min(T_1,T_2)/H_i] + Pr(u_\ell=H_1/H_i) \cdot E[\max(T_1,T_2)/H_i] \quad (5.18)$$

Thus, the problem reduces to finding the expected values of the minimum and the maximum of the local test lengths. Due to the functional complexity of the distribution of T_1 and T_2 under both hypotheses, an analytical solution does not seem feasible. However, the numerical computation is relatively simple and it can be performed via matrix multiplications [28,29].

ii) OR Fusion Rule: In this case, a global decision in favor of H_1 is declared if any of the local decisions is H_1, while H_0 is decided globally if and only if both local decisions are H_0. i.e.,

$$u_g \triangleq f(u_1,u_2) = \begin{cases} H_1 \text{ , if } u_1=H_1 \text{ or } u_2=H_1 \\ H_0 \text{ , if } u_1=H_0 \text{ and } u_2=H_0 \\ \text{continue, elsewhere} \end{cases} \quad (5.19)$$

As in the case of AND fusion rule, the OR fusion rule will reach a global decision no later than the time at which the second local decision reaches the fusion center. Since the local decisions are reached in a finite number of steps with probability one, it follows that a global

decision will be reached after a finite number of steps with probability one also. Therefore, we can express the global error probabilities in terms of the local error probabilities using (5.19) as follows

$$\begin{aligned} (1-\alpha) &= (1-\alpha_1) \bullet (1-\alpha_2) \\ \beta &= \beta_1 \bullet \beta_2 \end{aligned} \tag{5.20}$$

From (5.20), it is clear that the global power function $P_{Dg}(\theta_x,\theta_y)$ as defined earlier can be expressed as the product of the local power functions $P_{Dx}(\theta_x)$ and $P_{Dy}(\theta_y)$. In particular, we have

$$\{1 - P_{Dg}(\theta_x,\theta_y)\} = \{1 - P_{Dx}(\theta_x)\} \ . \{1 - P_{Dy}(\theta_y)\} \tag{5.21}$$

Therefore, the computation of the power functions is similar to the AND fusion rule and it can be done as in [1] at the local level. Equation (5.20) can be used to show that the OR fusion rule increases the local thresholds. Thus, the average local test length increases when H_1 is true, and decreases when H_0 is true.

The exact probability distribution function of the global test length T_g can be derived provided that the exact distributions of the local tests are known. In a manner similar to the AND fusion rule, we can show that

$$\begin{aligned} \Pr\{T_g=n/H_i\} &= \alpha_1^i(n)[1 - \tilde{\alpha}_1^i(n-1)] + \alpha_2^i(n)\ [1-\tilde{\alpha}_1^i(n)] \\ &\quad + \beta_1^i(n)\ \tilde{\beta}_2^i(n) + \beta_2^i(n)\tilde{\beta}_1^i(n-1) \end{aligned} \tag{5.22}$$

If we assume identical local detectors and use the fact that local decisions must satisfy the local error probabilities, we can easily show that the global test duration is given by (5.17). However, for the OR fusion rule $P_{min}(H_i) = \Pr(u_1 = H_1/H_i) = \Pr(u_2 = H_1/H_i) \stackrel{\Delta}{=} \Pr(u_\ell = H_1/H_i)$ and

$P_{max}(H_i) = Pr(u_1 = H_0/H_i) = Pr(u_2 = H_0/H_i) \stackrel{\Delta}{=} Pr(u_\ell = H_0/H_i)$. Hence, the average global test duration is given by

$$E[T_g/H_i] = Pr(u_\ell=H_0/H_i) \cdot E[\max(T_1,T_2)/H_i] + Pr(u_\ell=H_1/H_i) \cdot E[\min(T_1,T_2)/H_i] \quad (5.23)$$

which is similar to (5.18) except for the fact that H_0 is replaced by H_1 and vice versa.

iii) Three-Way Fusion Rule: According to this fusion rule, a global decision in favor of H_i is declared if and only if both local decisions are H_i, i=0,1. If the local decisions are not the same, the global decision maker can either use a randomized decision rule to stop the test, or it can repeat the local sequential tests without keeping the previous test results. We first consider the randomization option, and in this case the three-way fusion rule is given by

$$u_g \stackrel{\Delta}{=} f(u_1,u_2) = \begin{cases} H_1, \text{ with probability 1 if } u_1=u_2=H_1 \\ H_1, \text{ with probability } \gamma \text{ if } u_1 \neq u_2 \\ H_0, \text{ with probability } (1-\gamma) \text{ if } u_1 \neq u_2 \\ H_0, \text{ with probability 1 if } u_1=u_2=H_0 \\ \text{continue, elsewhere.} \end{cases} \quad (5.24)$$

Once again, because the global decision according to (5.24) is reached no later than the time at which the second incoming local decision is obtained, it follows that global decision will be obtained with probability one in a finite number of steps. The global error probabilities are expressed in terms of the local error probabilities as follows:

$$\alpha = \alpha_1\alpha_2 + \gamma\alpha_1(1-\alpha_2) + \gamma(1-\alpha_1)\alpha_2$$
$$\beta = \beta_1\beta_2 + (1-\gamma)(1-\beta_1)\beta_2 + (1-\gamma)\beta_1(1-\beta_2) \qquad (5.25)$$

Simplifying (5.25), we obtain

$$\alpha = (1-2\gamma)\alpha_1\alpha_2 + \gamma(\alpha_1+\alpha_2)$$
$$\beta = (2\gamma-1)\beta_1\beta_2 + (1-\gamma)(\beta_1+\beta_2) \qquad (5.26)$$

Therefore, for $\alpha_1=\alpha_2$ and $\gamma = 0.5$, we observe that each local detector must satisfy the global error probabilities. Because the global test can not terminate before both local decisions are obtained, it follows that the global average test duration is higher than that of a single sequential detector. In other words, the fusion rule in (5.24) will degrade the performance instead of improving it, so we don't consider it any further. Turning our attention to the second option, we can formulate the fusion rule as follows

$$u_g \triangleq f(u_1,u_2) = \begin{cases} H_1 & \text{if } u_1=H_1 \text{ and } u_2=H_1 \\ H_0 & \text{if } u_1=H_0 \text{ and } u_2=H_0 \\ \text{continue, elsewhere.} \end{cases} \qquad (5.27)$$

Let a trial be defined as the process of starting the local sequential tests until the two local decisions are reached. Note that if the two local decisions disagree, more than one trial will be required. The successive trials have identically distributed random lengths under the hypothesis H_i, i=0,1, as a consequence of the observations being identical. Each trial is guaranteed to terminate in a finite number of steps with probability one. However, the probability is one that a series of repeated trials will terminate within a finite number of trials,

it follows that the sequential process will terminate in a finite number of steps with probability one.

The global error probabilities can be obtained by summing their values at all trials. To this end, let p_i be the probability of a successful trial and $q_i = 1-p_i$ the probability of an unsuccessful trial when H_i is true. We obtain

$$\begin{aligned} p_0 &= \alpha_1\alpha_2 + (1-\alpha_1)(1-\alpha_2) \\ p_1 &= \beta_1\beta_2 + (1-\beta_1)(1-\beta_2) \end{aligned} \tag{5.28}$$

The probability of error of the first kind α is therefore given by

$$\alpha = \alpha_1\alpha_2 + \alpha_1\alpha_2 q_0 + \alpha_1\alpha_2 q_0^2 + \ldots = \frac{\alpha_1\alpha_2}{p_0} \tag{5.29}$$

Similarly, the probability of error of the second kind β is given by

$$\beta = \beta_1\beta_2 + \beta_1\beta_2 q_1 + \beta_1\beta_2 q_1^2 + \ldots = \frac{\beta_1\beta_2}{p_1} \tag{5.30}$$

The global power function $P_{Dg}(\theta_x,\theta_y)$ can be derived in the same way used in deriving (5.29) and (5.30). Recalling the definition of $P_{Dg}(\theta_x,\theta_y)$ and utilizing the assumed independence of local observations, we can express $P_{Dg}(\theta_x,\theta_y)$ in terms of $P_{Dx}(\theta_x)$ and $P_{Dy}(\theta_y)$ as follows

$$P_{Dg}(\theta_x,\theta_y) = \frac{P_{Dx}(\theta_x)\cdot P_{Dy}(\theta_y)}{P_{Dx}(\theta_x)\cdot P_{Dy}(\theta_y) + [1 - P_{Dx}(\theta_x)][1 - P_{Dy}(\theta_y)]} \tag{5.31}$$

where the denominator of (5.31) is simply the probability of a successful trial when θ_x and θ_y are the true values of the local parameters. From (5.31), it is evident that the computation of $P_{Dg}(\theta_x,\theta_y)$ is equivalent to

the computation of both $P_{Dx}(\theta_x)$ and $P_{Dy}(\theta_y)$ as described in detail by Wald in [1].

The expected duration of the global test when H_i is true is the next function to be derived. Let the random variable representing the number of observations in the mth trial when H_i is the true hypothesis be denoted by D_{im}, i=0,1; m≥1. Recall that the random variables D_{im}(m≥1) are iid with an expected duration $E[D_i]$ given by

$$E[D_i] = E[\max(T_1, T_2)/H_i] \tag{5.32}$$

where (5.32) is derived by observing that for a trial to reach an end, both local decisions must be reached. The conditional expected global test duration can be written as follows:

$$E[T_g/H_i, m] = m\, E[D_i] \tag{5.33}$$

However, $E[T_g/H_i, m]$ occurs with probability $q_i^{(m-1)} p_i$. The expected global test duration can be obtained by taking the expected value of (5.33), and the result is

$$E[T_g/H_i] = \sum_{m=1}^{\infty} m \cdot E[D_i]\, q_i^{(m-1)} p_i = \frac{E[D_i]}{p_i} \tag{5.34}$$

For sufficiently small error probabilities, we observe from (5.28), (5.29), and (5.30) that for $\alpha_1=\alpha_2=\alpha_\ell$ and $\beta_1=\beta_2=\beta_\ell$, we have $p_i \cong 1$, $\alpha_\ell > \alpha$, and $\beta_\ell > \beta$. Hence, unlike the two previous fusion rules, the effect of this fusion rule is to decrease the upper thresholds and to increase the lower thresholds which in turn decreases the average local test length under both hypotheses to a value smaller than that of a single sequential detector.

5.4. The Three Local Detectors Case

In this section, we consider extending the distributed system in Fig. 5.1 to include three local sequential detectors as shown in Fig. 5.2. The successive observations at the third sequential detector are denoted by $\omega_1, \omega_2, \ldots, \omega_n \triangleq \underline{\omega}_n (n \geq 1)$ and they are assumed to be iid random variables. We denote by $f(\omega, \theta_\omega)$ the pdf of the random variable ω representing the observations at the third local detector and assume that ω is independent of both x and y. Each local detector carries out an SPRT based on its own observations as described in Section 5.2. As is evident from the results of distributed FSS systems [11] and the discussion at the end of Section 5.2, the number of fusion rules increases quite rapidly as the number of local detectors increase. In our work, we will confine our analysis to a generalization of the three-way fusion rule which is symmetric with respect to the hypotheses. Therefore, a global decision in favor of the hypothesis H_i will be declared if and only if the three local detectors have reached their decisions and they are all favoring $H_i, i=0,1$. The fusion rule can therefore be formulated as follows:

$$u_g \triangleq f(u_1, u_2, u_3) = \begin{cases} H_1 \text{ , if } u_1=u_2=u_3=H_1 \\ H_0 \text{ , if } u_1=u_2=u_3=H_0 \\ \text{continue, elsewhere} \end{cases} \tag{5.35}$$

As in the case of two local detectors, we define a trial as the process of starting the local sequential tests until the three local decisions are reached. If the local decisions disagree, we begin another trial ignoring the results of previous trials. Let α_j and β_j be the error probabilities at the jth local detector, $j=1,2,3$. Let p_i be the

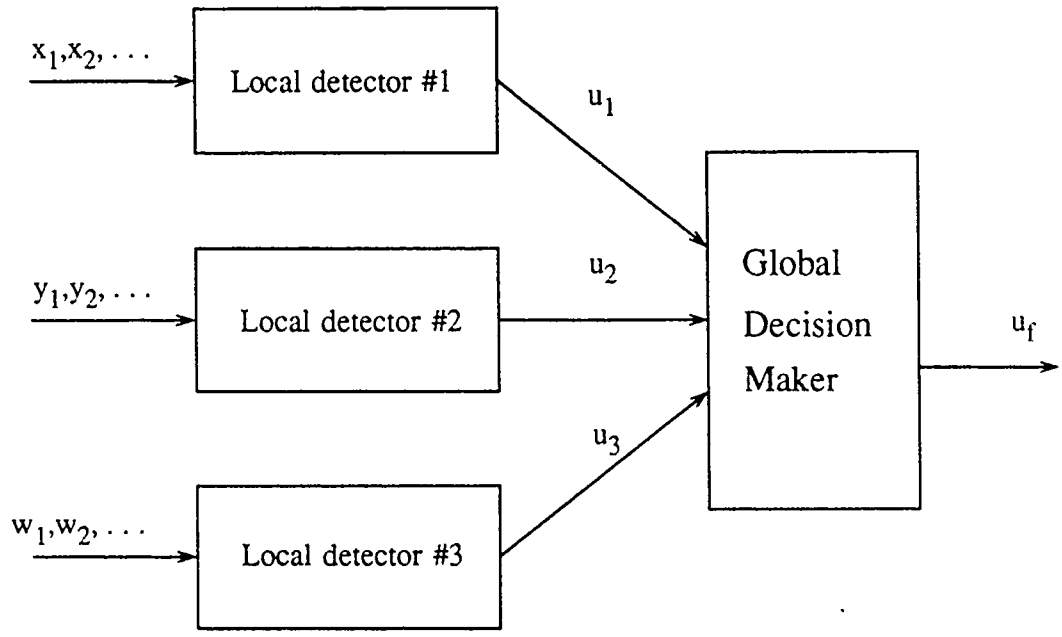

Fig. 5.2: A distributed system consisting of three local detectors and a global decision maker.

probability of achieving a successful trial and $q_i = 1-p_i$ when H_i is true, to obtain

$$p_0 = \prod_{j=1}^{3} \alpha_j + \prod_{j=1}^{3} (1 - \alpha_j)$$
$$p_1 = \prod_{j=1}^{3} \beta_j + \prod_{j=1}^{3} (1 - \beta_j) \tag{5.36}$$

The global error probabilities can be derived following the same approach used to derive (5.29) and (5.30), the results are given by the following equation

$$\alpha = \left(\prod_{j=1}^{3} \alpha_j \right) \; / \; p_0$$
$$\beta = \left(\prod_{j=1}^{3} \beta_j \right) \; / \; p_1 \tag{5.37}$$

The global power function $P_{Dg}(\theta_x, \theta_y, \theta_\omega)$ is a function of the local power functions $P_{Dg}(\theta_x)$, $P_{Dy}(\theta_y)$ and $P_{D\omega}(\theta_\omega)$ as given by

$$P_{Dg}(\theta_x, \theta_y, \theta_\omega) = \frac{P_{Dx}(\theta_x) \; P_{Dy}(\theta_y) \cdot P_{D\omega}(\theta_\omega)}{P_{Dx}(\theta_x) P_{Dy}(\theta_y) P_{D\omega}(\theta_\omega) + [1 - P_{Dx}(\theta_x)][1 - P_{Dy}(\theta_y)][1 - P_{D\omega}(\theta_\omega)]} \tag{5.38}$$

If we assume identical observations at all local detectors and equal local performance, i.e., $\alpha_j = \alpha_\ell$ and $\beta_j = \beta_\ell$, $j=1,2,3$. It follows that the local test lengths T_j, $j=1,2,3$, are identically distributed under the hypothesis H_i, $i=0,1$. Moreover, we can approximate

$$\alpha \cong \alpha_\ell^3 \quad \text{or} \quad \alpha_\ell \cong \sqrt[3]{\alpha}$$
$$\beta \cong \beta_\ell^3 \quad \text{or} \quad \beta_\ell \cong \sqrt[3]{\beta} \tag{5.39}$$

which can be used to show that $E[T_j/H_i]$, $j=1,2,3$; $i=0,1$, is approxi-

mately equal to one-third of the corresponding value in the centralized (one detector) case.

