**Università degli Studi di Salerno**

Dipartimento di Ingegneria Elettronica ed Ingegneria Informatica

Dottorato di Ricerca in Ingegneria dell'Informazione
X Ciclo – Nuova Serie

TESI DI DOTTORATO

# The Unlucky Broker

CANDIDATO: **FABIO MAZZARELLA**

TUTOR:          **PROF. STEFANO MARANO**

COORDINATORE:   **PROF. ANGELO MARCELLI**

Anno Accademico 2010 – 2011

*To Serena,*
*my dearest LOVE*

# Acknowledgments

This is the third time that I am faced with the writing of the acknowledgments of a thesis and I wish they looked as sincere as possible.

It might sound weird, but I am grateful to all those bad persons I have met during my PhD years, because they have helped me to realize that I am not as mediocre as I thought. I have learned that life and job are two different planets and "we are NOT the number of papers we write".

Then, I want to thank my advisor, Prof. Stefano Marano, and Dr. Vincenzo Matta, for their supervision and because they allowed me to spend three years of my life in the stimulating world of scientific research. I do not know if I have been a bright student, but I have tried to put all my energies, strengths and intellectual honesty in my job.

Special thanks go to Prof. Peter Willett for several reasons. He hosted me at the Department of Electrical & Computer Engineering at the University of Connecticut and dedicated much of his precious time to me. He allowed me to take part of the Conference on Information Fusion in Chicago and I will remember those moments for the rest of my life. I cannot forget the availability and the kindness of Ramona Georgescu and her husband, Ozgur Erdinc: they made my life in New England more enjoyable.

Thanks to Luigi Bruno, one of my lab mates, for having shared with me his worries, his experiences and, above all, his meals. I am glad to have met the other members of the *Laboratorio TETI* and of the Telecommunications group at the University of Salerno, that I mention here in a pure alphabetically order: Paolo Addesso,

# Contents

# List of Figures

# Chapter 1

# Introduction

The field of statistical signal processing mainly encompasses all those mathematical procedures that engineers and statisticians apply to draw inference from imperfect or incomplete measurements. The major domains of the discipline are detection, estimation and time series analysis. The theory of statistical signal processing relies on ideas from probability theory, mathematical statistics, linear algebra, Fourier analysis, system theory, and digital signal processing.

From an abstract point of view, statistical signal processing can be viewed as a theory that exploits experimental measurements to transform a prior model for a signal, or a phenomenon, into a posterior model that can be used to make inference, e.g., to make decisions. The quality of the decisions is measured by a loss function that depends on the ground truth and the decision taken. It is the fancy interplay between prior models, measurements schemes, loss functions, and decision rules that ensure to statistical signal processing its great appeal and wide variety. In this chapter and in the remainder of this dissertation, we focus our attention only on detection theory, also referred to as decision making.

*Statistical decision-making* is a rather generic expression referring to scenarios in which one or more individuals, or systems, deal with the task of deciding among a number of alternatives in

an uncertain environment, called *state of the nature*, so as to satisfy a given objective or a set of objectives. The decision makers know the probabilistic description of the state of the nature, and the decisions are based on an ensemble of measurements acquired about the unknown environment. If the number of alternatives is finite, such problems are commonly known as *hypothesis testing* problems.

While early applications of hypothesis testing mainly focused on radar/sonar signal processing, and on the design of digital communication receivers, newer areas of applications include image analysis and interpretation, document authentication, biometrics, sensor networks, human decision systems, social networks, and so forth. In all these cases, detection theory is used to select the physical or mathematical model, among a finite ensemble, that best describes the observed phenomena. This enlarged scope of application has also driven, in the last forty years, a non trivial evolution and a significant generalization of the methods and approaches originally exploited.

As a consequence, the theory of detection is now well assessed and understood and the line of investigation to many detection problems, including those arising from emerging areas, is well traced and available in standard references such as [36–41].

## 1.1  Motivation & State of the Art

Smart environments represent the current evolutionary step in building, utilities, industrial, home, shipboard, and transportation systems automation. Like any sentient organism, the smart environment relies first and foremost on sensory data from the real world, which are collected by multiple sensors placed in distributed locations. The challenges of detecting the relevant quantities, monitoring and collecting the data, assessing and evaluating the information, and taking decisions are enormous [1, 2].

Sensor networks are the key to gathering the information needed from smart environments, almost everywhere. Recent terrorist

and guerilla warfare countermeasures require distributed network of sensors that can be deployed using, e.g. aircraft, and have self-organizing capabilities. In such applications, running wires is usually impractical: a sensor network is required that is fast and easy to install and maintain. Moreover, recent advances in hardware technology have led to the development of small, low-power, mobile sensors with limited onboard processing and wireless communication capabilities. Therefore, Wireless Sensor Networks (WSNs) perfectly fit the smart environments' need of collecting information from the real word.

The distributed sensor systems, originally motivated by their applications in military surveillance, are thus now employed in a wide variety of applications. However, the study of WSNs is very challenging because they require an enormous breath of knowledge from an enormous variety of disciplines.

The much more growing diffusion of networks made of (tiny, cheap, low-power) sensors geographically dispersed, has spurred a huge research activity in decentralized detection. In classical approaches to multisensor detection, it is assumed that all the local sensors transmit their own data to a central processor that performs optimal detection by using the classical theory mentioned in the previous section. In decentralized detection, instead, some preliminary processing of data is carried out at each sensor and the resulting information is sent to a central processor, known as the fusion center. Some of the main advantages of the decentralized signal processing over the centralized one are reduced communication bandwidth requirement, reduced cost, and increased reliability. Unlike the central processor in centralized architectures, the fusion center of a decentralized system has only partial information as communicated by the sensors. This implies a loss of performance in decentralized systems, which can be made small by developing algorithms able to process the information at the local sensors in an optimally fashion. In general, a distributed sensor network has to address the issues of choices of topology, ability to reconfigure the structure in case of sensor/link failure, robustness of signal processing algorithms with respect to probability models,

jammers, and other external threats [26].

Existing architectures for decentralized detection can be fundamentally classified into the following categories [26–28]. The first class considers *parallel* architectures, in which all sensors transmit their local decision to the fusion center [4–12]. Subsequently, the central processor makes the final decision regarding the underlying phenomenon. The second class consists of *consensus-based* detection structures, in which there is no fusion center. In this case, sensors operate over two different phases. During the sensing phase, each sensor collects observations from the environment over a sufficiently long period of time. Subsequently, sensors exchange information with their neighbors and run a consensus algorithm to make their decisions [30]. The third class of distributed detection architectures has recently been proposed in [29], in which the phases of sensing and communication need not be mutually exclusive, i.e., sensing and communication may occur simultaneously. Similarly to the second class, there is no need for a fusion center. This class of detection problems is particularly relevant for ad-hoc networks. The last class is known as *serial* or *tandem* architecture [13–25], in which there is a cascade of sensors. The first one uses its own observations to derive the information to be passed to the second sensor which, in turn, can rely on its observations and the information coming from the previous sensor. The last sensor in the network decides between two or more alternatives and this represents the final decision of the system. Consequently, also in this case there is no need for a fusion center.

## 1.2   Our Contribution

This dissertation collects results of the work on the interpretation, characterization and quantification of a novel topic in the field of detection theory, that we call the *Unlucky Broker* problem, and its asymptotic extension, that can be linked to the well-known, wide and aforementioned literature on decentralized detection in serial schemes.

An example revealing the origin of the name *unlucky broker* can be found in everyday life. Consider a financial broker who aims to suggest some good investments to his customers. He has available two data sets: one in the public domain and the other one made of certain confidential information he has received. To recommend the appropriate investments, the broker must decide between a positive or a negative market trend. He makes his decision by setting the probability of wrongly predict a positive trend, while maximizing the corresponding probability of correctly making such prediction. Suppose that, later, he is asked to refine his own decision: he must assure the probability of wrongly predicting a positive trend to be much lower than that initially chosen. Unfortunately, the unlucky broker has lost the files containing the confidential information, but he reminds the decision previously made. What should the broker do? Should he simply retain the original decision, or should he ignore that and use only the available data set for a completely new decision? Or, what else?

From a statistical signal processing perspective, the same problem can be abstracted as follows. A certain entity, call it $\mathcal{A}$ (a sensor, a generic system, an individual), is faced with a standard binary statistical test between two hypotheses, such that a decision obeying a certain optimality criterion, and with a certain reliability, is made on the basis of the available data set. Another entity, call it $\mathcal{B}$, observing a portion of the original data and also receiving from the $\mathcal{A}$ its decision, aims to make an optimal decision, with a corresponding level of reliability.

From a signal processing point of view, the interest is in characterizing the *modus operandi* of the detector in correspondence of $\mathcal{B}$. In this connection, several questions arise. At one extreme, should $\mathcal{B}$ simply retain the original decision? One could naively guess that a decision based upon the full set of observations should be better than that available to $\mathcal{B}$, where only a fraction of measurements can be exploited. After a closer look, this choice exhibits a severe limitation, in that it does not leave to entity $\mathcal{B}$ the chance of changing the level of reliability originally set by $\mathcal{A}$. On the other hand: should $\mathcal{B}$ simply implement an optimal deci-

sion based only on the survived data? We address these issues in the first part of the dissertation, where we show that the optimal solution to the unlucky broker problem is not a single-threshold likelihood ratio test.

The second part of the thesis concerns the extension of the unlucky broker problem to further detection stages, in presence of multiple successive refinements. In particular, we are interested in the asymptotic performance of the system made of a very long chain of elementary detection units.

Note that a similar problem arises in decentralized detection with tandem (serial) architecture, in which each unit makes a decision on the basis of its own data and of the decision made by one neighbor. As mentioned in the above, an abundant literature is available on this topic, see, e.g., [13–25]. The observation structure, however, makes the problems very different since, oppositely to the classical assumptions in tandem detection systems, here the data sets available at each system are strongly dependent. Therefore, approaches and insights are different from those pertaining to the existing literature, and will be developed in this work.

The remainder of this dissertation is organized as follows. In Chapter 2, classic results on the hypothesis-testing problems are reviewed, by resorting to well-known results in [36–41].

Chapter 3 reviews basic results on distributed detection with multiple sensors. In particular, the focus is on the parallel [4–12] and serial configurations [13–25].

Chapter 4 deals with the unlucky broker problem. First of all, we present three different motivating examples. Then, the problem is formally stated and solved by exploiting classical tools from detection theory. In the same chapter, we show examples of applications and discuss the operative modalities of the optimal detectors.

Chapter 5 introduces an extension of the unlucky broker problem to an arbitrary number of stages, that we call the *Very Unlucky Broker* problem. We provide motivations for its study, formally state the problem and analyze the asymptotic behavior of the system. An anomalous exponent of the miss detection probability

turns out to characterize the asymptotic behavior of the decision system.

# Chapter 2

# Elements of Hypothesis Testing

In this chapter, classic results on the hypothesis-testing problems are reviewed, see e.g., [36–41]. In particular, the main problem is formulated in Section 2.1, while the most common optimum decision strategies are illustrated in Sections 2.2, 2.3 and 2.4. Section 2.5 shows how to evaluate the performance of a decision strategy, by introducing the notion of Receiver Operating Characteristic, while Section 2.6 summarizes the main findings and concludes the chapter.

## 2.1   Introduction

Statistical decision-making is a term referring to situations in which one or more entities are involved in the task of deciding among a number of alternatives, with the aim of satisfy a given requirement. The decision-makers operates in an uncertain environment (also called *state of the nature*), characterized by a probabilistic description. The decisions are based on the measurements acquired on the unknown state of the nature. Such problems are called *detection*, *decision* or *hypothesis-testing* problems.

The problem of $M-$ary hypothesis testing consists therefore of deciding among $M$ possible statistical situations describing the

available observations. For a given decision problem, the goal is to adopt a decision strategy that is optimum in some sense. There are several definitions of optimality, and in the following sections we consider the most common formulations (Bayes, Minimax, and Neyman-Pearson) and derive the corresponding optimum solutions. In the following, we limit the analysis to the binary case, in which $M = 2$.

The hypothesis-testing problems in which each of the two hypotheses corresponds to only a single distribution for the observations, are known as *simple hypothesis-testing problems*. Whenever there are many possible distributions that can occur under each of the hypotheses, we are in presence of a *composite hypothesis-testing problem*.

Let us consider the simple hypothesis-testing problem, where we assume that there are two possible hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, corresponding to two possible distributions $P_0$ and $P_1$, respectively, on the observation space $(\Gamma, \mathcal{G})^1$. $\Gamma$ represents the observation set, while $\mathcal{G}$ is a $\sigma$-algebra of subsets of $\Gamma$, i.e., is a class of subsets of $\Gamma$ to which we wish to assign probabilities. The sets in $\mathcal{G}$ are called observation events. In the following, we use upper case letters to denote random variable, with the corresponding lower case letters representing the associated realizations.

The considered problem can be written as

$$\mathcal{H}_0 : \quad \text{observations } x \text{ are drawn from } P_0,$$
$$\mathcal{H}_1 : \quad \text{observations } x \text{ are drawn from } P_1. \quad (2.1)$$

A *decision rule*, $\delta$, for the problem in (2.1), is any partition of $\Gamma$ into sets $\Gamma_1 \in \mathcal{G}$, the *rejection region*, and $\Gamma_0 = \Gamma_1^c$, the *acceptance region*, such that we choose for $\mathcal{H}_0$ (the null hypothesis) when $x \in \Gamma_0$ and for $\mathcal{H}_1$ (the alternative hypothesis) when $x \in \Gamma_1$. We can also think the decision rule $\delta$ as a function on $\Gamma$ given by

$$\delta(x) = \begin{cases} 1, & x \in \Gamma_1 \\ 0, & x \in \Gamma_1^c, \end{cases} \quad (2.2)$$

---

[1] Throughout this chapter we will be interested in the case $\Gamma = \Re^n$, the set of $n$-dimensional vectors with real components, and in the case in which $\Gamma$ is a countable set, $\Gamma = \{\gamma_1, \gamma_2, \ldots\}$.

so that the value of $\delta$ for a given $x \in \Gamma$ is the index of the hypothesis accepted by $\delta$.

## 2.2  Bayesian Hypothesis Testing

With the hypothesis-testing problem (2.1) in mind, the main goal is that of choosing the rejection region $\Gamma_1$ in some optimum way. To this end, let us assign costs to our decisions, so that we have positive numbers $C_{ij}$, with $i = 0, 1$ and $j = 0, 1$, representing the cost incurred by choosing hypothesis $\mathcal{H}_i$ when hypothesis $\mathcal{H}_j$ is true. We define the *conditional risk* for each hypothesis as the expected cost incurred by decision rule $\delta$ when that hypothesis is true; i.e.,

$$R_j(\delta) = C_{1j}P_j(\Gamma_1) + C_{0j}P_j(\Gamma_0), \quad j = 0, 1. \tag{2.3}$$

Suppose that we know the so-called *a priori* probabilities, $\pi_o$ and $\pi_1$, of hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. For given priors we can define the *Bayes risk* as the overall average cost incurred by decision rule $\delta$. This quantity is given by

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta). \tag{2.4}$$

By combining (2.3) and (2.4) we get

$$
\begin{aligned}
r(\delta) &= \sum_{j=0}^{1} \pi_j [C_{0j}(1 - P_j(\Gamma_1)) + C_{1j}(1 - P_j(\Gamma_1))] \\
&= \sum_{j=0}^{1} \pi_j C_{0j} + \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) P_j(\Gamma_1).
\end{aligned}
\tag{2.5}
$$

The *Bayes rule* for $\mathcal{H}_0$ versus $\mathcal{H}_1$ is the decision strategy minimizing the Bayes risk over all decision rules. From (2.5), assuming

that $P_j$ has density $p_j{}^2$ for $j = 0, 1$, we can write

$$
r(\delta) = \sum_{j=0}^{1} \pi_j C_{0j}
$$
$$
+ \int_{\Gamma_1} \left[ \sum_{j=0}^{1} \pi_j (C_{1j} - C_{0j}) p_j(x) \right] \mu(dx). \qquad (2.6)
$$

It is straightforward to see that the Bayes risk is minimized by the following rejection region

$$
\Gamma_1 = \{ x \in \Gamma : \pi_1 (C_{11} - C_{01}) p_1(x) \leq \pi_0 (C_{00} - C_{10}) p_0(x) \}. \quad (2.7)
$$

If we assume that $C_{11} < C_{01}$, (2.7) can be rewritten as

$$
\Gamma_1 = \{ x \in \Gamma : p_1(x) \geq \tau p_0(x) \}, \qquad (2.8)
$$

where

$$
\tau := \frac{\pi_0 (C_{10} - C_{00})}{\pi_1 (C_{01} - C_{11})}. \qquad (2.9)
$$

The quantity

$$
L(x) = \frac{p_1(x)}{p_0(x)}, \quad x \in \Gamma, \qquad (2.10)
$$

is known as the *likelihood ratio* between $\mathcal{H}_1$ and $\mathcal{H}_0$. Thus the Bayes decision rule corresponding to (2.8) consists of comparing the likelihood ratio for the observed value of $x$ to the threshold $\tau$, that is

$$
\delta_B(x) = \begin{cases} 1, & L(x) \geq \tau \\ 0, & L(x) < \tau. \end{cases} \qquad (2.11)
$$

A commonly used cost assignment is the *uniform* one, with $C_{ij} = 0$ if $i = j$ and $C_{ij} = 1$ if $i \neq j$. With this choice, the Bayes risk becomes

$$
r(\delta) = \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0), \qquad (2.12)
$$

---

[2]When the observation space is discrete, $p_j$ is a probability mass function. When $\Gamma = \Re^n$, $p_j$ is a probability density function. For simplicity, we will use the term density for both the probability mass function and probability density function, and we will use the notation $P_j(A) = \int_A p_j(x) \mu(dx)$ in both cases.

while the threshold is $\tau = \pi_0/\pi_1$. In this case, the Bayes risk represents the *average probability of error* incurred by the decision rule $\delta$, being $P_j(\Gamma_j)$ the probability of choosing $\mathcal{H}_j$ when $\mathcal{H}_i$ is true. Since the likelihood-ratio test with threshold $\tau$ minimizes $r(\delta)$, it is thus a *minimum-probability-of-error* decision strategy.

Bayes' formula implies that the conditional probability that $\mathcal{H}_j$ is true given that the random observation $X$ takes on value $x$ is given by

$$\pi_j(x) = P(H_j \ true | X = x) = \frac{p_j(x)\pi_j}{p(x)}, \qquad (2.13)$$

where $\pi_j(x)$ is the *a posteriori* probability for the hypothesis $\mathcal{H}_j$, and $p(x) = \pi_0 p_0(x) + \pi_1 p_1(x)$. By using (2.13), the rejection region of the Bayes rule can be rewritten as

$$\begin{aligned} \Gamma_1 &= \{x \in \Gamma : C_{10}\pi_0(x) + C_{11}\pi_1(x) \\ &\leq C_{00}\pi_0(x) + C_{01}\pi_1(x)\}, \end{aligned} \qquad (2.14)$$

where the quantity $C_{i0}\pi_0(x) + C_{i1}\pi_1(x)$ is the *posterior cost* of choosing $\mathcal{H}_i$ when the observation is $x$, and thus the Bayes rule makes its decision by choosing the hypothesis that yields the minimum posterior cost. Accordingly, for the uniform cost criterion, the Bayes rule becomes

$$\delta_B(x) = \begin{cases} 1, & \pi_1(x) \geq \pi_0(x) \\ 0, & \pi_1(x) < \pi_0(x). \end{cases} \qquad (2.15)$$

The decision rule in the above chooses the hypothesis that has the maximum a posteriori probability conditioned on $X = x$: it is sometimes known as the Maximum A posteriori Probability (MAP) decision rule for the binary hypothesis test (2.1).

In conclusion, when the costs $C_{ij}$ and priors $\pi_i$ are completely specified, the optimal decision rule according to the Bayes criterion is given by the likelihood ratio test (2.11).

## 2.3    Minimax Hypothesis Testing

The formulation in Section 2.2 requires the knowledge of the prior probabilities of the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. In many practical situations, the designer of a decision strategy may not have control of or access to the mechanism generating the state of the nature. In those cases, it is necessary to look for a criterion different from the Bayes strategy. One such solution is to seek a decision rule that minimizes, over all $\delta$, the maximum of the conditional risks, $R_0(\delta)$ and $R_1(\delta)$. A criterion minimizing the quantity $\max\{R_0(\delta), R_1(\delta)\}$ is known as a *minimax* rule. In the following we investigate the structure of such strategies.

For a given prior $\pi_0 \in [0, 1]$, let us consider the function $r(\pi_0, \delta)$ and decision rule $\delta$ as the average risk,

$$r(\pi_0, \delta) = \pi_0 R_0(\delta) + (1 - \pi_0)R_1(\delta). \qquad (2.16)$$

If we fix $\delta$, $r(\pi_0, \delta)$ is a straight line from $r(0, \delta) = R_1(\delta)$ to $r(1, \delta) = R_0(\delta)$. Thus, the maximum value of $r(\pi_0, \delta)$ occurs at either $\pi_0 = 0$ or $\pi_1 = 1$, and the maximum value is the maximum of $\{R_0(\delta), R_1(\delta)\}$. The problem becomes that of minimizing the quantity

$$\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) \qquad (2.17)$$

over $\delta$. For each prior $\pi_0 \in [0, 1]$, let $\delta_{\pi 0}$ denote a Bayes rule corresponding to that prior, and let $V(\pi_0) = r(\pi_0, \delta_{\pi 0})$, that is the minimum possible Bayes risk for the prior $\pi_0$. It is straightforward to show that $V(\pi_0)$ is a continuous concave function of $\pi_0 \in [0, 1]$, with $V(0) = C_{11}$ and $V(1) = C_{00}$. Fig. 2.1 depicts the typical behaviors of the curve $V(\pi_0)$, the straight line $r(\pi_0, \delta)$ and $r(\pi_0, \delta_{\pi_0'})$, that is both parallel to $r(\pi_0, \delta)$ and tangent to $V(\pi_0)$. It is interesting to observe that $\delta$ cannot be a minimax rule because the straight line $r(\pi_0, \delta_{\pi_0'})$ has a smaller maximum value. Since $r(\pi_0, \delta_{\pi_0'})$ touches $V(\pi_0)$ at $\pi_0 = \pi_0'$, $\delta_{\pi_0'}$ is a Bayes strategy for the a priori probability $\pi_0'$. Since a similar tangent line can be drawn for any decision rule $\delta$, it is straightforward to see that only Bayes rules can be minimax rules for this figure. Furthermore,
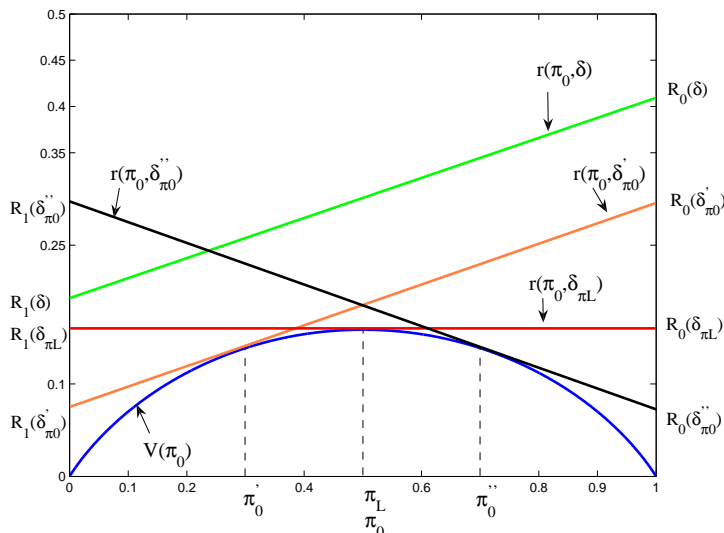
**Figure 2.1** Illustration of the minimax rule when the function $V$ has an interior maximum.

Fig. 2.1 shows that the minimax rule for this case is a Bayes rule for the prior $\pi_L$ that maximizes $V(\pi_0)$ over the whole range of $\pi_0$. For this particular prior we have that the straight line $r(\pi_0, \delta_{\pi L})$ is constant over $\pi_0$, thus $\max\{R_0(\delta_{\pi L}), R_1(\delta_{\pi L})\} = R_0(\delta_{\pi L}) = R_1(\delta_{\pi L})$. From the Fig. 2.1 follows that if $\pi_0' < \pi_L$, we have that $\max\{R_0(\delta_{\pi_0'}), R_1(\delta_{\pi_0'})\} = R_0(\delta_{\pi_0'}) > R_0(\delta_{\pi L})$. If, instead, $\pi_0'' > \pi_L$, we have that $\max\{R_0(\delta_{\pi_0''}), R_1(\delta_{\pi_0''})\} = R_1(\delta_{\pi_0''}) > R_1(\delta_{\pi L})$. Then, $\pi_L$ is a minimax rule.

The prior $\pi_L$ is called the *least-favorable prior* because it maximizes the minimum Bayes risk. In the above, we have not considered the possibility that $\max_{0 \le \pi_0 \le 1} V(\pi_0)$ may occur at $\pi_0 = 0$ or $\pi_0 = 1$, or that $V(\pi_0)$ may not be differentiable everywhere. However, even in these cases it is always true that the minimax rule is a Bayes rule for the least-favorable prior. The results in the following explain the general solution to the minimax hypothesis-testing problem, starting from a proposition [36] that summarizes the cases considered in the above.

**Proposition 2.3.1.** *Suppose that $\pi_L$ is a solution to $V(\pi_L) =$*

$\max_{0 \leq \pi_0 \leq 1} V(\pi_0)$. *Moreover, suppose that either $\pi_L = 0$, $\pi_L = 1$, or $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. Then, $\delta_{\pi_L}$ is a minimax rule.*

*Proof:* First, let us consider the case $R_1(\delta_{\pi_L}) = R_0(\delta_{\pi_L})$. Then, for any prior $\pi_0$ we have

$$\max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) = r(\pi_L, \delta_{\pi_L}) = r(\pi_0, \delta_{\pi_L}), \qquad (2.18)$$

where the first equality comes from the definition of $V$ and $\pi_L$, while the second equality follows from the fact that $r(\pi_0, \delta_{\pi_L})$ is constant in $\pi_0$. Accordingly, we can write that

$$\begin{aligned}
\max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) &= \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta_{\pi_L}) \\
&\geq \min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta). \qquad (2.19)
\end{aligned}$$

But for each $\delta$, we have

$$\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta_{\pi_L}) \geq \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta), \qquad (2.20)$$

implying that

$$\min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) \geq \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta). \qquad (2.21)$$

By combining (2.19) and (2.21), we get

$$\min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) = \max_{0 \leq \pi_0 \leq 1} \min_{\delta} r(\pi_0, \delta) \qquad (2.22)$$

and the left-hand equality of (2.18) implies

$$r(\pi_L, \delta_{\pi_L}) = \min_{\delta} \max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta) \qquad (2.23)$$

and the proposition is shown.

When $\pi_L = 0$ it is straightforward to prove that

$$\max_{0 \leq \pi_0 \leq 1} r(\pi_0, \delta_{\pi_L}) = R_1(\delta_{\pi_L}) = r(\pi_L, \delta_{\pi_L}). \qquad (2.24)$$

This fact and the argument of Eqs. (2.20) through (2.23) imply that $\delta_{\pi L}$ is a minimax rule. A similar argument holds for the case $\pi_L = 1$ and this completes the proof. $\qquad\qquad\bullet$

So far, we have analyzed minimax rules in all cases except those in which the function $V$ has an interior maximum at which it is not differentiable. In such cases, we define two decision rules by $\delta_{\pi L}^- = \lim_{\pi_0 \uparrow \pi_L} \delta_{\pi 0}$ and $\delta_{\pi L}^+ = \lim_{\pi_0 \downarrow \pi_L} \delta_{\pi 0}$. Note that $\delta_{\pi L}^-$ necessarily has rejection region

$$
\begin{aligned}
\Gamma_1^- = \quad & \{x \in \Gamma : (1 - \pi_L)(C_{11} - C_{01})p_1(x) \\
& \leq \pi_L(C_{00} - C_{10})p_0(x)\},
\end{aligned}
\tag{2.25}
$$

while $\delta_{\pi L}^+$ has rejection region

$$
\begin{aligned}
\Gamma_1^+ = \quad & \{x \in \Gamma : (1 - \pi_L)(C_{11} - C_{01})p_1(x) \\
& \leq \pi_L(C_{00} - C_{10})p_0(x)\},
\end{aligned}
\tag{2.26}
$$

regardless of which particular Bayes rules $\delta_{\pi 0}$ are used to define them. For a number $q \in [0, 1]$, consider the strategy $\tilde{\delta}_{\pi L}$ that uses $\Gamma_1^-$ with probability $q$ and the region $\Gamma_1^+$ with probability $(1 - q)$. The *conditional* risks can be written as

$$
R_j(\tilde{\delta}_{\pi L}) = qR_j(\delta_{\pi L}^-) + (1 - q)R_j(\delta_{\pi L}^+),
\tag{2.27}
$$

so that the condition $R_0(\tilde{\delta}_{\pi L})$ is achieved by choosing

$$
q = \frac{R_0(\delta_{\pi L}^+) - R_1(\delta_{\pi L}^+)}{R_0(\delta_{\pi L}^+) - R_1(\delta_{\pi L}^+) + R_1(\delta_{\pi L}^-) - R_0(\delta_{\pi L}^-)}.
\tag{2.28}
$$

Accordingly, as in Proposition 2.3.1, $\tilde{\delta}_{\pi L}$ with $q$ as in (2.28) is a minimax rule.

Note that, because $V$ is concave, it must have left- and right-hand derivatives at $\pi_L$, which we denote by $V'(\pi_L^-)$ and $V'(\pi_L^+)$, respectively. It is straightforward to see that $V'(\pi_L^+) = [R_0(\delta_{\pi L}^+) - R_1(\delta_{\pi L}^+)]$ and $V'(\pi_L^-) = [R_0(\delta_{\pi L}^-) - R_1(\delta_{\pi L}^-)]$, so that (2.28) becomes

$$
q = \frac{V'(\pi_L^+)}{V'(\pi_L^+) - V'(\pi_L^-)}.
\tag{2.29}
$$

The decision rule $\tilde{\delta}_{\pi L}$ is an example of *randomized decision rule*.