The average global test duration $E[T_g/H_i]$ when H_i is the true hypothesis can be derived in almost the same way used for the case of the local detectors. However, the derivation is more complicated due to the fact that our trials are of two different types each requiring, on the average, different number of observations. In particular, if the first two incoming local decisions are different, then a trial can be terminated requiring on the average $\tilde{T}_{2i}$ observations. On the other hand, if the first two incoming local decisions are the same, then a trial must continue until the third local decision is obtained. Let $\tilde{T}_{3i}$ denote the average number of observations in the trial in which the first and second incoming local decisions are identical. Let $\tilde{\gamma}_{2i}$ and $\tilde{\gamma}_{3i}$ denote the probability that when H_i is true, our trial is unsuccessful and requires $\tilde{T}_{2i}$ and $\tilde{T}_{3i}$ average number of observations respectively. Based on that, we can write

$$E[T_g/H_i] = \begin{cases} \tilde{T}_{3i}, \text{ with prob. } 1-\tilde{\gamma}_{2i}-\tilde{\gamma}_{3i} \\ \tilde{T}_{2i} + E[T_g/H_i] \text{ , with prob. } \tilde{\gamma}_{2i} \\ \tilde{T}_{3i} + E[T_g/H_i] \text{ , with prob. } \tilde{\gamma}_{3i} \end{cases} \qquad i=0,1 \qquad (5.40)$$

Simplifying (5.40), we obtain

$$E[T_g/H_i] = \tilde{T}_{3i}\left[1 + \frac{\tilde{\gamma}_{3i}}{1 - \tilde{\gamma}_{3i} - \tilde{\gamma}_{2i}}\right] + \tilde{T}_{2i}\left[\frac{\tilde{\gamma}_{2i}}{1 - \tilde{\gamma}_{3i} - \tilde{\gamma}_{2i}}\right] \qquad (5.41)$$

To simplify the numerical computation, we can derive an upper bound on $E[T_g/H_i]$. The upper bound is arrived at by equating $\tilde{T}_{2i}$ to $\tilde{T}_{3i}$ and observing that $\tilde{T}_{3i} \geq \tilde{T}_{2i}$ where equality holds if and only if the last two

incoming local decisions are obtained at the same time. By recognizing $\tilde{T}_{3i}$ to be the expected value of the maximum of the three local test lengths, we arrive at the following inequality

$$E[T_g/H_i] \leq \frac{E[\max(T_1, T_2, T_3)/H_i]}{p_i} \tag{5.42}$$

The inequality (5.42) is tight for sufficiently small local error probabilities; in this case both $\tilde{\gamma}_{3i}$ and $\tilde{\gamma}_{2i}$ are extremely small and the average global test length is dictated by the probability of a successful trial $p_i \approx 1$.

5.5. Numerical Example

We consider the case of two identical local detectors. The observations are assumed to be iid both spatially and temporally. Let the observation model be as follows:

$$H_0: \quad X_i \equiv Y_i \sim B(P_0)$$

$$P_1 = 1-P_0 \neq 0.5$$

$$H_1: \quad X_i \equiv Y_i \sim B(P_1)$$

where $B(P_i)$ is a Bernoulli random variable with parameter P_i. Therefore, the local test can be formulated as a random walk on the states numbered 0,1,2,...,N. The states 0 and N are absorbing states with local decisions H_0 and H_1 respectively while the remainder of the states are transient states. Starting the process at the kth state, it is known [29] that

$$Pr(u_\ell = H_0/H_i) = \frac{((1-P_i)/P_i)^k - ((1-P_i)/P_i)^N}{1 - ((1-P_i)/P_i)^N}$$

and

$$Pr(u_\ell = H_1/H_i) = 1 - Pr(u_\ell = H_0/H_i)$$

In the example we choose $P_1 = 1-P_0 = 0.6667$. The values of k and N are chosen to yield the desired values of the local error probabilities α_ℓ and β_ℓ. Let $\underline{R}$ = [00 ... 1 ... 0] be a row vector of size N+1 whose elements are all zero except at the kth position. Let $\underline{C}$ = [10 .. 001] be a column vector with elements that are all zero except for 1 at the 0th and the Nth positions. Let $\underline{P}_i$ be the transition probability matrix [29] when H_i is true, i=0,1. As was pointed out in [29], the probability that the jth local test will terminate before or at the nth stage is given by

$$Pr\{T_j \leq n/H_i\} = \underline{R}\ \underline{P}_i^n \underline{C} \quad ; \quad i=0,1;\ j=1,2$$

Using the concept of tail probabilities [29], the expected value of T_j when H_i is true can be derived as follows

$$E[T_j/H_i] = \sum_{n=0}^{\infty} Pr\{T_j > n/H_i\}$$

Therefore, we can write

$$E[T_j/H_i] = \sum_{n=0}^{\infty} \{1 - \underline{R}\ \underline{P}_i^n \underline{C}\}$$

which gives the ASN of the jth local test when H_i is the true hypothesis. To derive the ASN's of the $\min(T_1,T_2)$ and the $\max(T_1,T_2)$, we observe that T_1 and T_2 are iid, and hence

$$\Pr\{\min(T_1,T_2) > n/H_i\} = \{1 - \underline{R}\ \underline{P}_i^n\underline{C}\}^2$$

$$\Pr\{\max(T_1,T_2) < n/H_i\} = [\underline{R}\ \underline{P}_i^n\underline{C}]^2$$

From those two equations and using once again the tail probabilities, we obtain

$$E[\min(T_1,T_2)/H_i] = \sum_{n=0}^{\infty} \{1 - \underline{R}\ \underline{P}_i^n\underline{C}\}^2$$

$$E[\max(T_1,T_2)/H_i] = \sum_{n=0}^{\infty} \{1 - [\underline{R}\ \underline{P}_i^n\underline{C}]^2\}$$

The numerical results obtained for a single detector, the simple multi-sensor scheme in Chapter Four, and the various fusion rules are given below for some values of α and β.

	Single Detector	The Simple Scheme	AND	OR	Three-Way
	$\alpha = \beta = 3.891 \times 10^{-3}$				
$E[T_g/H_1]$	23.81	16.89	16.44	18.56	15.98(67%)
$E[T_g/H_0]$	23.81	16.89	18.56	16.44	15.98(67%)
	$\alpha = 2.43\times10^{-4}$, $\beta = 3.905\times10^{-3}$				
$E[T_g/H_1]$	35.76	26.97	23.75	27.85	24.18(67.6%)
$E[T_g/H_0]$	24	16.94	19.32	16.59	16.25(67.7%)
	$\alpha = \beta = 1.526 \times 10^{-5}$				
$E[T_g/H_1]$	48	37.5	31.05	40.04	31(64.5%)
$E[T_g/H_0]$	48	37.5	40.04	31.05	31(64.5%)
	$\alpha = 5.96\times10^{-8}$, $\beta = 1.526\times10^{-5}$				
$E[T_g/H_1]$	71.71	58.95	44.9	61.28	44.9(62.4%)
$E[T_g/H_0]$	48	37.53	40.17	31.00	31.03(64.6%)
	$\alpha = \beta = 2.33 \times 10^{-10}$				
$E[T_g/H_1]$	96	80.79	58.45	83.54	58.45(61%)
$E[T_g/H_0]$	96	80.79	83.54	58.45	58.45(61%)

5.6. Discussion

A distributed system consisting of two local sequential detectors and a global decision maker was considered. The global error probabilities are shown to be functions of the local error probabilities as well as of the fusion rule used. The global power function was shown to be a simple function of the local power functions for a given fusion rule. For two identical local detectors, the average global test length was derived in terms of the local test lengths. The case of three local detectors was considered briefly and an upper bound was derived for the average global test length. The numerical results obtained show that the AND fusion rule performs better than the OR fusion rule when H_1 is true, and vice versa when H_0 is true. On the other hand, the Three-Way fusion rule combines the advantages of both fusion rules and performs quite well under both hypotheses, with better performance for smaller error probabilities. Moreover, the Three-Way fusion rule has better performance than the simple sequential scheme presented in Chapter Four where no explicit fusion rule was employed.

CHAPTER SIX

A MODIFIED DISTRIBUTED SEQUENTIAL HYPOTHESIS TESTING PROCEDURE WITH DATA FUSION

6.1. Introduction

In binary signal detection theory, sequential hypothesis testing procedures [1,2] provide a significant advantage over fixed-sample-size (FSS) test procedures. For the same error probabilities α and β, the sequential procedures, on the average require substantially less number of observations. However, the sequential procedures are usually not easy to implement and require a random number of observations before they terminate. In other words, the sequential procedures do not give any definite upper bound on the number of observations required for decision. Such a situation will limit the practical use of sequential procedures and necessitate truncation. The truncation problem [31-33] is not easy to analyze except for the simple procedure given in [21].

The SPRT was generalized in [25,26] for the detection of M hypotheses with different means of the normal distribution. As was pointed out in [25], the generalized SPRT is not easy to implement, and its performance is difficult to evaluate. Furthermore, it should be mentioned that the sequential test in [25] is not guaranteed to be optimal. In [27], Fleisher et al. generalized the memoryless grouped-data sequential (MLGDS) test procedure of Lee and Thomas [21] to the case of multiple hypotheses with different means/variances of the normal distribution. The sequential testing procedure in [27] exhibits a performance superiority over the

optimal FSS test while maintaining a much simpler structure and analysis than the sequential test in [25].

In Chapters Four and Five, we have studied some decentralized sequential hypothesis testing schemes. The numerical results obtained show that the scheme in Chapter Five has a better performance than that in Chapter Four. However, the derivation of the average global test length is quite laborious and no simple truncation scheme is available. Motivated by this, we consider another decentralized sequential testing procedure. The distributed sequential testing procedure is a generalization of the MLGDS procedure in [21] to a distributed environment as described and analyzed in Section 6.2, for the binary hypothesis case. In Section 6.3, we consider the case of M hypotheses. Section 6.4 contains a description and analysis of the truncation scheme applied to the binary hypothesis case. Numerical results are presented in Section 6.5, and finally this chapter is concluded in Section 6.6 where we discuss our results.

6.2. The Modified Distributed Sequential Test for Binary Hypothesis Testing

We consider a distributed system consisting of two local detectors (LDs) and a global decision maker (GDM) as shown in Fig. 6.1. The LDs are not allowed to communicate with each other. Binary hypothesis testing problem is considered with equally probable hypotheses. At any observation time, each LD takes a package of observations of size N_o, and makes a decision based on its own observations. The local decision is either in favor of one of the two underlying hypotheses, or an indecision (ignorance) is declared. The two local decisions are communicated to the

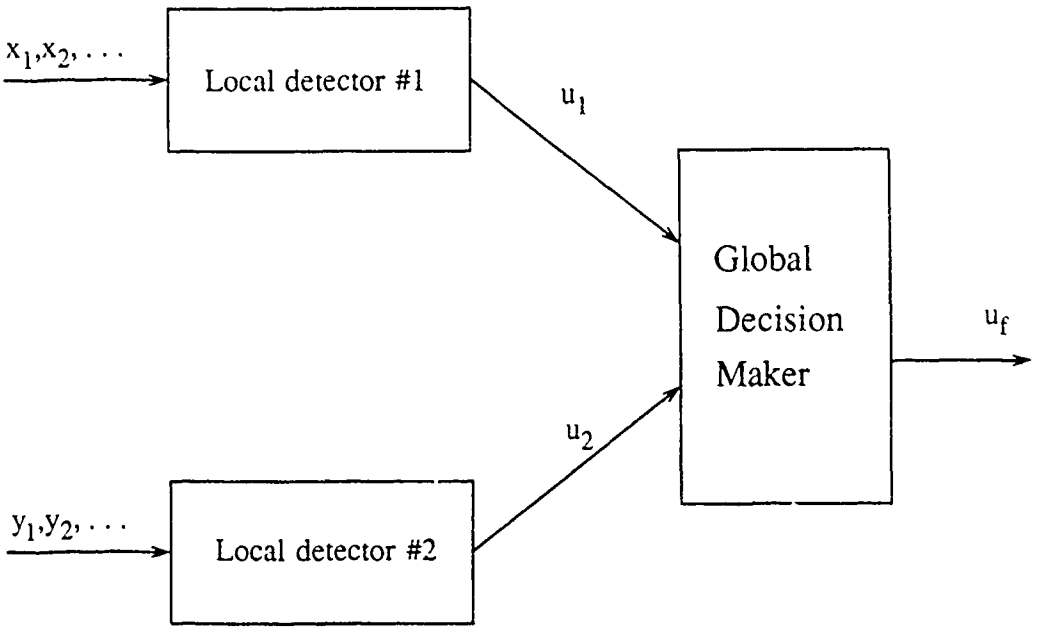

Fig. 6.1: A distributed system consisting of two local detectors and a global decision maker.

GDM, which in turn combines them according to the specified fusion rule to come up with the global decision. In this chapter, we are interested in sequential testing procedures. Therefore, the global decision is either one of the hypotheses or is to continue. When the global decision is in favor of one of the hypotheses, the local as well as the global tests are terminated. However, when the global decision is to continue, the local MLGDS tests are repeated discarding the results of all previous stages.