## 2.4 Neyman-Pearson Hypothesis Testing

In the Bayesian formulation of Section 2.2 optimality in testing (2.1) was based on the knowledge of the a priori probabilities of the hypotheses, $\pi_0$ and $\pi_1$, and costs $C_{ij}$, with $i, j = 0, 1$. In many physical situations it is very difficult to assign a realistic cost structure and a priori probabilities. A simple procedure to bypass this difficulty is to work with an alternative formulation, known as *Neyman-Pearson criterion*.

In the hypothesis-testing problem (2.1), two types of errors can be made: $\mathcal{H}_0$ can be falsely rejected or $\mathcal{H}_1$ can be falsely rejected. The former is known as a *Type I error* or a *false alarm*. The latter is called *Type II error* or a *miss detection*. For a decision rule $\delta$, the probability of a Type I error is known as the *size* or *false alarm probability* of $\delta$, an we will denote it by $P_F(\delta)$. The probability of a Type II error is called *miss detection probability* $P_M(\delta)$, even though we usually talk about the *detection probability*, $P_D(\delta) = 1 - P_M(\delta)$, also called the *power* of $\delta$.

Obviously, the design of a test for $\mathcal{H}_0$ versus $\mathcal{H}_1$ involves a trade-off between the probabilities of the two type of errors, because it is not possible to reduce both error probabilities simultaneously. The Neyman-Pearson criterion for making this trade-off is to set a bound on the false alarm probability and then to maximize the detection probability within this constraint, i.e., the Neyman-Pearson design strategy can be formulated as

$$\max_{\delta} P_D(\delta)$$
$$\textit{subject to } P_F(\delta) \leq \alpha, \qquad (2.30)$$

where $\alpha$ is known as the *level*, or *significance level* of the test. Thus, the Neyman-Pearson design goal is to find the most powerful $\alpha$-level test of $\mathcal{H}_0$ versus $\mathcal{H}_1$. The general solution to the Neyman-Pearson problem in (2.30) can be summarized in the following result [37, 38], in which we consider the observation space $\Gamma = \Re^n$, so that the observations are $n$-dimensional vectors, i.e., we have

$$\boldsymbol{x} = (x_1, \ldots, x_n).$$

**Proposition 2.4.1.** *The optimal solution of the problem (2.30) is*

$$\delta(\boldsymbol{x}) = \begin{cases} 1, & L(\boldsymbol{x}) \geq \gamma \\ 0, & L(\boldsymbol{x}) < \gamma, \end{cases} \tag{2.31}$$

*where the threshold $\gamma$ is found from*

$$P_F(\delta) = \int_{\boldsymbol{x}:L(\boldsymbol{x})\geq\gamma} p_0(\boldsymbol{x}) \, d\boldsymbol{x} = \alpha. \tag{2.32}$$

*Proof:* We use Lagrangian multipliers to solve the problem in (2.30). First of all, we form th Lagrangian as follows

$$\begin{aligned} F &= P_D + \lambda(P_F - \alpha) \\ &= \int_{\Gamma_1} p_1(\boldsymbol{x}) \, d\boldsymbol{x} + \lambda \left( \int_{\Gamma_1} p_0(\boldsymbol{x}) d\boldsymbol{x} - \alpha \right) \\ &= \int_{\Gamma_1} (p_1(\boldsymbol{x}) + \lambda p_0(\boldsymbol{x})) \, d\boldsymbol{x} - \lambda\alpha. \end{aligned} \tag{2.33}$$

To maximize $F$, we should include $x$ in $\Gamma_1$ if the integrand is positive for that value of $x$ or if

$$p_1(\boldsymbol{x}) + \lambda p_0(\boldsymbol{x}) > 0. \tag{2.34}$$

When $p_1(\boldsymbol{x}) + \lambda p_0(\boldsymbol{x}) = 0$, $\boldsymbol{x}$ may be included in either $\Gamma_1$ or $\Gamma_1$. Since the probability of this occurrence is zero (assuming the PDFs are continuous), we need not concern ourselves with this case. Hence, the $>$ sign in (2.34) and subsequent results can be replaced with $\geq$ if desired. We thus decide $\mathcal{H}_1$ if

$$\frac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} \geq -\lambda. \tag{2.35}$$

The Lagrangian multiplier is found from the constraint and must satisfy $\lambda < 0$. Otherwise, we decide $\mathcal{H}_1$ if the likelihood ratio

exceeds a negative number. Since the likelihood ratio is always nonnegative, we would always decide $\mathcal{H}_1$, irrespective of the hypothesis, resulting in $P_F = 1$. We let $\gamma = -\lambda$ so that finally we decide $\mathcal{H}_1$ if

$$\frac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} \geq \gamma, \tag{2.36}$$

where the threshold $\gamma > 0$ is found from $P_F = \alpha$.                    ●

In some cases, for example when the observations take on only integer values, the false alarm probability can take on only one of a countable set of values. Then it may happen that the preassigned value of the false alarm probability, $\alpha$, is not a member of that set. In order to achieve an arbitrary value of the false alarm probability, it is necessary to resort to what is called *randomized decision rule*, which can be formalized as follows

$$\tilde{\delta}(\boldsymbol{x}) = \begin{cases} 1, & L(\boldsymbol{x}) > \gamma \\ q, & L(\boldsymbol{x}) = \gamma \\ 0, & L(\boldsymbol{x}) < \gamma, \end{cases} \tag{2.37}$$

where $q \in [0, 1]$ with a certain probability distribution. With this definition, we see that a nonrandomized decision rule $\delta$ is a special case of a randomized decision rule $\tilde{\delta}$, with the difference that the value of $\delta$ is the index of the accepted hypothesis and the value of $\tilde{\delta}$ is the probability with which we accept $\mathcal{H}_1$. These coincide as long $\tilde{\delta}$ takes on only the two values 0 and 1.

The result above again indicates the optimality of the likelihood ratio test. The Neyman Pearson test for a given hypothesis pair differs from the Bayes and minimax tests only in the choice of threshold.

## 2.5   Performance: Receiver Operating Characteristic

To complete our discussion of the simple binary hypothesis-testing problem, we must evaluate the performance of the likelihood ratio

test. For a Neyman-Pearson criterion, the values of $P_F$ and $P_D$ completely specify the test performance. Looking at Eqs. (2.3) and (2.4), we see that the Bayes risk follows easily if the false alarm and detection probabilities are known. In the following, we suppose that $\Gamma = \Re^n$ and that the observations are the $n$-th vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$.

Let us consider the expressions of the false alarm probability

$$P_F = \int_{\Gamma_1(\gamma)} p_0(\boldsymbol{x}) \ d\boldsymbol{x} \tag{2.38}$$

and the detection probability for the Neyman-Pearson strategy

$$P_D = \int_{\Gamma_1(\gamma)} p_1(\boldsymbol{x}) \ d\boldsymbol{x}, \tag{2.39}$$

where the rejection region is $\Gamma_1 = \{\boldsymbol{x} \in \Re^n : L(\boldsymbol{x}) \geq \gamma\}$. Accordingly, the two probabilities are function of the threshold $\gamma$. By letting $\gamma$ range over $(-\infty, +\infty)$, we can plot the detection probability versus the false alarm probability: the resulting curve is called *Receiver Operating Characteristic*(ROC). It depends only on the probability density function of the observations under the two hypotheses and not on any costs or a priori probabilities. The ROC should always be above the $45^o$ straight line. This is because the $45^o$ ROC can be obtained by a detector that bases its decision on a coin flipping, ignoring the data at all. Fig. 2.2 depicts the typical behavior of a ROC. We now derive a few general properties of receiver operating characteristics. We confine our discussion to continuous likelihood ratio tests.

**Property 1.** All continuous likelihood ratio tests have ROC's that are concave downward. If they were not, a randomized test would be better and this would contradict our proof that a likelihood ratio test is optimum.

**Property 2.** All continuous likelihood ratio test have ROC's that are above the $P_D = P_F$ line. This is just a special case of Property 1, because the points $(P_F = 0, P_D = 0)$ and $(P_F = 1, P_D = 1)$ are contained on all ROC's.
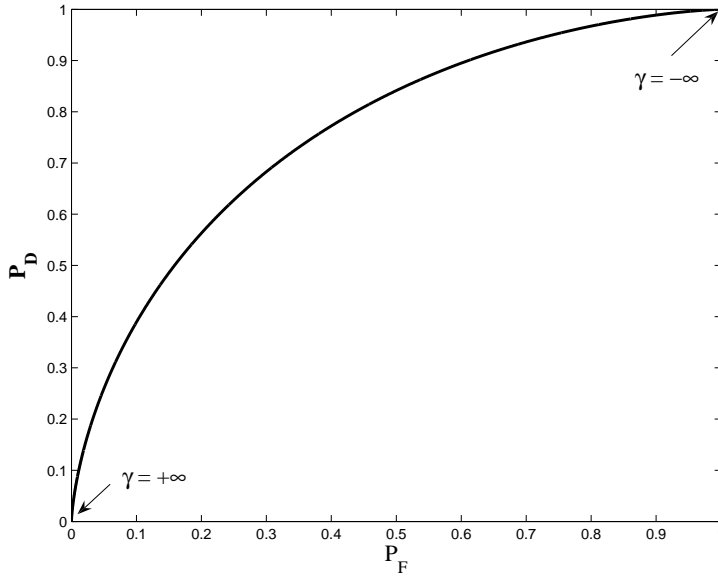
**Figure 2.2** Typical behavior of a Receiver operating characteristic.

**Property 3.** The slope of a curve in a ROC at a particular point is equal to the value of the threshold $\gamma$ required to achieve the $P_D$ and $P_F$ at that point.

*Proof:* The false alarm and detection probabilities can be written, respectively, as follows,

$$
\begin{aligned}
P_F &= \int_{\gamma}^{\infty} p_{L0}(l)dl \\
P_D &= \int_{\gamma}^{\infty} p_{L1}(l)dl,
\end{aligned} \tag{2.40}
$$

where $p_{Lj}(l)$ is the pdf of the likelihood ratio under the hypothesis $\mathcal{H}_j$, for $j = 0, 1$. Differentiating both expressions with respect to $\gamma$ and writing the results as a quotient, we have

$$
\frac{dP_D/d\gamma}{dP_F/d\gamma} = \frac{-p_{L1}(\gamma)}{-p_{L0}(\gamma)} = \frac{dP_D}{dP_F}. \tag{2.41}
$$

We now show that

$$
\frac{p_{L1}(\gamma)}{p_{L0}(\gamma)} = \gamma. \tag{2.42}
$$

Let

$$\Gamma_1(\gamma) = \{\boldsymbol{x} \in \Re^n : L(\boldsymbol{x}) \geq \gamma\}. \qquad (2.43)$$

Then, we have

$$P_D(\gamma) = Pr\{L(\boldsymbol{x}) \geq \gamma\} \;=\; \int_{\Gamma_1(\gamma)} p_1(\boldsymbol{x}) \; d\boldsymbol{x}$$

$$=\; \int_{\Gamma_1(\gamma)} L(\boldsymbol{x}) p_0(\boldsymbol{x}) \; d\boldsymbol{x}, \quad (2.44)$$

Using the definition or the region $\Gamma_1(\gamma)$, we can rewrite the last integral as follows

$$P_D(\gamma) = \int_\gamma^\infty X p_{L0}(X) \; dX. \qquad (2.45)$$

Differentiating (2.45) respect to $\gamma$, we get

$$\frac{dP_D(\gamma)}{d\gamma} = -\gamma p_{L0}(\gamma). \qquad (2.46)$$

Equating the expression for $dP_D(\gamma)/d\gamma$ in the numerator of (2.41) to the right side of (2.46) gives the desired result. $\qquad \bullet$

**Property 4.** Whenever the maximum value of the Bayes risk is interior to the interval $(0, 1)$, the minimax operating point is the intersection of the line

$$(C_{11} - C_{00}) + (C_{01} - C_{11})(1 - P_D) - (C_{10} - C_{00})P_F = 0 \quad (2.47)$$

and the appropriate ROC.

## 2.6 Summary

In this chapter we have reviewed the essential detection theory results that provides the basis for our work in the remainder of this dissertation. We began our discussion by considering the simple binary hypothesis-testing problem. Using either a Bayes criterion or a Neyman-Pearson criterion, we found that the optimum test

is a likelihood ratio test. Thus, regardless of the dimensionality of the observation space, the test consists of comparing a scalar variable to a threshold, whose choice depends on the adopted optimality criterion.

Moreover, a complete description of the likelihood ratio test performance was obtained by plotting the detection probabilities versus the false alarm probabilities as the threshold of the test was varied. The resulting curve, the Receiver Operating Characteristic (ROC), can be used to calculate the Bayes risk for any cost assignment.

# Chapter 3

# Distributed Detection with Multiple Sensors

In this chapter, basic results on distributed detection with multiple sensors are reviewed. In particular, the parallel and serial configurations are considered in Section 3.2 and 3.3, respectively, paying attention to the decision rules obtained from their optimization in the Neyman-Pearson (NP) formulation. We show that, for conditionally independent sensor observations, the likelihood ratio test made at all the sensors is optimum. A brief comparison between the two different architectures is presented in Section 3.4, while Section 3.5 concludes the chapter.

## 3.1    Introduction

Signal processing with distributed sensors has attracted a lot of attention during the last decades. The low cost of sensors, the availability of high speed communication networks, and increased computational capability have spurred great research interest in this topic [35]. Distributed sensor systems were originally motivated by their applications in military surveillance [4, 5] but they are now employed in a broad range of civilian applications. When distributed sensor systems are considered, it is possible to classify them on the basis of the degree of processing made by the single

sensors.

Consider thus a set of sensors that receive observations from the environment and transmit messages to a central processor, known as *fusion center*, that aims to make a final decision between two or more alternatives. In a *centralized* scheme, each sensor transmits its raw observation to the fusion center that solves a classical hypothesis-testing problem of the type analyzed in Chapter 2. In a *decentralized* scenario, introduced in [4, 5], each sensor sends to the fusion center a summary of its own observations. The fusion center is again involved in a classical hypothesis-testing problem, but a new question arises: How should each sensor decide what message to send?

The schemes described in the above differ in many respects. First of all, the fusion center in a decentralized scheme can rely on less information than in a centralized scheme, which results in reduced performance. On the other hand, the decentralized scheme offers the possibility for drastic reductions in the bandwidth requirements for communication between the sensors and the fusion center. As for the off-line computation, the only computation that it is needed in a centralized paradigm is the determination of the threshold to be used in a likelihood ratio test by the fusion center. The situation is much more complex with a decentralized architecture because, as will be seen later, a set of thresholds must be computed for each sensor. Finally, the on-line computational requirements of centralized and decentralized schemes are comparable.

A decentralized scheme is definitely worth considering in applications involving geographically distributed sensors. Another application can be found in the context of failure detection, where different sensors monitor different pieces of equipment and transmit small messages to a central processor that makes the final decision whether a failure has occurred or not. A further applicative context is represented by human decision-making problems [12].

The existing architectures for distributed detection can be essentially classified as follows. The first class considers *parallel* architectures, in which all sensors transmit their local decision

to the fusion center [4–12]. Subsequently, the central processor makes the final decision regarding the underlying phenomenon. The second class consists of *consensus-based* detection structures, in which there is no fusion center. In this case, sensors operate over two different phases. During the sensing phase, each sensor collects observations from the environment over a sufficiently long period of time. Subsequently, sensors exchange information with their neighbors and run a consensus algorithm to make their decisions [30]. The third class of distributed detection architectures has recently been proposed in [29], in which the phases of sensing and communication need not be mutually exclusive, i.e., sensing and communication occur simultaneously. Similarly to the second class, there is no need for a fusion center. This class of detection problems is particularly relevant for ad-hoc networks. The last class is known as *serial* or *tandem* architecture [13–25], in which there is a cascade of sensors. The first one uses its own observations to derive the information to be passed to the second sensor which, in turn, can rely on its observations and the information coming from the previous sensor. The last sensor in the network decides between two or more alternatives and this represents the final decision of the system. Consequently, also in this case there is no need for a fusion center.

In the following, we present fundamental results from the Neyman-Pearson (NP) formulation of the distributed detection problem. Parallel and serial configurations are considered in detail and we assume a binary hypothesis-testing problem in which the observations at all the sensors belongs to the state of the nature $\mathcal{H}_1$ or to the state of the nature $\mathcal{H}_0$.

## 3.2 Parallel Configuration

Let us consider the parallel configuration of $N$ sensors depicted in Fig. 3.1. Suppose that the sensors do not communicate with each other and that there is no feedback from the fusion center to any sensor. Let $x_i$ denote the observation available to the $i$-th

sensor, which employs the mapping rule $u_i = g_i(x_i)$ and transmits the information $u_i$ to the fusion center. By exploiting the received information $\boldsymbol{u} = (u_1, \ldots, u_N)$, the fusion center makes the global decision $u_0 = g_0(\boldsymbol{u})$ that favors either $\mathcal{H}_1$ ($u_0 = 1$) or $\mathcal{H}_0$ ($u_0 = 0$).

Tenney and Sandell [4] have treated the Bayesian detection problem with distributed sensors. However, they did not consider the design of data fusion algorithms. Sadjadi [11] has considered the hypothesis-testing problem in a distributed environment and has provided a solution in terms of a number of coupled nonlinear equations. Chair and Varshney [5] have solved the problem of data fusion when the a priori probabilities of the tested hypotheses are known and the likelihood ratio test can be implemented at the receiver. Thomopoulos, Viswanathan and Bougoulias [8] have derived the optimal fusion rule for unknown a priori probabilities in terms of the NP test. In the following, we will present a general proof that the optimal fusion rule for the parallel detection problem of Fig. 3.1 involves an NP test at the fusion center and likelihood ratio tests at all sensors [9].

### 3.2.1   Global Optimality

In order to derive the globally optimal fusion rule we assume that the received data $x_i$, for $i = 1, \ldots, N$, are statistically independent, conditioned on each hypothesis. This implies that the received decisions at the fusion center are independent conditioned on each hypothesis. The NP formulation of the problem can be stated as follows: given a desired level of the false alarm probability at the fusion center, $P_{F,0} = \alpha_0$, find optimum local and global decision rules that maximize the detection probability at the fusion center, $P_{D,0}$. Next, we prove that the optimal solution to the fusion problem involves an NP test at the fusion center and likelihood ratio tests at the sensors.

Let us consider the scenario in which the $\{u_i = 1, \ldots, N\}$ are binary-valued. According to the standard NP lemma, the optimal
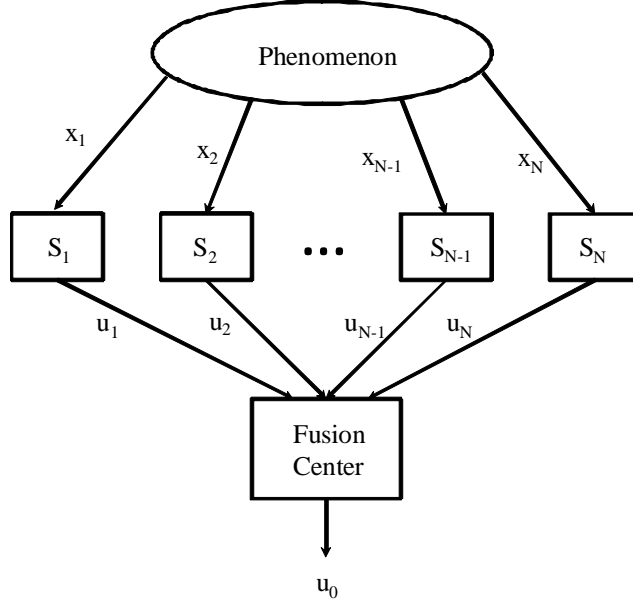
**Figure 3.1** Parallel topology with fusion center.

fusion center test is given by

$$L(\boldsymbol{u}) = \frac{P(\boldsymbol{u}; \mathcal{H}_1)}{P(\boldsymbol{u}; \mathcal{H}_0)} \begin{cases} > \gamma_0 \Rightarrow u_0 = 1 \\ = \gamma_0 \Rightarrow u_0 = 1 \text{ with probability } q \\ < \gamma_0 \Rightarrow u_0 = 0, \end{cases} \quad (3.1)$$

where the threshold $\gamma_0$ and the randomization constant $q$ are chosen to fulfill the constraint on the false alarm probability. Thus, the optimal fusion center test is a likelihood ratio test (LRT). Let $g_0(\boldsymbol{u})$ be the binary decision function rule at the fusion center. Since $g_0(\boldsymbol{u})$ is either 0 or 1, and all the possible combinations of decisions $\{u_1, \ldots, u_N\}$ is $2^N$, the set of all possible decision functions contain $2^{2^N}$ functions. However, not all these functions $g_0(\boldsymbol{u})$ can be optimal as the following lemma states [9].

**Lemma 3.2.1.** *Let the sensors individual decisions $u_i$, for $i = 1, \ldots, N$, be independent from each other conditioned on each hypothesis. Let $P_{F,i} = P(u_i = 1; \mathcal{H}_0)$ and $P_{D,i} = P(u_i = 1; \mathcal{H}_1)$ be*

*the false alarm and detection probabilities at the i-th sensor. As-*
*sume, without loss of generality, that for every sensor $P_{D,i} \geq P_{F,i}$.*
*Then, for a given vector $\boldsymbol{u}^*$ such that $L(\boldsymbol{u}^*) > \gamma_0$, and any other*
*vector $\boldsymbol{u}^+$ such that $u_i^+ \geq u_i^*$ for all i, $L(\boldsymbol{u}^+) > \gamma_0$.*

*Proof:* The conditional independence assumption implies that

$$L(\boldsymbol{u}) = \frac{P(\boldsymbol{u}; \mathcal{H}_1)}{P(\boldsymbol{u}; \mathcal{H}_0)} = \prod_{i=1}^{N} \frac{P(u_i; \mathcal{H}_1)}{P(u_i; \mathcal{H}_0)}. \tag{3.2}$$

The lemma becomes apparent when one exploits the relation $P_{D,i} \geq P_{F,i}$ in (3.2). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\bullet$

If $P_{D,i} = P_{F,i}$ for all sensors, the likelihood ratio at the fu-
sion center degenerates to one, identically for any combination of
the peripheral decisions [8]. Then, for any likelihood ratio test,
the false alarm probability $P_{F,0}$ and the detection probability $P_{D,0}$
at the fusion center are either a) both one, if the threshold is
less or equal to one, or b) both zero, if the threshold is greater
than one. In the first case, the fusion rule always favors the al-
ternative hypothesis, implying that $u_0 = 1$ for all $\boldsymbol{u}$, which is
a monotone increasing function satisfying Lemma 3.2.1. In the
second case, the fusion center always decides in favor of the null
hypothesis, independent of the combination of sensor decisions,
i.e., $u_0 = 0$ for all $\boldsymbol{u}$, which is a monotone increasing function
satisfying Lemma 3.2.1.

Whenever $P_{D,i} \leq P_{F,i}$, Lemma 3.2.1 still holds. Finally, if
for some sensors $P_{D,i} \geq P_{F,i}$ while for some others $P_{D,i} \leq P_{F,i}$,
Lemma 3.2.1 does not hold.

Since the decision variables $u_i$ are binary-valued, a likelihood
ratio test of the form (3.1) is equivalent to the fusion center deci-
sion $u_0 = g_0(\boldsymbol{u})$ being a Boolean function. Since $\boldsymbol{u}$ can assume $2^N$
possible values, the number of possible Boolean functions are $2^N$.
However, for an optimal solution of the NP problem, the fusion
has to satisfy Lemma 3.2.1. A Boolean function that satisfies the
monotonicity property given in Lemma 3.2.1 is called a *positive
unate function*.

Before to establish the optimality of the LRT, consider the following results [9].

**Lemma 3.2.2.** *For any fixed threshold $\gamma_0$ and any fixed monotonic function $t(u_1, \dots u_N)$, $P_{D,0}$ is an increasing function of the $P_{D,i}s$, for $i = 1, \dots, N$.*

*Proof:* The decision function that corresponds to the likelihood test at the fusion center is contained in the set of monotonic functions of $N$ variables. Consider one such monotone increasing function, $d(u_1, \dots, u_N)$. Being the random variables $U_1, \dots, U_N$ statistically independent, it is possible to compute $P_{D,0}$ knowing the $P_{D,i}s$, as shown in [8]. Taking partial derivatives of the $P_{D,0}$ with respect to $P_{D,i}s$, one obtains that $(\partial P_{D,0}/\partial P_{D,i}) > 0$. As an illustration, consider the function $d(u_1, u_2, u_3) = u_1 + u_2 u_3$. For this function, we have $P_{D,0} = P_{D,1} + P_{D,2}P_{D,3}$, form which $(\partial P_{D,0}/\partial P_{D,i}) > 0$, for $i = 1, 2, 3$. ●

The Optimality of the LRT can now be stated through the following theorem [9].

**Theorem 3.2.1.** *Under the assumption of statistical independence of the sensor decisions conditioned on each hypothesis, the optimal decision fusion rule for the parallel sensor topology consists of an NP test (or a randomized NP test) at the fusion center and a likelihood ratio test at all sensors.*

*Proof:* Consider a set of decision rules $G^* = (g_0^*, g_1^*, \dots, g_N^*)$ that achieves the desired $P_{F,0} = \alpha_0$ and at the same time achieves the detection probability $P_{D,0}^*$. Suppose that, for this strategy, local sensors have false alarm and detection probabilities $P_{F,i}^*$ and $P_{D,i}^*$, respectively. Consider another set of decision rules $G = (g_0, g_1, \dots, g_N)$, with $g_0 = g_0^*$, with the same local false alarms $P_{F,i}^*$ but different detection probabilities. The NP strategy implies that the LRT achieves the largest possible detection probability for a given false alarm level. Thus, if $G$ is such that each local decision is a LRT, then $P_{D,i} \geq P_{D,i}^*$. Since the optimal fusion rule $g_0^*$ has to be

a monotone rule, lemma 3.2.2 implies that $P_{D,0} \geq P_{D,0}^*$. Thus an optimal solution to the NP distributed detection problem should employ LRT's at the local sensors.                                        •

We have shown that both the local and global decision rules solving the parallel distributed detection problem are LRT's, but it is quite difficult to find the actual LRT's. This is because the threshold $\gamma_0$ in (3.1) as well as the local thresholds $\gamma_i$, for $i = 1, \ldots, N$ involved in the local tests

$$L(x_i) = \frac{p(x_i; \mathcal{H}_1)}{p(x_i; \mathcal{H}_0)} \begin{cases} > \gamma_i \Rightarrow u_i = 1 \\ = \gamma_i \Rightarrow u_i = 1 \ with \ probability \ q_i \\ < \gamma_i \Rightarrow u_i = 0, \end{cases} \quad (3.3)$$

need to be determined so as to maximize $P_{D,0}$ for a given $P_{F,0} = \alpha_0$. However, if the likelihood ratio in (3.3) is a continuous random variable, then the randomization is unnecessary and $q_i$ can be put to zero without losing optimality. Now, by considering that (3.1) is a monotone fusion rule, it is possible to solve for the set of optimal local thresholds $\{\gamma_1, \ldots, \gamma_N\}$ for a given monotone fusion rule and compute the corresponding $P_{D,0}$. Then, it is possible consider other monotone fusion rules and get the corresponding detection probabilities. Finally, the optimal solution is the one monotone fusion rule and the corresponding local decision rules (3.3) that provide the largest detection probability for the fusion center. This way of solving the problem is possible only for very small values of $N$. The complexity increases with the number of sensors because the number of monotone rules grows exponentially with $N$ and, moreover, finding the optimal set of local thresholds for a given fusion rule is an optimization problem involving an $(N-1)$ dimensional search.

## 3.2.2   Asymptotic Considerations

In the NP formulation of the parallel detection problem, as stated in the Subsection 3.2.1, we wish to maximize the detection probability at the fusion center, $P_{D,0}$, while keeping fixed its false alarm probability to a value $P_{F,0} = \alpha_0$. Then, we keep $\alpha_0$ fixed and
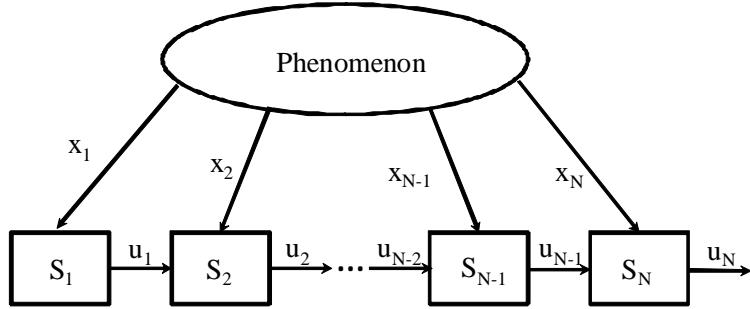
**Figure 3.2** Serial topology.

let the number of sensors $N$ diverge. If all the local sensors are using the same decision rule $\delta \in \Delta$, then, the detection probability converges to one exponentially fast and we have the following result [6, 7]

$$\lim_{N \to \infty} \frac{1}{N} \ln[1 - P_{D,0}] = -K(\delta), \qquad (3.4)$$

where $K(\delta)$ is the Kullback-Leibler information distance [42, 43] between the distributions of the random variable $\delta$ under the two hypotheses. Thus, finding an asymptotically optimal strategy, subject to every sensor to use the same decision rule, is equivalent to choose the decision rule so as to maximize $K(\delta)$. Moreover, it can be shown [7] that the restriction to identical decision strategies does not affect the exponent that rule the convergence rate of the detection probability to one.

## 3.3 Serial Configuration

Consider the serial or tandem configuration of $N$ distributed sensors shown in Fig. 3.2. The $(j-1)$-th sensor transmits its information $u_{j-1}$, function of its own observations, to the $j$-th sensor which generates its information based on its own observed data and the information received from the previous sensor. The first sensor in the system uses only the available observations to de-

rive the information $u_1$. The last sensor makes a decision, $u_N$, as to which one of two possible hypotheses the observations at the sensors correspond to. When the observations at the sensors are conditionally independent, the optimal solution to the NP problem can be found in [15, 18, 19]. The problem is intractable when the conditional independence assumption is not valid and when the number of sensors is $N > 2$.

The general serial array problem originates from a topic known as *learning with finite memory* [20–22]. In this scenario, the authors consider a sensor with a memory of $b$ bit, so that at each observation time it remembers only one of $2^b$ values. In practice, at the observation time $i$ the system observes new data, say $x_i \in \mathcal{X}$, and the previous opinion, $u_{i-1}$. Then, it generates a new opinion, $u_i$, by exploiting an appropriate mapping rule $g_i : \{1, \ldots, 2^b\} \times \mathcal{X} \to \{1, \ldots, 2^b\}$. In the binary case ($b = 1$) and for long sequences, we expect the detection performance to be severely degraded when compared to the infinite memory systems. Cover demonstrated in [20] that when the observations are independent and identically distributed and the likelihood ratio of the data is unbounded, then there exist functions $g_i$ such that the error probabilities vanish to zero.

Let us consider the case in which each sensor makes a binary decision, i.e., $u_i \in \{0, 1\}$, for $i = 1, \ldots, N$. Denoting the false alarm and the detection probabilities of the $j$-th stage as $P_{F,j}$ and $P_{D,j}$, respectively, the NP problem can be formulated as follows: subject to the constraint $P_{F,N} \leq \alpha_N$, find the decision rules at all the sensors so that these rules together achieve the maximum possible $P_{D,N}$.

We assume that the data at the sensors, conditioned on each hypothesis, are statistically independent. This implies that $X_j$ and $U_j$ are also conditionally independent. Suppose that the $j$-th sensor employs an NP test using data $(X_j, U_{j-1})$. The optimality of this assumption is analyzed in the next subsection. The

likelihood ratio at the $j$-th sensor becomes [15]

$$
\begin{aligned}
L(x_j, u_{j-1}) &= \frac{p(x_j; \mathcal{H}_1)}{p(x_j; \mathcal{H}_0)} \\
&\times \frac{P_{D,j-1}\mathcal{I}(u_{j-1}=1) + (1-P_{D,j-1})\mathcal{I}(u_{j-1}=0)}{P_{F,j-1}\mathcal{I}(u_{j-1}=1) + (1-P_{F,j-1})\mathcal{I}(u_{j-1}=0)},
\end{aligned}
\tag{3.5}
$$

where $p(x_j; \mathcal{H}_i)$, for $i = \{0, 1\}$, is the pdf of $X_j$ under hypothesis $\mathcal{H}_i$, and $\mathcal{I}(x = a)$ is the indicator function of the event $x = a$, i.e., $\mathcal{I}(x = a) = 1$ if and only if the event $x = a$ occurs; otherwise, $\mathcal{I}(x = a) = 0$. Therefore, if $u_{j-1} = 1$ the test at the $j$-th sensor is given by

$$
L(x_j) = \frac{p(x_j; \mathcal{H}_1)}{p(x_j; \mathcal{H}_0)} \frac{P_{D,j-1}}{P_{F,j-1}} \begin{cases} \geq \tau \Rightarrow u_j = 1 \\ < \tau \Rightarrow u_j = 0. \end{cases}
\tag{3.6}
$$

When $u_{j-1} = 0$ the test at the $j$-th sensor is given by

$$
L(x_j) = \frac{p(x_j; \mathcal{H}_1)}{p(x_j; \mathcal{H}_0)} \frac{1 - P_{D,j-1}}{1 - P_{F,j-1}} \begin{cases} \geq \tau \Rightarrow u_j = 1 \\ < \tau \Rightarrow u_j = 0. \end{cases}
\tag{3.7}
$$

It is convenient to use the log likelihood ratio, $\Delta(x_j) = \ln L(x_j)$. The local tests can be thus rewritten as

$$
\Delta(x_j) \begin{cases} \geq t_{j,i} \Rightarrow u_j = 1 \\ < t_{j,i} \Rightarrow u_j = 0, \end{cases}
\tag{3.8}
$$

where $i = 1$ when $u_{j-1} = 1$ and $i = 0$ when $u_{j-1} = 0$. Moreover,

$$
\Delta(x_j) = \ln \frac{p(x_j; \mathcal{H}_1)}{p(x_j; \mathcal{H}_0)}
\tag{3.9}
$$

and

$$
t_{j,1} = t_{j,0} + \ln \left( \frac{P_{F,j-1}}{1 - P_{F,j-1}} \frac{1 - P_{D,j-1}}{P_{D,j-1}} \right),
\tag{3.10}
$$

for $j = 2, \ldots, N$ and for the first stage $t_{1,1} = t_{1,0}$.

The false alarm probability at the $j$-th stage is given by

$$
\begin{aligned}
P_{F,j} = \quad & P \quad (\Delta(x_j) \geq t_{j,0}|\mathcal{H}_0, u_{j-1} = 0)P(u_{j-1} = 0|\mathcal{H}_0) \\
+ \quad & P(\Delta(x_j) \geq t_{j,1}|\mathcal{H}_0, u_{j-1} = 1) \\
\times \quad & P(u_{j-1} = 1|\mathcal{H}_0).
\end{aligned}
\tag{3.11}
$$

By using (3.11) and the conditional independence assumption, we get

$$
\begin{aligned}
P_{F,j} &= a_j(1 - P_{F,j-1}) + b_j P_{F,j-1} \\
P_{D,j} &= c_j(1 - P_{D,j-1}) + d_j P_{D,j-1},
\end{aligned}
\tag{3.12}
$$

where

$$
\begin{aligned}
a_j &= P(\Delta(x_j) \geq t_{j,0}|\mathcal{H}_0) \\
b_j &= P(\Delta(x_j) \geq t_{j,1}|\mathcal{H}_0) \\
c_j &= P(\Delta(x_j) \geq t_{j,0}|\mathcal{H}_1) \\
d_j &= P(\Delta(x_j) \geq t_{j,1}|\mathcal{H}_1).
\end{aligned}
\tag{3.13}
$$

Equations (3.8), (3.11), (3.13) can be used to evaluate the local detection probabilities recursively, provided that the false alarm probabilities are specified. However, for a given $P_{F,N}$ this procedure does not guarantee a maximum $P_{D,N}$. In order to globally maximize $P_{D,N}$ for a given $P_{F,N}$, we need a multidimensional search with respect to the variables $P_{F,j}s$, for $j = 1, \ldots, (N-1)$.

## 3.3.1   Global Optimality

The global optimization problem consists of finding the local tests at the sensors such that the detection probability of the last sensor, $P_{D,N}$, is maximized for a given false alarm probability $P_{F,N} = \alpha_N$. The following theorem [15] shows that the global optimality is achieved when each sensor adopts the NP test, justifying the assumption made previously.

**Theorem 3.3.1.** *Given that the observations at each stage in a serial distributed detection environment with N sensors are independent identically distributed (IID), the probability of detection is*

*maximized for a given probability of false alarm, at the N-th stage, when each sensor employs the NP test.*

*Proof:* At the last sensor of the system in Fig. 3.2, the NP test based on the dataset $(x_N, u_{N-1})$ maximizes $P_{D,N}$ for a fixed $P_{F,N}$ [36–38]. The log likelihood is given by

$$L(x_N, u_{N-1}) = \ln \frac{p(x_N, u_{N-1}; \mathcal{H}_1)}{p(x_N, u_{N-1}; \mathcal{H}_0)}, \tag{3.14}$$

while the log likelihood based only on the observations is

$$\Delta(x_N) = \ln \frac{p(x_N; \mathcal{H}_1)}{p(x_N; \mathcal{H}_0)}. \tag{3.15}$$

Then, we can write that

$$
\begin{aligned}
P(L < \tau; \mathcal{H}_1) = \quad & P_{D,N-1}\, P\left(\Delta + \ln\left(\frac{P_{D,N-1}}{P_{F,N-1}}\right) < \tau; \mathcal{H}_1\right) \\
& + (1 - P_{D,N-1}) \\
& \times P\left(\Delta + \ln\left(\frac{1 - P_{D,N-1}}{1 - P_{F,N-1}}\right) < \tau; \mathcal{H}_1\right).
\end{aligned}
\tag{3.16}
$$

By introducing the cumulative distributions and the density functions $F_1()$, $f_1()$ and $F_0()$, $f_0()$ of $\Delta$ under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively, we get

$$
\begin{aligned}
1 - P_{D,N} = \quad & P_{D,N-1} F_1\left(\tau - \ln\left(\frac{P_{D,N-1}}{P_{F,N-1}}\right)\right) \\
& + (1 - P_{D,N-1}) F_1\left(\tau - \ln\left(\frac{1 - P_{D,N-1}}{1 - P_{F,N-1}}\right)\right).
\end{aligned}
\tag{3.17}
$$

Similarly, we obtain

$$
\begin{aligned}
1 - P_{F,N} = \quad & P_{F,N-1} F_0\left(\tau - \ln\left(\frac{P_{D,N-1}}{P_{F,N-1}}\right)\right) \\
& + (1 - P_{F,N-1}) F_0\left(\tau - \ln\left(\frac{1 - P_{D,N-1}}{1 - P_{F,N-1}}\right)\right).
\end{aligned}
\tag{3.18}
$$

Given the value of $P_{F,N}$ and for any arbitrary but fixed value of $P_{F,N-1}$, we require the $P_{D,N}$ to be a monotonic increasing function of $P_{D,N-1}$. Taking the derivative of (3.18) with respect to $P_{D,N-1}$ and equating the result to zero, we get

$$\frac{d\tau}{dP_{D,N-1}} = \frac{\frac{P_{F,N-1}}{P_{D,N-1}} f_0(x_1) - \frac{1-P_{F,N-1}}{1-P_{D,N-1}} f_0(x_2)}{P_{F,N-1} f_0(x_1)(1 - P_{F,N-1}) f_0(x_2)}, \tag{3.19}$$

where

$$\begin{aligned} x_1 &= \tau - \ln(P_{D,N-1}/P_{F,N-1}) \\ x_2 &= \tau - \ln\left(\frac{1 - P_{D,N-1}}{1 - P_{F,N-1}}\right). \end{aligned} \tag{3.20}$$

In the same way, we can get

$$\begin{aligned} \frac{d(1 - P_{D,N})}{dP_{D,N-1}} = \quad & F_1(x_1) - F_1(x_2) \\ & + \left[ P_{D,N-1} f_1(x_1) \left(\frac{d\tau}{dP_{D,N-1}} - \frac{1}{P_{D,N-1}}\right) \right. \\ & + (1 - P_{D,N-1}) f_1(x_2) \\ & \left. \times \left(\frac{d\tau}{dP_{D,N-1}} - \frac{1}{1 - P_{D,N-1}}\right) \right]. \end{aligned} \tag{3.21}$$

A reasonable requirement is $P_{D,N-1} > P_{F,N-1}$ and thus $F_1(x_1) - F_2(x_2) < 0$. Consequently, a sufficient condition for $\frac{P_{D,N}}{P_{D,N-1}} > 0$ is that the term in the brackets in (3.21) be less than or equal to zero. By using (3.20), we get the following sufficiency condition

$$\frac{\frac{f_1(x_2)}{f_0(x_2)}}{\frac{f_1(x_1)}{f_0(x_1)}} \le e^{x_2 - x_1}. \tag{3.22}$$

But the likelihood ratio of the likelihood ratio is the likelihood ratio itself [37], then it follows that (3.22) is satisfied with equality.
●

As in the parallel case considered in Section 3.2, all the local tests made by single sensors are likelihood ratio tests, but solving for the optimal thresholds is, in general, quite difficult. For

example, for $N = 2$ and $N = 3$, Viswanathan, Thomopoulos and Tumuluri [15] solve the problem and present ROC curves for the Gaussian and Rayleigh noise cases. Unfortunately, the complexity of solving the simultaneous equations for the thresholds, limits the system lengths considered by these authors. However, algorithms that can obtain optimal thresholds with complexity that is linear in $N$ have been introduced [25].

## 3.3.2 Asymptotic Considerations

The NP formulation of the serial detection problem, as stated in Subsection 3.3.1, consists of maximizing the detection probability of the last sensor in the system, $P_{D,N}$, subject to the constraint on its false alarm probability, $P_{F,N} = \alpha_N$. We keep fixed the value of $\alpha_N$ and let the number of sensors $N$ grow to infinity. It has been conjectured in [6, 18, 19, 26] that the detection probability goes to one with a rate that is subexponential. Only recently this idea has been formally demonstrated in [23], where the authors study the Bayesian decentralized binary hypothesis-testing problem in a network of sensor arranged in a tandem. In practice, they consider the behavior of the probability of error at the sensor $N$, defined as $P_{E,N} = \pi_0 P_{F,N} + \pi_1(1 - P_{D,N})$, where $\pi_0$ and $\pi_1$ are the a priori probabilities of the hypotheses. The first result derived in [23] is that the rate of decay of the error probability is always subexponential, i.e.

$$\lim_{N \to \infty} \frac{1}{N} P_{EN}^* = 0, \tag{3.23}$$

where $P_{E,N}^* = \inf P_{E,N}$, with the infimum taken over all possible strategies. Under the additional assumption of bounded Kullback-Leibler divergences, it is shown in [23] that for all $d > 1/2$, the error probability is $\Omega(e^{-cNd})$, where $c$ is a positive constant[1]

Under a NP formulation, the picture is less complete. In [24], the authors establish that the Type II error probability decays sub-

---

[1]For two nonnegative functions $f$ and $g$, we write $f(n) = \Omega(g(n))$ (resp. $f(n) = O(g(n))$) if, for all $n$ sufficiently large, there exists a positive constant $c$ such that $f(n) > cg(n)$ (resp. $f(n) \leq cg(n)$).

exponentially, by considering the case in which the message sent by each node of the network is a NP optimal decision at that node. In particular, they suppose that each node $i$ can rely on a random variable, say $V_i$, which acts like a randomization variable. Given the received message, $U_{i-1}$, the randomization variable $V_i$, and its own observation $X_i$, the $i$-th node makes its decision $U_i$ so as to minimize its miss detection probability, subject to a constraint on its false alarm probability. Such a strategy is known as a *myopic* one. The case of general strategies is an interesting open problem.

## 3.4   Serial vs. Parallel

The serial configuration has serious reliability problems. First of all, since each sensor in the system has to wait for results from the previous stage, a delays problem exists. This problem can be overcome by modifying the communication structure as in [15]. A second and more serious problem is that the performance degrade considerably if the link is broken at an intermediate stage.

When comparing the serial and parallel architecture, a very interesting question arises: can serial networks provide a better detection performance than the parallel network in the absence of any failures? For distributed detection networks made of two sensors, it can be shown [6, 15, 35] that the tandem architecture can perform as well as the optimal parallel structure.

Similar results on the relative performance of serial and parallel networks consisting of more than two detectors are not available. For networks in which sensors generate binary decisions, a related result is that there always exists a better serial rule than a parallel fusion rule that is implementable as a sequence of two-input and one-output Boolean rules [15]. However, it is possible that that an optimal parallel fusion rule does not belong to the class of fusion rules that are implementable as a sequence of two-input and one-output rules and parallel network might considerably outperforms the serial network. In fact, asymptotically, as compared to the parallel scheme, the probability of miss detection for a serial

network goes to a zero at a much slower rate.

## 3.5   Summary

In this chapter we have discussed some basic issues concerning the problem of distributed detection with multiple sensors.

Two main sensor architectures, parallel and serial, were considered in some detail and in a Neyman-Pearson framework. A fundamental result is that for conditionally statistically independent observation at the sensors, the optimal tests at the sensors and at the fusion center, are likelihood ratio threshold tests. However, finding the actual tests implies the determination of the single thresholds, that generally passes through a set of coupled integral equations, a task that is computationally complex, especially for systems made of a large number of sensors.

Furthermore, we have reported some considerations on the behavior of both the architectures for a large number of sensors. In particular, it can be shown that, asymptotically with $N$, the probability of miss detection for a serial network goes to zero at a much slower rate than a parallel system.

# Chapter 4

# The Unlucky Broker Problem

In this chapter we formulate, analyze, and solve a novel topic in the framework of detection theory, here referred to as the *unlucky broker problem*. In particular, suppose that at a given time instant a standard statistical hypothesis test is formulated, leading to a binary decision made by using a certain data set. Later, suppose that part of the data set is lost, and a refinement of the test is required by exploiting both the surviving data and the previous decision. The unlucky broker problem faces the question: What is the best one can do in this scenario?

Motivating examples are presented in Section 4.1. In Section 4.2, the problem is formally stated in the NP formulation and it is faced by standard tools from detection theory. In Section 4.3, we afford the general form of the optimal detectors, and discuss their operative modalities. Examples of applications are showed in Section 4.4, while the Section 4.5 analyzes the Bayesian formulation of the problem. In Section 4.6 we investigate what happens when the observations originally available were correlated. Finally, Section 4.7 summarizes the main results and concludes the chapter.
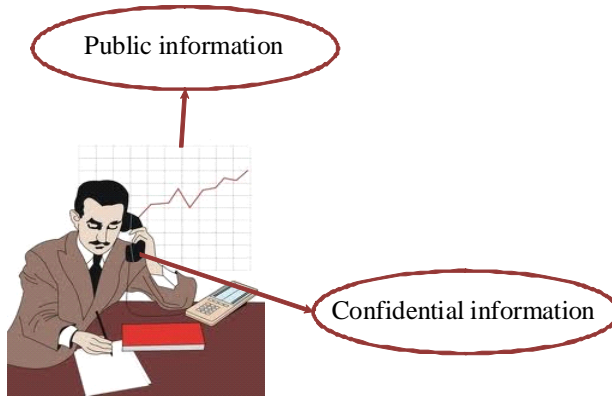
**Figure 4.1**  Depiction of the Unlucky Broker motivating example.

# 4.1    Motivating Examples

In this section, we present three different examples which moti-
vate the study of the unlucky broker problem. They allow us to
recognize it as a completely novel problem in the area of detec-
tion theory, with a broad range of applications in several fields
(economics, medicine, wireless sensor networks, etc...).

## 4.1.1    The Unlucky Broker

A first motivating example, that reveals the origin of the name
*unlucky broker*, can be found in everyday life. Bernard is a broker
and his job consists of suggesting an appropriate portfolio assess-
ment to his customers by using two data sets: one in the public
domain and another one made of certain confidential information
he has received. To suggest the appropriate investments, Bernard
must decide between, say, a positive or a negative market trend
during this week, see Fig. 4.1. His decision criterion is to maximize
the probability of correctly predicting a positive trend, subject to
a constraint on the corresponding probability of wrongly making
such prediction. Bernard makes his decision. Just before visiting
his customers to communicate his decision and to propose them

**Figure 4.2** A medical doctor involved in making a decision about surgery.

the appropriate course of action, Bernard is informed that his colleague Charlie lost his job as a consequence of having proposed many bad investments. Charlie, indeed, wrongly predicted a positive trend of the market the week before. At this point, Bernard wants to revise his own decision: he must ensure that the probability of wrongly predicting a positive trend be much lower than that initially thought. Clearly, he understands that this would unavoidably imply also a much lower probability of correctly predicting a positive trend and, consequently, that his potential income will be substantially decreased. However, saving his job is Bernard's main priority. Unfortunately, the unlucky Bernard has lost the files containing the confidential information. His new decision about the weekly trend must be based only on the public domain data set.

What should Bernard do? Should he simply retain the previous decision, or should he ignore that, and use only the currently available data set for a completely new decision? Or, what else?

## 4.1.2 The Medical Doctor

Mark is a medical doctor and Paula is one of his patients. Based on the ensemble of certain clinical parameters, Mark should make a decision about two alternatives: to operate or not to operate on Paula, see Fig. 4.2. As in the previous case, the decision is made so as to maximize the probability of operating when it is appropriate
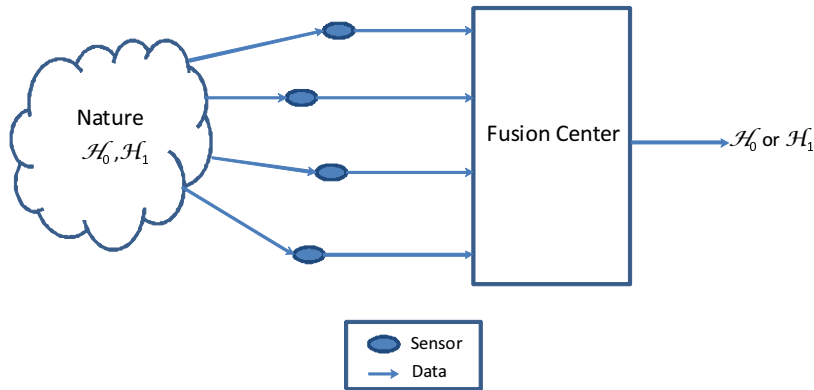
**Figure 4.3**  A Wireless Sensor Network engaged in a binary hypothesis-testing problem.

while limiting the probability of an unnecessary operation. Mark makes his decision about the surgery.

The day after, Mark is informed that many patients operated upon for the same disease that is suspected to affect Paula have incurred heart troubles. At this point, Mark wants to revise his original decision in order to cope with a much smaller probability of making an unnecessary surgery. Unfortunately, part of the clinical tests is no longer available. Can we suggest what Mark should do at this point? He remembers his original decision, but has now available only a fraction of the clinical tests made by Paula. Is it better to maintain the original decision (after all, that was made using a more complete clinical picture), or is the original decision useless and a completely new decision based on the currently available data must be conceived? Perhaps, Mark's intuition suggests that neither of these two extremes is the best option. But it is not obvious how to proceed.

## 4.1.3   Wireless Sensor Network

A Wireless Sensor Network is engaged in a binary detection task. Each node of the system collects measurements about the state of

the nature ($\mathcal{H}_0$ or $\mathcal{H}_1$) to be scrutinized, and delivers these data to a common fusion center where the final decision is made, as depicted in Fig. 4.3. The fusion center, upon receiving the whole set of data, implements an optimal NP test that maximizes the detection probability subject to the constraint that the false alarm level does not exceed a given value.

Later, the system is required to modify its operating modality: the required optimal NP decision must be now compatible with a false alarm level different from that used for the original decision. The fusion center did not store neither the data nor a sufficient statistic thereof and, consequently, it queries the sensors to retransmit their measurements. Unfortunately, part of the sensors are out of order at the time of this request, so that the fusion center collects only a subset of the data originally available to it. The problem we want to address is how an optimal NP decision can be made, based on both the surviving data and the original decision, that is the optimal NP decision made upon observing the full data set.

# 4.2  Problem Formalization

The examples discussed so far pose a common inference problem that we call that of the *unlucky broker*. In this section, we provide a suitable formalization of the described situations and we show how the unlucky broker problem can be linked to existing literature. Then, the unlucky broker problem is solved by exploiting standard tools from statistical decision theory, and its special features are emphasized. The main focus of this chapter is, in fact, on the insights provided by the optimal solution to the unlucky broker problem in different settings of practical interest.

## 4.2.1  Abstraction

In statistical terminology, by abstracting the matter, the unlucky broker problem can be described as follows.

Initially, one is faced with a standard binary statistical test between two simple hypotheses, such that a decision $\delta$ obeying the NP optimality criterion is made, on the basis of the available data set $(\boldsymbol{x}, \boldsymbol{y})$, made of independent, identically distributed (i.i.d.) observations. Then, the vector $\boldsymbol{y}$ is lost and one wants to make a new decision $\delta_{\mathcal{B}}$ between the binary state of the nature using the pair $(\boldsymbol{x}, \delta)$, see Fig. 4.4.

Note that a similar problem arises in decentralized detection with tandem (serial) architecture, analyzed in Chapter 3, in which each unit makes a decision on the basis of its own data and of the decision made by one neighbor; an abundant literature is available on this topic, see, e.g., [13–25]. In our case, however, the decision $\delta$ does depend upon $\boldsymbol{x}$, and this makes the problem essentially different from that considered in the literature.

If the decision $\delta_{\mathcal{B}}$ must be taken at the same false alarm level as $\delta$, then data $\boldsymbol{x}$ become irrelevant and the best one can do is to retain the original inference: $\delta_{\mathcal{B}} = \delta$. This, as we will show, is well known and understood. On the other hand, it is interesting to investigate what happens when a different false alarm level is desired. Since $\delta$ and $\delta_{\mathcal{B}}$ are both binary variables, the issue can be rephrased by saying that our interest is in understanding when and why the original decision $\delta$ should be retained, or flipped to $1 - \delta$. We will address such issues in this chapter.

We use upper case letters to denote random variables, with the corresponding lower case letters representing the associated realizations; vectors are displayed in bold.

Let $\boldsymbol{X} = [X_1, X_2, \ldots, X_{N_x}]$ and $\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_{N_y}]$ be two independent continuous-valued random vectors. The following binary hypothesis test is to be solved:

$$
\begin{array}{llll}
\mathcal{H}_0 & : & X_i \sim f_X(x; \mathcal{H}_0), & Y_j \sim f_Y(y; \mathcal{H}_0), \\
\mathcal{H}_1 & : & X_i \sim f_X(x; \mathcal{H}_1), & Y_j \sim f_Y(y; \mathcal{H}_1),
\end{array}
\tag{4.1}
$$

where $i = 1, 2, \ldots, N_x$, and $j = 1, 2, \ldots, N_y$.

In the above, $f_X(x; \mathcal{H}_0)$ is the marginal probability density function (pdf, for short) of the variable $X_i$, under hypothesis $\mathcal{H}_0$. Similarly, $f_X(x; \mathcal{H}_1)$ is the pdf under $\mathcal{H}_1$, while $f_Y(y; \mathcal{H}_1)$ and
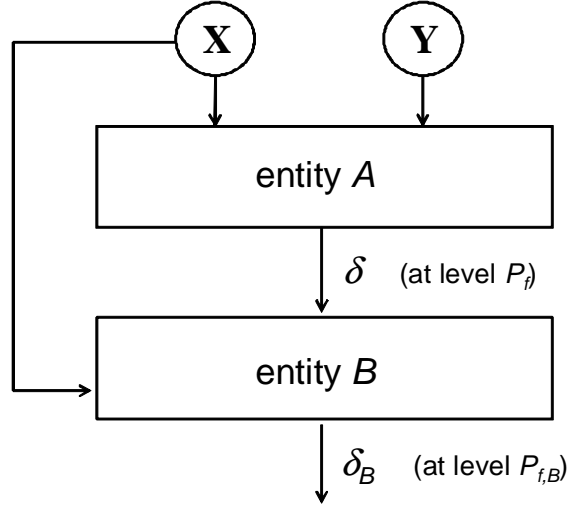
**Figure 4.4** Conceptual scheme of the unlucky broker problem.

$f_Y(y; \mathcal{H}_0)$ are the corresponding quantities pertaining to $Y_j$. We assume throughout the chapter that both $\boldsymbol{X}$ and $\boldsymbol{Y}$ have mutually i.i.d. entries.

The optimal NP strategy (see, e.g., [36–38]) amounts to comparing the log-likelihood ratio

$$L(\boldsymbol{x}, \boldsymbol{y}) \;=\; \sum_{i=1}^{N_x} \ln \frac{f_X(x_i; \mathcal{H}_1)}{f_X(x_i; \mathcal{H}_0)} + \sum_{j=1}^{N_y} \ln \frac{f_Y(y_j; \mathcal{H}_1)}{f_Y(y_j; \mathcal{H}_0)}$$
$$=\; L_x(\boldsymbol{x}) + L_y(\boldsymbol{y}), \tag{4.2}$$

to a threshold $\gamma$, yielding the decision rule $\delta$:

$$\delta(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 1, & L(\boldsymbol{x}, \boldsymbol{y}) \geq \gamma, \\ 0, & L(\boldsymbol{x}, \boldsymbol{y}) < \gamma. \end{cases}$$

Borrowing standard terminology from detection theory, we introduce the false alarm and detection probabilities:

$$P_f \;=\; \Pr\{L(\boldsymbol{X}, \boldsymbol{Y}) \geq \gamma; \mathcal{H}_0\},$$
$$P_d \;=\; \Pr\{L(\boldsymbol{X}, \boldsymbol{Y}) \geq \gamma; \mathcal{H}_1\}.$$

With reference to Fig. 4.4, the basic problem addressed in this chapter is as follows. A certain entity (say, $\mathcal{A}$) observes the data $(\boldsymbol{x}, \boldsymbol{y})$, and accordingly implements an optimal NP test at a prescribed false alarm level $P_f$, ending up with a decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ at (the best) detection probability level $P_d$. Another entity, say $\mathcal{B}$, observes a portion of the data, namely, the vector $\boldsymbol{x}$, and also receives from $\mathcal{A}$ its decision $\delta(\boldsymbol{x}, \boldsymbol{y})$. Note that $\mathcal{B}$ has no access to the data vector $\boldsymbol{y}$. The goal is to make the best NP decision, say $\delta_{\mathcal{B}}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$ (the symbol "$\mathcal{B}$" is mnemonic of "broker"), based upon the data accessible at site $\mathcal{B}$, i.e., the observation vector $\boldsymbol{x}$ and the binary decision $\delta(\boldsymbol{x}, \boldsymbol{y})$.

The road toward deriving the optimal detector based upon the set of information $(\boldsymbol{x}, \delta)$ available to $\mathcal{B}$ is well traced and understood: it amounts to computing the likelihood-ratio of the pair $(\boldsymbol{x}, \delta)$ and comparing it to a threshold. From a signal processing perspective, the interest is in characterizing the real *modus operandi* of the detector, and, in this connection, several questions arise. At one extreme, should $\mathcal{B}$ simply retain the original decision $\delta$? One could naively guess that a decision based upon $(\boldsymbol{x}, \boldsymbol{y})$ should be better than that available to $\mathcal{B}$, where only the reduced data set $\boldsymbol{x}$ can be exploited. After a closer look, this choice exhibits a severe limitation, in that it does not leave to entity $\mathcal{B}$ the chance of changing the false alarm probability originally set by $\mathcal{A}$.

At the other extreme, should $\mathcal{B}$ simply ignore $\delta(\boldsymbol{x}, \boldsymbol{y})$ and implement an optimal NP decision based on $\boldsymbol{x}$, i.e., a likelihood ratio threshold test on these data? While it is true that the data $\boldsymbol{y}$ are irremediably lost, the (one-bit) quantized information packed into $\delta(\boldsymbol{x}, \boldsymbol{y})$ acts as a *side information* [42, 44] (or, more specifically, side decision) for the detector $\mathcal{B}$, integrating the data $\boldsymbol{x}$ still available, and, as such, it should be taken into account. Accordingly, in general, the answer is expected to be at neither of these extremes.