Let $x_1, x_2, \ldots$ be a sequence of iid random variables which represent the successive observations at local detector #1. Similarly, let $y_1, y_2, \ldots$ be a sequence of iid observations at local detector #2. The observations are assumed to be spatially iid. The hypothesis H_0 represents the absence of a signal while the alternative H_1 represents the presence of the signal. The observation model is given by

$$\begin{aligned} H_1: \quad & x_i = \theta + n_i \\ & y_i = \theta + n_i \qquad i \geq 1 \qquad (6.1) \\ H_0: \quad & x_i = n_i \\ & y_i = n_i \end{aligned}$$

where θ is the term representing the signal and n_i is the additive noise component. We assume that the noise is normally distributed with mean zero and variance σ^2, i.e., $n_i \sim N(0,\sigma^2)$. The observations are taken at the LDs in groups of size N_o each. Therefore, the test statistics at LD # 1 and LD # 2 are given by

$$T_{N_o}(\underline{x}_{N_o}) = \sum_{i=(n-1)N_o+1}^{nN_o} x_i \qquad (6.2)$$

and

$$T_{N_o}(\underline{y}_{N_o}) = \sum_{i=(n-1)N_o+1}^{nN_o} y_i \qquad (6.3)$$

respectively. Based on the local sufficient test statistics $T_{N_o}(\underline{x}_{N_o})$ and $T_{N_o}(\underline{y}_{N_o})$, the hypothesis testing problem is modeled by:

$$\begin{aligned} H_1: \quad & T_{N_o}(\underline{x}_{N_o}) \sim N(N_o\theta \ , \ N_o\sigma^2) \\ & T_{N_o}(\underline{y}_{N_o}) \sim N(N_o\theta \ , \ N_o\sigma^2) \\ H_0: \quad & T_{N_o}(\underline{x}_{N_o}) \sim N(0 \quad , \ N_o\sigma^2) \\ & T_{N_o}(\underline{y}_{N_o}) \sim N(0 \quad , \ N_o\sigma^2) \end{aligned} \tag{6.4}$$

For equal global error probabilities α and β (where α is the probability of deciding H_1 when H_0 is true, and β is the probability of deciding H_0 when H_1 is the true hypothesis), it follows from the symmetry of the problem that the local thresholds are symmetric around the point $N_o\theta/2$. Let A_j and B_j, j=1,2, be the upper and lower thresholds at the jth LD respectively. Therefore, we can write

$$\begin{aligned} A_j &= N_o\theta/2 + C_j(N_o) \\ & \qquad\qquad j = 1 \ , \ 2 \\ B_j &= N_o\theta/2 - C_j(N_o) \end{aligned} \tag{6.5}$$

where $C_j(N_o)$ is a constant which depends on the package size N_o for specified error probabilities. Following the convention in the literature [7-13], we assume identical local detectors. Under the above assumption, equation (6.5) can be rewritten as follows

$$\begin{aligned} A = A_1 = A_2 &= N_o\theta/2 + C(N_o) \\ B = B_1 = B_2 &= N_o\theta/2 - C(N_o) \end{aligned} \tag{6.6}$$

Let $p_i \triangleq \Pr\{T_{N_o}(\underline{x}_n) \geq A/H_i\}$ and $q_i \triangleq \Pr\{T_{N_o}(\underline{x}_n) \leq B/H_i\}$ for i=0,1. It follows that

$$p_1 = q_0 = \Phi(B/\sqrt{N_o}\,\sigma) \qquad (6.7)$$
$$p_0 = q_1 = 1 - \Phi\, A/\sqrt{N_o}\,\sigma)$$

where $\Phi(\bullet)$ is the cdf of the unit normal random variable N(0,1). At any stage of the test, LD #1 computes its own test statistic using only its current group of observations and performs the following test:

$$T_{N_o}(\underline{x}_{N_o}) \begin{cases} \geq A\,, & u_1 = 1 \\ \leq B\,, & u_1 = 0 \\ \text{otherwise}\,, & u_1 = I \end{cases} \qquad (6.8)$$

where u_1 denotes the decision of LD #1, and I represents indecision in the three level quantizer. The procedure at LD #2 is identical to that performed at LD #1. Therefore, the two local decisions are iid discrete random variables. In particular, the jth local decision u_j, j=1,2, under the hypothesis H_i, i = 0,1, is distributed as follows:

$$H_i : u_j = \begin{cases} 1\,, & \text{with prob. } p_i \\ 0\,, & \text{with prob. } q_i \\ I\,, & \text{with prob. } r_i \end{cases} \qquad \begin{matrix} j=1,2 \\ i=0,1 \end{matrix} \qquad (6.9)$$

where $r_i = 1 - (p_i + q_i)$. The GDM observes only the set of incoming local decisions and combines them according to some fusion rule. The fusion rule $g(u_1, u_2)$ must satisfy the monotone property of the global likelihood ratio (LR) function and admit three different courses of action (decide H_0, H_1, or continue). Motivated by the results of the previous chapter, we propose to use the following symmetric fusion rule which satisfies the above requirements

$$u_g \triangleq g(u_1,u_2) \begin{cases} u_1=u_2=1 \text{ , stop and decide } H_1 \\ u_1=u_2=0 \text{ , stop and decide } H_0 \\ \text{otherwise, discard all previous} \\ \quad \text{local decisions and continue.} \end{cases} \tag{6.10}$$

where u_g denotes the global (final) decision.

The probability of error of the first kind, α, can be computed by summing the probabilities of deciding H_1 at all stages when H_0 is the true hypothesis. The successive stages are identically distributed each with the probability of a hypothesis decision (deciding either H_0 or H_1) equal to $(p_0^2 + q_0^2)$ and a probability of continue equal to $1 - (p_0^2 + q_0^2)$. Therefore, we can write

$$\alpha = \sum_{k=1}^{\infty} \Pr\{u_g=H_1 \text{ at the kth stage}/H_0 \text{ is true}\} \cdot \Pr\{\text{the kth stage is reached}\}.$$

Simplifying, we obtain

$$\alpha = \sum_{k=1}^{\infty} p_0^2[1 - p_0^2 + q_0^2)]^{k-1} = \frac{p_0^2}{p_0^2 + q_0^2} = \frac{q_1^2}{p_0^2 + q_0^2} \tag{6.11}$$

Let Pr(e) be defined as the average probability of error, i.e.,

$$\Pr(e) = \alpha \Pr\{H_0 \text{ is true}\} + \beta \Pr\{H_1 \text{ is true}\}$$

Since $\alpha=\beta$, it follows that $\Pr(e) = \alpha = \beta$, and the probability of correct decision is $1 - \Pr(e) = [q_0^2/(p_0^2 + q_0^2)]$. Inspecting (6.11) closely, we observe that in order to minimize $\alpha = \Pr(e)$ and maximize $P_D \triangleq 1-\beta=1-\Pr(e)$ simultaneously, we must minimize their ratio, i.e., $(p_0/q_0)^2 = (p_0/p_1)^2$. Similarly, we can minimize β and maximize $(1-\alpha) = 1-\Pr(e)$ simultaneously by minimizing the ratio $(q_1/q_0)^2$. However, the quantity $(p_0/p_1)^2$ or equivalently (p_0/p_1) is minimized in the region R_1 of the local observation space such that the local LR is maximum. Similarly the quantity $(q_1/q_0)^2$ or equivalently (q_1/q_0) is minimized in the region R_0 of the

local observation space where the local LR is minimum. Because the local decisions are ternary by definition, it follows that R_I is the remainder of the observation space as shown in Fig. 6.2. Therefore, we conclude that an optimal local test must necessarily be a likelihood ratio test that partitions the observation space into three mutually exclusive regions R_0, R_I, and R_1 such that $LR(R_0) < LR(R_I) < LR(R_1)$. When the observation falls in R_i, $i = 0, I, 1$, the corresponding local decision is $u_\ell = i$; $\ell = 1,2$. The independence assumption implies that the joint LR function is the product of the local likelihood ratios LR_1 and LR_2 at LD #1 and LD #2 respectively. Moreover, the global decision maker observes only the set of incoming local decisions. Therefore, we conclude that the optimal global test (fusion rule) is a LR test performed on the set of local decisions as given by (6.10). At this stage, it should be emphasized that because the observations are normal, the tests in (6.5) and (6.6) are likelihood ratio tests, and hence, they are optimal.

Returning to equation (6.11), we observe that when H_1 is the true hypothesis, the probability that the global test will terminate at the kth stage ($k \geq 1$) is given by $[1 - (p_i^2 + q_i^2)]^{k-1} (p_i^2 + q_i^2)$. Therefore, the average sample number (ASN) is the same under both hypotheses and is given by

$$ASN = N_o \sum_{k=1}^{\infty} k[1 - (p_0^2 + q_0^2)]^{k-1} (p_0^2 + q_0^2) = \frac{N_o}{(p_0^2 + q_0^2)} \qquad (6.12)$$

It is clear from (6.11) that $p_0^2 < \alpha$, or equivalently $p_0 < \sqrt{\alpha}$ for any value of α. Moreover, (6.11) can be written as follows:

$$p_0 = \sqrt{\alpha/(1-\alpha)}\ q_0 = \sqrt{\alpha/(1-\alpha)}\ p_1 \qquad (6.13)$$

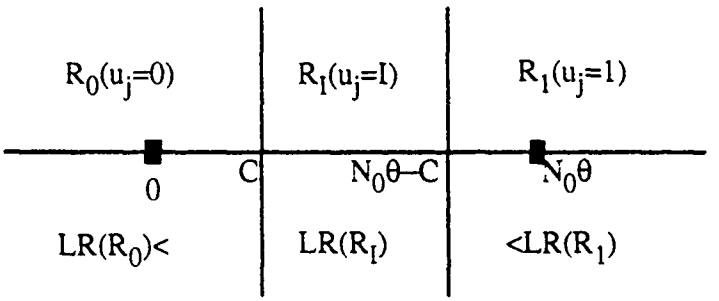

Fig. 6.2 : Partitioning of the j^{th} local observation space to minimize error probabilities.

For any value of p_0, the package size N_o is [21] given by

$$N_o = \left(\frac{\sigma}{\theta}\right)^2 \left(\Phi^{-1}(1-p_0) + \Phi^{-1}\left(\sqrt{\alpha/(1-\alpha)}\ p_0\right)\right)^2 \qquad (6.14)$$

or equivalently

$$N_o = \left(\frac{\sigma}{\theta}\right)^2 \left(\Phi^{-1}\left(\sqrt{\alpha/(1-\alpha)}\,p_0\right) - \Phi^{-1}(p_0)\right)^2 \qquad (6.15)$$

where $\Phi^{-1}(\cdot)$ is the inverse of $\Phi(\cdot)$. Substituting for N_o from (6.15) in (6.12), we obtain

$$N_o = \alpha\left(\frac{\sigma}{\theta}\right)^2 \left(\Phi^{-1}\left(\sqrt{\alpha/(1-\alpha)}\,p_0\right) - \Phi^{-1}(p_0)\right)^2 / p_0^2 \qquad (6.16)$$

The package size N_o is a monotonically increasing function of p_0 and therefore, for each value of N_o there exists a unique value of p_0 that will satisfy the error probabilities. To show that, we differentiate (6.15) with respect to p_0 to obtain

$$\frac{dN_o}{dp_0} = 2\left(\frac{\sigma}{\theta}\right)^2 \left(\Phi^{-1}\left(\sqrt{\alpha/(1-\alpha)}\ p_0 - \Phi^{-1}(p_0)\right)\right) \left(\sqrt{(1-\alpha)/\alpha} \cdot \frac{1}{\phi[\Phi^{-1}(\sqrt{(1-\alpha)/\alpha}\ p_0)]} - \frac{1}{\phi[\Phi^{-1}(p_0)]}\right) \qquad (6.17)$$

where ϕ is the unit normal density function. The term $[\Phi^{-1}\sqrt{(1-\alpha)/\alpha}\ p_0 - \Phi^{-1}(p_0)]$ is positive because $\sqrt{(1-\alpha)/\alpha}\ p_0 > p_0$ for values of α of interest ($\alpha < 0.5$). Rewriting (6.17), we obtain

$$\frac{dN_o}{dp_0} = \left\{2\left(\frac{\sigma}{\theta}\right)^2 \left(\Phi^{-1}\left(\sqrt{\alpha/(1-\alpha)}\ p_0 - \Phi^{-1}(p_0)\right)\right)\right\} \Big/ \left\{\phi[\Phi^{-1}(p_0)]\right\} \cdot \left\{\sqrt{\alpha/(1-\alpha)} \cdot \frac{\phi[\Phi^{-1}(p_0)]}{\phi[\Phi^{-1}(\sqrt{(1-\alpha)/\alpha}\ p_0)]} - 1\right\} \qquad (6.18)$$

The first term on the right hand side of (6.18) is positive because the numerator and denominator are both positive. The term $\phi[\Phi^{-1}(p_0)]/\phi[\Phi^{-1}\sqrt{(1-\alpha)/\alpha}\, p_0)]$ is recognized as the inverse of the LR function at the upper threshold A. The LR function is strictly monotonically increasing function of its argument. Therefore, it follows that

$$\frac{p_1}{p_0} \triangleq \frac{Pr\{T_{N_o}(\underline{x}_{N_o}) \geq A/H_1\}}{Pr\{T_{N_o}(\underline{x}_{N_o}) \geq A/H_0\}} > \frac{\phi[\Phi^{-1}(\sqrt{(1-\alpha)/\alpha}\, p_0)]}{\phi[\Phi^{-1}(p_0)]} \tag{6.19}$$

From (6.13), we have $p_1/p_0 = \sqrt{(1-\alpha)/\alpha}$, and thus

$$\sqrt{\alpha/(1-\alpha)} \cdot \frac{\phi[\Phi^{-1}(p_0)]}{\phi[\Phi^{-1}(\sqrt{(1-\alpha)/\alpha}\, p_0)]} - 1 > 0 \tag{6.20}$$

We conclude from (6.20) and (6.18) that $\frac{dN_o}{dp_0}$ is strictly positive for all $p_0 < \alpha$. Consequently N_o is a monotonically increasing function of p_0. The ASN as given in (6.16) is also a function of p_0 which can be written simply as follows

$$ASN = \frac{\alpha N_o}{p_0^2} \tag{6.21}$$

It is clear that an optimal choice of the package size N_o or equivalently p_0 will minimize the ASN. Let N_o^* and p_0^* be the optimal package size and its corresponding value of p_0. Differentiating (6.21) with respect to p_0 and setting the result equal to zero, we obtain

$$\frac{dN_o}{dp_0} - \frac{2N_o}{p_0} = 0 \tag{6.22}$$

Solving (6.22), we obtain the optimal package size N_o^*. However, an analytical solution for N_o^* does not appear to be feasible, and one needs to consider numerical search. In [21], it was shown numerically, that for

$10^{-18} \leq \alpha \leq 10^{-2}$, the choice $N_o \cong N_{FSS}/3$ is nearly optimal, where N_{FSS} is the number of observations required by the optimal FSS detector to satisfy the same error rates. In our case, since two local detectors are being used, we can restrict the numerical search to the set of integers $N_{FSS}/6 \leq N_o \leq N_{FSS}/3$. The upper limit $N_{FSS}/3$ represents the case when the decentralized system is equivalent to one detector (no gain in performance), while the lower limit $N_{FSS}/6$ represents the case when the decentralized system is equivalent to the centralized case (transmission of the analog observations).