It should be clear that, as long as $\mathcal{B}$ works at the same false alarm level chosen by $\mathcal{A}$, then $\delta_{\mathcal{B}} = \delta$ is the best possible final decision. On the other hand, when the false alarm constraint is different, the decision may have to be refined. Thus, to avoid misunderstanding, the decision *refinement*, as used in this chapter,

refers to the possibility allowed to $\mathcal{B}$ of working at a different false alarm level than that used at site $\mathcal{A}$ and the consequent potential change in the optimum decision made by $\mathcal{B}$ based on the available data $\boldsymbol{x}$ and decision $\delta$ obtained at $\mathcal{A}$. It remains true that the ROC (Receiver Operating Characteristic) $P_{d,\mathcal{B}} = P_{d,\mathcal{B}}(P_{f,\mathcal{B}})$ achievable by $\mathcal{B}$ must be upper bounded by the ROC $P_d = P_d(P_f)$ available at site $\mathcal{A}$. Therefore, here refinement does not refer to the final operating characteristic, but rather to the possibility of arbitrarily modifying the optimization constraint. It should be clear that the notion of decision refinement is not to be confused with that of sample-space refinement used in information theory, see e.g., [45].

## 4.3   Problem Solution

This section explores the solution to the unlucky broker problem described in the above. In Subsection 4.3.1, we find the structure of the optimal decision maker, while a closer look to the solution is given in Subsection 4.3.2. Finally, the expressions for the false alarm and detection probabilities of the unlucky broker are provided in Subsection 4.3.3.

### 4.3.1   Structure of the optimal decision maker

The data set available at site $\mathcal{B}$ is the vector $\boldsymbol{x}$ and the decision $\delta = \delta(\boldsymbol{x}, \boldsymbol{y})$, a binary variable, taken at site $\mathcal{A}$. The corresponding detection statistic can thus be written as the ratio of the two probability measures [36]

$$T(\boldsymbol{x}, \delta) = \ln \frac{P(\boldsymbol{x}, \delta; \mathcal{H}_1)}{P(\boldsymbol{x}, \delta; \mathcal{H}_0)}, \tag{4.3}$$

where $P(\boldsymbol{x}, \delta; \mathcal{H}_0)$ and $P(\boldsymbol{x}, \delta; \mathcal{H}_1)$ are the joint probabilities of the pair $(\boldsymbol{X}, \delta(\boldsymbol{X}, \boldsymbol{Y}))$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. These can be explicitly written, for $h = 0, 1$, as:

$$P(\boldsymbol{x}, \delta; \mathcal{H}_h) = \Pr\{\delta(\boldsymbol{X}, \boldsymbol{Y}) | \boldsymbol{X} = \boldsymbol{x}; \mathcal{H}_h\} \, f_X(\boldsymbol{x}; \mathcal{H}_h), \tag{4.4}$$

and also further expanded as

$$P(\boldsymbol{x}, 1; \mathcal{H}_h) = \Pr\{L_y(\boldsymbol{Y}) \geq \gamma - L_x(\boldsymbol{x}); \mathcal{H}_h\} f_X(\boldsymbol{x}; \mathcal{H}_h),$$
$$P(\boldsymbol{x}, 0; \mathcal{H}_h) = \Pr\{L_y(\boldsymbol{Y}) < \gamma - L_x(\boldsymbol{x}); \mathcal{H}_h\} f_X(\boldsymbol{x}; \mathcal{H}_h),$$

where conditioning has been removed thanks to the assumed independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Let $P_{fy}(z) = \Pr\{L_y(\boldsymbol{Y}) \geq z; \mathcal{H}_0\}$ and $P_{dy}(z) = \Pr\{L_y(\boldsymbol{Y}) \geq z; \mathcal{H}_1\}$ be, respectively, the false alarm and detection probabilities of an optimal NP test based upon $\boldsymbol{y}$. We obtain

$$
\begin{aligned}
P(\boldsymbol{x}, \delta; \mathcal{H}_0) &= f_X(\boldsymbol{x}; \mathcal{H}_0) \left\{\delta\, P_{fy}(\gamma - L_x(\boldsymbol{x}))\right. \\
&\quad + (1-\delta)[1 - P_{fy}(\gamma - L_x(\boldsymbol{x}))]\}, \\
P(\boldsymbol{x}, \delta; \mathcal{H}_1) &= f_X(\boldsymbol{x}; \mathcal{H}_1) \left\{\delta\, P_{dy}(\gamma - L_x(\boldsymbol{x}))\right. \\
&\quad + (1-\delta)[1 - P_{dy}(\gamma - L_x(\boldsymbol{x}))]\}.
\end{aligned}
\tag{4.5}
$$

By considering separately the two cases $\delta = 0, 1$, it is readily seen that the log-likelihood ratio can be written as

$$
\begin{aligned}
T(\boldsymbol{x}, \delta) &= L_x(\boldsymbol{x}) + \delta \ln \frac{P_{dy}(\gamma - L_x(\boldsymbol{x}))}{P_{fy}(\gamma - L_x(\boldsymbol{x}))} \\
&\quad + (1-\delta) \ln \frac{1 - P_{dy}(\gamma - L_x(\boldsymbol{x}))}{1 - P_{fy}(\gamma - L_x(\boldsymbol{x}))},
\end{aligned}
\tag{4.6}
$$

where, by convention, we set $0 \log(0/0) = 0$.

As can be seen, (4.6) is indeterminate either when $P_{fy}(\cdot) = P_{dy}(\cdot) = 0$, or when $P_{fy}(\cdot) = P_{dy}(\cdot) = 1$. However, note that, according to (4.4), $P_{fy}(\gamma - L_x(\boldsymbol{x})) = P_{dy}(\gamma - L_x(\boldsymbol{x})) = 0$ means $\Pr[\delta = 0 | L_x(\boldsymbol{x}); \mathcal{H}_0] = \Pr[\delta = 0 | L_x(\boldsymbol{x}); \mathcal{H}_1] = 1$, and, similarly, $P_{fy}(\gamma - L_x(\boldsymbol{x})) = P_{dy}(\gamma - L_x(\boldsymbol{x})) = 1$ means $\Pr[\delta = 1 | L_x(\boldsymbol{x}); \mathcal{H}_0] = \Pr[\delta = 1 | L_x(\boldsymbol{x}); \mathcal{H}_1] = 1$. Roughly speaking, whenever $P_{fy}(\cdot) = P_{dy}(\cdot) = 0$, $\delta = 0$, and whenever $P_{fy}(\cdot) = P_{dy}(\cdot) = 1$, $\delta = 1$. In these cases, eqs. (4.5) tell that $T(\boldsymbol{x}, \delta) = L_x(\boldsymbol{x})$, such that formula (4.6) can be safely used by employing the indicated convention.

Equation (4.6) emphasizes that the detection statistic is only a function of the decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ and of the log-likelihood ratio of data $\boldsymbol{X}$, namely $L_x(\boldsymbol{X})$. This suggests defining, with a slight

abuse of notation, the random variable $L_x = L_x(\boldsymbol{X})$. Now, by further defining

$$
\begin{aligned}
t_1(l_x) &= l_x + \ln \frac{P_{dy}(\gamma - l_x)}{P_{fy}(\gamma - l_x)}, \\
t_0(l_x) &= l_x + \ln \frac{1 - P_{dy}(\gamma - l_x)}{1 - P_{fy}(\gamma - l_x)},
\end{aligned}
\tag{4.7}
$$

we get

$$
T(l_x, \delta) = \delta \, t_1(l_x) + (1 - \delta) \, t_0(l_x).
\tag{4.8}
$$

Note that, in general, $P_{dy}(z)$ and $P_{fy}(z)$ may grow at different rates with their argument $z$, and the functions $t_{0,1}(l_x)$ in (4.8) are not necessarily monotone with $l_x$.

The optimal decision rule for entity $\mathcal{B}$ is well known to be

$$
\delta_{\mathcal{B}}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y})) =
\begin{cases}
1, & T(l_x, \delta) \geq \gamma_{\mathcal{B}}, \\
0, & T(l_x, \delta) < \gamma_{\mathcal{B}},
\end{cases}
\tag{4.9}
$$

where $\gamma_{\mathcal{B}}$ is a threshold set to ensure a false alarm probability of $P_{f,\mathcal{B}}$.

All throughout the chapter, we shall assume that the random variable $T(l_x, \delta)$ has no point masses, a condition grounded on the original working hypothesis of continuous-valued $X_i$'s and $Y_i$'s. Note also that the functions in (4.7) are continuous.

### 4.3.2  Modus Operandi

It is of interest to understand how the detector at site $\mathcal{B}$ works, in comparison with the decision maker at site $\mathcal{A}$. In practice, we want to understand when and why $\mathcal{B}$ changes the decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ made by $\mathcal{A}$. As a matter of fact, the decision rule has some physical interpretation and implications which can be grasped by exploring the properties of the functions $t_1(\cdot)$ and $t_0(\cdot)$ defined in (4.7). By the concavity of the ROC $P_{dy} = P_{dy}(P_{fy})$, it follows [37]

$$
\frac{P_{dy}(\gamma - l_x)}{P_{fy}(\gamma - l_x)} \geq e^{\gamma - l_x}, \quad \frac{1 - P_{dy}(\gamma - l_x)}{1 - P_{fy}(\gamma - l_x)} \leq e^{\gamma - l_x},
\tag{4.10}
$$

yielding

$$t_1(l_x) \geq \gamma, \quad t_0(l_x) \leq \gamma. \tag{4.11}$$

To prove the inequalities (4.10), let us consider the definition of the false alarm probability $P_{dy}$

$$
\begin{aligned}
P_{dy}(t) &= \Pr\{L_y(\boldsymbol{Y}) \geq t; \mathcal{H}_1\} \\
&= \int_{l_y \geq t} g_1(l_y)\, dl_y \\
&= \int_{l_y \geq t} \frac{g_1(l_y)}{g_0(l_y)}\, g_0(l_y)\, dl_y \\
&\geq e^t \int_{l_y \geq t} g_0(l_y)\, dl_y \\
&= e^t P_{fy}(t), \tag{4.12}
\end{aligned}
$$

where $g_1(l_y)$ and $g_1(l_y)$ represent the pdfs of the log-likelihood ratio $L_y(\boldsymbol{Y})$, under the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. Analogously, we obtain that $1 - P_{dy}(t) \leq e^t[1 - P_{fy}(t)]$ and the proof of inequalities (4.10) is complete.

Now, let us choose a value for the detection threshold $\gamma_\mathcal{B}$ larger than the threshold used by $\mathcal{A}$: $\gamma_\mathcal{B} > \gamma$. Suppose that a decision $\delta = 0$ has been made by entity $\mathcal{A}$. In this case, from the second inequality in (4.11) we get $t_0(l_x) \leq \gamma < \gamma_\mathcal{B}$. The immediate implication, in view of (4.9), is that the decision $\delta = 0$ is never to be changed. The opposite case that $\delta = 1$, implies comparing the detection statistic $t_1(x)$ with the threshold $\gamma_\mathcal{B}$. Looking at the first of (4.11), we have $t_1(x) \geq \gamma$. This, however, does not help, in general, for knowing whether $t_1(x) \geq \gamma_\mathcal{B}$, since $\gamma_\mathcal{B} > \gamma$.

The rule employed by the detector can be summarized by stating that, in the region $\gamma_\mathcal{B} > \gamma$, a decision in favor of $\mathcal{H}_0$ must be retained, while a decision in favor of $\mathcal{H}_1$ needs a further check: it is retained only if $t_1(l_x)$ is larger than, or equal to, the new threshold $\gamma_\mathcal{B}$. Similar arguments apply to the dual case $\gamma_\mathcal{B} < \gamma$: a decision in favor of $\mathcal{H}_1$ is always retained, while a decision in favor of $\mathcal{H}_0$ is retained only if $t_0(l_x)$ is less than the threshold $\gamma_\mathcal{B}$. Finally, for $\gamma_\mathcal{B} = \gamma$, the original decision is retained with probability one:

the unlucky broker ignores $\boldsymbol{x}$, and makes decisions with the same detection and false alarm probabilities of entity $\mathcal{A}$.

The behavior of the detector $\mathcal{B}$, which optimally solves the unlucky broker problem, is summarized in the following definition, where $u(\cdot)$ denotes the unit step function, and where the arguments of $\delta_{\mathcal{B}}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$ and $\delta(\boldsymbol{x}, \boldsymbol{y})$ are omitted for notational simplicity.

**Definition 4.3.1.** $\delta_{\mathcal{B}}$ *is a "multiple-threshold" detector with side decision if*

$$\delta_{\mathcal{B}} = \begin{cases} \delta \, u \left( t_1(l_x) - \gamma_{\mathcal{B}} \right), & \text{for } \gamma_{\mathcal{B}} > \gamma, \\ (1 - \delta) \, u \left( t_0(l_x) - \gamma_{\mathcal{B}} \right) + \delta, & \text{for } \gamma_{\mathcal{B}} \leq \gamma. \end{cases} \quad (4.13)$$

●

Suppose now that the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are invertible and strictly increasing. Then comparing $t_1(l_x)$ with $\gamma_{\mathcal{B}}$ is tantamount to comparing the log-likelihood $l_x$ to $t_1^{-1}(\gamma_{\mathcal{B}})$, and the same holds when comparing $t_0(l_x)$ with $\gamma_{\mathcal{B}}$. Let:

$$s_1 = t_1^{-1}(\gamma_{\mathcal{B}}), \quad s_0 = t_0^{-1}(\gamma_{\mathcal{B}}). \quad (4.14)$$

The detector then simplifies to the following.

**Definition 4.3.2.** $\delta_{\mathcal{B}}$ *is a "single-threshold" detector with side decision if*

$$\delta_{\mathcal{B}} = \begin{cases} \delta \, u \left( l_x - s_1 \right), & \text{for } \gamma_{\mathcal{B}} > \gamma, \\ (1 - \delta) \, u \left( l_x - s_0 \right) + \delta, & \text{for } \gamma_{\mathcal{B}} \leq \gamma. \end{cases} \quad (4.15)$$

●

The terminology should be clear. In both cases, detector $\delta_{\mathcal{B}}$ exploits $\delta(\boldsymbol{x}, \boldsymbol{y})$ (whence the reference to the side decision) and data $\boldsymbol{x}$, the latter entering in the computation only through their log-likelihood $l_x(\boldsymbol{x})$. In the single-threshold case, $l_x(\boldsymbol{x})$ is compared to a single threshold, i.e., $\delta_{\mathcal{B}}$ works, as regard to $\boldsymbol{x}$, just as a

log-likelihood threshold test. The multiple-threshold detector, instead, may involve much more tricky log-likelihood comparisons: it requires checking whether $l_x(\boldsymbol{x})$ belongs to some arbitrarily shaped subset of the real line, that typically amounts to comparing $l_x(\boldsymbol{x})$ to a set of different thresholds defining these regions.

At first glance, the existence of multiple-threshold detectors with side decision might appear counterintuitive. Recall however that a single-threshold detector requires monotonicity of the terms $t_{0,1}(l_x)$ that is not always guaranteed, as stated in the comment just below eq. (4.8). Later, we shall return to this point.

### 4.3.3   Performance evaluation

The performances pertaining to entity $\mathcal{B}$ can be evaluated by appealing to eq. (4.13). Let us first consider the case $\gamma_{\mathcal{B}} > \gamma$. We have

$$
\begin{aligned}
P_{d,\mathcal{B}} &= \Pr\{\delta_{\mathcal{B}} = 1; \mathcal{H}_1\} \\
&= \Pr\{\delta = 1, \delta_{\mathcal{B}} = 1; \mathcal{H}_1\} \\
&= \int_{t_1(l_x) \geq \gamma_{\mathcal{B}}} P_{dy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_1) dl_x \\
&= P_d - \int_{t_1(l_x) < \gamma_{\mathcal{B}}} P_{dy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_1) dl_x, \quad (4.16)
\end{aligned}
$$

where $f_{L_x}(l_x; \mathcal{H}_1)$ is the pdf of the random variable $L_x$ under the alternative hypothesis. The integral in (4.16) represents the probability that the original decision (although correct) is changed by entity $\mathcal{B}$; recall that $P_d$ is the detection probability of the detector operating on the full data set.

In a similar fashion, for the false alarm probability one gets

$$
P_{f,\mathcal{B}} = P_f - \int_{t_1(l_x) < \gamma_{\mathcal{B}}} P_{fy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_0) dl_x. \quad (4.17)
$$

In the opposite case of $\gamma_{\mathcal{B}} \leq \gamma$, the same arguments yield

$$
\begin{aligned}
P_{d,\mathcal{B}} &= P_d + \int_{t_0(l_x) \geq \gamma_{\mathcal{B}}} [1 - P_{dy}(\gamma - l_x)] f_{L_x}(l_x; \mathcal{H}_1) dl_x, \\
P_{f,\mathcal{B}} &= P_f + \int_{t_0(l_x) \geq \gamma_{\mathcal{B}}} [1 - P_{fy}(\gamma - l_x)] f_{L_x}(l_x; \mathcal{H}_0) dl_x.
\end{aligned}
\tag{4.18}
$$

The knowledge of the false alarm and detection probabilities of the unlucky broker allows us to evaluate the performance of the detector in terms of ROC. Moreover, from Eqs. (4.16), (4.17), we get that when $\gamma_{\mathcal{B}} > \gamma$ a decrease of false alarm probability occurs, i.e., $P_{f,\mathcal{B}} < P_f$. Similarly, from Eqs. (4.18) we see that when $\gamma_{\mathcal{B}} > \gamma$ an increase of false alarm probability occurs, i.e., $P_{f,\mathcal{B}} > P_f$.

## 4.4 Examples

We now discuss two specific hypothesis-testing problem, commonly encountered in practice: the Gaussian shift-in-mean, and the exponential shift-in-scale. Then, we show what happens for arbitrary distributions.

### 4.4.1 Exponential shift-in-scale

Let us consider the following hypothesis-testing problem

$$
\begin{aligned}
\mathcal{H}_0 &: \quad X_i, Y_j \sim \mathcal{E}(\lambda_0), \\
\mathcal{H}_1 &: \quad X_i, Y_j \sim \mathcal{E}(\lambda_1),
\end{aligned}
\tag{4.19}
$$

where $i = 1, 2, \ldots, N_x$, $j = 1, 2, \ldots, N_y$, and $\mathcal{E}(\lambda)$ is our shortcut for an exponential pdf with expectation $1/\lambda$. Without loss of generality, we assume $\lambda_0 > \lambda_1$.

**Proposition 4.4.1.** *For the exponential shift-in-scale hypothesis-testing problem in (4.19), the optimal solution to the unlucky broker problem is a single-threshold detector $\delta_{\mathcal{B}}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$.* •

To prove this claim, let us start by considering the pertinent log-likelihood ratios:

$$
\begin{aligned}
L_x(\boldsymbol{x}) &= N_x \ln\left(\frac{\lambda_1}{\lambda_0}\right) + (\lambda_0 - \lambda_1)\sum_{i=1}^{N_x} x_i, \\
L_y(\boldsymbol{y}) &= N_y \ln\left(\frac{\lambda_1}{\lambda_0}\right) + (\lambda_0 - \lambda_1)\sum_{j=1}^{N_y} y_j.
\end{aligned}
\tag{4.20}
$$

Let $\Delta_0 = \lambda_0/(\lambda_0 - \lambda_1)$, $\Delta_1 = \lambda_1/(\lambda_0 - \lambda_1)$, $\rho_x = N_x \ln(\lambda_1/\lambda_0) < 0$, $\rho_y = N_y \ln(\lambda_1/\lambda_0) < 0$. Also, let $\Gamma(N, \Delta)$ be a Gamma pdf with shape parameter $N$ and scale parameter $\Delta$ [49]. The random variables $L_x$ and $L_y$ are distributed as follows

$$
L_x \sim \rho_x + \Gamma(N_x, \Delta_h), \quad L_y \sim \rho_y + \Gamma(N_y, \Delta_h),
\tag{4.21}
$$

where $h = 0, 1$, denotes the hypothesis[1]. Introducing now the function $G(z; N, \Delta) = 1$, for $z < 0$, and

$$
G(z; N, \Delta) = \Pr\{\Gamma(N, \Delta) > z\} = \sum_{k=0}^{N-1} \frac{(\Delta z)^k}{k!} e^{-\Delta z}
\tag{4.22}
$$

otherwise [50, 51], we get

$$
P_{fy}(\gamma - l_x) = \begin{cases} G(\gamma - \rho_y - l_x; N_y, \Delta_0), & \rho_x \le l_x \le \gamma - \rho_y, \\ 1, & l_x > \gamma - \rho_y, \end{cases}
\tag{4.23}
$$

and

$$
P_{dy}(\gamma - l_x) = \begin{cases} G(\gamma - \rho_y - l_x; N_y, \Delta_1), & \rho_x \le l_x \le \gamma - \rho_y, \\ 1, & l_x > \gamma - \rho_y, \end{cases}
\tag{4.24}
$$

---

[1]With expressions like $L \sim \rho + \Gamma(N, \Delta)$ we mean that $L$ is a random variable distributed as the sum of the constant $\rho$ plus a random variable with distribution $\Gamma(N, \Delta)$. Similarly, later we shall write $\Pr\{\Gamma(N, \Delta) > a\}$ to denote the probability that a random variable distributed according to $\Gamma(N, \Delta)$ exceeds $a$.

where we used the fact that $l_x \geq \rho_x$ and that, to avoid trivialities, $\gamma - \rho_y > \rho_x$. We note explicitly that, complying with the discussion in footnote 1, for all $l_x > \gamma - \rho_y$, since $P_{fy}(\gamma - l_x) = P_{dy}(\gamma - l_x) = 1$, we have $t_1(l_x) = l_x$, and it makes no sense to consider $t_0(l_x)$ outside the range $(\rho_x, \gamma - \rho_y)$.

Using these relationships, the proof of Proposition 4.4.1, which amounts to showing that the (continuous) functions $t_1(l_x)$ and $t_0(l_x)$ for the exponential shift-in-scale problem (4.19) are strictly increasing, is detailed in Appendix A.1. There, it is also shown that at the extremes of their domain, $t_1(l_x)$ and $t_0(l_x)$ verify:

$$
\begin{aligned}
t_1(\rho_x) > \gamma, & \qquad \lim_{l_x \to +\infty} t_1(l_x) = +\infty, \\
t_0(\rho_x) > -\infty, & \qquad \lim_{l_x \to \gamma - \rho_y} t_0(l_x) = \gamma.
\end{aligned}
\tag{4.25}
$$

One implication is that the threshold $\gamma_\mathcal{B}$ is to be chosen, in the region $\gamma_\mathcal{B} > \gamma$, as $\gamma_\mathcal{B} \in [t_1(\rho_x), \infty)$. Note the gap between the original threshold $\gamma$ and the minimum of $t_1(l_x)$ at which new threshold $\gamma_\mathcal{B}$ begins to make a difference. Similarly, in the region $\gamma_\mathcal{B} \leq \gamma$, the interesting region is $\gamma_\mathcal{B} \in [t_0(\rho_x), \gamma)$. For these meaningful choices, we can state the following proposition, whose proof is deferred to Appendix A.2. Let

$$
\begin{aligned}
I(\Delta; s) = {} & \frac{\Delta^{N_x} e^{-\Delta (\gamma - \rho_x - \rho_y)}}{(N_x - 1)!} \sum_{k=0}^{N_y - 1} \Delta^k \\
& \times \sum_{i=0}^{k} \frac{(\gamma - \rho_x - \rho_y)^{k-i}}{(k-i)!} \frac{(-1)^i}{i!} \frac{(s - \rho_x)^{N_x + i}}{N_x + i}.
\end{aligned}
\tag{4.26}
$$

**Proposition 4.4.2.** *For the exponential shift-in-scale hypothesis-testing problem in (4.19), the performance of the single-threshold detector $\delta_\mathcal{B}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$ is as follows. In the region $\gamma_\mathcal{B} > \gamma$,*

$$
P_{d,\mathcal{B}} = \begin{cases} P_d - I(\Delta_1; s_1) & s_1 \leq \gamma - \rho_y, \\ G(s_1 - \rho_x; N_x, \Delta_1) & s_1 > \gamma - \rho_y, \end{cases}
\tag{4.27}
$$

$$
P_{f,\mathcal{B}} = \begin{cases} P_f - I(\Delta_0; s_1) & s_1 \leq \gamma - \rho_y, \\ G(s_1 - \rho_x; N_x, \Delta_0) & s_1 > \gamma - \rho_y. \end{cases}
\tag{4.28}
$$
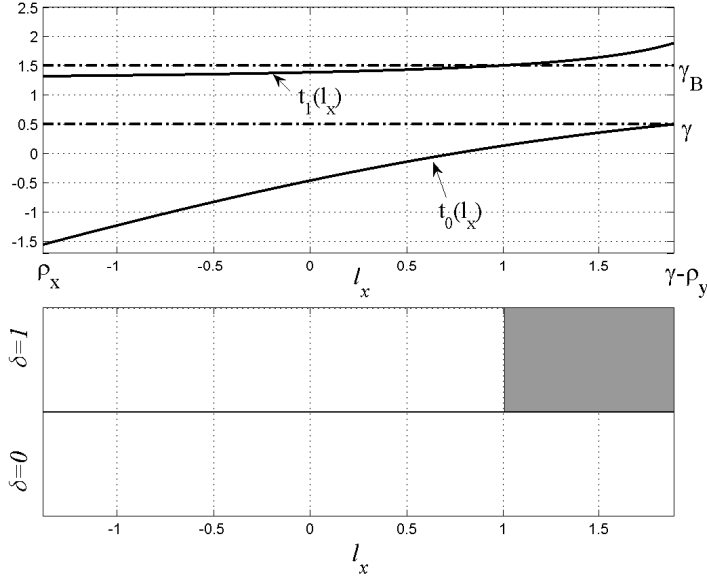
**Figure 4.5** Functions $t_1(l_x)$ and $t_0(l_x)$ (upper plot) and final decisions in the "plane" $(l_x, \delta)$ (lower plot) for the exponential shift-in-scale problem with $N_x = N_y = 2$, $\lambda_0 = 2$, $\lambda_1 = 1$ and $\gamma = 0.5$. The grey regions in the lower plot mean that $\delta_{\mathcal{B}} = 1$, i.e., the unlucky broker there decides for $\mathcal{H}_1$; the decision is $\delta_{\mathcal{B}} = 0$ in correspondence of the white regions.

*In the region $\gamma_{\mathcal{B}} \leq \gamma$,*

$$
\begin{aligned}
P_{d,\mathcal{B}} &= G(s_0 - \rho_x; N_x, \Delta_1) + I(\Delta_1, s_0), \\
P_{f,\mathcal{B}} &= G(s_0 - \rho_x; N_x, \Delta_0) + I(\Delta_0, s_0).
\end{aligned}
\tag{4.29}
$$

• 

In the above, $s_0$ and $s_1$ are defined as in (4.14). In the upper plot of Fig. 4.5 the functions $t_1(l_x)$ and $t_0(l_x)$ are displayed for the case study $N_x = N_y = 2$, $\lambda_0 = 2$, $\lambda_1 = 1$, and $\gamma = 0.5$. Note how the horizontal dashed line corresponding to the value of $\gamma$, represents a marked separation between the two curves. We see that both functions are strictly increasing, and that their limiting values agree with the predictions of (4.25). The lower plot in Fig. 4.5 illustrates how the detector works in practice. In the white regions the final decision $\delta_{\mathcal{B}}$ is in favor of $\mathcal{H}_0$, while in the

grey region a decision in favor of $\mathcal{H}_1$ is taken. We see that if the original decision (given on the vertical axis) is $\delta = 0$, no matter what $l_x$ is, the final decision will be $\delta_{\mathcal{B}} = 0$. On the other hand, a decision $\delta = 1$ is retained only if $l_x$ is large enough. Note also that if $\gamma_{\mathcal{B}}$ approaches $\gamma$, then the original decisions are always retained (with probability one).

In Fig. 4.6, we compare three different systems, namely: the optimal NP detector having access to the full data $(\boldsymbol{x}, \boldsymbol{y})$; the optimal NP detector having access to the data subset $\boldsymbol{x}$; the detector that is the subject of this chapter, that having access to the pair $(\boldsymbol{x}, \delta)$. The ROCs of these systems are easily computed for the first two detectors, and have been derived in Proposition 4.4.2 for the unlucky broker case. As must be, the optimal detector using the pair $(\boldsymbol{x}, \delta)$ outperforms that exploiting only $\boldsymbol{x}$, and it is outperformed by the detector that uses the dataset $(\boldsymbol{x}, \boldsymbol{y})$. Note also how the unlucky broker's ROC intersects the upper bound at the point whose abscissa is the false alarm level selected by entity $\mathcal{A}$. In this point the thresholds of the two detectors are the same: $\gamma_{\mathcal{B}} = \gamma$.

It is also of interest to see what happens in an unbalanced situation where $N_x$ and $N_y$ are different. In Fig. 4.7 we display the ROCs for this scenario: in the upper plot $N_x = 1$ and $N_y = 5$, and the information contained in $\delta$ is expected to have a major role in the determining the final decision $\delta_{\mathcal{B}}$, as compared to the role of $\boldsymbol{x}$. Indeed, the behavior of detector $\mathcal{B}$ approaches a limiting case where the only information used for the final decision $\delta_{\mathcal{B}}$ is the decision $\delta$ taken by $\mathcal{A}$, while $\boldsymbol{x}$ becomes irrelevant. In such a limit, in order to vary the false alarm rate, the decision maker $\mathcal{B}$ has only the choice of randomizing the test among the only three available pairs $(0, 0)$, $(P_f, P_d)$, $(1, 1)$. The first and the third are always part of any ROC, while the pair $(P_f, P_d)$ [approximately $(0.06, 0.8)$ in the figure] is that pertaining to the decision made by $\mathcal{A}$.