6.3. The Modified Sequential Test for M-ary Hypothesis Testing

Once again we consider the distributed system shown in Fig. 6.1. The observations are as defined in the previous section except that there are M equally probable hypotheses. In particular, the observation model is as follows:

$$H_k : \begin{cases} x_i = k\theta + n_i \\ y_i = k\theta + n_i \end{cases} \quad i \geq 1 \; ; \; k = 0,1,\ldots,M-1 \qquad (6.23)$$

where $k\theta$ is the constant signal component under the hypothesis H_k and n_i is as defined earlier. The local test procedures are identical, therefore, we describe only the test procedure at LD #1. At any stage of the test, a package of N_o observations is taken and a sufficient statistic $T_{N_o}(\underline{x}_{N_o})$ is computed. Let the pdf of the test statistic $T_{N_o}(\underline{x}_{N_o})$ be denoted by $f(t/H_k)$ when H_k is the true hypothesis. The local test is obtained by generalizing the test for the binary hypotheses case as given by (6.8). It is known [34] that for the general case, where the hypotheses are not ordered, we require a set of $\frac{M(M-1)}{2}$ pairwise likelihood ratio tests. However, the hypotheses in (6.23) are ordered and, there-

fore, we need to test between adjacent hypotheses only. The resulting test can be summarized as follows: accept the hypothesis H_k if both of the following two likelihood ratio tests,

$$\Lambda_{k+1} \triangleq \frac{f(t/H_{k+1})}{f(t/H_k)} \begin{cases} \geq \tau_u = \tau & , \text{ accept } H_{k+1} \\ \leq \tau_\ell = \tau^{-1} & , \text{ accept } H_k \end{cases}$$
$$\Lambda_k \triangleq \frac{f(t/H_k)}{f(t/H_{k-1})} \begin{cases} \geq \tau_u = \tau & , \text{ accept } H_k \\ \leq \tau_\ell = \tau^{-1} & , \text{ accept } H_{k-1} \end{cases} \qquad (6.24)$$

lead to the acceptance of H_k. The choice of the thresholds in (6.24) is intended to maintain the symmetry of the local test, i.e., the local test weighs all the hypotheses equally. Let α_k and β_k be the probabilities of deciding H_{k+1} and H_{k-1} respectively, when H_k is the true hypothesis. It is clear from (6.23) and (6.24) that α_k and β_k are the same for all values of k and that $\alpha_k = \beta_k$ is maintained. For sufficiently small values of the average error probability, we can neglect the errors that can occur between nonadjacent hypotheses. In this case, it can be shown that the test in (6.24) is optimal and indeed minimizes the local average probability of error. Observe that the thresholds in (6.24) are different from those in [27]. The test in [27] uses thresholds that are identical to those of the sequential test in [25,26] and, therefore, they are not necessarily optimal. Simplifying equation (6.24), we arrive at the following equivalent tests

$$T_{N_o}(\underline{x}_{N_o}) \triangleq \sum_{i=1}^{N_o} x_i \begin{cases} \geq \frac{N_o(2k+1)\theta}{2} + \tilde{c}(N_o) & , \text{ accept } H_{k+1} \\ \leq \frac{N_o(2k+1)\theta}{2} - \tilde{c}(N_o) & , \text{ accept } H_k \end{cases}$$
$$T_{N_o}(\underline{x}_{N_o}) \triangleq \sum_{i=1}^{N_o} x_i \begin{cases} \geq \frac{N_o(2k-1)\theta}{2} + \tilde{c}(N_o) & , \text{ accept } H_k \\ \leq \frac{N_o(2k-1)\theta}{2} - \tilde{c}(N_o) & , \text{ accept } H_{k-1} \end{cases} \qquad (6.25)$$

From (6.25), it is clear that the local test is simply to determine the region of the observation space in which $T_{N_o}(\underline{x}_{N_o})$ falls and decide $H_k(u_1 = k)$ if $T_{N_o}(\underline{x}_{N_o}) \varepsilon R_k$. The regions R_k; k=0,1,...,M-1, are as summarized below:

$$\begin{aligned} R_k &= [N_o k\theta - c \ , \ N_o k\theta + c] \ ; \quad k=1,2,\ldots,M-1 \\ R_0 &= (-\infty, \ c] \\ R_{M-1} &= [(M-1) \ N_o\theta - c \ , \ \infty) \\ I_{k,k+1} &= (kN_o\theta + c \ , \ (k+1)N_o\theta - c) \ ; \quad k = 1,2,\ldots,M-2 \end{aligned} \tag{6.26}$$

where $c = N_o\theta/2 - \tilde{c}(N_o)$ is a real number whose value depends on N_o and satisfies $c \leq N_o\theta/2$. We define the indecision region I to be the union of all indecision regions $I_{k,k+1}$; k = 0,1,...,M-2. Let e(j,ℓ) be the probability that $T_{N_o}(\underline{x}_{N_o}) \varepsilon R_j$ when H_ℓ is the true hypothesis, i.e.

$$e(j,\ell) = Pr\{u_1 = j/H_\ell\} \quad j,\ell = 0,1,\ldots,M-1 \tag{6.27}$$

For small error probabilities, it follows that

$$e(j,j) >> e(j,j-1) = e(j-1,j) >> e(j,j-2) = e(j-2,j) >> \ldots \tag{6.28}$$

which states that if an error occurs, it is most likely to occur between adjacent hypotheses. A straightforward generalization of the fusion rule given by (6.10) leads to the following symmetric fusion rule

$$u_g \triangleq g(u_1,u_2) \equiv \begin{cases} u_1=u_2=k \ , & \text{stop and decide } H_k \\ \text{otherwise,} & \text{discard all previous stages and continue.} \end{cases} \tag{6.29}$$

From (6.29), it follows that the probability of error when H_k is true is given by

$$\Pr(e/H_k) = \sum_{\substack{j=0 \\ j\neq k}}^{M-1} e^2(j,k) \Big/ \sum_{j=0}^{M-1} e^2(j,k) \qquad k = 0,1,\ldots,M-1 \quad (6.30)$$

where $\sum_{j=0}^{M-1} e^2(j,k)$ is the probability of obtaining two idetical local decisions at any stage. Neglecting all errors between nonadjacent hypotheses, we can approximate $\Pr(e/H_k)$ for a middle hypothesis by

$$\Pr(e/H_k) \cong 2e^2(k-1,k)/\{e^2(k,k) + 2e^2(k-1,k)\}, \quad k=1,2,\ldots,M-2 \qquad (6.31)$$

and for the end hypotheses H_0 and H_{M-1}, we can write

$$\Pr(e/H_0) = \Pr(e/H_{M-1}) \cong e^2(1,0)/\{e^2(0,0) + e^2(1,0)\} \qquad (6.32)$$

Equations (6.31) and (6.32), show that the error probability for a middle hypothesis is approximately twice that of a boundary hypothesis which is in agreement with the results of centralized detection under the same criterion [27].

The test procedure described above is clearly memoryless because the local decisions are ignored if a hypothesis decision is not reached. Denoting by $P_T(k)$ the probability of global test termination at any stage provided that it has not terminated at any prior stage and H_k is the true hypothesis. It follows that

$$P_T(k) = \sum_{j=0}^{M-1} e^2(j,k) \ ,$$

which can be approximated as follows:

$$P_T(k) \cong \begin{cases} e^2(k,k) + 2e^2(k-1,k) & , \quad k=1,2,\ldots,M-2 \\ e^2(0,0) + e^2(1,0) & , \quad k=0 \\ e^2(M-1,M-1) + e^2(M-2,M-1) & , \quad k=M-1 \end{cases} \tag{6.33}$$

From (6.33), the average number of observations ASN(k) required for termination when H_k is true can be easily derived. The derivation is similar to the binary hypothesis case and the result is given by

$$ASN(k) = N_o/P_T(k) \tag{6.34}$$

The average error probability Pr(e) and the average sample number ASN can be derived by averaging $Pr(e/H_k)$ and ASN(k) respectively to obtain

$$Pr(e) = \sum_{k=0}^{M-1} Pr\{H_k \text{ is true}\} \cdot Pr(e/H_k) = \frac{1}{M} \sum_{k=0}^{M-1} Pr(e/H_k) \tag{6.35}$$

$$ASN = \sum_{k=0}^{M-1} Pr\{H_k \text{ is true}\} \cdot ASN(k) = \frac{N_o}{M} \sum_{k=0}^{M-1} \frac{1}{P_T(k)} \tag{6.36}$$

As in the binary hypothesis case, it seems that we cannot find the optimal package size N_o^* analytically and numerical techniques must be considered. However, it can be shown that $P_T(k)$ is a monotonically increasing function of N_o. Therefore, given the value of Pr(e), we can find the optimal package size N_o^* by keeping Pr(e) fixed while minimizing ASN over the set of possible package sizes. As is evident from the numerical results obtained, the optimal package size is less than $N_{FSS}/4$ for $Pr(e) \leq 10^{-6}$. Moreover, the ASN does not change considerably as a result of small variations in the package size, i.e., the ASN is a relatively flat concave function of N_o.

6.4. Truncation of the Test Procedure

As was pointed out in Sections 6.2 and 6.3, the number of observation

packages necessary for the termination of the global test is a geometrically distributed random variable. Therefore, the GDM will reach a decision in a finite number of trials (packages). However, the GDM may require an excessively large number of trials before a decision is reached. In other words, the number of trials required is an unbounded random variable. To avoid such undesirable situation, a reasonable truncation scheme is usually necessary. The analysis of the truncation problem associated with Wald's SPRT is difficult [31-33] and either approximate expressions or bounds are available for both the error probabilities and the ASN. The truncation problem for the generalized SPRT in [25] is even more difficult and no truncation analysis is reported. For the memoryless sequential procedures under consideration, the exact distribution of the global test duration given the hypotheses is known and simple. Therefore, the resulting truncation problem is simpler to analyze. In this section, we propose a truncation scheme and analyze its performance for the decentralized binary hypothesis testing problem. The M hypotheses case can be treated similarly.

The proposed truncation scheme can be described as follows: At any stage $k(k \leq m)$, a package of N_o observations is taken at each LD. The resulting local sufficient statistics $T_{N_o}(\underline{x}_{N_o})$ and $T_{N_o}(\underline{y}_{N_o})$ are compared with the thresholds A* and B*, where $A^* > A$ and $B^* < B$ (see definitions of A and B in equation (6.6)). The monotonicity of the LR function of the sufficient statistics guarantees lower error probabilities than α and β. The global test is carried out sequentially in the usual manner. If no decision is reached up to the mth stage, each local detector is allowed to take one more package of N_o observations. The choice of equal package sizes at all stages is necessary only to

simplify the implementation of the test. At the truncation stage, the local test statistics are computed and single threshold tests are performed as follows:

$$T_{N_o}(\underline{x}_{N_o}) \underset{u_1=0}{\overset{u_1=1}{\gtrless}} t$$

$$T_{N_o}(\underline{y}_{N_o}) \underset{u_2=0}{\overset{u_2=1}{\gtrless}} t \qquad (6.37)$$

where $u_j = i$ means that the jth, j=1,2, local decision is in favor of H_i, i=0,1. The fusion rules possible at the truncation stage are AND and OR rules. The symmetry of the problem implies that these fusion rules have identical performance at α_T and β_T, where α_T and β_T are the error probabilities at the truncation stage with $\alpha_T = \beta_T > \alpha = \beta$. We assume the AND fusion rule here, therefore, the global decision u_g is H_1 if and only if $u_1 = u_2 = 1$ and u_g is H_0 otherwise. Consequently, we may write

$$\alpha_T = \alpha_1 \cdot \alpha_2 = \alpha_\ell^2 = \{1 - \Phi\left(\frac{t}{\sqrt{N_o}\,\sigma}\right)\}^2$$

and

$$1-\beta_T = (1-\beta_1) \cdot (1-\beta_2) = (1-\beta_\ell)^2 = \Phi^2\left(\frac{N_o\theta - t}{\sqrt{N_o}\,\sigma}\right)$$

Equating α_T and β_T, it follows that the threshold t in (6.37) is the unique solution of the equation

$$\{1 - \Phi\left(\frac{t}{\sqrt{N_o}\,\sigma}\right)\}^2 = 1 - \Phi^2\left(\frac{N_o\theta - t}{\sqrt{N_o}\,\sigma}\right) \qquad (6.38)$$

Similarly, in the case of an OR fusion rule, it can be easily shown that the threshold t' is the unique solution of the equation

$$1 - \Phi^2\left(\frac{t'}{\sqrt{N_o}\,\sigma}\right) = \left\{1 - \Phi\left(\frac{N_o\theta - t'}{\sqrt{N_o}\,\sigma}\right)\right\}^2 \qquad (6.39)$$

Let $\sqrt{N_o}\theta - t' = t$, we obtain

$$1 - \Phi^2\left(\frac{N_o\theta - t}{\sqrt{N_o}\,\sigma}\right) = \left\{1 - \Phi\left(\frac{t}{\sqrt{N_o}\,\sigma}\right)\right\}^2 \qquad (6.40)$$

which is the same as equation (6.38). Therefore, we conclude that the performance of the AND and OR fusion rules is the same and that their thresholds are related by

$$t_o = \sqrt{N_o}\theta - t_a \qquad (6.41)$$

where t_o and t_a are the thresholds associated with the OR and AND fusion rules at $\alpha_T = \beta_T$ respectively.