In the lower plot of Fig. 4.7 the situation is reversed, with $N_x = 5$ and $N_y = 1$. Here we see that the performances of the systems are quite close to each other, and the unlucky broker's
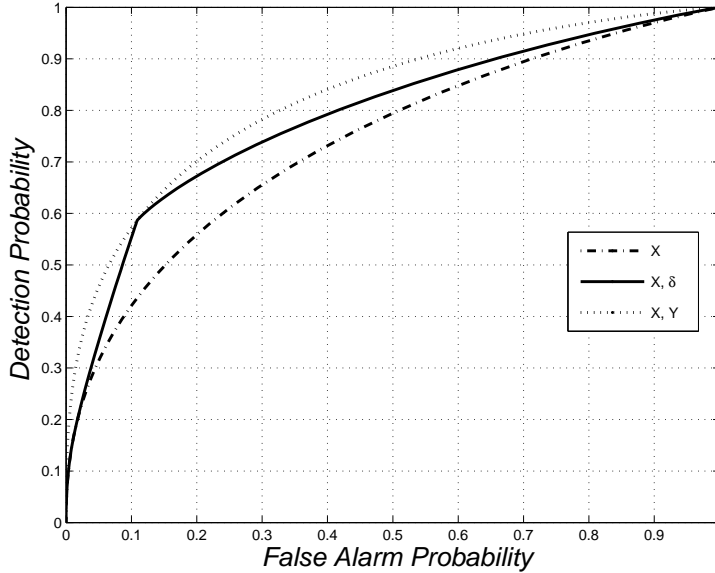
**Figure 4.6** ROCs for the exponential shift-in-scale problem with $N_x = N_y = 2$, $\lambda_0 = 2$, $\lambda_1 = 1$ and $\gamma = 0.5$. Three optimal NP detectors are considered, whose decisions are based on the dataset specified in the legend.

ROC closely approaches that of the detector operating on the data $\boldsymbol{x}$, for false alarm levels just slightly different from that selected by entity $\mathcal{A}$. In this situation, the initial decision $\delta$ is of minor relevance and the best option (except at false alarm rate $P_f$) is to base the decision $\delta_\mathcal{B}$ almost exclusively on $\boldsymbol{x}$.

## 4.4.2 Gaussian shift-in-mean

Let us consider the following hypothesis test

$$
\begin{array}{llll}
\mathcal{H}_0 & : & X_i \sim \mathcal{N}(0, \sigma_x^2), & Y_j \sim \mathcal{N}(0, \sigma_y^2), \\
\mathcal{H}_1 & : & X_i \sim \mathcal{N}(\mu_x, \sigma_x^2), & Y_j \sim \mathcal{N}(\mu_y, \sigma_y^2),
\end{array} \tag{4.30}
$$

where $i = 1, 2, \ldots, N_x$, $j = 1, 2, \ldots, N_y$, and $\mathcal{N}(\mu, \sigma^2)$ stands for a Gaussian pdf with mean $\mu$ and variance $\sigma^2$.

**Proposition 4.4.3.** *For the Gaussian shift-in-mean hypothesis-testing problem in (4.30), the optimal solution to the unlucky broker problem is the single-threshold detector $\delta_\mathcal{B}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$.* ●
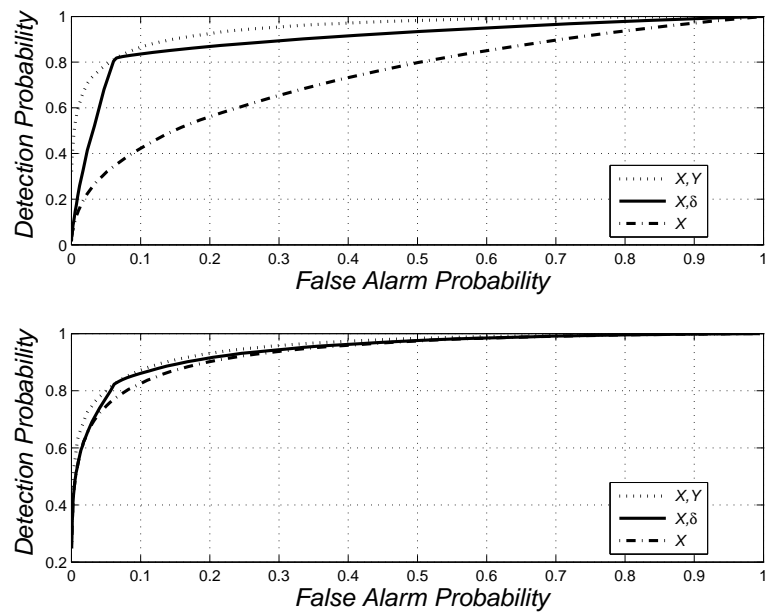
**Figure 4.7** Comparison between ROCs evaluated in the exponential shift-in-scale problem with $N_x = 1$, $N_y = 5$ (upper plot) and $N_x = 5$, $N_y = 1$ (lower plot), $\lambda_0 = 2$, $\lambda_1 = 1$ and $\gamma = 0.5$.

As for Proposition 4.4.1, the proof is deferred to Appendix A.3, and amounts to showing that the functions $t_1(l_x)$ and $t_0(l_x)$ for the Gaussian shift-in-mean problem (4.30) are strictly increasing. To characterize these functions, let us start by considering the log-likelihood ratios, which are

$$L_x(\boldsymbol{x}) = \frac{\mu_x}{\sigma_x^2} \sum_{i=1}^{N_x} \left( x_i - \frac{\mu_x}{2} \right),$$

$$L_y(\boldsymbol{y}) = \frac{\mu_y}{\sigma_y^2} \sum_{j=1}^{N_y} \left( y_j - \frac{\mu_y}{2} \right). \tag{4.31}$$

By defining $d_x = N_x \mu_x^2/(2\sigma_x^2)$ and $d_y = N_y \mu_y^2/(2\sigma_y^2)$, the random variables $L_x$ and $L_y$ are distributed as

$$L_x \sim \mathcal{N}\left(\mp d_x, 2\, d_x\right), \quad L_y \sim \mathcal{N}\left(\mp d_y, 2\, d_y\right), \tag{4.32}$$

where the negative sign refers to $\mathcal{H}_0$ and the positive sign to $\mathcal{H}_1$. The functions $t_1(l_x)$ and $t_0(l_x)$ can be evaluated explicitly in terms of the standard Gaussian exceedance function $Q(\cdot)$:

$$t_{1,0}(l_x) = l_x + \ln \frac{Q\left(\pm \frac{\gamma - l_x - d_y}{\sqrt{2\, d_y}}\right)}{Q\left(\pm \frac{\gamma - l_x + d_y}{\sqrt{2\, d_y}}\right)}, \tag{4.33}$$

where the positive sign applies to $t_1(l_x)$ and the negative sign refers to $t_0(l_x)$. The above expressions are used in Appendix A.3 to prove Proposition 4.4.3. There, it is also shown that

$$\lim_{l_x \to -\infty} t_1(l_x) = \gamma, \qquad \lim_{l_x \to +\infty} t_1(l_x) = +\infty,$$
$$\lim_{l_x \to -\infty} t_0(l_x) = -\infty, \qquad \lim_{l_x \to +\infty} t_0(l_x) = \gamma.$$

In Fig. 4.8 two case studies are considered. In both $N_x = N_y = 2$, $\mu_x = \mu_y = 1$, and $\gamma = 0.5$. The difference is that in the upper plot we set $\sigma_x = 2\,\sigma_y$, while the lower plot refers to $\sigma_y = 2\,\sigma_x$.
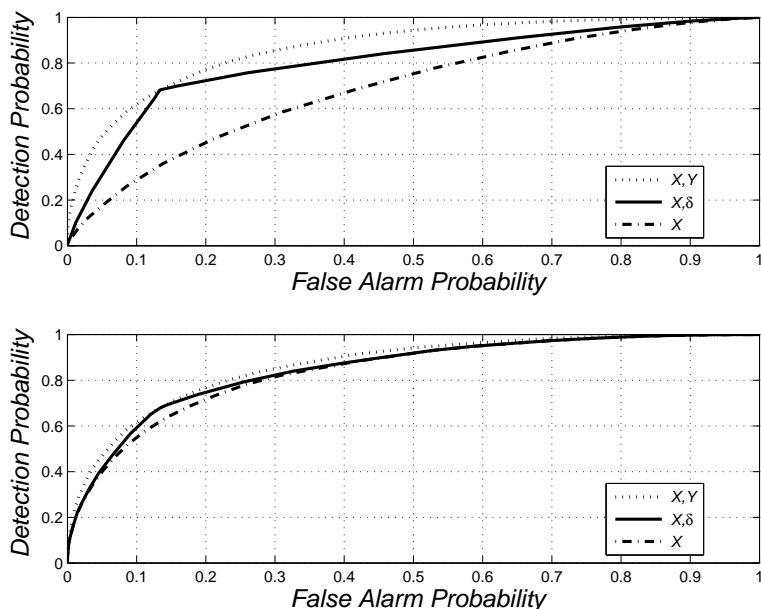
**Figure 4.8** Comparison between ROCs evaluated in the Gaussian shift-in-mean problem with $N_x = N_y = 2$, $\sigma_x = 2\sigma_y$ (upper plot) and $\sigma_y = 2\sigma_x$ (lower plot) and $\mu_x = \mu_y = 1$.

Comments are similar to those made in connection with Fig. 4.7: in the upper plot the surviving data $\boldsymbol{x}$ are "too noisy" and the original decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ is more informative, while in the lower plot the opposite is true.

### 4.4.3 Arbitrary distributions

For the observation models (4.19) and (4.30), proving that the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are strictly monotone is by no means trivial. Nonetheless, one might guess that this difficulty is only a matter of algebra, and that the monotone property may hold under fairly general observation scenarios. We have, however, the following proposition.

**Proposition 4.4.4.** *Under general observation models, the optimal solution to the unlucky broker problem is the multiple-threshold*
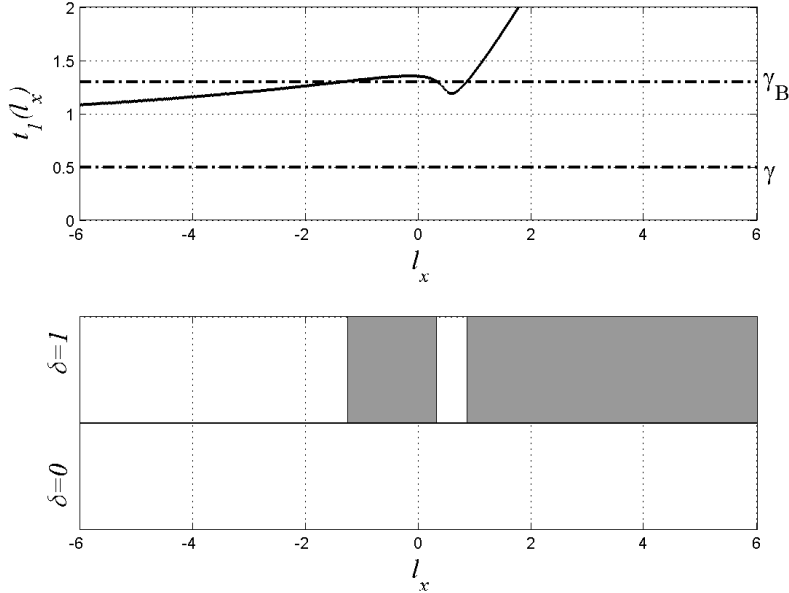
**Figure 4.9** Function $t_1(l_x)$ (upper plot) and final decisions in the "plane" $(l_x, \delta)$ (lower plot), for the generalized Gaussian shift-in-mean problem with $N_x = N_y = 1$, $\mu_x = \mu_y = 1$, $\sigma_x = \sigma_y = 1$ and $\gamma = 0.5$. The grey (resp. white) regions in the lower plot mean that $\delta_{\mathcal{B}} = 1$ (resp. $\delta_{\mathcal{B}} = 0$).

*detector* $\delta_{\mathcal{B}}(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$.                                        ●

The meaning of this claim is that the simplification to a single-threshold detector is a special case and does not hold in general. The proof simply consists in providing an example in which the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are not monotone.

To this end, we consider now a generalized Gaussian detection problem [40]. Let us introduce the generalized Gaussian distribution:

$$H(x, \sigma, k) = \frac{k}{2V(k, \sigma)\Gamma(\frac{1}{k})} \exp\left[-\frac{|x|}{V(k, \sigma)}\right]^k, \qquad (4.34)$$

with $V(k, \sigma) = \sigma\sqrt{\Gamma(1/k)/\Gamma(3/k)}$. The corresponding shift-in-

mean hypothesis test can be formalized as:

$$
\begin{aligned}
\mathcal{H}_0 \quad &: \quad f_X(x; \mathcal{H}_0) = H(x, \sigma_x, k), \\
&\qquad f_Y(y; \mathcal{H}_0) = H(y, \sigma_y, k), \\
\mathcal{H}_1 \quad &: \quad f_X(x; \mathcal{H}_1) = H(x - \mu_x, \sigma_x, k), \\
&\qquad f_Y(y; \mathcal{H}_1) = H(y - \mu_y, \sigma_y, k).
\end{aligned}
$$

Suppose that $k = 3$, $N_x = N_y = 1$, $\mu_x = \mu_y = 1$, and $\sigma_x = \sigma_y = 1$. In the upper plot of Fig. 4.9 the function $t_1(l_x)$, obtained numerically, is depicted and we see that it is no longer monotone; the same can be shown to hold for $t_0(l_x)$.

This non-monotone behavior has a strong impact on the decision rule, as it can be appreciated by looking at the lower plot in Fig. 4.9. The figure shows the final decisions taken by the detector, with a threshold $\gamma_\mathcal{B} > \gamma$. The decisions are plotted on the "plane" $(l_x, \delta)$: in the white regions the final decisions are for $\mathcal{H}_0$, while in the grey regions a final decision for $\mathcal{H}_1$ is made. As predicted by theory, since $\gamma_\mathcal{B} > \gamma$, all the decisions $\delta = 0$ are retained. However, the rule for retaining the decisions $\delta = 1$ is markedly different from that observed in the previous cases. Indeed, the grey region is no longer simply connected, meaning that the detector structure does *not* simply involve the comparison of the log-likelihood ratio $l_x$ with a single threshold.

The generalized Gaussian distribution has smooth shape, and it is enough to show that the general approach to the unlucky broker problem requires a multiple-threshold detector. On the other hand, in specific applications involving less regular distributions, the operating modality of the multiple-threshold detector can be shown to be even more complicated than that observed in Fig. 4.9.

As an example, let us consider detection problem (4.1) in which the variables $X_i$'s and $Y_i$'s are zero-mean Gaussian with variance $\sigma_0^2$ under the null hypothesis $\mathcal{H}_0$, while under the alternative hypothesis they come from a Gaussian population, with expectation uniformly selected among $m$ allowable values. Therefore, under $\mathcal{H}_1$ the variables are identically distributed as a balanced homoscedastic mixture of $m$ Gaussian random variables with means
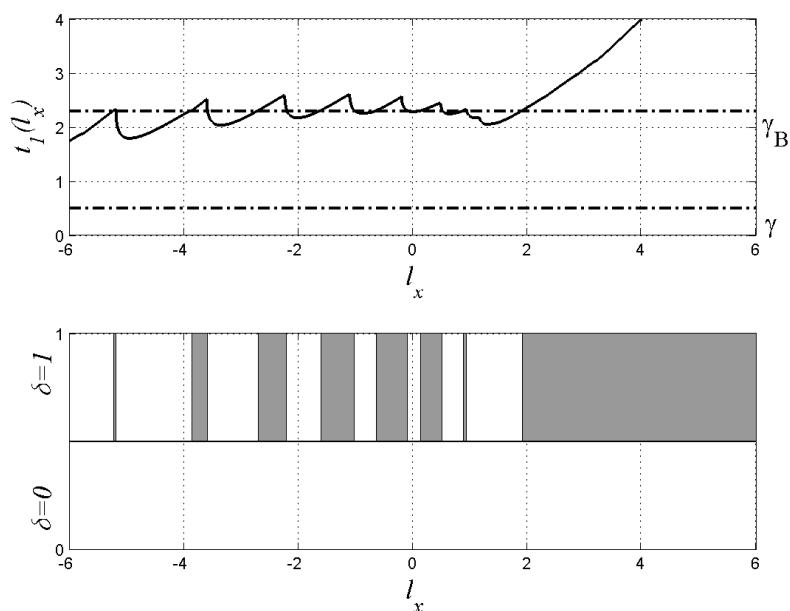
**Figure 4.10** Function $t_1(l_x)$ (upper plot) and final decisions in the plane $(l_x, \delta)$ (lower plot), for the example involving the balanced mixture of Gaussians; values of the parameters are detailed in the main text. As for the previous figures, the grey (resp. white) regions in the lower plot mean that $\delta_\mathcal{B} = 1$ (resp. $\delta_\mathcal{B} = 0$).

$\mu_1, \mu_2, \ldots, \mu_m$, and common variance $\sigma_1^2$, formally:

$$\begin{aligned} \mathcal{H}_0 &: \quad X_i \text{ and } Y_i \sim \mathcal{N}(0, \sigma_0^2), \\ \mathcal{H}_1 &: \quad X_i \text{ and } Y_i \sim \frac{1}{m} \sum_{i=1}^{m} \mathcal{N}(\mu_i, \sigma_1^2). \end{aligned}$$

To get insight about the detector structure, let us consider the above problem for $N_x = N_y = 1$, $m = 20$, $\sigma_0^2 = 2$, $\sigma_1^2 = 0.2$, $\gamma = 0.5$, and the means $\mu_i$'s selected as equally spaced points in the range $[-9, 9]$. The upper plot in Fig. 4.10 shows the function $t_1(l_x)$, while in the lower plot we display the corresponding decision regions at site $\mathcal{B}$. It can be noticed that, as in the generalized Gaussian problem, the region for retaining the decision is not simply connected, giving rise to a multiple-threshold detector. In this case, however, it can be appreciated how the detection regions exhibit a more complex behavior, corroborating the physical relevance of Proposition 4.4.4.

## 4.5 The Bayesian Unlucky Broker

In this section we introduce a variation of the unlucky broker problem considered so far, making reference to the Bayesian paradigm where a-priori probabilities of the hypotheses to be tested are given. The problem is formally rephrased in Subsection 4.5.1, while the optimal detector and its modus operandi are investigated in Subsection 4.5.2. Subsection 4.5.3 provides two examples of practical interest.

### 4.5.1 Problem Formulation

In statistical terms, the scenario considered in the previous sections can be re-abstracted as in Fig. 4.11. A certain entity, we call it $S_A$, observes the data $(\boldsymbol{x}, \boldsymbol{y})$ and implements an optimal Bayesian test, exploiting the a-priori probabilities of the hypotheses, $\pi_{0A}$ and $\pi_{1A}$, and ending up with a decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ at (the best) Bayes risk level $r_A$. Another entity, $S_B$, which has a refined version of priors, $\pi_{0B}$ and $\pi_{1B}$, observes a portion of the data (the
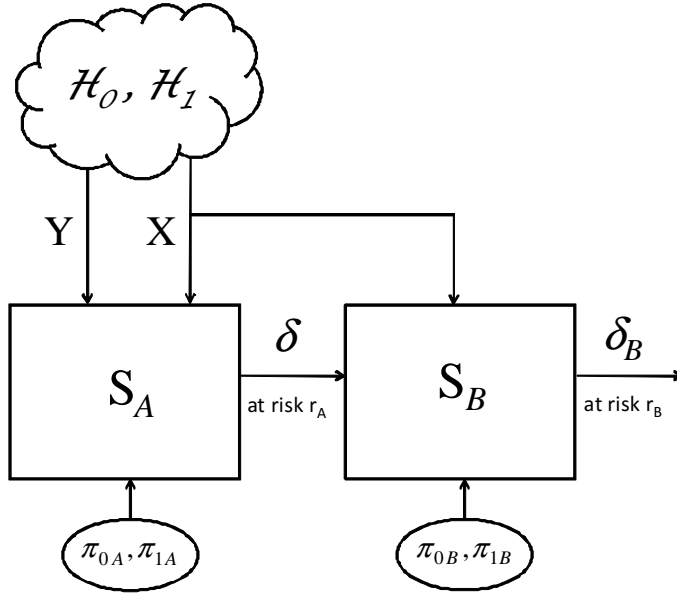
**Figure 4.11** Notional scheme of the unlucky broker problem in the Bayesian framework.

vector $\boldsymbol{x}$), receives from $S_A$ its decision $\delta(\boldsymbol{x}, \boldsymbol{y})$ and makes the best Bayes decision $\delta_B(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y}))$ at a Bayes risk $r_B$.

In the described system architecture, as made fo the NP formulation, we address the following basic questions. What is the best decision $S_B$ can make, by exploiting the observation vector $\boldsymbol{x}$ and the binary decision $\delta(\boldsymbol{x}, \boldsymbol{y})$? What about the behavior of the optimal detector? Should it simply retain the previous decision $\delta$, or should it ignore that, and use only the currently available data set for a completely new decision? Or, what else?

Starting from the formulation presented in Section 4.2, we have that the strategy in (4.3) is optimal for $S_A$, in the Bayesian sense, if the threshold $\gamma$ is given by

$$\gamma = \ln \frac{\pi_{0A}(C_{10} - C_{00})}{\pi_{1A}(C_{01} - C_{11})}, \tag{4.35}$$

where $C_{ij}$ is the cost incurred by choosing the hypothesis $\mathcal{H}_i$ when $\mathcal{H}_j$ is true and $\pi_{1A} = 1 - \pi_{0A}$. Let us consider an uniform cost

assignment, that is $C_{ij} = 0$ if $i = j$ and $C_{ij} = 1$ if $i \neq j$. Thus, the threshold in (4.35) becomes $\gamma = \ln \pi_{0A}/(1 - \pi_{0A})$, and the Bayes risk $r_A$ is [36]

$$r_A = \pi_{0A} P_f + (1 - \pi_{0A})(1 - P_d), \qquad (4.36)$$

where $P_f$ and $P_d$ are the false alarm and detection probabilities of the optimal Bayes test based upon $(\boldsymbol{X}, \boldsymbol{Y})$, respectively.

The optimal decision rule for the unlucky broker in (4.9) can be reformulated in a Bayesian framework in the same way, with the corresponding threshold $\gamma_B = \ln \pi_{0B}/(1 - \pi_{0B})$. The Bayes risk can be defined as

$$r_B = \pi_{0B} P_{f,B} + (1 - \pi_{0B})(1 - P_{d,B}), \qquad (4.37)$$

where $P_{f,B}$ and $P_{d,B}$ are the associated detection and false alarm probabilities.

## 4.5.2 Optimal Solution

The definitions 4.3.1 and 4.3.2 in Section 4.3 becomes, in the Bayesian formulation, the following

**Definition 4.5.1.** $\delta_{\mathcal{B}}$ *is a "multiple-threshold" detector with side decision if*

$$\delta_{\mathcal{B}} = \begin{cases} \delta \, u \, (t_1(l_x) - \gamma_{\mathcal{B}}), & \text{for } \pi_{0\mathcal{B}} > \pi_{0\mathcal{A}}, \\ (1 - \delta) \, u \, (t_0(l_x) - \gamma_{\mathcal{B}}) + \delta, & \text{for } \pi_{0\mathcal{B}} \leq \pi_{0\mathcal{A}}. \end{cases} \qquad (4.38)$$

$\bullet$

Suppose now that the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are invertible and strictly increasing. Then comparing $t_1(l_x)$ with $\gamma_{\mathcal{B}}$ is tantamount to comparing the log-likelihood $l_x$ to $t_1^{-1}(\gamma_{\mathcal{B}})$, and the same holds when comparing $t_0(l_x)$ with $\gamma_{\mathcal{B}}$. Let:

$$s_1 = t_1^{-1}(\gamma_{\mathcal{B}}), \quad s_0 = t_0^{-1}(\gamma_{\mathcal{B}}). \qquad (4.39)$$

The detector then simplifies to the following.

**Definition 4.5.2.**   $\delta_{\mathcal{B}}$ *is a "single-threshold" detector with side decision if*

$$
\delta_{\mathcal{B}} = \left\{ \begin{array}{ll} \delta\, u\,(l_x - s_1)\,, & \text{for } \pi_{0\mathcal{B}} > \pi_{0\mathcal{A}}, \\ (1-\delta)\, u\,(l_x - s_0) + \delta, & \text{for } \pi_{0\mathcal{B}} \leq \pi_{0\mathcal{A}}. \end{array} \right.
\qquad (4.40)
$$

●

The above definitions allow us to identify the modus operandi of the optimal detector solving the unlucky broker problem in the Bayesian scenario. The behavior of the unlucky broker (entity $S_B$) is summarized as follows.

When there is no refinement of the a-priori probabilities ($\pi_{0B} = \pi_{0A}$), the best solution for the unlucky broker is, clearly, to retain the original decision. It can be shown that this result is indeed embodied in the statement of the previous definitions.

When $\pi_{0B} > \pi_{0A}$, an original decision $\delta = 0$ (in favor of $\mathcal{H}_0$) is always retained by $S_B$, while a decision $\delta = 1$ (i.e., for $\mathcal{H}_1$) needs a double check: it is kept only if the function $t_1(l_x)$ is larger than, or equal to, the new threshold $\gamma_B$.

Conversely, in the case $\pi_{0B} < \pi_{0A}$, a decision in favor of $\mathcal{H}_1$ is always retained, while a decision in favor of $\mathcal{H}_0$ is retained only if $t_0(l_x)$ is less than $\gamma_B$.

In summary, when the refined prior tells that, say, $\mathcal{H}_0$ is becoming more likely ($\pi_{0B} > \pi_{0A}$), the unlucky broker should accordingly tip the balance toward the null hypothesis. The optimal way to do this is to change some of the decisions in favor of $\mathcal{H}_1$, based upon a suitable detection statistic of the available data set $\boldsymbol{x}$. The situation clearly reverses when $\pi_{0B} < \pi_{0A}$.

Also, the above definitions imply the following. Whenever the functions $t_1(l_x)$ and $t_0(l_x)$ are invertible and strictly increasing, the required double check results in comparing $l_x(\boldsymbol{x})$ to a single threshold. Otherwise stated, *given* that the original decision is $\delta = 1$ (when $\pi_{0B} > \pi_{0A}$) or $\delta = 0$ (when $\pi_{0B} \leq \pi_{0A}$), the optimal test for $S_B$ behaves as a (likelihood) single threshold test, based upon the available data set $\boldsymbol{x}$.

In the general case, instead, the double check may involve much more tricky log-likelihood comparisons: it requires checking whether $l_x(\boldsymbol{x})$ belongs to some arbitrarily shaped subset of the real line, not necessarily simply connected. This may in fact amount to compare $l_x(\boldsymbol{x})$ to multiple thresholds.

Before ending this section, we report the analytical expressions for system performances evaluation. The Bayes risk $r_B$ can be evaluated by resorting to eq. (4.38). As made in Section 4.3, when $\pi_{0B} > \pi_{0A}$ we have

$$
\begin{aligned}
P_{d,B} &= P_d - \int_{t_1(l_x) < \gamma_B} P_{dy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_1) dl_x, \\
P_{f,B} &= P_f - \int_{t_1(l_x) < \gamma_B} P_{fy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_0) dl_x, \quad (4.41)
\end{aligned}
$$

where $f_{L_x}(l_x; \mathcal{H}_1)$ is the pdf of the random variable $L_x$ under the alternative hypothesis. In the opposite case of $\pi_{0B} \leq \pi_{0A}$, we get

$$
\begin{aligned}
P_{d,B} &= P_d + \int_{t_0(l_x) \geq \gamma_B} [1 - P_{dy}(\gamma - l_x)] f_{L_x}(l_x; \mathcal{H}_1) dl_x, \\
P_{f,B} &= P_f + \int_{t_0(l_x) \geq \gamma_B} [1 - P_{fy}(\gamma - l_x)] f_{L_x}(l_x; \mathcal{H}_0) dl_x.
\end{aligned}
$$

### 4.5.3 Examples

In this section, the above theoretical framework is applied to two specific decision problems. We start by considering the classical Gaussian shift-in-mean hypothesis-testing problem.

$$
\begin{aligned}
\mathcal{H}_0 &: \quad X_i, Y_j \sim \mathcal{N}(0, \sigma^2), \\
\mathcal{H}_1 &: \quad X_i, Y_j \sim \mathcal{N}(\mu, \sigma^2).
\end{aligned} \quad (4.42)
$$

The upper plot in Fig. 4.12 shows the Bayes risk as a function of the refined prior $\pi_{0B}$, when $\pi_{0A} = 0.3$, for three different systems. We consider the optimal Bayes detector having access to the full data set $(\boldsymbol{x}, \boldsymbol{y})$ and the detector $S_B$ that exploits the pair $(\boldsymbol{x}, \delta)$. For comparison purposes, we also show the risk pertaining to a
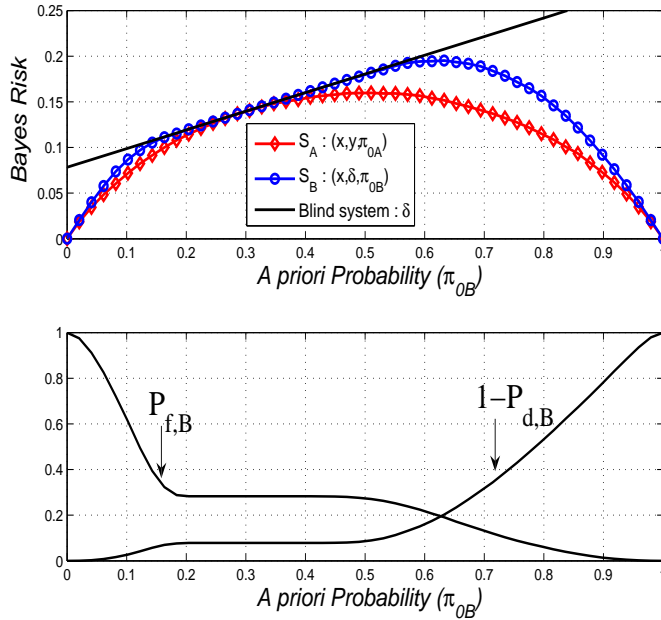
**Figure 4.12** Bayes risk (upper plot) and probabilities $P_{f,B}$, $1 - P_{d,B}$ (lower plot) versus a-priori probability $\pi_{0B}$ for the Gaussian shift-in-mean problem with $N_x = N_y = 2$, $\sigma = 1$, $\mu = 1$ and $\pi_{0A} = 0.3$.

detector which always retains the original decision, ignoring thus the availability of a refined prior. The error probabilities of this latter would clearly coincide with $P_f$ and $1 - P_d$, yielding a linear behavior with $\pi_{0B}$. We can observe that the curves get in contact when the refined priors are equal to those used in the original test. Furthermore, a precise ordering relationship exists: $S_A$ uniformly outperforms $S_B$, which in turn uniformly outperforms the "blind" system which ignores the refined prior availability.

The lower plot in Fig. 4.12 depicts the behavior of the probabilities $P_{f,B}$, $1 - P_{d,B}$ as functions of the refined prior $\pi_{0B}$ and allows us to describe more in detail what happens in terms of Bayes risk. In the considered example, the original a priori probability is $\pi_{0A} = 0.3$, giving $\gamma = -0.85$. If the priors selected by the broker (entity $S_B$) coincide with these initial values, the best that one can do is to retain the original decision, ignoring the surviving

**Figure 4.13** Function $t_1(l_x)$ (upper plot) and final decisions in the plane $(l_x, \delta)$ (lower plot) when $\gamma_B = -0.41$. This refers to the Gaussian example.

data. Accordingly, $P_{f,B} = P_f$, $P_{d,B} = P_d$, and the curves in the upper plot of Fig. 4.12 get in contact just for $\pi_{0A} = \pi_{0B} = 0.3$, when the priors are actually not refined at all.

Now, let us choose $\pi_{0B} \neq 0.3$, but sufficiently close to that. Assume, for instance, $\pi_{0B} = 0.4$, implying $\gamma_b = -0.41$. In this case we have $\pi_{0B} > \pi_{0A}$, so we must compare (see the Definition 3 in the above) $t_1(l_x)$ to $\gamma_B$. It can be shown that the function $t_1(l_x)$, plotted in Fig. 4.13, crosses the threshold $\gamma_B$ when $l_x = -4.6$, and we can compute the unlucky broker's detection probability thanks to the first of eqs. (4.16). By numerically integration, we find that the integral term in the first of eqs. (4.16) is much smaller ($\approx 10^{-7}$) than the detection probability $P_d$ (0.92 in the example) and, accordingly, $1 - P_{d,B}$ is almost equal to $1 - P_d$. Similarly, we have $P_{f,B} \approx P_f$ (0.28 in the example). This basically means that, in a neighborhood of $\pi_{0A}$, the original decision is very often retained ($\delta_B \approx \delta$). Looking at Fig. 4.12, this immediately explains the similarity between the performance of $S_B$ and that of the system
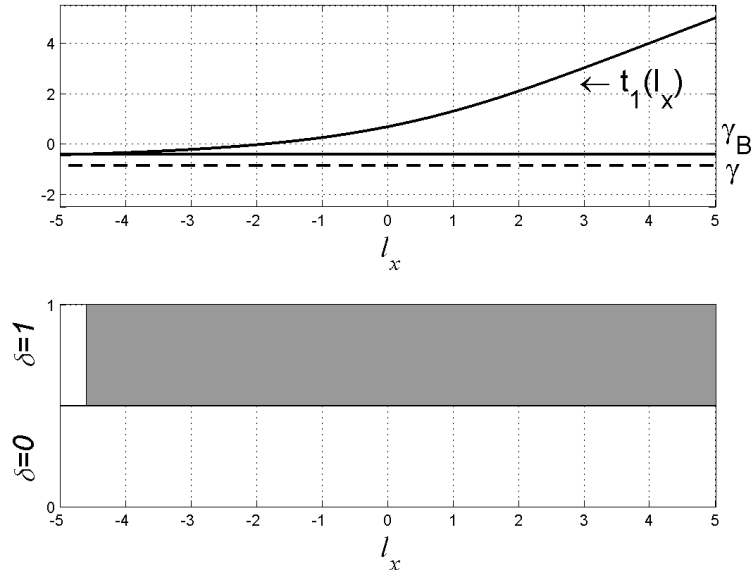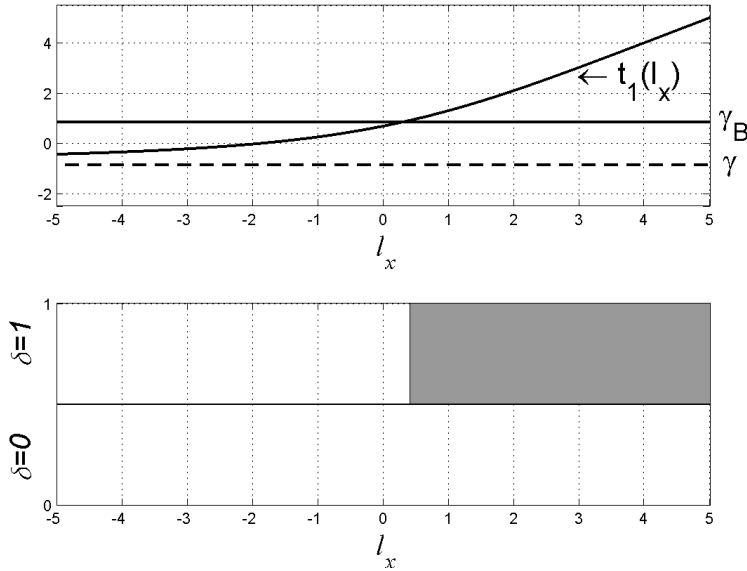
**Figure 4.14** Function $t_1(l_x)$ (upper plot) and final decisions in the plane $(l_x, \delta)$ (lower plot) when $\gamma_B = 0.85$. This refers to the Gaussian example.

which always retains the decision $\delta$, in the range where $\pi_{0B}$ is sufficiently close to $\pi_{0A}$.

Figure 4.14 shows, instead, what happens when $\pi_{0B} = 0.7$, implying $\gamma_B = 0.85$. The intersection between $t_1(l_x)$ and $\gamma_B$ occurs when $l_x = 0.4$. The integral term in the first of eqs. (4.16) is now comparable ($\approx 0.27$) to $P_d$, and $1 - P_{d,B}$ grows as shown in Fig. 4.12. In a similar manner, we find that the integral term in the second of eqs. (4.16) is comparable ($\approx 0.16$) to $P_f$, and $P_{f,B}$ decreases as shown in the lower plot of Fig. 4.12.

We can conclude that for small variations of the priors around the original values, the unlucky broker changes the original decision with probability much smaller than the false alarm and detection probabilities of the original decisions. Hence, in this range, the Bayes risk of the broker is almost linear. A physical interpretation is that in this interval the unlucky broker essentially exploits only the information provided by the original decision.

From the upper plot of Fig. 4.14 we see that the function $t_1(l_x)$

for the case study we are considering is strictly increasing, as formally proved in Appendix A.1. The same result holds for the function $t_0(l_x)$. Thus, we are in the setting of Definition 4, and we can state that the optimal solution to the unlucky broker problem is provided by a detector that compares the log-likelihood to a single threshold. This is what one usually expects by a detector.

The lower plot in Fig. 4.14 is to be interpreted as follows. In the white regions the final decision is in favor of $\mathcal{H}_0$, while in the grey region a decision in favor of $\mathcal{H}_1$ is taken. We see that if the original decision (given on the vertical axis) is $\delta = 0$, no matter what $l_x$ is, the final decision will be $\delta_B = 0$. On the other hand, a decision $\delta = 1$ is retained only if $l_x$ is large enough. Finally, we can observe that if the two thresholds approach each other, then the original decisions are always retained. Summarizing, when the functions $t_1(l_x)$ and $t_0(l_x)$ are invertible and strictly increasing, the optimal detector solving the unlucky broker problem *works like a log-likelihood threshold test*.

For the observation model (4.19), proving when the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are strictly monotone is by no means trivial. Nonetheless, aside from mathematical difficulties, the reader might guess that the monotone property may hold very in general. Instead, the unlucky broker problem does not lead, in general, to a simple single-threshold detector. Usually, the behavior is more complex, since typically the functions $t_1(l_x)$ and $t_0(l_x)$ defined in (4.7) are not monotone. As an example, consider the following detection problem.

$$
\begin{aligned}
\mathcal{H}_0 \quad &: \quad X_i \text{ and } Y_i \sim \mathcal{N}(0, \sigma_0^2), \\
\mathcal{H}_1 \quad &: \quad X_i \text{ and } Y_i \sim \tfrac{1}{m} \sum_{i=1}^{m} \mathcal{N}(\mu_i, \sigma_1^2),
\end{aligned}
$$

where the variables $X_i$'s and $Y_i$'s are zero-mean Gaussian with variance $\sigma_0^2$ under $\mathcal{H}_0$, while under the alternative hypothesis they are identically distributed as a balanced mixture of $m$ Gaussian random variables with mean values $\mu_1, \mu_2, \ldots, \mu_m$, and common variance $\sigma_1^2$.

To get insights about the detector structure, let us consider the above problem for $N_x = N_y = 1$, $m = 20$, $\sigma_0^2 = 2$, $\sigma_1^2 = 0.2$,
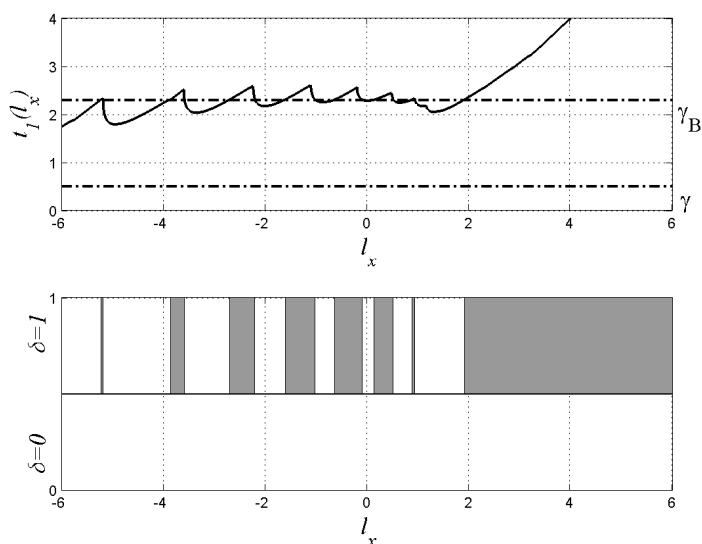
**Figure 4.15**  Function $t_1(l_x)$ (upper plot) and final decisions in the plane $(l_x, \delta)$ (lower plot), for the example involving the balanced mixture of Gaussians; values of the parameters are detailed in the main text. As for the previous figures, the grey (resp. white) regions in the lower plot mean that $\delta_B = 1$ (resp. $\delta_B = 0$).

$\gamma = 0.5$, and the mean values $\mu_i$'s selected as equally spaced points in the range $[-9, 9]$. In the upper plot of Fig. 4.10, obtained numerically, the function $t_1(l_x)$ is shown and we see that it is no longer monotone.

This non-monotone behavior has a strong impact on the decision rule, as it can be appreciated by looking at the lower plot in Fig. 4.10. The figure shows the final decisions taken by the detector, with $\pi_{0B} > \pi_{0A}$. The decisions are plotted on the "plane" $(l_x, \delta)$: in the white regions the final decisions are for $\mathcal{H}_0$, while in the grey regions a final decision for $\mathcal{H}_1$ is made. As predicted by the theory, since $\pi_{0B} > \pi_{0A}$, all the decisions $\delta = 0$ are retained. However, the rule for retaining the decisions $\delta = 1$ is markedly different from that observed in the previous case. Indeed, the grey region is no longer simply connected, meaning that the detector structure does *not* simply involve the comparison of the log-likelihood ratio $l_x$ with a single threshold.

In general, therefore, the detection regions have a complicated shape and the unlucky broker task cannot be reduced to a single-threshold comparison. The behavior of the functions $t_1(\cdot)$ and $t_0(\cdot)$ that rules the optimal detector, indeed, implies multiply-connected and irregularly-shaped optimal decision regions.

## 4.6 The Unlucky Broker with correlated data

In this section we study the unlucky broker problem in presence of correlation between the full data set originally available and the data set survived to the data losing. In particular, in Subsection 4.6.1 the problem is formulated and the optimal solution is showed. In Subsection 4.6.2 the theoretical framework is applied to an example of practical interest.

### 4.6.1   Problem Solution

Consider the unlucky broker problem in Fig. 4.4 and the binary hypothesis-testing problem (4.1). Let us assume that vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are made of i.i.d. but mutually correlated entries. In the NP formulation, the optimal strategy for the entity $\mathcal{A}$ amounts to comparing the log-likelihood ratio

$$
\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{y}) &= \ln \frac{f(\boldsymbol{x}, \boldsymbol{y}; \mathcal{H}_1)}{f(\boldsymbol{x}, \boldsymbol{y}; \mathcal{H}_0)} \\
&= \ln \frac{f(\boldsymbol{x}; \mathcal{H}_1)}{f(\boldsymbol{x}; \mathcal{H}_0)} + \ln \frac{f(\boldsymbol{y}|\boldsymbol{x}; \mathcal{H}_1)}{f(\boldsymbol{y}|\boldsymbol{x}; \mathcal{H}_0)} \\
&= L_x(\boldsymbol{x}) + L_{y|x}(\boldsymbol{y}|\boldsymbol{x}), \quad\quad (4.43)
\end{aligned}
$$

to a threshold $\gamma$, yielding the decision rule $\delta$:

$$
\delta(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 1, & L(\boldsymbol{x}, \boldsymbol{y}) \geq \gamma, \\ 0, & L(\boldsymbol{x}, \boldsymbol{y}) < \gamma. \end{cases}
$$

At site $\mathcal{B}$, the probabilities involved in the computation of the log-likelihood ratio (4.3), can be expanded as

$$
\begin{aligned}
P(\boldsymbol{x}, 1; \mathcal{H}_h) &= \Pr\{L_{y|x}(\boldsymbol{Y}|\boldsymbol{X}) \geq \gamma - L_x(\boldsymbol{x})|\boldsymbol{X} = \boldsymbol{x}; \mathcal{H}_h\} \\
&\times \ f_X(\boldsymbol{x}; \mathcal{H}_h) \\
P(\boldsymbol{x}, 0; \mathcal{H}_h) &= \Pr\{L_{y|x}(\boldsymbol{Y}|\boldsymbol{X}) < \gamma - L_x(\boldsymbol{x})|\boldsymbol{X} = \boldsymbol{x}; \mathcal{H}_h\} \\
&\times \ f_X(\boldsymbol{x}; \mathcal{H}_h),
\end{aligned}
$$

$$(4.44)$$

where $h = 0, 1$ denotes the hypothesis. In this case, the conditioning cannot be removed because of correlation between the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $P_{fy|x}(z, \boldsymbol{x}) = \Pr\{L_{y|x}(\boldsymbol{Y}|\boldsymbol{X}) \geq z|\boldsymbol{X} = \boldsymbol{x}; \mathcal{H}_0\}$ and $P_{dy|x}(z, \boldsymbol{x}) = \Pr\{L_{y|x}(\boldsymbol{Y}|\boldsymbol{X}) \geq z|\boldsymbol{X} = \boldsymbol{x}; \mathcal{H}_1\}$ be, respectively, the false alarm and detection probabilities of an optimal NP test based upon $\boldsymbol{y}|\boldsymbol{x}$. We obtain

$$
\begin{aligned}
P(\boldsymbol{x}, \delta; \mathcal{H}_0) &= f_X(\boldsymbol{x}; \mathcal{H}_0) \left\{ \delta \, P_{fy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x}) \right. \\
&+ \ (1 - \delta)[1 - P_{fy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})] \Big\} , \\
P(\boldsymbol{x}, \delta; \mathcal{H}_1) &= f_X(\boldsymbol{x}; \mathcal{H}_1) \left\{ \delta \, P_{dy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x}) \right. \\
&+ \ (1 - \delta)[1 - P_{dy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})] \Big\} .
\end{aligned}
$$

$$(4.45)$$

By considering separately the two cases $\delta = 0, 1$, it is readily seen that the log-likelihood ratio in (4.3) can be rewritten as

$$
\begin{aligned}
T(\boldsymbol{x}, \delta(\boldsymbol{x}, \boldsymbol{y})) \;=\; & L_x(\boldsymbol{x}) + \delta \, \ln \frac{P_{dy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})}{P_{fy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})} \\
& + \; (1 - \delta) \ln \frac{1 - P_{dy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})}{1 - P_{fy|x}(\gamma - L_x(\boldsymbol{x}), \boldsymbol{x})}.
\end{aligned}
$$

$$(4.46)$$

Then, in presence of correlation between the data set available at site $\mathcal{A}$, the functions in (4.7) become

$$
\begin{aligned}
t_1(l_x, \boldsymbol{x}) \;&=\; l_x + \ln \frac{P_{dy|x}(\gamma - l_x, \boldsymbol{x})}{P_{fy|x}(\gamma - l_x, \boldsymbol{x})}, \\
t_0(l_x, \boldsymbol{x}) \;&=\; l_x + \ln \frac{1 - P_{dy|x}(\gamma - l_x, \boldsymbol{x})}{1 - P_{fy|x}(\gamma - l_x, \boldsymbol{x})},
\end{aligned}
$$

$$(4.47)$$

and, finally, the decision statistic used by the unlucky broker is

$$
T(l_x, \boldsymbol{x}, \delta) = \delta \, t_1(l_x, \boldsymbol{x}) + (1 - \delta) \, t_0(l_x, \boldsymbol{x}).
\qquad (4.48)
$$

By substituting the log-likelihhod ratio (4.48) in (4.9) we get the optimal decision rule for entity $\mathcal{B}$. Since the ROC $P_{dy|x} = P_{dy|x}(P_{fy|x})$ satisfies again the conditions

$$
\frac{P_{dy|x}(\gamma - l_x)}{P_{fy|x}(\gamma - l_x)} \geq e^{\gamma - l_x}, \quad
\frac{1 - P_{dy|x}(\gamma - l_x)}{1 - P_{fy|x}(\gamma - l_x)} \leq e^{\gamma - l_x},
$$

the behavior of the detector $\mathcal{B}$ is the same of that summarized by eqs. (4.13) and (4.15), where the only difference lies in the definitions of functions $t_0$ and $t_1$.

## 4.6.2   Example

To corroborate the considerations in the above, let us consider the binary hypothesis-testing problem in (4.1), with $N_x = N_y = 1$, $X, Y \sim \mathcal{N}(\boldsymbol{0}; \boldsymbol{\Sigma})$ under the null hypothesis and $X, Y \sim \mathcal{N}(\mu; \boldsymbol{\Sigma})$

under the hypothesis $\mathcal{H}_1$. We have $\mu = [\mu_{\mathbf{x}} \ \ \mu_{\mathbf{y}}]$, while the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho \\ \rho & \sigma_y^2 \end{pmatrix}. \tag{4.49}$$

Then, the log-likelihood on the data set $x$ involved in (4.43) is

$$\begin{aligned} L_x(x) &= \frac{\mu_x}{\sigma_x^2} x - \frac{\mu_x^2}{2\sigma_x^2} \\ &= Ax + B. \end{aligned} \tag{4.50}$$

Consequently, we have that $L_x(X) \sim \mathcal{N}(B; A^2 \sigma_x^2)$ under the null hypothesis and $L_x(X) \sim \mathcal{N}(A\mu_x + B; A^2\sigma_x^2)$ under the alternative hypothesis. If we define the random variable $Z = (Y|X = x)$, we have that $Z \sim \mathcal{N}(\mu_i; \sigma^2)$, with $i = 0, 1$, $\mu_0 = \sigma_y/\sigma_x \rho x$ and $\mu_1 = \mu_y + \sigma_y/\sigma_x \rho(x - \mu_x)$, $\sigma^2 = (1 - \rho^2)\sigma_y^2$. The log-likelihood ratio of $Z$ is given by

$$\begin{aligned} L_z(z) &= (\mu_1 - \mu_0)\frac{z}{\sigma^2} + \frac{1}{2\sigma^2}(\mu_0^2 - \mu_1^2) \\ &= Cz + D. \end{aligned} \tag{4.51}$$

Thus, we get that $L_z(Z) \sim \mathcal{N}(C\mu_i + D; C^2\sigma^2)$, with $i = 0, 1$. The functions $t_1(l_x)$ and $t_0(l_x)$ can be evaluated explicitly in terms of the standard Gaussian exceedance function $Q(\cdot)$ as

$$t_{1,0}(l_x) = l_x + \ln \frac{Q\left(\pm\frac{\gamma - l_x - C\mu_1 - D}{C\sigma}\right)}{Q\left(\pm\frac{\gamma - l_x - C\mu_0 - D}{C\sigma}\right)}, \tag{4.52}$$

where the positive sign applies to $t_1(l_x)$ and the negative sign refers to $t_0(l_x)$. We have considered a Gaussian shift-in-mean problem with $\mu_x = \mu_y = 1$, $\sigma_x = \sigma_y = 1$, $\rho = 0.5$, $\gamma = 0.5$ and $\gamma_B = 2$. In Fig. 4.16 the ROCs of three different systems are compared, namely: the optimal NP detector having access to the full data $(x, y)$; the optimal NP detector having access to the data subset $x$; the unlucky broker, that having access to the pair $(x, \delta)$.

In the upper plot of Fig. 4.17 the function $t_1(l_x)$ is displayed for the case study $\mu_x = \mu_y = 1$, $\sigma_x = \sigma_y = 1$, $\rho = 0.5$, $\gamma = 0.5$
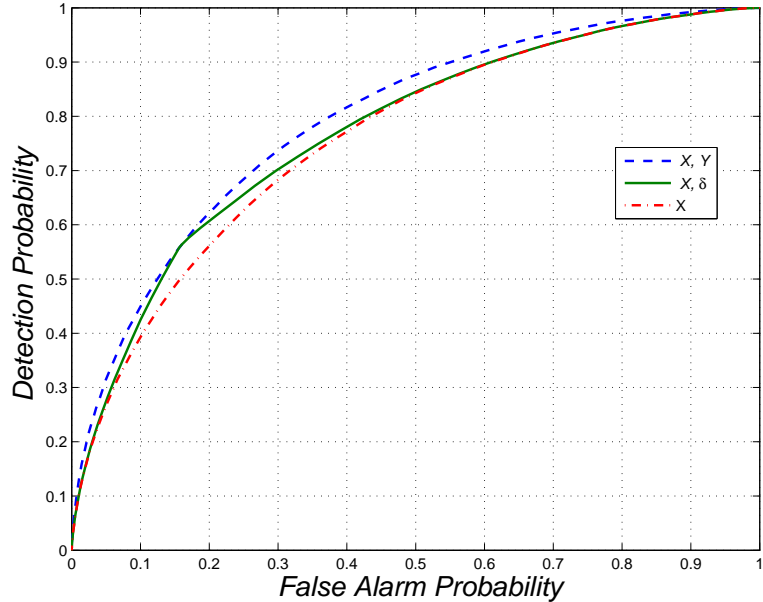
**Figure 4.16** ROCS for the Gaussian shift-in-mean problem with $\mu_x = \mu_y = 1$, $\sigma_x = \sigma_y = 1$, $\rho = 0.5$, $\gamma = 0.5$ and $\gamma_B = 2$.

and $\gamma_B = 2$. As we have already discussed for the unlucky broker problem with uncorrelated data, when $\gamma_B > \gamma$ and the original decision is $\delta = 0$, the final decision will be $\gamma_B = 0$, no matter what $l_x$ is. On the other hand, a decision $\delta = 1$ is retained only if $l_x$ is large enough.

## 4.7 Summary

In this chapter, we have considered a novel topic in detection theory, called the unlucky broker problem, with a broad range of potential applications (economics, medicine, wireless sensor networks, etc...).

The problem can be stated as follows. Consider a statistical test between two hypotheses, exploiting a certain data set $(\boldsymbol{x}, \boldsymbol{y})$ and leading to a decision $\delta$ obeying the NP optimality criterion. Then, suppose that the vector $\boldsymbol{y}$ is lost and one wants to make a

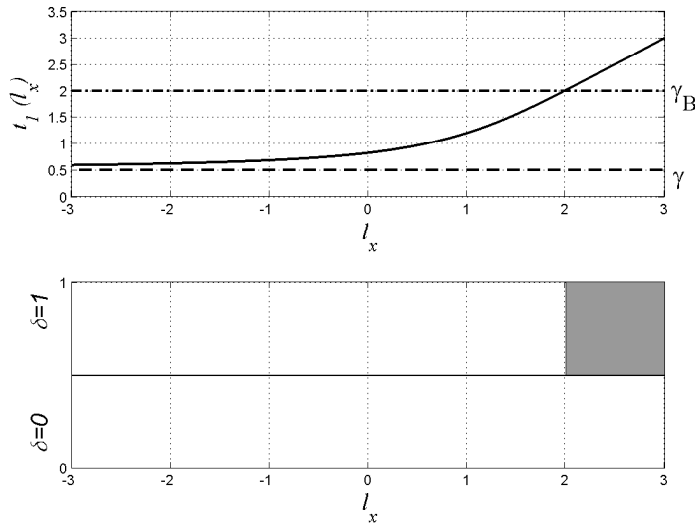**Figure 4.17** Function $t_1(l_x)$ (upper plot) and final decisions in the plane $(l_x, \delta)$ (lower plot), for the Gaussian shift-in-mean problem with $\mu_x = \mu_y = 1$, $\sigma_x = \sigma_y = 1$, $\rho = 0.5$, $\gamma = 0.5$ and $\gamma_B = 2$.

new decision $\delta_\mathcal{B}$ using the pair $(\boldsymbol{x}, \delta)$. If the decision $\delta_\mathcal{B}$ is made at the same false alarm level as the original decision $\delta$, then $\boldsymbol{x}$ is irrelevant and setting $\delta_\mathcal{B} = \delta$ is NP-optimal.

When the desired false alarm level for the final decision is different, however, some decisions can be safely retained (i.e., $\delta_\mathcal{B} = \delta$), but other requires a deeper analysis. As one might expect, we find that the sufficient statistic for the final decision is the pair $(\delta, L_x(\boldsymbol{x}))$, where $L_x(\boldsymbol{x})$ is the log-likelihood of vector $\boldsymbol{x}$: both the original decision $\delta$, and the data $\boldsymbol{x}$ influence the final decision $\delta_\mathcal{B}$, with data $\boldsymbol{x}$ playing their role only through the related log-likelihood $L_x(\boldsymbol{x})$.

Intuition perhaps might also suggest that the final detection structure amounts to comparing $L_x(\boldsymbol{x})$ with a suitable threshold level (what is commonly called a threshold test). This, however, is not true. In general, we show that the optimal decision consists of checking whether or not $L_x(\boldsymbol{x})$ belongs to some subset of the real axis having a complicated structure.

We have mainly addressed the NP framework, but the Bayesian counterpart has also been considered. As to this latter, we note that the refinement is not necessary due to improvement in the knowledge of the priors, as assumed in Section 4.5. Of practical interest is also the case where the priors are held fixed and the refinement involves instead a different assignment of the costs. Since this amounts to a different threshold setting, the two scenarios can be addressed exactly in the same way.

Finally, we have considered what happens when the full data set $(\boldsymbol{x}, \boldsymbol{y})$ is made of correlated data. We have shown that the behavior of the optimal detector solving the unlucky broker problem is the same of that solving the same problem but in presence of uncorrelated observations.

# Chapter 5

# The Very Unlucky Broker

In this chapter we consider a detection problem in a system made
of multiple sensors connected in series (or in tandem), referred
to as the *Very Unlucky Broker* problem. In particular, the data
available to the system are modeled as a vector of observations
and the first sensor of the series has fully access to this vector.
Successive sensors can only observe a fraction of the data, i.e.,
a subset of the vector entries. Each sensor makes its own local
decision about the state of the nature by exploiting its own data
set and the binary decision from the previous sensor. The final
decision of the system is that of the last sensor of the chain and
we are interested in characterizing the system behavior when the
sensor chain is very long.

It is straightforward to observe that the unlucky broker prob-
lem analyzed in Chapter 4 is essentially the very unlucky broker
problem made of two sensors.

In Section 5.1, we provide motivations for the study of the very
unlucky broker by considering three different scenarios of practical
interest. In Section 5.2 we formulate the problem, while in Sec-
tion 5.3 we analyze the asymptotic performance of the system and
we illustrate the main contribution of the chapter. In Section 5.4
numerical results corroborate theoretical findings and Section 5.5
summarizes the results and concludes the chapter.
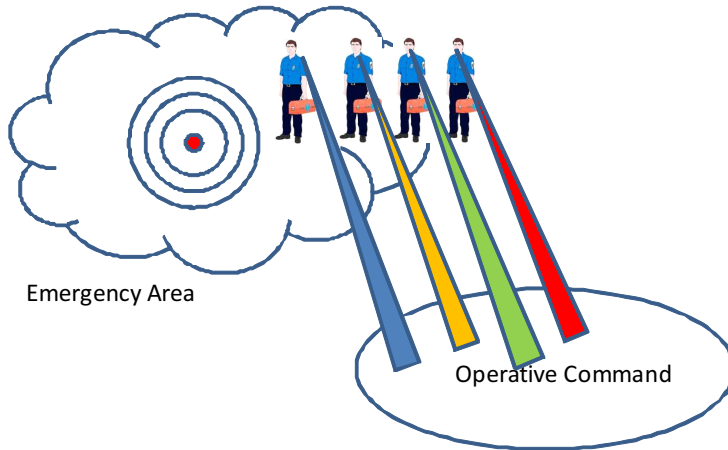
**Figure 5.1** Rescuers rushing to an emergency area where some disaster has just happened.

## 5.1   Introduction

In this section, we present three different scenarios which motivates the study of the very unlucky broker problem. Moreover, they allow us to figure out the links and the differences with the existing literature.

### 5.1.1   Rescue Operations

Consider a group of rescuers rushing to an emergency area where some disaster has just happened, as depicted in Fig. 5.1. They proceeds in single line and as they penetrate the damaged area approaching the core of the event, they experience a gradual reduction of available information from the outside world, since the communication with the outside is partially impaired, visibility is poor, people are confused, the psychological situation of the rescuer worsen, and so forth. Thus, the amount of information available to any individual of the rescue team depends on his position in the line and decreases progressively. Any rescuer, based on the information he has available and on the decision made by the individual behind him, makes a decision about how to oper-
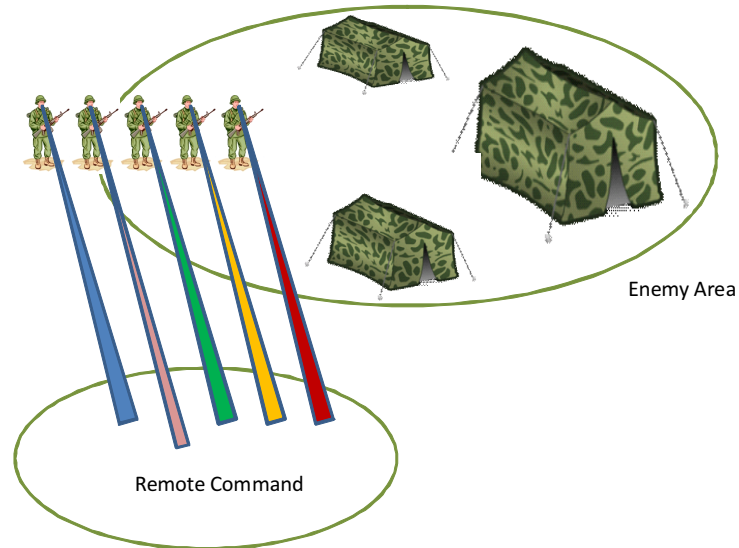
**Figure 5.2** Soldiers marching to an enemy area.

ate. The reliability of the decision depends upon the position of the rescuer in the line, as well: the more one is close to the core of the disaster, the more critical the decision becomes and, therefore, a more reliable decision is required. How accurate is the decision of the last rescuer in a very long line?