The expressions of the global error probabilities $\bar{\alpha}$ and $\bar{\beta}$ for the proposed truncation scheme are obtained by averaging their values at all stages as follows:

$$\bar{\alpha} = \bar{\beta} = \alpha^* \cdot \Pr\{T_g \le mN_o/H_0\} + \alpha_T \cdot \Pr\{T_g > mN_o/H_0\} \qquad (6.42)$$

where T_g denotes the global test duration. Equation (6.42) can be written as

$$\bar{\alpha} = \bar{\beta} = \alpha^*[1 - (1-p_0^{*2}-p_1^{*2})^m] + \alpha_T\,[1-p_0^{*2}-p_1^{*2}]^m \qquad (6.43a)$$

$$\bar{\alpha} = \bar{\beta} = \alpha^* + (\alpha_T-\alpha^*) + [1-p_0^{*2}-p_1^{*2}]^m \qquad (6.43b)$$

from which it is obvious that by increasing m, the coefficient of $(\alpha_T-\alpha^*)$ decreases monotonically. Consequently, there exists an integer m such that $\bar{\alpha} = \bar{\beta} \le \alpha = \beta$.

The ASN of the truncated sequential procedure, denoted by ASN^T, has the same value under both hypotheses given by

$$ASN^T = N_o \sum_{k=0}^{m} k[1 - p_0^{*2} - p_1^{*2})^{k-1}] + [1 - p_0^{*2} - p_1^{*2}) \qquad (6.44)$$

Simplifying equation (6.44), we obtain

$$ASN^T = N_o \frac{1 - (1 - p_0^{*2} - p_1^{*2})^m}{(p_0^{*2} + p_1^{*2})} \qquad (6.45)$$

where m is the truncation stage. From (6.45), it is evident that ASN^T depends on the choice of N_o as well as the thresholds A* and B*. The optimization of the package size N_o for a given A* and B* is simple and can be done in the same way used for the untruncated test. On the other hand, if we fix the package size N_o, then an optimal choice of A* and B* is not simple due to the continuous nature of the variables A* and B*. The minimization of ASN^T requires an optimal choice of both N_o and the thresholds A* and B*, which is not easy to accomplish. However, it should be emphasized that the superiority of the sequential test procedure over the optimal FSS test procedure implies that the truncated sequential procedure must have a higher average sample number than its untruncated counterpart, i.e., $ASN^T > ASN$. This is so because the truncated test is actually a mixture of both the sequential and FSS test procedures. Consequently, a reasonable truncated sequential procedure is obtained when $ASN^T \cong ASN$ as demonstrated in the numerical results.

6.5. Numerical Results

In this section, we present some of the numerical results obtained.

The observation model for the binary hypothesis case is as given by (6.1) with $\theta = 0.2$ and $\sigma^2 = 1$. We denote by N_{FSS1} and N_{FSS2} the number of observations required by the optimal fixed-sample-size distributed system consisting of one and two local detectors respectively. Similarly, we denote by ASN_1 and ASN_2 the average number of observations required for the termination of the global test when one and two local detectors are employed respectively. The numerical results obtained for different values of error probabilities are summarized in Table 6.1. These results indicate that for small values of error probabilities ($\alpha=\beta$), our distributed sequential procedure is far more efficient than the distributed FSS system and that its efficiency (ASN_2/N_{FSS2}) improves monotonically as the error probabilities decrease. Moreover, our distributed procedure requires less than 60% of the ASN required by the Lee-Thomas [21] MLGDS procedure.

The observation model for the M hypotheses case is as given by (6.23) with $\theta = 0.2$, $\sigma^2 = 1$, and $M = 10$. In Table 6.2, E(n) denotes the average sample number of the sequential test in [25]. The numerical results obtained for some values of the error probability are given in Table 6.2. From these results, it is evident that the increase in the ASN due to both grouping of the observations and the memroyless nature of the test procedure is less than 25% of E(n). Moreover, the distributed system requires approximately 60% of the average number of observations (ASN_1) required for the termination of the single sensor scheme in [27], which is quite an interesting result knowing that the centralized processing necessarily requires 50% of ASN_1.

Finally, the proposed truncation scheme is applied to the binary

hypothesis case with the same parameters used in obtaining Tables 6.1. The optimization is performed by first choosing slightly smaller values of α^* and β^* ($\alpha^* = \beta^*$) than the specified values of α and β ($\alpha=\beta$). The package size N_o is then varied while adjusting the thresholds A* and B* such that α^* and β^* are fixed for all values of N_o. The resulting ASN_2^T is calculated for all values of N_o and the minimization is obtained by selecting the package size that yields the minimum of ASN_2^T. In Table 6.3, we give the values of α^*, α_T, $\bar{\alpha}$, the termination stage, ASN_2, and ASN_2^T for some values of α and β. Comparing the values of ASN_2^T and ASN_2, we observe that they are almost equal for all values of α and β ($\alpha=\beta$). The maximum difference $\delta \stackrel{\Delta}{=} ASN_2^T - ASN_2$ is obtained for $\alpha=\beta=10^{-6}$ and is equal to 2.2 observations, which corresponds to a percentage increase $(\delta/ASN_2) \times 100\% \cong 0.37\%$.

Table 6.1. Results for the binary hypotheses case, $\theta=0.2$, $\sigma^2=1$.

$\alpha=\beta$	N_{FSS1}	N_{FSS2}	ASN_1	ASN_2	ASN_1/N_{FSS1}	ASN_2/N_{FSS2}	ASN_2/ASN_1
10^{-4}	1383	967	664.2	405.7	0.48	0.42	0.611
10^{-6}	2259	1593	993.5	598	0.44	0.375	0.602
10^{-8}	3149	2231	1310.5	780.1	0.416	0.35	0.595
10^{-10}	4046	2875	1619.8	955.3	0.4	0.332	0.59
10^{-12}	4948	3529	1922.7	1125.8	0.39	0.32	0.585
10^{-14}	5853	4183	2220.8	1295.1	0.38	0.31	0.583

Table 6.2. Results for the M hypotheses case, M=10, $\theta=0.2$ and $\sigma^2=1$.

Pr(e)	N_{FSS1}	ASN_1	ASN_2	E(n)	ASN_1/N_{FSS1}	$ASN_1/E(n)$	ASN_2/ASN_1
10^{-4}	1494	823	511.6	666.5	0.551	1.235	0.622
10^{-6}	2373	1162.1	711.1	934.3	0.49	1.244	0.612
10^{-8}	3264	1489.8	899.2	2296.4	0.546	1.245	0.604
10^{-10}	4162	1806.0	1080.7	1454.8	0.434	1.241	0.598
10^{-12}	5064	2115.8	1255.8	1710.6	0.418	1.237	0.594
10^{-14}	5969	2420.3	2427.2	1964.4	0.405	1.232	0.589

Table 6.3. Results of the proposed truncation scheme, M=2,θ=0.2 and σ^2=1.

$\alpha^*=\beta^*$	$\alpha_T = \beta_T$	$\bar{\alpha} = \bar{\beta}$	Truncation Stage m	ASN_2	ASN_2^T
9.5×10^{-5}	1.755×10^{-2}	0.996×10^{-4}	6	405.7	406.8
9.473×10^{-7}	4.05×10^{-3}	0.988×10^{-6}	7	598	600.2
9.591×10^{-9}	1.22×10^{-3}	0.99×10^{-8}	9	780.1	781.5
9.573×10^{-11}	3.55×10^{-4}	0.967×10^{-10}	11	955.3	957.1
9.534×10^{-13}	1.08×10^{-4}	0.971×10^{-12}	12	1125.8	1128.2
9.478×10^{-15}	3.83×10^{-5}	0.966×10^{-14}	14	1295.1	1296.8

6.6 Discussion

In this chapter, we have extended the results of Lee and Thomas [21] to a distributed system consisting of two local detectors. The resulting sequential procedure exhibits the simplicity of the FSS test procedures while maintaining the performance superiority of the sequential procedures as is clear from Table 6.1. In addition, we have considered the case of M hypotheses without increasing the complexity of the test over the binary hypothesis case. The numerical results obtained for M hypotheses show clearly a substantial saving in the average sample number at the local level as a result of decentralization. Consequently, on the average, the decision (detection) process is much faster. The numerical results obtained for the truncation scheme indicate that its effect on the average sample number is very insignificant. Therefore, the test is practically realizable in the sense that its duration is bounded. Fin-

ally, it is worth mentioning that the distributed system can be generalized to have more than two local detectors and different fusion rules.

CHAPTER SEVEN

NONPARAMETRIC SEQUENTIAL DETECTION BASED ON MULTISENSOR DATA

7.1. Introduction

Several attempts have been made in the literature to combine the desirable properties of nonparametric fixed-sample-size (FSS) tests and the advantages of sequential tests. The Kassam-Thomas dead zone limiter (DZL) [35] is a generalization of the classical FSS nonparametric sign detector [36,37]. The DZL in [35] exhibits performance superiority over the FSS sign detector. This performance gain is achieved at the expense of additional implementation complexity. A read-only-memory (ROM) is required for the implementation of the test in [35]. For small values of error probabilities α and β, where α is the probability of error of the first kind and β is the probability of error of the second kind, the size of the ROM required is large and hence, the complexity increases as α and β decrease. The performance of the DZL in [35] has been further improved by Shin and Kassam [22] by means of sequential testing. The sequential test in [22] called conditional sequential DZL is similar to the sequential probability ratio test (SPRT) of Wald [1], and it requires ROM accessing and a two threshold test after each individual observation. This greatly increases the complexity of the test. This added complexity motivated Tantaratana and Poor [24] to consider a two- stage version of the conditional sequential DZL test in [22] to reduce the implementation complexity while retaining the performance superiority of the sequential tests. Tantaratana [23] considered the formulation of the conditional sequential DZL in [22] as a simple random walk between two fixed absorbing boundaries. His formulation does not require the table look-up oper-

ation and, therefore, the implementation of the test is much simpler. The restriction of the random walk formulation to the simple case in which the positive step is equal to the magnitude of the negative step is not suitable and leads to undesirable increase in the average sample numbers (ASN'S). In fact, it can be easily shown from the theory of random walk [29] that for small number of states as in [23], a truncation is necessary to limit α from becoming much larger than its prespecified value. However, the truncation scheme in [23] is based on the number of observations whose magnitude is larger than a fixed real number. Therefore, the truncation in [23] has no effect on limiting the actual number of observations from being excessively large. Moreover, it should be mentioned that if no truncation is employed, then the simple random walk formulation [29] requires a large number of states thereby leading to a substantial increase in the average number of observtions required for the test termination under both hypotheses.

In this chapter, we show that the conditional sequential DZL in [22] can be implemented without the need for the table look-up operation. In other words, we can sequentially test against two fixed thresholds which can be chosen prior to the test. The threshold design is the same as that of the SPRT [1]. In Section 7.3, we study the formulation of the test as a random walk without a restriction on the steps to be equal. In Section 7.4, we generalize the nonparametric sequential sign test to a distributed system consisting of two local sensors and a global decision maker as shown in Fig. 7.1. In Section 7.5, we derive a conditional sequential DZL detector for the distributed system in Fig. 7.1. It is shown that if the observations are spatially independent and have identical distributions, and that no excess over the thresholds is assumed, the resulting tests

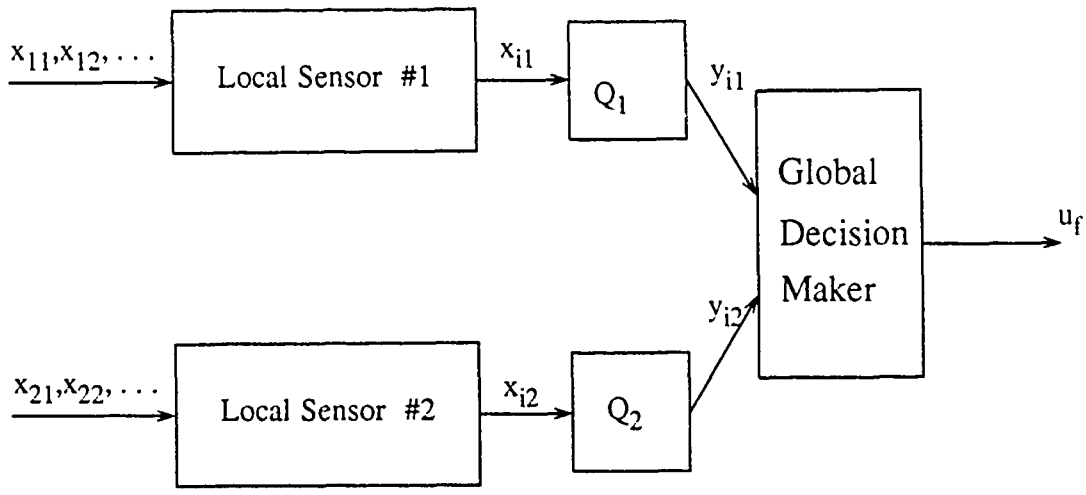

Fig. 7.1: A distributed system consisting of two local sensors followed by two local quantizers and a global decision maker.

require half the number of observations required for the single detector in [22]. In Section 7.6, we present a numerical example for the nonparametric sequential sign test. Finally, in Section 7.7, we conclude the chapter with a discussion of the results obtained.

7.2. The Nonparametric Sequential Sign and Conditional Sign Tests

Consider the problem of detecting a constant signal θ corrupted by an additive noise. The noise pdf is assumed to be symmetric, continuous, and with zero mean. Let $X_1, X_2, \ldots$ be a sequence of iid ramdom variables which represent the successive observations. This detection problem is a problem of testing the null hypothesis H_0 against the location alternative H_1, where

$$H_0 : X_i \sim f(x)$$
$$H_1 : X_i \sim f(x-\theta) , \quad \theta > 0 \tag{7.1}$$

with $f(x) = f(-x)$. We denote by $x_1, x_2, \ldots$ a realization of the observations $X_1, X_2, \ldots$. Let $\underline{x}_n \triangleq [x_1\ x_2\ \ldots\ x_n]$ be a vector of observations of size n.

The sequential sign test in [36] is a SPRT performed on the sign of the observations. Each observation x_i $(i \geq 1)$ is passed through a hard limiter whose output is given by

$$Y_i = \begin{cases} +1 & \text{if } x_i \geq 0 \\ -1 & \text{if } x_i < 0 \end{cases} \tag{7.2}$$

Let

$$p_\theta = Pr(Y_i = +1/\theta) = 1 - F(-\theta) \tag{7.3a}$$

$$q_\theta = Pr(Y_i = -1/\theta) = F(-\theta) \tag{7.3b}$$

where $F(x) = \int_{-\infty}^{x} f(t)dt$. It follows that for any pdf $f(x)$ as defined earlier, the hypothesis testing problem is now given by

$$H_0 : Y_i \sim B(p_\theta = 1/2)$$

$$H_1 : Y_i \sim B(p_\theta > 1/2) \tag{7.4}$$

where $B(p)$ is a Bernoulli random variable with probability of success (Y_i=+1) equal to p, and probability of failure (Y_i= -1) equal to 1-p=q. When $H_0(\theta = 0)$ is the true hypothesis, the distribution of the quantized observations is invariant in the class C of continuous, symmetric, and zero mean pdf's. It follows that a SPRT based on Y_i's is nonparametric [36], in the sense that it has a fixed α for any X εC. Following Wald [1], the SPRT based on the Y_i's is as follows.

$$\sum_{i=1}^{n} Z_i \begin{cases} \geq \log (1-\beta)/\alpha \triangleq \log t_u \text{ , decide } H_1 \\ \leq \log \beta/(1-\alpha) \triangleq \log t_\ell \text{ , decide } H_0 \\ \text{otherwise, continue.} \end{cases} \tag{7.5}$$

where $Z_i = \log \dfrac{f(y_i/H_1)}{f(y_i/H_0)}$, $f(y_i/H_j)$ is the probability density function of the discrete random variable Y_i when the hypothesis H_j, j=0,1, is true, and t_u and t_ℓ are the upper and lower thresholds respectively. Let n_p be the number of positive observations at stage n, and let $\tilde{p}_\theta$ be an estimate of p_θ which we use in designing the test. The random variable Z_i in (7.5) assumes the value $\log(\tilde{p}_\theta/(1/2))$ when Y_i=+1 and the value $\log((1-\tilde{p}_\theta)/(1/2))$ when y_i = -1. Therefore, equation (7.5) can be written explicitly as

$$n_p \log(2\tilde{p}_\theta) + (n-n_p) \log 2(1-\tilde{p}_\theta) \begin{cases} \geq \log t_u \text{ , decide } H_1 \\ \leq \log t_\ell \text{ , decide } H_0 \\ \text{otherwise, continue.} \end{cases} \tag{7.6}$$

where n_p is the number of observations such that Y_i is positive and n is the total number of observations taken so far. Neglecting any excess over the test thresholds upon termination. The average sample number (ASN_θ), defined as the average number of observations required for termination when θ is the true parameter is approximated [1] by

$$ASN_\theta \cong \frac{L(\theta)\ \log t_\ell + [1-L(\theta)]\log t_u}{E[Z_i/\theta]} \tag{7.7}$$

where $L(\theta)$ is the operating characteristic (OC) function of the test, and $1-L(\theta)$ is the power function of the test. To obtain $L(\theta)$, we solve the parametric equation

$$\sum_{y_i} \left[\frac{f(y_i/H_1)}{f(y_i/H_0)}\right]^{h(\theta)} f(y_i/\theta) = 1 \tag{7.8}$$

where the summation is over all possible values of y_i, and $f(y_i/\theta)$ is the pdf when θ is the true parameter. Wald [1] has shown that $h(\theta)$ is unique and that $L(\theta)$ is given in terms of $h(\theta)$ by

$$L(\theta) = \frac{t_u^{h(\theta)} - 1}{t_u^{h(\theta)} - t_\ell^{h(\theta)}} \tag{7.9}$$

In our particular case, the random variable Y_i assumes the values ± 1 only. Therefore, we can write equation (7.8) as follows:

$$p_\theta(2\tilde{p}_\theta)^{h(\theta)} + q_\theta(2(1-\tilde{p}_\theta))^{h(\theta)} = 1 \tag{7.10}$$

where $p_\theta = Pr[Y_i = 1/\theta]$ and $q_\theta = 1-p_\theta$. From the definition of Z_i in (7.5), it is clear that $E[Z_i/\theta)]$ is given by

$$E[Z_i/\theta)] = p_\theta \cdot \log 2\tilde{p}_\theta + q_\theta \log 2(1-\tilde{p}_\theta) \qquad (7.11)$$

when H_0 is true ($\theta=0$), it follows that $p_0 = q_0 = 1/2$. Therefore, $h(0)=1$ and $L(0)=1-\alpha$. Similarly, when $p_\theta = \tilde{p}_\theta$ (our estimated value), it follows that $h(\tilde{\theta}) = -1$ and $L(\tilde{\theta}) = \beta$, which means that the power of the test is $1-\beta$ as required.

The conditional sequential DZL detector is similar to the sequential sign detector. The only difference is that it employs the following DZL quantizer

$$Y_i = \begin{cases} +1, & \text{if } x_i > c \\ 0, & \text{if } -c \le x_i \le c \\ -1, & \text{elsewhere} \end{cases} \qquad (7.12)$$

Let

$$p_\theta = Pr(Y_i = +1/\theta) = 1 - F(c-\theta) \qquad (7.13a)$$

$$q_\theta = Pr(Y_i = -1/\theta) = F(-c-\theta) \qquad (7.13b)$$

$$r_\theta = Pr(Y_i = 0/\theta) = 1 - p_\theta - q_\theta \qquad (7.13c)$$

When H_0 is tr e, ($\theta=0$), it follows that $p_0 = q_0$ for all $X \varepsilon C$. However, the actual value of p_0 depends on $f(x)$ as given by (7.13a). In order to obtain a nonparametric test we must base our test on a random variable whose distribution is invariant to $f(x)$ when H_0 is true. Let us define W_i as follows

$$W_i = sgn(Y_i/Y_i = \pm 1) \qquad (7.14)$$

Clearly, W_i is a random variable defined on the set of observations such that $Y_i = \pm 1$. Given that an observation Y_i is in that set, the conditional distribution of W_i given the hypothesis is as follows

$$H_0 : W_i \sim B(p_0' = 1/2)$$

$$H_1 : W_i \sim B(p_\theta' > 1/2) \qquad (7.15)$$

where $p_\theta' = p_\theta/(p_\theta+q_\theta)$. When H_0 is true ($\theta=0$), it is clear that $p_0/(p_0+q_0) = 1/2$ and consequently, the test based on W_i is nonparametric. The SPRT based on W_i's is similar to the sign test, the only difference is that when $Y_i = 0$, the observation is ignored. Let m be the number of observations of W_i, i.e., m is the number of observations such that $Y_i = \pm 1$. It follows that the ASN_θ expressed in terms of m is given by [1]

$$ASN_\theta(m) = \frac{L(\theta)\ \log t_\ell + [1-L(\theta)]\log t_u}{E[Z_i/\theta]} \qquad (7.16)$$

where $Z_i = \log \frac{f(w_i/H_1)}{f(w_i/H_0)}$, and $L(\theta)$ is computed using (7.7) and (7.9) after replacing y_i with w_i. Observe that the observations $Y_i = 0$ have no effect on the power of the test simply because they are not included in the computation. However, those observations have a major effect on the ASN_θ when computed on the actual (unconditional) observations. Observe that $ASN_\theta(m)$ is the average sample number conditioned on the event $Y_i=\pm1$. To derive the actual average sample number ASN_θ, we observe that the event $Y_i=\pm1$ has a probability of occurrence equal to $(p_\theta + q_\theta)$ for any individual observation. Therefore, for any vlue of $ASN_\theta(m)$, say k, the distribution of the actual number of observations N is the Pascal (negative binomial) distribution as given by

$$Pr(N=n/ASN_\theta(m)=k) = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} (p_\theta+q_\theta)^k\ (1-p_\theta-q_\theta)^{n-k},\ n=k,k+1\ldots \qquad (7.17)$$

The actual average sample number ASN_θ is obtained by averaging over all possible values of N. The result is known and is given by

$$ASN_\theta = \frac{ASN_\theta(m)}{(p_\theta + q_\theta)} \tag{7.18}$$

7.3. Random Walk Formulation of the Tests

Both the nonparametric sequential sign and conditional sign tests described in Section 7.2 employ a two-valued discrete random variable for testing as is evident from (7.4) and (7.15). It follows that these sequential tests can be easily implemented as a random walk on a finite number of states as described in Chapter Two. Tantaratana [24] considered the case in which the nonparametric sequential conditional sign test is formulated as a simple random walk on the states $S_1,\ldots,S_N$ as shown in Fig. 7.2. The states S_0 and S_N are absorbing states with associated decisions H_0 and H_1 respectively. The states $S_0, S_1,\ldots,S_{N-1}$ are all transient with an associated decision of continue in an analogy with Wald's SPRT [1]. Initially, the process starts at state S_Z, $0<Z<N$, and it goes one step higher ($\Delta^+ = 1$) if the current observation is positive and one step lower ($\Delta^- = 1$) if the observation is negative. As soon as the process reaches one of the absorbing states S_0 or S_N, the test terminates with the acceptance of the hypothesis associated with that state. The results for equal step random walk [29] can be used in conjunction with the observation model of (7.4) and (7.15). It is well known [29] that for a process starting at S_Z, the probability of absorption in state S_0 is given by

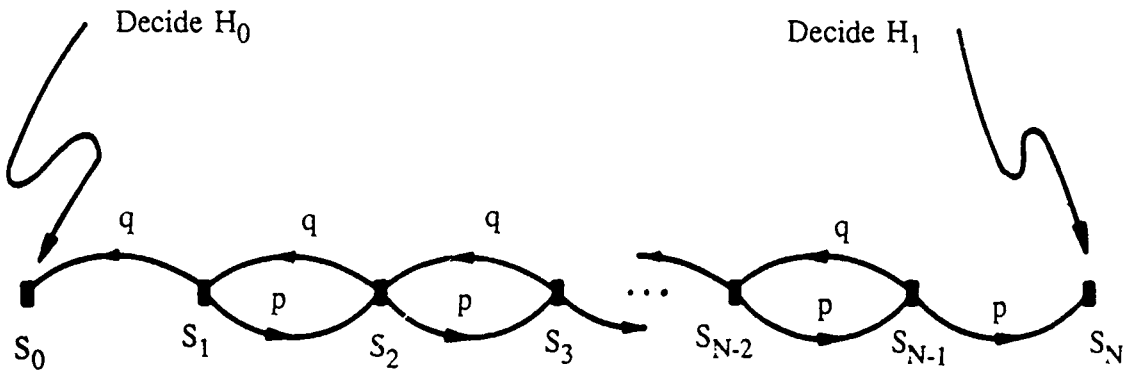

Fig. 7.2: An (N+1) state random walk process with $\Delta^+ = \Delta^- = 1$.

$$u_Z = \frac{(q/p)^Z - (q/p)^N}{1 - (q/p)^N} \tag{7.19}$$

where $q = Pr\{\Delta^-\}$ and $p = Pr\{\Delta^+\}$. When H_0 is true, we have $q_0 = p_0 = 1/2$, therefore, we obtain

$$\alpha = 1 - \lim_{q/p \to 1} u_Z = \frac{Z}{N} \tag{7.20}$$

On the other hand, when H_1 is true, we substitute for q and p their estimated (nominal) values $\tilde{q}_\theta$ and $\tilde{p}_\theta$ from (7.6), to obtain

$$\beta = \frac{(\tilde{q}_\theta/\tilde{p}_\theta)^Z - (\tilde{q}_\theta/\tilde{p}_\theta)^N}{1 - (\tilde{q}_\theta/\tilde{p}_\theta)^N} \tag{7.21}$$

The random walk formulation of the test requires the specification of both Z and N such that α and β are satisfied at the nominal parameter values. However, it is obvious from (7.20) that for Z=1, we must have $N=\alpha^{-1}$, therefore, the number of states necessary to satisfy α is extremely large for small values of α. On the other hand, for sufficiently large N, we can neglect the term $(\tilde{q}_\theta/\tilde{p}_\theta)^N$ in equation (7.21) to obtain

$$\beta \cong (\tilde{q}_\theta/\tilde{p}_\theta)^Z \tag{7.22}$$

It follows from (7.22) that $Z \cong \log(\beta^{-1})/\log(\tilde{q}_\theta/\tilde{p}_\theta)$, and thus, Z must be greater than one, in general, yielding a further increase in the number (N+1) of required states. Consequently, the average sample number, which is directly related to N, is very large. To overcome this highly undesirable property while still maintaining implementation and analysis

simplicity, we propose to allow the magnitude of the negative step to be twice that of the positive step, i.e., $\Delta^- = 2$, and $\Delta^+ = 1$ as shown in Fig. 7.3. According to Fig. 7.3, the states S_0 and S_1 are absorbing states and associated with H_0 decision while S_N is an absorbing state associated with H_1 decision.

Starting the process in state S_j, $1 < j < N$, we can express the probability of absorption in state S_N as follows:

$$\begin{aligned} u_j &= pu_{j+1} + qu_{j-2} \\ u_N &= 1 \\ u_0 &= u_1 = 0 \end{aligned} \tag{7.23}$$

where u_j, $0 < j < N$, is the probability of absorption in state S_N starting from the state S_j and $p(q)$ is the probability of positive (negative) step. Using the method of particular solutions, we obtain

$$\begin{aligned} u_j &= A + B(x_1)^j + C(x_2)^j \\ u_N &= 1 \\ u_0 &= u_1 = 0 \end{aligned} \tag{7.24}$$

where A,B, and C are constants to be determined from the boundary conditions u_0, u_1, and u_N, while x_1 and x_2 are the roots of the second order homogeneous equation

$$px^2 - qx - q = 0 \tag{7.25}$$

Solving (7.25), we obtain $x_1=(q + \sqrt{q^2 + 4pq})/2p$ and $x_2=(q - \sqrt{q^2 + 4pq})/2p$. Substituting the boundary conditions in (7.24), we obtain for j=2 that

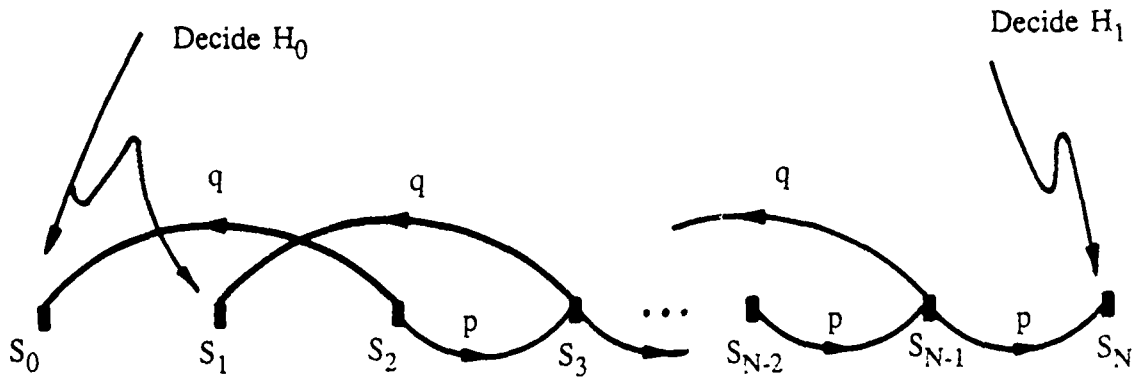

Fig. 7.3: An (N+1) state random walk process with $\Delta^+ = 1$ and $\Delta^- = 2$.