## 5.1.2 Soldiers

Consider a platoon of soldiers marching in single line toward an enemy campsite, see Fig. 5.2. As more in depth they enter into the enemy area, less contact they have with the remote command and, at the same time, more dangerous is their position: soldiers at the head of the line have less information but they need to make decisions with smaller error probability. Suppose that each soldier receives from the previous one a decision made about some course of action to take. On this basis, and by exploiting the information he receives from the command, his own decision is made and delivered to the next soldier in the line, and so on.
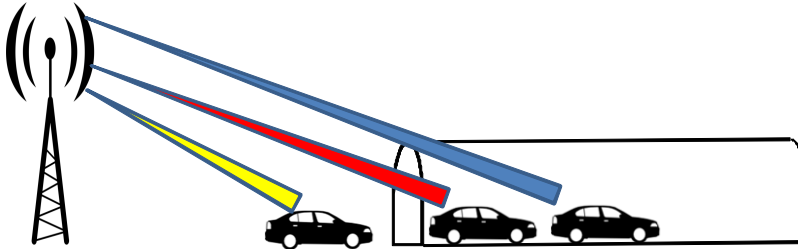
**Figure 5.3**  Queue of cars entering into a long tunnel.

We want to study the behavior of such decision system, when the number of soldiers is very large.

### 5.1.3   Vehicles in Tunnel

Imagine a queue of cars entering into a long tunnel, as depicted in Fig. 5.3. Each vehicle driver receives information from the outside, for instance by listening a devoted radio program, concerning the environmental situation (e.g., traffic) and he wants to decide between two possible courses of action. Each driver observes the decision made by the previous car in the queue, and he can only listen a fraction of the transmitted traffic information, depending on how much he is inside the tunnel, where radio communication is problematic. Given that the more inside the tunnel the car is, the more important the decision is, the driver selects the risk of making a wrong decision as a function of the road traveled inside the tunnel. What can one say about the statistical behavior of the decision made by a driver that is very far away from the tunnel entrance?

## 5.2   Problem Formulation

A number of practical situations can be assimilated to the examples given in the previous section and, to attack the problem in a systematic way, a suitable abstraction of the matter is provided in Section 5.2.1. In Section 5.2.2 we propose a reasonable deci-
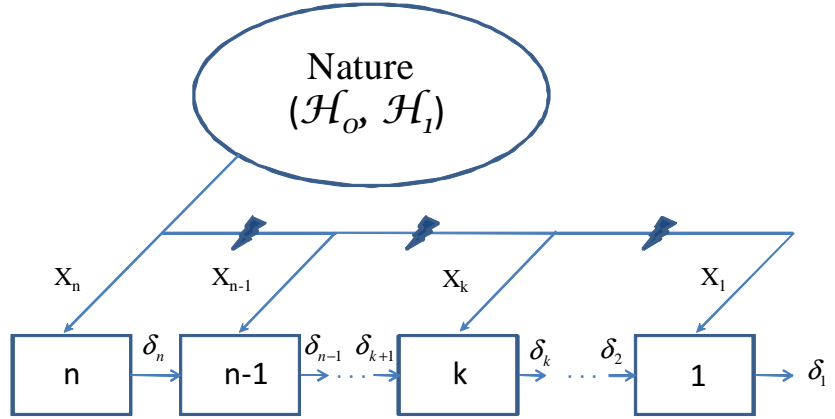
**Figure 5.4** Notional sketch of the addressed problem.

sion strategy to be used for the analysis of the system depicted in Fig. 5.4 .

## 5.2.1 Abstraction

The scenarios considered in the above can be sketched as done in Fig. 5.4, where we consider $n$ nodes (or sensors) connected in series $S_n \rightarrow S_{n-1} \rightarrow \cdots \rightarrow S_1$. Each node is engaged in a detection problem consisting of deciding between two mutually exclusive hypotheses

$$
\begin{aligned}
\mathcal{H}_0 : & \quad \text{data } x \text{ are drawn from } f_0(x), \\
\mathcal{H}_1 : & \quad \text{data } x \text{ are drawn from } f_1(x), \quad\quad (5.1)
\end{aligned}
$$

where $f_0$ and $f_1$ represent two probability density functions.

The first node has available a set of data modeled as a random vector $\boldsymbol{X}_n := [X_1, X_2, \ldots, X_n]$, with entries $X_i$ that are mutually independent random variables with identical distribution (henceforth i.i.d.). While sensor $S_n$ observes $\boldsymbol{X}_n$, the node $S_{n-1}$ has available the subvector $\boldsymbol{X}_{n-1} := [X_1, X_2, \ldots, X_{n-1}]$, and similarly for successive nodes: the $k^{th}$ sensor observes $\boldsymbol{X}_k := [X_1, X_1 \ldots, X_k]$, and node $S_1$ only $\boldsymbol{X}_1 := X_1$.

Each node makes a binary decision about the hypothesis-testing problem in (5.1), with the $k^{th}$ decision being a random variable taking values in $\{0, 1\}$. To make this decision, node $S_k$ exploits the available data set $\boldsymbol{X}_k$ and knowledge of $\delta_{k-1}$, the decision made by the previous node. The first node, clearly, has no previous decision available and it only exploits $\boldsymbol{X}_n$ for its decision.

Let us consider the two error probabilities for the decisions made at the $k^{th}$ node. We impose the Type I error (false alarm) probability at node $S_k$ to be smaller than that at the previous node $S_{k+1}$ and we want to investigate the asymptotic behavior (for $n \to \infty$) of the Type II (miss detection) error probability experienced at $S_1$, subject to a constraint on its false alarm rate. A number of features characterize the stated problem.

First of all, it is interesting to observe that the system architecture depicted in Fig. 5.4 is the same of the well-studied serial (or tandem) detection systems [13–25]. The observation structure, however, makes the problems very different since, oppositely to the classical assumptions in tandem detection systems, here the data sets available at successive nodes are strongly dependent. Essentially, there exists a unique data set $\boldsymbol{X}_n$ of size $n$, but successive nodes observe only a fraction of that vector. We are faced with a progressive reduction in knowledge, combined with a tandem network architecture in which local decisions are propagated sequentially. The inspiration for this work actually comes instead from the so-called unlucky broker problem studied in Chapter 4 (see also [31, 32]). Recall that there we consider the optimal investment strategy proposed by a broker to his customers, and the focus is on how the broker's strategy should be adjusted if later on ($i$) he remembers the original optimal strategy (decision), but loses part of the data initially available (whence the adjective unlucky), and ($ii$) he wants to reduce the risk of bad investments with respect to the risk level of the originally proposed portfolio assessment. The unlucky broker problem is essentially the problem we are considering here, but with $n = 2$. For this reason we will also refer to the scenario described in the above as the Very Unlucky Broker (VUB) problem, where "very" refers to the

progressive and continuous loss of data in the system.

Second, the sensors of the system must be thought of as devices programmed in advance to make their task. Once deployed, there is no possibility of modify them and the only tasks they can do are $i$) to take the decision from the previous node and the data available to them, $ii$) to implement some threshold test to decide between the hypotheses, and $iii$) to forward the resulting decision to the next node.

It can be shown that, even with only two nodes, the optimal test according to typical optimization criteria including Neyman-Pearson, Bayes, minimax, is very difficult to be implemented. It has a multiple threshold structure, it does not admit a simple expression for the decision statistics, and it is not amenable to a precise statistical characterization [31, 32]. The situation is worse for $n > 2$ and even worse for $n$ very large, which is the regime of interest here.

The previous consideration implies that, in order to make the system actually implementable in practice and to ensure analytical tractability, we must give up any foolish ambition of designing a detector that cannot be ever beat. Rather, some "reasonable" detector must be conceived, and this requires clever non-optimal local tests at individual sensors.

## 5.2.2 Decision strategy

With reference to Fig. 5.4, we have a series connection of $n$ entities $S_n \to S_{n-1} \to \cdots \to S_1$, where node $S_k$ has available the decision $\delta_{k+1}$ made by node $k + 1$, and the observation vector $\boldsymbol{X}_k = [X_1, X_2, \ldots, X_k]$.[1] Using that, $S_k$ makes its own decision about a binary hypothesis test and then it forwards that to the next node.

We consider simple hypothesis test of the form

$$\mathcal{H}_0 : P \equiv P_0, \quad \text{vs.} \quad \mathcal{H}_1 : P \equiv P_1,$$

---

[1]We use capital letter to denote random variables and the correspondent lowercase to indicate their realizations.

where $P_1$ and $P_0$ are two probability measures on $(\Re, \mathcal{B})$, where $\Re$ is the real line and $\mathcal{B}$ represents the smallest $\sigma$-algebra of its subsets that contains all half-lines. We assume throughout the chapter that $P_1$ is absolutely continuous with respect to $P_0$, and that the two distributions admit Radon-Nikodym derivatives, $f_1$ and $f_0$, respectively, both with respect to the usual Lebesgue measure. Clearly, the random variable $X_k$ corresponds to the measure probability $P_0$ (resp. $P_1$) and the vector $\boldsymbol{X}_n$ to the $n$-fold product measure originated by $P_0$ (resp. $P_1$). Therefore, its density can be written as $f_j(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_j(x_i)$, $j = 0, 1$.

As to the decision updating rule, it is assumed that node $k$ is constrained to send a binary decision to the node $k-1$, of the form $\delta_k = g_{k,n}(\delta_{k+1}, \boldsymbol{x}_k)$, where $g_{k,n} : \{0,1\} \times \Re^k \mapsto \{0,1\}$.

Let us introduce the local Type I and Type II error probabilities at the node $k$:

$$\begin{aligned}
\alpha_{k,n} &= \Pr\{\delta_k = 1; \mathcal{H}_0\} \\
\beta_{k,n} &= \Pr\{\delta_k = 0; \mathcal{H}_1\},
\end{aligned} \qquad (5.2)$$

and let us refer to an NP setting [36]. The focus here is to characterize the asymptotic ($n \to \infty$) behavior of $\beta_{1,n}$, subject to a constraint on the false alarm probability $\alpha_{1,n}$.

In order to maximize the system performance at each stage of the chain, the optimal NP test should be implemented. Unfortunately, we know from [31, 32] and Chapter 4 that such an optimal detector cannot be easily implemented, neither for $n = 2$, and that there is no hope to compute analytically its asymptotic performance. On the other hand, as it is shown in the Appendix B.1, the optimal way to make the false alarm probability decreasing along the system is to retain all the decisions in favor of the null hypothesis $\mathcal{H}_0$ and to double-check the decisions in favor of $\mathcal{H}_1$ by using an intractable statistic.

Then, our approach is as follows. First, we account for the fact that a non increasing false alarm rate along the chain is required, i.e., we want $\alpha_{k+1,n} \geq \alpha_{k,n}$. A simple way to obtain that is to retain all the decisions in favor of $\mathcal{H}_0$, by setting:

$$g_{k,n}(0, \boldsymbol{x}_k) = 0, \qquad \forall \boldsymbol{x}_k.$$

Conversely, whenever the previous decision is in favor of $\mathcal{H}_1$, i.e., $\delta_{k+1} = 1$, the local decision is based on a function of the log-likelihood ratio of the available vector $\boldsymbol{x}_k$:

$$L_k(\boldsymbol{x}_k) := \log \frac{f_1(\boldsymbol{x}_k)}{f_0(\boldsymbol{x}_k)} = \sum_{i=1}^{k} l_i(x_i),$$

where $l_i(x_i) = \log \frac{f_1(x_i)}{f_0(x_i)}$ represents the marginal log-likelihood ratio of the single observation $x_i$. Specifically, given that the previous decision favors $\mathcal{H}_1$, the current decision is made by comparing $L_k(\boldsymbol{x}_k)$ to a suitable threshold level.

To specify exactly the proposed approach we need some preliminary considerations. In order to investigate the asymptotic regime of $n \to \infty$, we essentially rely, as will be clear later, upon the convergence of the regularized version of the log-likelihood $L_k(\boldsymbol{x}_k)$ to a Wiener process defined in the interval $t \in [0,1]$, with an appropriate drift, under the hypothesis $\mathcal{H}_0$. This requires the introduction of the following quantities.

Let $\eta(t)$ be an arbitrary function, continuous, integrable and bounded in $(0,1)$, with

$$\max_{t \in [0,1]} |\eta(t)| = \zeta < \infty. \tag{5.3}$$

We define a *drift function* $d : [0,1] \mapsto \Re$ as

$$d(t) = \int_0^t \eta(\tau) d\tau, \tag{5.4}$$

with $d(0) = 0$. From the integral mean value theorem, we have

$$d\left(\frac{k}{n}\right) = \sum_{i=1}^{k} \int_{(i-1)/n}^{i/n} \eta(\tau) d\tau = \frac{1}{n} \sum_{i=1}^{k} \eta_{i,n}, \quad k \le n,$$

where $\eta_{i,n} = \eta(i/n) - \eta((i-1)/n)$ are the pertinent centroids. Using the above we define the following scaled and shifted version of the log-likelihood, which refers to $k$ out of $n$ total data:

$$L_{k,n}(\boldsymbol{x}_k) = \frac{\sum_{i=1}^{k}(l_i(x_i) + D_{01})}{\sigma_{01}\sqrt{n}} - d\left(\frac{k}{n}\right), \tag{5.5}$$

where $D_{01}$ is the usual Kullback-Leibler distance of $f_0$ from $f_1$ [42], i.e., the expected value (but for a sign) of $l_i(x_i)$, while $\sigma_{01}$ is the standard deviation of $l_i(x_i)$, both under measure $P_0$.

Expression (5.5) is equivalent to

$$
L_{k,n}(\boldsymbol{x}_k) = \frac{\sum_{i=1}^{k}\left(l_i(x_i) + D_{01} - \frac{\sigma_{01}}{\sqrt{n}}\eta_{i,n}\right)}{\sigma_{01}\sqrt{n}}, \qquad (5.6)
$$

which represents the statistics used at the $k^{th}$ node of the chain, and which is to be compared to one and the same threshold value, say $-\gamma$.

Accordingly, the decision strategy at the $k^{th}$ node is

$$
g_{k,n}(\delta_{k+1}, \boldsymbol{x}_k) = \begin{cases} 0, & \text{if } \delta_{k+1} = 0, \\ \mathfrak{I}\left(L_{k,n}(\boldsymbol{x}_k) > -\gamma\right), & \text{if } \delta_{k+1} = 1. \end{cases}
$$

where $\mathfrak{I}(\cdot)$ is the indicator function.

Note that the decision made at the last stage of the chain will be in favor of $\mathcal{H}_1$ if and only if all the stages of the chain have decided for $\mathcal{H}_1$, namely if

$$
T_n(\boldsymbol{x}_n) := \min_{k \leq n} L_{k,n}(\boldsymbol{x}_k) \qquad (5.7)
$$

lies above $-\gamma$. Thus, the error probabilities we are interested in are

$$
\begin{aligned}
\alpha_n &:= \alpha_{1,n} = \Pr\{T_n(\boldsymbol{X}_n) > -\gamma; \mathcal{H}_0\}. & (5.8) \\
\beta_n &:= \beta_{1,n} = \Pr\{T_n(\boldsymbol{X}_n) \leq -\gamma; \mathcal{H}_1\}. & (5.9)
\end{aligned}
$$

In the next section we show how to set $\lim_{n\to\infty}\alpha_n$ to a desired value, and how to compute the corresponding $\beta_n$ for $n \to \infty$. For later use, we also consider the NP-optimal performance of an ideal detector that optimally processes the whole set of data $\boldsymbol{X}_n$:

$$
\begin{aligned}
\alpha_n^{opt} &:= \alpha_{n,n}, & (5.10) \\
\beta_n^{opt} &:= \beta_{n,n}. & (5.11)
\end{aligned}
$$

# 5.3 Asymptotic Performance of VUB

In this section, we study the asymptotic behavior of the system described in the above. We start by stating the following result, which is essentially Donsker's theorem combined with Slutsky's theorem [46], and therefore requires no proof.

**Proposition 5.3.1.** *Let $W_n(t) - \delta_n(t)$, with $t \in (0,1)$, be a continuous-time random process such that $W_n(t) - \delta_n(t) = L_{k,n}(\boldsymbol{X}_k)$ for $k = \lfloor nt \rfloor$, and for $k \neq \lfloor nt \rfloor$ is defined as the linear interpolation of those values. Donsker's theorem implies that, for $n \to \infty$, $W_n(t) - \delta_n(t)$ converges in distribution to $W(t) - d(t)$ where $W(t)$ is a standard Wiener process and $d(t)$ is the deterministic drifting function in (5.4).* ●

The result in the Proposition 5.3.1 provides an asymptotic correspondence between the discrete-time process $L_{k,n}(\boldsymbol{X}_k)$ where $k = n, n-1, n-2 \dots, 1$, and $W(t) - d(t)$, a continuous-time Wiener process with drift, defined in $t \in [0,1]$, under the null hypothesis $\mathcal{H}_0$. In this chapter, $t = 1$ refers to the presence of the largest data set, while $t = 0$ corresponds to the last sensor with the smallest data set. It is obvious that $t = 0$ is achieved only asymptotically. Straightforward byproducts are now derived. First, we see that $\gamma$ must be selected as a positive number. Indeed,

$$
\begin{aligned}
\lim_{n\to\infty} \alpha_n &= \lim_{n\to\infty} \Pr\{\min_{k\le n} L_{k,n}(\boldsymbol{x}_k) > -\gamma; \mathcal{H}_0\} \\
&= \Pr\{\min_{t\in[0,1]} [W(t) - d(t)] > -\gamma; \mathcal{H}_0\} \\
&\le \Pr\{W(0) - d(0) > -\gamma; \mathcal{H}_0\},
\end{aligned}
$$

which is certainly zero for any $\gamma \le 0$, being $W(0) - d(0) = 0$.

Second, we show how to set the detection threshold $-\gamma$ as a function of the false alarm probability of the optimal system:

$$
\begin{aligned}
\lim_{n\to\infty} \alpha_n^{opt} &= \lim_{n\to\infty} \Pr\{L_{n,n}(\boldsymbol{x}_n) > -\gamma; \mathcal{H}_0\} \\
&= \Pr\{W(1) - d(1) > -\gamma; \mathcal{H}_0\} \\
&= Q\left[-\gamma + d(1)\right], \tag{5.12}
\end{aligned}
$$

where $Q(\cdot)$ is the standard Gaussian exceedance function. Let $Q^{-1}(\cdot)$ be its inverse function. Assuming $\lim_n \alpha_n^{opt} < 0.5$, and choosing a drifting function such that

$$d(1) > Q^{-1}\left(\lim_{n\to\infty} \alpha_n^{opt}\right), \tag{5.13}$$

we get

$$\gamma = d(1) - Q^{-1}\left(\lim_{n\to\infty} \alpha_n^{opt}\right). \tag{5.14}$$

The asymptotic false alarm probability of the VUB can be predicted by exploiting some well-known results on the barrier-crossing probabilities of Wiener paths, see e.g., [33,34]. In particular, we can write

$$
\begin{aligned}
\lim_{n\to\infty} \alpha_n &= \Pr\{\min_{t\in[0,1]}[W(t) - d(t)] > -\gamma; \mathcal{H}_0\} \\
&= \Pr\{\max_{t\in[0,1]}[W(t) + d(t)] \le \gamma; \mathcal{H}_0\} \\
&= 1 - \Pr\{\max_{t\in[0,1]}[W(t) + d(t)] > \gamma; \mathcal{H}_0\} \\
&= 1 - \Pr\{\max_{t\in[0,1]}[W(t) - f(t)] > 0; \mathcal{H}_0\}, \quad (5.15)
\end{aligned}
$$

where $f(t) = -d(t) + \gamma$. For the evaluation of the second term on the right hand side, we can use the following theorem [34]

**Theorem 5.3.1.** *Let $SC[a,b]$ denote the class of all sectionally continuous functions[2] $f(t)$ on $[a,b]$ such that at each point $t_0$ of discontinuity $f(t_0) = \min\{f(t_0-), f(t_0+)\}$. Let $f(t) \in SC[0,T]$. Then the probability*

$$F(T) \equiv P\{\sup_{t\in[0,T]}[W(t) - f(t)] > 0\} \tag{5.16}$$

---

[2]A function $f(x)$ is said to be sectionally continuous (or piecewise continuous) on an interval $[a,b]$ if the interval can be subdivided into a finite number of intervals in each of which the function is continuous and has finite left and right limits.

*is one if $f(0) \leq 0$, and otherwise it satisfies the integral equation*

$$
\begin{aligned}
F(T) &= Q[f(T)/T^{1/2}] \\
&+ \int_0^T 1 - Q([f(T) - f(t)]/(T - t)^{1/2})dF(t).
\end{aligned}
$$
(5.17)

●

By using eq. (5.17) in (5.15), we obtain

$$
\begin{aligned}
\lim_{n \to \infty} \alpha_n &= 1 - Q[f(1)] \\
&- \int_0^1 1 - Q([f(1) - f(t)]/(1 - t)^{1/2})dF(t).
\end{aligned}
$$
(5.18)

From eq. (5.18) it is clear that if we want to choose the threshold $\gamma$ by setting the asymptotic false alarm of the VUB to a desired value, we should be able to solve the integral equation in (5.18).

Let us denote by $l(X_i)$ the log-likelihood of the scalar random variable $X_i$ (any $i$), and consider now the random variable $-l(X_i) - D_{01}$, which under $P_0$ has zero mean and standard deviation $\sigma_{01}$, while under $P_1$ has negative mean. Let $\Psi(\theta)$ be its semi-invariant (or cumulant) generating function [47]:

$$
\Psi(\theta) = \log \mathrm{E}_1 \left[ e^{-\theta\,(l(X_i)+D_{01})} \right], \quad \theta \in \Re, \qquad (5.19)
$$

where the expectation is computed under $P_1$.

It is assumed throughout the chapter that $\Psi(\theta)$ is finite in an open interval $(\theta_l, \theta_u)$ including the origin[3]. Then $\Psi(0) = 0$, $\Psi'(0) = \mathrm{E}_1[-l(X_i) - D_{01}] < 0$, $\Psi''(0) = \mathrm{VAR}_1[-l(X_i) - D_{01}] > 0$, $\Psi(\theta)$ is strictly convex, and its derivative is strictly increasing (see

---

[3]This implies that $l(X_i)$ has moments of all orders and its density approaches zero at infinity at least exponentially.

e.g., [47]). We further assume the existence of $\theta^*$ as detailed in the following.

ASSUMPTION. *$\Psi(\theta)$ is finite in an open interval $(\theta_l, \theta_u)$ with $\theta_l < 0 < \theta_u$ and there exists $\theta^* < \theta_u$, which is the unique positive solution of the equation $\Psi(\theta) = 0$.* ●

The main result of the chapter is now stated.

**Theorem 5.3.2.** *For any desired value of the false alarm probability $0 < \lim_n \alpha_n^{opt} < 0.5$, the miss detection probability of the VUB verifies*

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \log \beta_n = -\theta^* \sigma_{01} \gamma, \tag{5.20}$$

*where $\gamma$ is given by (5.14) and $\theta^*$ is the unique positive solution of $\Psi(\theta) = 0$.* ●

*Proof:* Consider the miss detection probability as defined in eq. (5.9). This is clearly equivalent, for any positive real number $\theta$, to[4]

$$\beta_n = P_1 \left\{ \max_{k \leq n} e^{-\theta L_{k,n}(\boldsymbol{X}_k)} \geq e^{\theta \gamma} \right\}. \tag{5.21}$$

We now proceed to bound the above quantity from above and from below. Let us start with the upper bound.

Thanks to eq. (5.3), we have

$$L_{k,n}(\boldsymbol{X}_k) \geq \frac{\sum_{i=1}^{k} \left( l_i(X_i) + D_{01} - \frac{\sigma_{01}\zeta}{\sqrt{n}} \right)}{\sigma_{01}\sqrt{n}} \tag{5.22}$$

almost surely (a.s.), yielding

$$\beta_n \leq P_1 \left\{ \max_{k \leq n} e^{-\theta \sum_{i=1}^{k} \left( l_i(X_i) + D_{01} - \frac{\sigma_{01}\zeta}{\sqrt{n}} \right)} \geq e^{\sqrt{n}\theta\sigma_{01}\gamma} \right\}. \tag{5.23}$$

Next, consider the semi-invariant generating function as given in (5.19). We form the equation

$$\Psi(\theta) = -\theta \frac{\sigma_{01}\zeta}{\sqrt{n}}, \tag{5.24}$$

---

[4]In the following, we use the notation $\mathrm{Pr}\{ \ \cdot \ ; \ \mathcal{H}_j \} = P_j(\cdot)$.

and denote its positive solution by $\theta = \theta_n^*$. Thanks to the assumptions made about the function $\Psi(\theta)$, such value exists at least for $n \geq n_0$ where $n_0$ is sufficiently large. Assuming such values of $n$, and using $\theta_n^*$, we define the following random variables (dependence upon $n$ is omitted for notational convenience)

$$
\begin{aligned}
Z_k &= e^{-\theta_n^* \sum_{i=1}^k \left( l_i(X_i) + D_{01} - \frac{\sigma_{01}\varsigma}{\sqrt{n}} \right)} \\
&= Z_{k-1}\, e^{-\theta_n^* \left( l(X_k) + D_{01} - \frac{\sigma_{01}\varsigma}{\sqrt{n}} \right)}.
\end{aligned} \tag{5.25}
$$

It is easy to see that $Z_k$ is a unit-mean martingale under measure $P_1$, *viz.*

$$
\mathrm{E}_1[Z_k | Z_{k-1}, Z_{k-2}, \ldots, Z_1] = Z_{k-1}. \tag{5.26}
$$

Kolmogorov's sub-martingale inequality [47] then implies

$$
P_1 \left\{ \max_{k \leq n} Z_k \geq a \right\} \leq \frac{\mathrm{E}_1[Z_n]}{a}
$$

for any $a > 0$, thus yielding

$$
\beta_n \leq \frac{\mathrm{E}_1[Z_n]}{e^{\sqrt{n}\,\theta_n^*\,\sigma_{01}\gamma}} = e^{-\sqrt{n}\,\theta_n^*\,\sigma_{01}\gamma}. \tag{5.27}
$$

Since $\theta_n^* \overset{n \to \infty}{\longrightarrow} \theta^*$, where $\theta^*$ is the unique positive solution of $\Psi(\theta) = 0$, we finally get

$$
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \log \beta_n \leq -\theta^* \sigma_{01}\, \gamma. \tag{5.28}
$$

The computation of a lower bound for (5.21) is now in order. First, the miss detection probability can be easily lower bounded by restricting the set where the maximum is taken, that is, for any $0 < \epsilon < 1$,

$$
\begin{aligned}
\beta_n &= P_1 \left\{ \max_{k \leq n} -L_{k,n}\left(\boldsymbol{X}_k\right) \geq \gamma \right\} \\
&\geq P_1 \left\{ \max_{k \leq \epsilon n} -L_{k,n}\left(\boldsymbol{X}_k\right) \geq \gamma \right\} \tag{5.29}
\end{aligned}
$$

Then, exploiting again condition (5.3), we can write

$$
L_{k,n}\left(\boldsymbol{X}_k\right) \leq \frac{\sum_{i=1}^k \left(l_i\left(X_i\right) + D_{01} + \frac{\sigma_{01}}{\sqrt{n}}\zeta\right)}{\sigma_{01}\sqrt{n}}. \tag{5.30}
$$

By using the restriction of the set where the maximum is taken, i.e., $k \leq \epsilon n$, we get

$$
L_{k,n}\left(\boldsymbol{X}_k\right) \leq \frac{\sum_{i=1}^k \left(l_i\left(X_i\right) + D_{01}\right)}{\sigma_{01}\sqrt{n}} + \bar{\epsilon}, \tag{5.31}
$$

where we have set $\bar{\epsilon} := \epsilon\,\zeta$. Using (5.31) in (5.29) yields

$$
\beta_n \geq P_1\left\{\max_{k \leq \epsilon n} e^{-\theta^* \sum_{i=1}^k (l_i(X_i)+D_{01})} \geq e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar{\epsilon})}\right\}.
$$

By defining $Y_k = e^{-\theta^* \sum_{i=1}^k (l_i(X_i)+D_{01})}$ and introducing the stopping time

$$
N = \left\{ \begin{array}{ll} \lfloor \epsilon\,n \rfloor, & \text{if } Y_k < e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar{\epsilon})} \quad \forall k \leq \epsilon\,n, \\ \min\{k : Y_k \geq e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar{\epsilon})}, & k \leq \epsilon\,n\}, \end{array} \right.
$$

we can write

$$
\begin{aligned}
\beta_n &\geq& P_1\left\{\max_{k \leq \epsilon n} Y_k \geq e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar{\epsilon})}\right\} \\
&=& P_1\left\{Y_N \geq e^{\theta^*\sigma_{01}\sqrt{n}(\gamma+\bar{\epsilon})}\right\}, \\
&\geq& P_1\{Y_N \in \mathcal{A}\},
\end{aligned}
$$

where in the last inequality the interval

$$
\mathcal{A} := \left[e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar{\epsilon})}, e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})}\right]
$$

has been defined.