$$u_Z = \frac{(x_1-x_2)/(1-x_1) - ((1-x_2)/(1-x_1))\ x_1^Z + x_2^Z}{(x_1-x_2)/(1-x_1) - ((1-x_2)/(1-x_1))\ x_1^N + x_2^N} \qquad (7.26)$$

Let ξ_i, i=0,1,N, denote the probability of absorption in the state S_i starting from the state S_Z. Therefore, ξ_N is equal to u_Z as given by (7.26). The probability of absorption in the state S_0 is obtained by solving the following stochastic difference equation

$$\begin{aligned} u_j &= pu_{j+1} + qu_{j-2} \\ u_0 &= 1 \\ u_1 &= u_N = 0 \end{aligned} \qquad (7.27)$$

whose solution is similar to that of (7.23) and the result is given by

$$\xi_0 = \frac{x_2(x_1^{N-1}-x_2^{N-1})/(1-x_1^{N-1}) - ((x_2)/(x_1))(1-x_2^{N-1})/(1-x_1^{N-1})x_1^Z + x_2^Z}{x_2(x_1^{N-1}-x_2^{N-1})/(1-x_1^{N-1}) - ((x_2)/(x_1))(1-x_2^{N-1})/((1-x_1^{N-1}) + 1} \qquad (7.28)$$

Because the process will ultimately be absorbed in one of the absorbing states, it follows that

$$\xi_1 = 1 - \xi_0 - \xi_N \qquad (7.29)$$

The average number of steps until absorption when θ is the actual parameter is denoted by ASN_θ and is given by [30] by

$$ASN_\theta = \frac{\sum_i \xi_i(i-Z)}{p_\theta\Delta^+ - q_\theta\Delta^-} = \frac{\sum_i \xi_i(i-Z)}{3p_\theta - 2} \ ;\ i=0,\ 1,N \qquad (7.30)$$

where p_θ and q_θ are as defined in (7.13)

The number of states N+1 and the starting point Z must be chosen such that when H_0 is true ($p_\theta = 1/2$), we have $u_N \leq \alpha$ and when H_1 is true ($p_\theta = \tilde{p}_\theta$), we have $\xi_N \geq 1-\beta$. Although we have so far considered a single

detector case, our results can be extended to the case of two detectors in a straightforward manner. The only problem encountered is that the roots of the stochastic difference equation governing the random walk can not be obtained analytically and, therefore, numerical solutions must be considered. Equation (7.17) shows that the actual ASN_θ is a function of the pdf of the observations and the parameter of the DZL, which can be designed to satisfy some criterion of optimality [22] without affecting the nonparametric property of the test. It should be emphasized that the test is performed sequentially on the set of observations $|X_i| > c$ in exactly the same way the sign test is performed, without any need for the table look-up as it was originally required by Shin and Kassam in [22].

7.4. Distributed Nonparametric Sequential Sign Test

In this section, we generalize the nonparametric sequential sign test in [22] to a two-sensor distributed environment. The system is as shown in Fig. 7.1. It is assumed that the observations at the kth sensor, k=1,2, are iid random variables with a continuous, symmetric, and zero mean pdf $f_k(x)$. The observations are independent from sensor to sensor, i.e., X_{i1} is independent of $X_{\ell 2}$ for all i,ℓ. The detection problem is, therefore, a hypothesis testing problem in which the hypothesis H_0 is tested versus the hypothesis H_1 according to the following model

$$\begin{aligned} H_0 &: X_{ik} \sim f_k(x) \\ & \qquad\qquad k=1,2 \qquad\qquad (7.31) \\ H_1 &: X_{ik} \sim f_k(x-\theta_k) \end{aligned}$$

with $f_k(x) = f_k(-x)$, and $\theta_k > 0$, k=1,2. Each local sensor quantizes its

own observation into two levels as given by equation (7.2), and communicates the quantized output to the central detector. The central detector combines the quantized observations and performs the SPRT. From the central processor point of view, the hypothesis testing problem in (7.31) is as follows:

$$H_0 : Y_{ik} \sim B(1/2)$$

$$k=1,2 \qquad (7.32)$$

$$H_1 : Y_{ik} \sim B(p_k)$$

Let

$$p_k = Pr(Y_{ik} = +1/\theta_k) = 1 - F_k(-\theta_k) \qquad (7.33a)$$

$$q_k = Pr(Y_{ik} = -1/\theta_k) = 1 - F_k(-\theta_k) \qquad (7.33b)$$

where $F_k(x) = \int_{-\infty}^{x} f_k(t)dt$. Observe that for all $X_k \varepsilon C$, the distribution Y_{ik} is invariant under the hypothesis H_0. Let $\underline{V}_i$ be an observation vector with Y_{i1} and Y_{i2} as elements, i.e.,

$$\underline{V}_i^T = (Y_{i1}\ Y_{i2}) \qquad (7.34)$$

Let $f(\underline{v}_i/H_j)$ be the pdf of $\underline{V}_i$ when H_j, j=0,1 is true. Then, we have

$$f(\underline{v}_i/H_j) = f(y_{i1}/H_j) \cdot f(y_{i2}/H_j) \quad , \ j=0,1 \qquad (7.35)$$

It should be emphasized that the distribution of $\underline{V}_i$ when H_0 is true is also invariant to changes in the pdf's of the observations X_{ik}, k=1,2 as long as $X_{ik} \varepsilon C$. Consequently the SPRT performed on $\underline{V}_i$ is nonparametric (constant α). Let $\underline{V}_1$, $\underline{V}_2$,... be a sequence of successive central obser-

vations. The SPRT at the central level in [1] as follows:

$$\sum_{i=1}^{n} Z_i \begin{cases} \geq \log t_u \text{ , decide } H_1 \\ \leq \log t_\ell \text{ , decide } H_0 \\ \text{otherwise, continue.} \end{cases} \tag{7.36}$$

where $Z_i \triangleq \log [f(\underline{v}_i/H_1)/f(\underline{v}_i/H_0)]$. From (7.36), it follows that

$$Z_i = \sum_{k=1}^{n} \log[f(y_{ik}/H_1)/f(y_{ik}/H_0)] \tag{7.37a}$$

or,

$$Z_i = Z_{i1} + Z_{i2} \tag{7.37b}$$

From (7.37), we observe that

$$E[Z_i/\theta_1,\theta_2] = E[Z_{i1}/\theta_1] + E[Z_{i2}/\theta_2] \tag{7.38}$$

Following Wald [1] in assuming no excess over the thresholds at termination, the $ASN(\theta_1,\theta_2)$ is given by

$$ASN(\theta_1,\theta_2) \cong \frac{L(\theta_1,\theta_2) \log t_\ell + [1-L(\theta_1,\theta_2)] \log t_u}{E[Z_i/\theta_1,\theta_2]} \tag{7.39}$$

where $L(\theta_1,\theta_2)$ is the OC function, which is now a function of θ_1 and θ_2. To obtain $L(\theta_1,\theta_2)$, we need to solve equation (7.9) for $h(\theta) \triangleq h(\theta_1,\theta_2) \triangleq h$. However, $h(\theta_1,\theta_2) \triangleq h$ is obtained as the unique nonzero solution of equation (7.8) which can be written explicily in terms of the four possible values of Z_i as follows:

$$p_1p_2[4\tilde{p}_1\tilde{p}_2]^h + p_1q_2[4\tilde{p}_1\tilde{q}_2]^h + q_1p_2[4\tilde{q}_1\tilde{p}_2]^h + q_1q_2[4\tilde{q}_1\tilde{q}_2]^h = 1 \tag{7.40}$$

where $\tilde{p}_k$ and $\tilde{q}_k \triangleq 1-\tilde{p}_k$ are the estimated values used in designing the steps of the random process $\sum_{i=1}^{n} Z_i$, while p_k and $q_k = 1-p_k$ are the

actual parameters. Once $h(\theta_1,\theta_2)$ is computed, we can solve (7.9) to obtain $L(\theta_1,\theta_2)$. From (7.40) we observe when H_0 is true ($\theta_1=\theta_2=0$), $h(0,0) = 1$. Therefore, $L(0,0) = 1-\alpha$ as expected. Similarly when $\theta_1=\tilde{\theta}_1$ and $\theta_2=\tilde{\theta}_2$ (the actual parameters equal to their estimates), then $h(\tilde{\theta}_1,\tilde{\theta}_2) = 1$ Therefore, $h(\tilde{\theta}_1,\tilde{\theta}_2) = \beta$.

Finally, if the observations X_{i1} and $X_{i2}(i \geq 1)$ are independent and identical, then equation (7.38) is as follows

$$E[Z_i/\theta_1=\theta_2] = 2\ E[Z_{i1}/\theta_1] \tag{7.41}$$

Consequently, it follows from (7.41) that for the same power of the test, the ASN $(\theta_1,\theta_2:\theta_1 = \theta_2)$ is exactly half the ASN_θ in the case of one sensor (detector). However, this result is drawn under the assumption of no excess over the thresholds which is true only for vanishingly small signal strengths $(\theta_1=\theta_2 \cong 0)$. Moreover, it should be mentioned that when X_{i1} and X_{i2} are iid, the computation of the OC function is identical to the single detector case as given by (7.8).

7.5. Distributed Conditional Sequential Sign Test

We consider generalizing the conditional sequential DZL detector of Shin and Kassam [22] to the distributed system of two sensors shown in Fig. 7.1. The observations are as defined earlier in Section 7.4. The only difference from the sign detector is that the quantizers are now DZL's as defined in (7.12). For clarity, we rewrite (7.12) and (7.13) to obtain.

$$Y_{ik} = \begin{cases} +1\ , & \text{if } x_{ik} > c \\ 0\ , & \text{if } -c \leq x_{ik} \leq c\ , \quad k=1,2 \\ -1\ , & \text{elsewhere} \end{cases} \tag{7.42}$$

and

$$p_{\theta k} = Pr\{Y_{ik} = +1/\theta_k\} = 1 - F_k(c-\theta_k) \tag{7.43a}$$

$$q_{\theta k} = Pr\{Y_{ik} = -1/\theta_k\} = F_k(c-\theta_k) \tag{7.43b}$$

$$r_{\theta k} = Pr\{Y_{ik} = 0/\theta_k\} = 1 - p_{\theta k} - q_{\theta k} \tag{7.43c}$$

When H_0 is true $\theta_1=0$, $\theta_2=0$, it is clear that $p_{01} = q_{01}$ and $p_{02} = q_{02}$ for all $X \in C$, with actual values being functions of the actual pdf's as is obvious from (7.43). In order to obtain a nonparametric test, we need to think in terms of conditional tests. To this end, we arrange the observations of the central processor into three groups as follows:

i) Group 1: contains all possible combinations of Y_{i1} and Y_{i2} such that $Y_{i1} = \pm 1$ and $Y_{i2} = \pm 1$. In this group, it follows from the independence assumption that

$$Pr(Y_{i1} = \eta,\ Y_{i2} = \nu/Y_{i1} = \pm 1 \text{ and } Y_{i2} = \pm 1) = \tag{7.44}$$

$$Pr(Y_{i1} = \eta/Y_{i1} = \pm 1) \cdot Pr(Y_{i2} = \nu/Y_{i2} = \pm 1)$$

where $\eta = \pm 1$ and $\nu = \pm 1$. When H_0 is the true hypothesis, equation (7.44) reduces to

$$Pr(Y_{i1} = \eta,\ Y_{i2} = \nu/Y_{i1} = \pm 1 \text{ and } Y_{i2} = \pm 1) = 1/4 \tag{7.45}$$

for any of the four possible events in the group. Similarly, when H_1 is true, we have

$$Pr(Y_{i1} = \eta,\ Y_{i2} = \nu/Y_{i1} = \pm 1 \text{ and } Y_{i2} = \pm 1) =$$

$$\frac{Pr(Y_{i1} = \eta/H_1)}{\tilde{p}_{\theta 1} + \tilde{q}_{\theta 1}} \cdot \frac{Pr(Y_{i2} = \nu/H_1)}{\tilde{p}_{\theta 2} + \tilde{q}_{\theta 2}} \tag{7.46}$$

where $\tilde{p}_{\theta k}$ and $\tilde{q}_{\theta k}$ are the estimated values of the parameters $p_{\theta k}$ and $q_{\theta k}$ as defined in (7.43).

ii) Group 2: contains all possible combinations of Y_{i1} and Y_{i2} such that one and only one of them is zero. When $Y_{i2} = 0$, we obtain

$$\Pr(Y_{i1} = \eta,\ Y_{i2} = 0 / Y_{i1} = \pm 1 \text{ and } Y_{i2} = 0) = \Pr(Y_{i1} = \eta / Y_{i1} = \pm 1) \qquad (7.47)$$

Similarly, when $Y_{i1} = 0$, we obtain

$$\Pr(Y_{i1} = 0,\ Y_{i2} = \nu / Y_{i1} = 0 \text{ and } Y_{i2} = \pm 1) = \Pr(Y_{i2} = \nu / Y_{i2} = \pm 1) \qquad (7.48)$$

It is interesting to observe that when one of the sensors has its observation in the dead-zone region i.e. $|X_{ik}| < c$, we obtain the desired conditional distribution by simply not counting that sensor (ignoring its current observation). From the results of the single detector case [22], the test is still nonparametric.

iii) Group 3: contains the observation $Y_{i1} = Y_{i2} = 0$, which must be ignored.

From (7.44) - (7.48), it follows that the increments by which the logarithm of the likelihood ratio function is updated depend on the random variables Y_{i1} and Y_{i2}. When Y_{i1} and Y_{i2} are in group 1, the increment is a random variable given by

$$z_i = \log \frac{2\Pr(Y_{i1} = \eta / H_1)}{\tilde{p}_{\theta 1} + \tilde{q}_{\theta 1}} + \log \frac{2\Pr(Y_{i2} = \eta / H_1)}{\tilde{p}_{\theta 2} + \tilde{q}_{\theta 2}} \quad \text{when } Y_{i1} = \eta \text{ and } Y_{i2} = \nu \qquad (7.49)$$

However, when $Y_{i1} = 0$ and $Y_{i2} = \nu$, the increment is given by the second term in (7.49) only, and when $Y_{i1} = \eta$ and $Y_{i2} = 0$, the increment is given by the first term of (7.49) only. Therefore, the observations as seen at the central level are not identical at all stages of the test. Consequently, it is difficult if not impossible [3] to solve for the $ASN(\theta_1,\theta_2)$. Moreover the computation of the OC function is also difficult, because upon termination we do not know the exact number of usable observations obtained by any sensor. To overcome this difficulty, we assume that the observations are identical at both sensors, and propose to perform the central test by considering Y_{i1} first, and then Y_{i2}, at all stages ($i \geq 1$). In this case, the OC function is the same as for the single detector case previously described. The ASN_θ is given by the following inequality

$$\frac{1}{2} ASN_\theta^1 \leq ASN_\theta \leq \frac{1}{2} [ASN_\theta^1 + 1] \qquad (7.50)$$

where ASN_θ^1 is the average sample number for a single detector when θ is the actual parameter.

7.6. Numerical Results

We first present an example of the random walk formulation of the sequential test discussed in Section 7.3. The nominal observation model is given by:

$$H_0 : X_i \sim N(0, \sigma^2)$$

$$, \quad i \geq 1$$

$$H_1 : X_i \sim N(\theta, \sigma^2)$$

where $N(\theta, \sigma^2)$ is a Gaussian random variable with mean θ and variance σ^2, The nominal value of θ is assumed to be $\tilde{\theta} = 0.676$ while the variance σ^2 is assumed to be unity. The quantizer nonlinearity is as given by (7.12)

with an optimal [22] value of c = 0.6, which yields a nominal value of p'_θ given by $\tilde{p}'_\theta = 0.84$. Using Wald's approximation to the thresholds, we obtain

$$N-Z \cong \ln\left(\frac{1-\beta}{\alpha}\right)/\ln 2\tilde{p}'_\theta$$

and

$$Z \cong \ln\left(\frac{\beta}{1-\alpha}\right) \Big/ \frac{1}{2} \ln 2(1-\tilde{p}'_\theta)$$

For $\alpha = \beta = 10^{-3}$, we obtain $N-Z \cong 13.3$ and $Z = 12.12$. The actual values required to satisfy $\alpha \leq 10^{-3}$ and $\beta \leq 10^{-3}$ at the nominal value of θ are $N = 27$ and $Z = 12$, which are obtained by an exact solution of (7.24). Presented in Table 7.1 are the numerical values of the power function $P_D(\theta)$ and the average sample number $ASN(\theta)$ obtained for some value of θ in the range $0 \leq \theta \leq \tilde{\theta}$, i.e., $0.5 \leq p'_\theta \leq \tilde{p}'_\theta$.

Table 7.1: Performance of the sequential DZL detector when implemented as a random walk

θ	p'_θ	$P_D(\theta)$	ASN(θ)
0.676	0.84	0.99903	45.62
0.619	0.82	0.9974	52.44
0.567	0.8	0.9936	60.95
0.518	0.78	0.985	71.7
0.45	0.75	0.951	93.38
0.348	0.7	0.735	140.33
0.167	0.6	0.0481	91.04
0.1	0.56	0.0094	63.12
0.066	0.54	0.00404	53.96
0.034	0.52	0.00173	46.94
0.0	0.5	0.00073	41.43

Next, we present some numerical results for the nonparametric sequential sign test described in Section 7.4. The observation model is as follows:

$$H_0 : X_{ik} \sim N(0, \sigma_k^2)$$

$$H_1 : X_{ik} \sim N(\theta_k, \sigma_k^2)$$

, $i \geq 1$; $k=1,2$

where $N(\theta_k, \sigma_k^2)$ is as defined earlier. The nominal design values are $\tilde{\theta}_1 = 0.525$, $\tilde{\theta}_2 = 0.824$, and $\sigma_k^2 = 1$, $k=1,2$. The resulting values of $\tilde{p}_\theta$ are $\tilde{p}_{\theta 1} = 0.7$ and $\tilde{p}_{\theta 2} = 0.8$. The required value of α is set at $\alpha = 10^{-3}$,

while the power of the test is required to be $P_D(\theta_1,\theta_2) = 0.999$ at the nominal (estimated) values of the parameters.

In Table 7.2, the parameters $p_{\theta 1}$ and $p_{\theta 2}$ are varied such that their effects on the power of the test are identical. In other words, the test has the same power as if it is performed based on the observations of sensor one alone or sensor two alone. In this case, it is meaningful to compare the resulting ASN's as given in Table 7.2 below, where $ASN_{\theta k}$ denotes the ASN considering only the kth sensor observations.

Table 7.2. Performance of the distributed nonparametric sequential sign test.

$p_{\theta 1}$	$p_{\theta 2}$	$P_D(\theta_1,\theta_2)$	$ASN_{\theta 1}$	$ASN_{\theta 2}$	$ASN(\theta_1,\theta_2)$
0.5	0.5	1×10^{-3}	79.07	30.89	22.21
0.5103	0.516	2×10^{-3}	87.62	34.23	24.62
0.5309	0.5486	7.89×10^{-3}	111.44	43.64	31.36
0.623	0.6915	0.8	243.18	97.89	69.8
0.6814	0.7753	0.996	103	43.24	30.45
0.7	0.8	0.999	83.77	35.76	25.06
0.744	0.854	0.99997	57.77	25.81	17.84

In Table 7.3, the parameters $p_{\theta 1}$ and $p_{\theta 2}$ are allowed to change such that their effect on the test power is not necessarily the same. We denote by P_{Dk}, k=1,2, the power of the test performed based on the observations of the kth sensor alone. Observe that this situation is not encountered in the single detector case [22] where we have a single parame-

ter to deal with. This case is intended to demonstrate the effect of the parameter changes on the central test power.

Table 7.3. Performance of the distributed nonparametric sequential sign test with unequal power functions.

$P_{\theta 1}$	$P_{\theta 2}$	P_{D1}	P_{D2}	$P_D(\theta_1,\theta_2)$	$ASN(\theta_1,\theta_2)$
0.65	0.85	0.9633	0.99996	0.9997	22.86
0.75	0.75	0.99998	0.9854	0.9975	27.7
0.7	0.85	0.999	0.999996	0.9999	20.05
0.75	0.8	0.99998	0.999	0.9997	21.75
0.6	0.7	0.515	0.857	0.77	72.18
0.65	0.8	0.9633	0.999	0.9972	29.52

7.7. Discussion

The nonparametric sequential conditional sign test of Shin and Kassam was investigated and it was shown that it can be implemented without the table look-up operation as previously required. In addition, the above test was formulated as a random walk on a finite number of states and exact analytical expressions were derived for both the test power and average sample number functions. Both the sequential sign and conditional sign nonparametric tests were generalized to a distributed system of two sensors and shown to maintain the desired nonparametric property. It was shown that when the observations of the two sensors are iid, the

distributed system requires half the number of observations required by a single sensor. Thus, it has the desirable advantage of shorter decision time in addition to the well known advantages of distributed detection systems like reliability and survivability.

CHAPTER EIGHT

SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

8.1. Summary

In this report, we have considered some sequential hypothesis testing problems where the detection network consists of a number of local sensors and/or detectors. We have studied a centralized SPRT based on quantized mutlisensor data, and shown its performance superiority over the single sensor case. Moreover, we have studied the issues of optimal quantization of local observations, channel errors, and random walk formulation of the central SPRT.

A simple multi-sensor decentralized sequential detection procedure has been investigated when no explicit fusion rule is employed. It has been shown that its performance improves monotonically with the number of local detectors. Next a distributed sequential detection system employing explicit fusion rule has been considered.

The SPRT of Wald has been generalized to a distributed system consisting of two local sequential detectors and a global decision maker. The global error probabilities are derived in terms of the local error probabilities and the fusion rule employed at the global decision maker. Moreover, the global ASN is obtained in terms of the local test lengths and the fusion rule. In addition, we have generalized the Lee-Thomas MLGDS detection procedure to the distributed system described above. The results obtained in both cases show the performance superiority of these distributed tests over their single detector counterparts.

Finally, we have generalized the nonparametric sequential sign and conditional sequential sign tests of Shin and Kassam to a distributed system of two local sensors. We have shown that the distributed system exhibits an improved performance over the single detector case and maintains the desired nonparametric property.

8.2. Suggestions for Future Research

Throughout this report, we have assumed that the observations are statistically independent and identically distributed at the local sensors. However, in practice, the observations can be dependent both spatially and temporally. Therefore, one fruitful area for research is to develop suitable sequential detection schemes under the appropriate dependent observation models.

Another possibility is to investigate sequential detection schemes for different network structures such as serial and tree topologies. Appropriate fusion schemes should be developed and system performance should be evaluated.

REFERENCES

[1] A. Wald. Sequential Analysis, New York: Wiley, 1947.

[2] H. L. Van Trees, Detection, Estimation and Modulation Theory, John Wiley and Sons, New York, 1968.

[3] J. J. Bussgang and D. Middleton, "Optimal sequential detection of signals in noise," IRE Trans. Inform. Theory, vol. IT-1, pp. 5-18, December 1955.

[4] S. Tantaratana and J. B. Thomas, "Relative efficiency of sequential probability ratio test in signal detection," IEEE Trans. Infor. Theory, vol. IT-24, No. 1, pp. 22-31, Jan. 1978.

[5] H. Blasbalg, "Experimental results in sequential detection," IRE Trans. Inform. Theory, vol. IT-5, pp. 41-51, June 1959.

[6] B. K. Ghosh, Sequential Tests of Statistical Hypothesis, Reading, MA; Addison-Wesley, 1970.

[7] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," IEEE Trans. Aerospace and Elec. Syst., vol. AES-17, pp. 501-510, July 1981.

[8] H. J. Kushner and A. Pacut, "A simulation study of decentralized detection problem," IEEE Trans. Automatic Control, vol. AC-27, pp. 1116-1119, Oct. 1982.

[9] Z. Chair and P. K. Varshney, "Optimum data fusion in multiple sensor detection system," IEEE Trans. Aerospace and Elec. Syst., vol. AES-22, pp. 98-101, Jan. 1986.

[10] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," IEEE Trans. Information Theory. vol.,IT-35, pp. 995-1001, Sept. 1989.

[11] A.R. Reibman and L.W. Nolte,"Design and performance comparison of distributed detection networks,IEEE Trans. Aerospace and Elec. Syst., vol. AES-23, pp. 789-797, Nov. 1987.

[12] R. Srinivasan,"A theory of distributed detection," Signal Processing, vol. 11, No. 4, pp. 319-327, Dec. 1986.

[13] I. Y. Hoballah and P. K. Varshney, "Neyman-Pearson detection with distributed sensors," Proceeding of the 25th Conference on Decision and Control, Athens, Greece, Dec. 1986.

[14] S.C.A. Thomopoulos and N. N. Okello, "Distributed detection with consulting sensors and communication cost," Proceedings of SPIE Conference, Orlando, FL, April 1988.

[15] C. C. Lee and J. J. Chao, "Optimum local decision space partitioning for distributed detection," IEEE Trans. Aerospace and Elec. Syst., vol. AES-25, pp. 536-543, July 1989.

[16] A. M. Al-Bassiouni, Optimum signal detection in distributed sensor system, Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, 1987.

[17] M. Longo, T. Lookabaugh, and R. M. Gray, Quantization for decentralized hypothesis testing under communication constraints, Preprint Stanford University.

[18] D. Teneketzis, "The decentralized Wald problem," IEEE International Large-Scale Symposium, Virginia Beach, VA, Oct. 1982.

[19] H. R. Hashemi and I. B. Rhodes, "Distributed sequential detection," IEEE Trans. on Inform. Theory, vol. IT-35, pp. 509-520, May 1989.

[20] Z. Chair, I. Y. Hoballah, and P. K. Varshney, "The Decentralized sequential probability ratio test (DSPRT)," Twentieth Asilomar Conf. on Signals, Systems, and Computers, Pacific-Grove, CA, Nov. 10-12, 1986.

[21] C. C. Lee and J. B. Thomas, "A modified sequential detection procedure," IEEE Trans. Inform. Theory, vol. IT-30, No. 2, pp. 16-23, Jan. 1984.

[22] J. G. Shin and S. A. Kassam, "Sequential detection based on quantization and conditional testing," Proc. 1979 Conf. on Inform. Sci. and Syst., John Hopkins University, pp. 487-492, 1979.

[23] S. Tantaratana and H. V. Poor, "A two stage version of the Kassam-Thomas nonparametric dead-zone limiter system," J. Acoust. Soc. of Amer. vol. 71, pp. 110-115, Jan. 1982.

[24] S. Tantaratana, "Nonparametric sequential conditional sign tests," Proc. 1988 Conf. on Inform. Sci. and Syst., Princeton University, 1988.

[25] S. Fleisher, H. Singh, and E. Shwedyk, "A generalized two threshold detection procedure," IEEE Trans. Inform. Theory, vol. IT- 34, No. 2, pp. 347-352, March 1988.

[26] M. Hecht and M. Schwartz, "M-ary sequential detection for amplitude modulated signals in one and two dimensions," IEEE Trans. Commun. Technol, vol. COM-16, No. 5, pp. 669-675, Oct. 1968.

[27] S. Fleisher and E. Shwedyk, "A sequential multiple hypotheses test for the unknown parameters of a Gaussian distribution," IEEE Trans. Inform. Theory vol. IT-26, No. 2, pp. 255-259, March 1980.

[28] J. G. Proakis, "Exact distribution functions of test length for sequential processors with discrete input data," IEEE Trans. Inform. Theory, vol. IT-9, No. 5, pp. 182-191, July 1963.

[29] H. M. Taylor and S. Karlin, An Introduction to Stochastic Modeling, Academic Press, Orlando, FL. 1984.

[30] M. A. Girshick, "Contributions to the theory of sequential analysis, II, III," Am. Math. Statistics, vol. 17, pp. 282-298, Sept. 1946.

[31] S. Tantaratana and J. B. Thomas, "Truncated sequential probability ratio test," Inform. Sci., vol. 13, pp. 283-300, 1977.

[32] J. Bussgang and M. Marcus, "Truncated sequential hypothesis tests," IEEE Trans. Inform. Theory, vol. IT-13, No. 3, pp. 512-516, July 1967.

[33] T. W. Anderson, "A modification of the sequential probability ratio test to reduce the sample size," Ann, Math. Statist. vol. 31, pp. 165-197, 1960.

[34] P. Armitage, "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," Journal of the Royal Statistical Society, Ser. B. vol. 12, pp. 137-144, Nov. 1950.

[35] S. A. Kassam and J. B. Thomas, "Dead-zone limiter: an application of conditional tests in nonparametric detection," J. Acoust. Soc. of Amer., vol. 60, pp. 857-862, 1976.

[36] J. B. Thomas, "Nonparametric detection," IEEE Proceedings, vol. 58, pp. 623-631, 1970.

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.