Therefore, we have

$$
\beta_n \geq P_1\{Y_N \in \mathcal{A}\} = \sum_{k=1}^{\lfloor \epsilon n \rfloor} P_1\{Y_k \in \mathcal{A}, N = k\}.
$$

Let us focus on the single terms of the sum. We define the set

$$\mathcal{B}_k = \{\boldsymbol{x}^k : Y_k \in \mathcal{A}, N = k\},$$

such that $P_1\{Y_k \in \mathcal{A}, N = k\}$ can be written as

$$\frac{e^{-\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})}}{e^{-\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})}} \int_{\mathcal{B}_k} f_1(\boldsymbol{x}_k)d\boldsymbol{x}_k. \tag{5.32}$$

Now, $Y_k \in \mathcal{A}$ implies $Y_k \leq e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})}$ and, by definition of the set $\mathcal{B}_k$, we know that $N = k$. Therefore

$$P_1\{Y_k \in \mathcal{A}, N = k\} \geq e^{-\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})} \int_{\mathcal{B}_k} f_1(\boldsymbol{x}_k)\, Y_k d\boldsymbol{x}_k$$

yielding

$$\beta_n \geq e^{-\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})} \int_{\mathcal{B}_k} \prod_{i=1}^{k} f_1(x_i)e^{-\theta^*(l(X_i)+D_{01})}d\boldsymbol{x}_k. \tag{5.33}$$

Let us now define

$$q(x) = f_1(x)e^{-\theta^*(l(x)+D_{01})},$$

that, as it is easy to show by the definition of $\theta^*$, is a valid probability density function. In terms of the new measure $P_q$, the bound in (5.33) can be recast as

$$\beta_n \geq e^{-\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar{\epsilon})}\, P_q\{Y_N \in \mathcal{A}\}. \tag{5.34}$$

At this point, note that

$$\log \frac{q(x)}{f_1(x)} = -\theta^*(l(x) + D_{01}),$$

such that we can define the random walk

$$\log Y_k = \sum_{i=1}^{k} \log \frac{q(x_i)}{f_1(x_i)}. \tag{5.35}$$

We also define the *relaxed* stopping time

$$
\begin{aligned}
N_0 &= \min\{k : Y_k \geq e^{\theta^* \sigma_{01} \sqrt{n}(\gamma + \bar{\epsilon})}, \quad k > 0\}, \\
&= \min\{k : \log Y_k \geq \theta^* \sigma_{01} \sqrt{n}(\gamma + \bar{\epsilon}), \quad k > 0\},
\end{aligned}
$$

that, a.s. verifies $N_0 \geq N$.

Let us focus on the term $P_q\{Y_N \in \mathcal{A}\}$ in (5.34). We have

$$
\begin{aligned}
P_q\{Y_N \in \mathcal{A}\} &= 1 - P_q\left\{Y_N < e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + \bar{\epsilon})}\right\} \\
&\qquad - P_q\left\{Y_N > e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + 2\bar{\epsilon})}\right\}.
\end{aligned} \tag{5.36}
$$

As to the first term, we have, by definition of the stopping time $N$:

$$
\begin{aligned}
&P_q\left\{Y_N < e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + \bar{\epsilon})}\right\} \\
&= P_q\left\{Y_{n\epsilon} < e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + \bar{\epsilon})}, N = n\epsilon\right\} \\
&\leq P_q\left\{\sum_{i=1}^{n\epsilon} \log \frac{q(x_i)}{f_1(x_i)} < \sqrt{n}\,\theta^* \sigma_{01}(\gamma + \bar{\epsilon})\right\} \\
&= P_q\left\{\frac{1}{n\epsilon} \sum_{i=1}^{n\epsilon} \log \frac{q(x_i)}{f_1(x_i)} < \frac{\theta^* \sigma_{01}}{\sqrt{n}} \frac{(\gamma + \bar{\epsilon})}{\epsilon}.\right\}
\end{aligned} \tag{5.37}
$$

Now, the expected value of $\log \frac{q(x)}{f_1(x)}$ *under* $q(x)$ is positive, such that, by the Weak Law of Large Numbers (WLLN), the last probability goes to zero as $n$ diverges, for each $0 < \epsilon < 1$.

Switching now to the term

$$
P_q\left\{Y_N > e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + 2\bar{\epsilon})}\right\}, \tag{5.38}
$$

we have, by definition of the relaxed stopping time $N_0$:

$$
P_q\left\{Y_N \geq e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + 2\bar{\epsilon})}\right\} = P_q\left\{Y_{N_0} \geq e^{\sqrt{n}\,\theta^* \sigma_{01}(\gamma + 2\bar{\epsilon})}\right\}. \tag{5.39}
$$

To evaluate the asymptotic behavior of the term on the right hand side of (5.39) we can invoke the following theorem(see [48], page 89).

**Theorem 5.3.3.** *Let $\{S_n,\ n > 0\}$ be a random walk with positive drift. Moreover, let $\nu(m)$ be the first passage time beyond the level $m$, that is $\nu(m) = \min\{n : S_n > m\}$, with $m > 0$. Then, the following strong law for the stopped sum and the stopping summand holds*

$$\frac{S_{\nu(m)}}{m} \xrightarrow{a.s.} 1 \tag{5.40}$$

*as $m \to \infty$.* ●

By applying this theorem to the positive drifted random walk $\log Y_k$ and to the relaxed stopped time $N_0$, we obtain

$$\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} \xrightarrow{a.s.} 1, \tag{5.41}$$

as $n \to \infty$. Almost surely convergence implies convergence in probability. Accordingly, we have that for each $\varsigma > 0$:

$$\lim_{n\to\infty} P_q\left\{\left|\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1\right| > \varsigma\right\} = 0. \tag{5.42}$$

We can clearly write that

$$P_q\left\{\left|\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1\right| > \varsigma\right\}$$
$$= P_q\left\{\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1 < -\varsigma\right\}$$
$$+ P_q\left\{\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1 > \varsigma\right\}$$
$$\geq P_q\left\{\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1 > \varsigma\right\}. \tag{5.43}$$

As $n$ diverges

$$\lim_{n\to\infty} P_q\left\{\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} > 1 + \varsigma\right\}$$
$$\leq \lim_{n\to\infty} P_q\left\{\left|\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma + \bar{\epsilon})} - 1\right| > \varsigma\right\} = 0. \tag{5.44}$$

By invoking the Theorem in the above, we get

$$
\begin{aligned}
& P_q\left\{Y_{N_0} > e^{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar\epsilon)}\right\} \\
=\ & P_q\left\{\log Y_{N_0} > \sqrt{n}\,\theta^*\sigma_{01}(\gamma+2\bar\epsilon)\right\} \\
=\ & P_q\left\{\frac{\log Y_{N_0}}{\sqrt{n}\,\theta^*\sigma_{01}(\gamma+\bar\epsilon)} > 1 + \frac{\bar\epsilon}{\gamma+\bar\epsilon}\right\},
\end{aligned}
\tag{5.45}
$$

which, from what just stated, vanishes as $n \to \infty$.

Putting the last two results together, from (5.36) we conclude

$$
\lim_{n\to\infty} P_q\{Y_N \in \mathcal{A}\} = 1,
$$

implying, in view of (5.34)

$$
\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \log \beta_n \geq -\theta^*\sigma_{01}\gamma - 2\theta^*\sigma_{01}\bar\epsilon
$$

Since $\bar\epsilon$ can be made as small as desired by an appropriate choice of $\epsilon$, we finally get

$$
\liminf_{n\to\infty} \frac{1}{\sqrt{n}} \log \beta_n \geq -\theta^*\sigma_{01}\gamma.
$$

This, along with the previous result on $\limsup$ in (5.28), gives the desired result

$$
\lim_{n\to\infty} \frac{1}{\sqrt{n}} \log \beta_n = -\theta^*\sigma_{01}\gamma
$$

and the proof is complete. ●

## 5.4   Examples

In this section, the theoretical framework presented in the previous sections is applied to specific decision problems of practical interest.

### 5.4.1 Gaussian shift-in-mean

Let us consider the following decision problem

$$
\begin{aligned}
\mathcal{H}_0 &: \quad X_i \sim \mathcal{N}(0, \sigma^2), \\
\mathcal{H}_1 &: \quad X_i \sim \mathcal{N}(\mu, \sigma^2),
\end{aligned}
\tag{5.46}
$$

where $i = 1, \ldots, n$ and $\mathcal{N}(\mu, \sigma^2)$ stands for a Gaussian pdf with mean $\mu$ and variance $\sigma^2$. In this case, the marginal log-likelihoods $l_i \sim \mathcal{N}(\pm D, 2D)$, $i = 1, \ldots, n$, where $D := D_{01} = D_{10} = \mu^2/2\sigma^2$, while the negative and positive signs are referred to the null and alternative hypotheses, respectively. The semi-invariant generating function of the random variable $-l_i - D$ in (5.19) is $D\theta(\theta - 2)$. Fig. 5.5 depicts the behavior of the function $\Psi(\theta)$ in the considered Gaussian shift-in-mean problem. The unique positive solution of the equation $\Psi(\theta) = 0$ is $\theta^* = 2$. Then, the error exponent becomes

$$
\lim_{n \longrightarrow \infty} \frac{1}{\sqrt{n}} \log \beta_n = -2\sqrt{2D} \left[ d(1) - Q^{-1}(\alpha_n^{opt}) \right].
\tag{5.47}
$$

Fig. 5.6 shows the asymptotic performance of the system for the Gaussian shift-in mean problem with $\mu = 0.3$, $\sigma^2 = 1$, $\lim_{n \to \infty} \alpha_n^{opt} = 0.3$ and a linear drift function $d(t) = d(1)t$, with $d(1) = 0.6$.

The upper plot shows the behavior of the asymptotic false alarm probability versus the ratio $\rho = 1 - k/n$, which represents the fraction of data lost along the chain. Accordingly, at the node $k = n$, where all the data set is available, $\rho = 0$, while the node $k = 1$ can exploit only the observation $X_1$ and $\rho = 1$. We see that, for $n$ sufficiently large, the first node in the chain fulfills the requirements on the false alarm rate (0.3). The asymptotic value of the false alarm probability of the last node in the chain can be predicted by exploiting the convergence of the decision statistic to a Wiener process, under the hypothesis $\mathcal{H}_0$. When the drift is linear, the asymptotic false alarm probability of the VUB can be found by using the well-known result in [33]. In particular

$$
\begin{aligned}
\lim_{n \to \infty} \alpha_n &= 1 - Q(-d(1) + \gamma) \\
&\quad + \exp(2\delta(1)\gamma)[1 - Q(-d(1) - \gamma)]
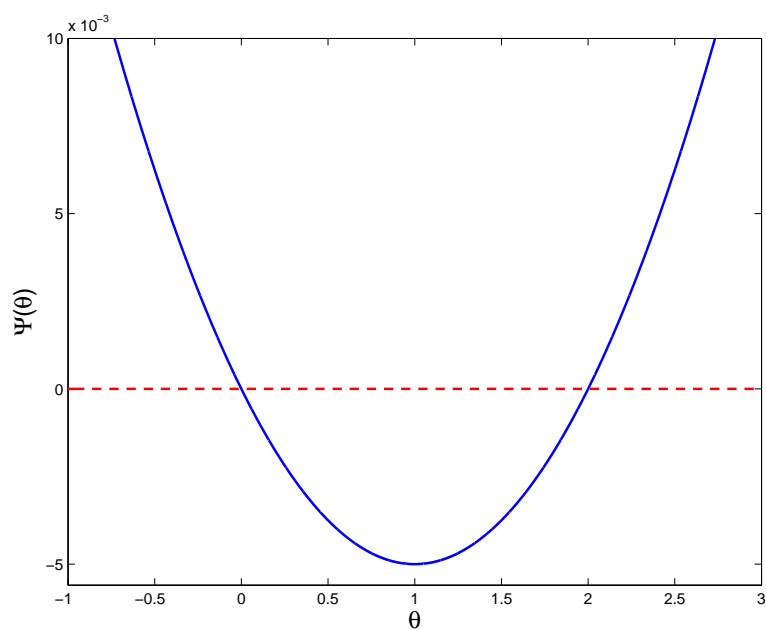\end{aligned}
\tag{5.48}
$$

**Figure 5.5** Semi-invariant moment generating function $\Psi(\theta)$ for the random variable $-l_i - D$ in the Gaussian shift-in-mean problem.
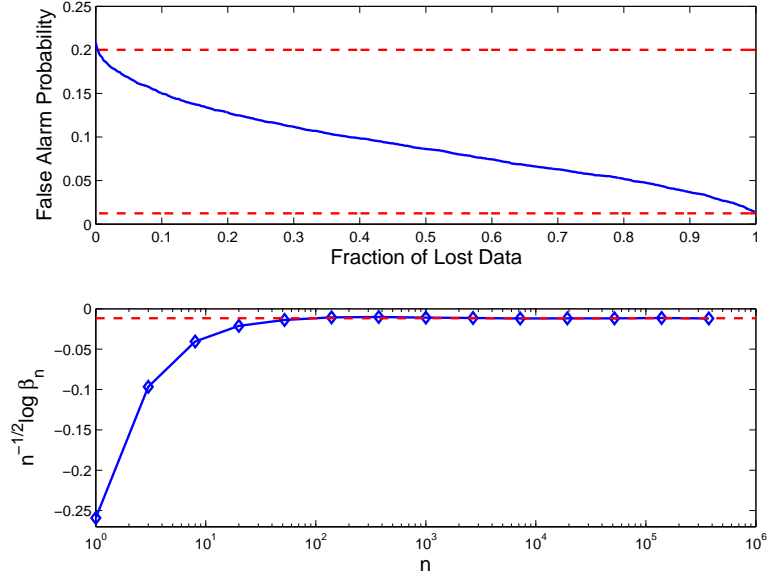
**Figure 5.6** The upper plot shows the asymptotic false alarm probability versus the fraction of lost data. The lower plot represents the asymptotic miss detection probability versus the number of nodes in the system.

The lower red dashed line in the upper plot of Fig. 5.6 represents the numerical value calculated by using (5.48).

Finally, the lower plot shows the behavior of the miss detection probability versus the number of sensors in the system. The dashed red line is the theoretical value of the Type-II error probability's exponent, which represents how fast the Type-II error probability decreases to zero.

## 5.4.2 Exponential shift-in-scale

Let us consider the following hypothesis test

$$
\begin{aligned}
\mathcal{H}_0 &: \quad X_i \sim \mathcal{E}(\lambda_0), \\
\mathcal{H}_1 &: \quad X_i \sim \mathcal{E}(\lambda_1),
\end{aligned}
\tag{5.49}
$$

where $i = 1, \ldots, n$ and $\mathcal{E}(\lambda)$ is our shortcut for an exponential pdf with expectation $1/\lambda$. Without loss of generality, we assume

$\lambda_0 > \lambda_1$. The corresponding log-likelihood ratio is

$$l(x_i) = \ln(\lambda_1/\lambda_0) + (\lambda_0 - \lambda_1)x_i. \tag{5.50}$$

If we define $\rho = \ln(\lambda_1/\lambda_0)$, $\Delta_h = \lambda_h/(\lambda_0 - \lambda_1)$, where $h = 0, 1$ denotes the hypotheses, the log-likelihood is distributed as $l(X_i) \sim \rho + \mathcal{E}(\Delta_h)$. Accordingly, $D_{01} = -\rho - 1/\Delta_0$ and $\sigma_{01} = 1/\Delta_0$. Under the hypothesis $\mathcal{H}_1$, we have that $-l(X_i) - D_{01} \sim 1/\Delta_0 - \mathcal{E}(\Delta_1)$. By recalling that the semi-invariant generating function of an exponential random variable with mean $1/\lambda$ is $\Psi(\theta) = \log(\lambda/(\lambda - \theta))$ with $\theta < \lambda$, it is straightforward to get that the semi-invariant generating function in (5.19) is

$$\begin{aligned} \Psi(\theta) &= \log\left[e^{\theta/\Delta_0}\frac{\Delta_1}{\Delta_1 + \theta}\right] \\ &= \frac{\theta}{\Delta_0} + \log\left[\frac{\Delta_1}{\Delta_1 + \theta}\right], \end{aligned} \tag{5.51}$$

with $\theta > \Delta_1$. Fig. 5.7 shows the behavior of the function $\Psi(\theta)$ for the Exponential shift-in-scale problem with $\lambda_0 = 3$, $\lambda_1 = 1$, $d(1) = 0.6$, $\lim_{n \to \infty} \alpha_n^{opt} = 0.3$, $d(t) = d(1)\ t$. The unique positive solution of the equation $\Psi(\theta) = 0$ is $\theta^* = 2.86$, which can be found by solving the equation $\Psi(\theta) = 0$ numerically. Fig. 5.8 shows the behavior of the false alarm probability versus the fraction $\rho = 1 - k/n$ of lost data (upper plot), and the behavior of the miss detection probability of the first node of the system, versus the number of sensors in the chain (lower plot).

## 5.5   Summary

In this chapter, with the unlucky broker problem of Chapter 4 in mind, we have considered a detection problem in a system made of multiple sensors connected in tandem. In particular, the first sensor observes a certain data set of, say, $n$ entries. Successive sensors in the chain can rely only on a fraction of the vector entries. Moreover, each sensor makes a decision on the state of the nature

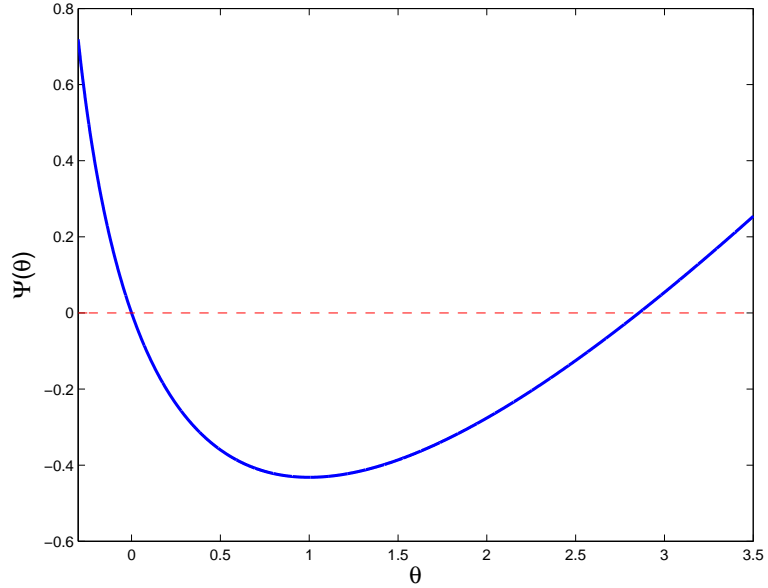**Figure 5.7** Semi-invariant moment generating function $\Psi(\theta)$ for the random variable $-l_i - D$ in the Exponential shift-in-scale problem.
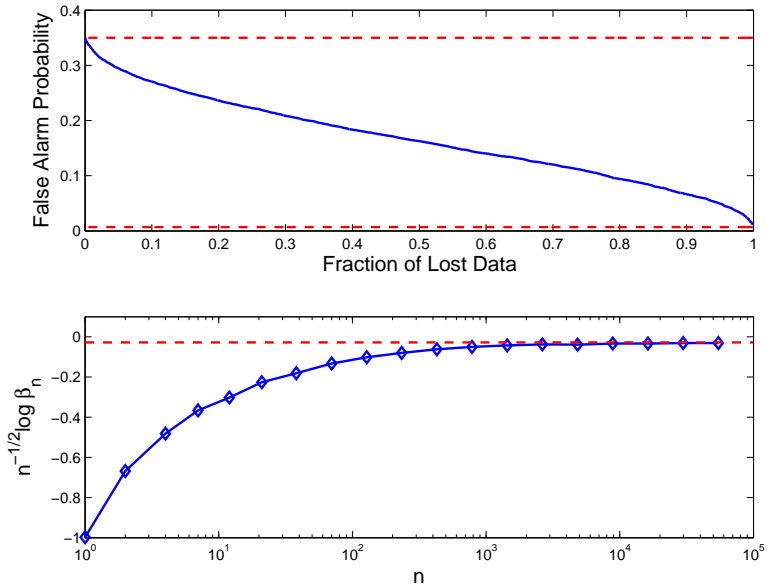


**Figure 5.8** The upper plot shows the asymptotic false alarm probability versus the fraction of lost data. The lower plot represents the asymptotic miss detection probability versus the number of nodes in the system.

by exploiting its own observations and the binary decision coming form the prior node in the system and by using a non optimal, but reasonable, decision strategy. We have studied the asymptotic detection performance of the last sensor, the one having access to the smallest data set, subject to a constraint on its false alarm probability. We have shown that the miss detection probability decays exponentially fast with the square root of the number of nodes in the system. We have furthermore calculated the closed-form expression of the miss detection probability's rate of decay.

# Chapter 6

# Conclusions

Wireless Sensor Networks have received worldwide attention in recent years, particularly with the proliferation in Micro-Electro-Mechanical Systems (MEMS) technology which has facilitated the development of the so-called *smart sensors*. These kinds of sensors are small, with limited processing and computing resources, and less expensive than the traditional ones. They are able to sense, measure and collect measurements from the surrounding environment and, based on some local decision process, they can deliver the sensed data either to an intender user or to a fusion center. Smart sensor nodes are low power devices equipped with one or more sensors, a processor, a memory, a power supply, a radio and an actuator  [1–3].

Wireless Sensor Networks consist of a number of sensor nodes (few tens to thousands) working together to monitor a region to obtain data about the environment. They have great potential for many applications in scenarios such as target tracking, surveillance, natural disaster relief, biomedical health monitoring, hazardous environment exploration, seismic sensing, and so forth. At the same time, the flexibility, fault tolerance, high sensing fidelity, low cost and rapid deployment characteristics of sensor networks, have created many new and exciting research activities. In particular, in Chapter 3 we have focused our interest on the research area of distributed detection of a global event in sensor networks.

There, we have reviewed the basic results on distributed detection with multiple sensors, with emphasis on the parallel [4–12] and serial [13–25] architectures, that are relevant for structured Wireless Sensor Networks, where all or some of the sensor nodes are deployed in a pre-planned manner.

In Chapters 4 and 5 we have proposed, analyzed and solved a novel topic in the frame of detection theory, called *the unlucky broker* problem [31,32], and its asymptotic extension, *the very unlucky broker* problem. The studies problems have a wide range of potential applications in economics, medicine, distributed inference in sensor networks, decision-making in emergency scenarios, and military surveillance for citizen safety.

In Chapter 4 we have addressed the main issues arising from the unlucky broker problem, that has been abstracted as follows. Consider a statistical test between two hypotheses, based on the data set $(\boldsymbol{x}, \boldsymbol{y})$, and leading to a binary decision $\delta$ obeying the Neyman-Pearson optimality criterion with a certain level of the false alarm probability. Suppose that, later, the data set available is made only of $\boldsymbol{x}$ and the problem is that of making a new, optimal decision, $\delta_B$, by exploiting the decision $\delta$ and the surviving data $\boldsymbol{x}$, at a certain false alarm probability. The main efforts have been in *(i)* deriving the optimal detector solving the problem and *(ii)* understanding its *modus operandi* or, in other words, the physical interpretations and implications underlying the optimal decision rule.

We have proved that, as intuition would suggest, if the decision $\delta_B$ is made at the same false alarm level as the original one, then the surviving data are irrelevant and retaining the decision $\delta$ is NP-optimal.

When a refinement is asked, i.e., the desired false alarm level for the final decision is different, some decisions can be retained, while others require a double-check. We have found that the sufficient statistic for the final decision is made of the available data set $\boldsymbol{x}$ and its log-likelihood ratio, namely $L(\boldsymbol{x})$. The surprisingly finding is that the aforementioned double-check *does not* amount to compare the log-likelihood $L(\boldsymbol{x})$ to a single threshold level, as

it usually happens in classic detection theory with the so-called *single threshold tests*. In general, we have showed that the optimal solution to the unlucky broker problem consists of checking whether or not the log-likelihood ratio $L(\boldsymbol{x})$ falls in some sets of the real axis having an intricate structure, not necessarily simply connected.

We have mainly addressed the problem in the NP framework, but the Bayesian paradigm has also been considered. In that case, the idea of the refinement could be linked either to an improvement in the knowledge of the priors or to the scenario in which the priors are held fixed and the refinement involves a different assignment of the costs. Both of these situations imply a different threshold setting and thus they can be addressed in the same way.

In Chapter 5 we have considered a detection problem in a system made of multiple sensors connected in series (or in tandem), referred to as the *Very Unlucky Broker* problem. In particular, the data available to the system are modeled as a vector of observations and the first sensor of the series has fully access to this vector. Successive sensors can only observe a fraction of the data, i.e., a subset of the vector entries. Each sensor makes its own local decision about the state of the nature by exploiting its own data set and the binary decision from the previous sensor. The final decision of the system is that of the last sensor of the chain and we have focused on characterizing the system behavior when the sensor chain was very long. It is straightforward to observe that the unlucky broker problem analyzed in Chapter 4 is essentially the very unlucky broker problem in presence of two sensors.

In Chapter 4 we have found that the optimal test is very difficult to be implemented even with only two nodes, because of the strong correlation between the available data and the decision coming from the previous node. Then, to make the designed detector implementable in practice and to ensure analytical tractability, we have considered non optimal, but reasonable, local tests. In this scenario, we have studied the asymptotic detection performance of the last sensor, the one having access to the smallest data set, subject to a constraint on its false alarm probability. We have

shown that the miss detection probability decays exponentially fast with the square root of the number of nodes in the system, thus exhibiting an anomalous scaling law. We have furthermore provided a closed-form expression for the rate of decay of the miss detection probability.

It is worth noting that the system architectures arising in the unlucky and in the very unlucky broker problems are the same of the well-studied serial detection schemes in presence of two or more detection stages, respectively [13–25]. The problems proposed in this dissertation can thus be linked and placed into the existing and broad literature of decentralized detection in tandem structures. At the same time, however, the problems studied in this thesis require rather different approaches and tools, as consequence of the inherent statistical dependence of the data available to the elementary units of the system. Therefore, while formulating and solving the unlucky broker and its asymptotic version is certainly the main contribution of this work, providing general guidelines and insights about detection problems based upon strongly dependent observations is, in our opinion, a second important contribution of the dissertation.

# Appendix A

## A.1 Proof of Proposition 4.4.1

The proof requires only some algebra. Let us prove the claim for $t_1(l_x)$, that for $t_0(l_x)$ being afforded similarly. It is expedient to work in terms of

$$h(l_x) = e^{t_1(l_x)} = \frac{P_{dy}(\gamma - l_x)}{P_{fy}(\gamma - l_x)} \, e^{l_x},$$

which, in view of eqs. (4.22), (4.23) and (4.24), becomes

$$h(l_x) = \frac{\sum_{k=0}^{N_y-1} \frac{\Delta_1^k \, (\gamma - \rho_y - l_x)^k}{k!} \, e^{-\Delta_1 \, (\gamma - \rho_y - l_x)}}{\sum_{k=0}^{N_y-1} \frac{\Delta_0^k \, (\gamma - \rho_y - l_x)^k}{k!} \, e^{-\Delta_0 \, (\gamma - \rho_y - l_x)}} \, e^{l_x}, \qquad (A.1)$$

for $l_x \leq \gamma - \rho_y$, while $h(l_x) = e^{l_x}$, $\forall l_x > \gamma - \rho_y$. Being $h(\gamma - \rho_y) = e^{\gamma - \rho_y}$, by continuity arguments, it will thus suffice to show that the function is strictly increasing in the range $l_x < \gamma - \rho_y$. To this aim, let us set $z = \gamma - \rho_y - l_x$, and focus on the range $z > 0$. With a slight abuse of notation, we can write $h(z) \propto \sum_{k=0}^{N_y-1} \frac{\Delta_1^k z^k}{k!} / \sum_{k=0}^{N_y-1} \frac{(1+\Delta_1)^k z^k}{k!}$, where we used $\Delta_0 = 1 + \Delta_1$, and the proportionality factor is positive. In view of the definition of $z$, we must show that the above $h(z)$ is strictly decreasing. The

first derivative, $h'(z)$, is proportional (with a positive factor) to

$$\sum_{k=0}^{N_y-1} k \frac{\Delta_1^k z^{k-1}}{k!} \sum_{m=0}^{N_y-1} (1+\Delta_1)^m \frac{z^m}{m!}$$

$$-\sum_{m=0}^{N_y-1} m \frac{(1+\Delta_1)^m z^{m-1}}{m!} \sum_{k=0}^{N_y-1} \Delta_1^k \frac{z^k}{k!},$$

$$= \sum_{k,m=0}^{N_y-1} \frac{z^{k+m-1}}{k!m!} \Delta_1^k (1+\Delta_1)^m (k-m)$$

$$= \sum_{k=0}^{N_y-1} \sum_{m<k} \frac{z^{k+m-1}}{k!m!} \left[ \Delta_1^k (1+\Delta_1)^m - \Delta_1^m (1+\Delta_1)^k \right] (k-m).$$

The term in brackets is equivalent to

$$(1+\Delta_1)^k (1+\Delta_1)^m \left[ \left( \frac{\Delta_1}{1+\Delta_1} \right)^k - \left( \frac{\Delta_1}{1+\Delta_1} \right)^m \right],$$

which is negative since $\Delta_1 > 0$ and $m < k$. This completes the proof.

As to the properties detailed in eqs. (4.25), $\lim_{l_x \to +\infty} t_1(l_x) = +\infty$ trivially follows by observing that $t_1(l_x) = l_x$ for $l_x \geq \gamma - \rho_y$. As to the right limit of $t_0(l_x)$, we can write

$$\lim_{l_x \to \gamma-\rho_y} t_0(l_x) = \gamma - \rho_y + \lim_{z \to 0^+} \ln \frac{1 - \sum_{k=0}^{N_y-1} \frac{(\Delta_1 z)^k}{k!} e^{-\Delta_1 z}}{1 - \sum_{k=0}^{N_y-1} \frac{(\Delta_0 z)^k}{k!} e^{-\Delta_0 z}}.$$

Applying the l'Hopital's rule we obtain

$$\lim_{l_x \to \gamma-\rho_y} t_0(l_x) = \gamma - \rho_y + \lim_{z \to 0^+} \ln \frac{\Delta_1^{N_y} z^{N_y-1}}{\Delta_0^{N_y} z^{N_y-1}},$$

where the last term on the right hand side exactly computes to $\rho_y$ and $\lim_{l_x \to \gamma-\rho_y} t_0(l_x) = \gamma$. The inequalities at the left extreme

$l_x = \rho_x$ can be verified as follows. The relationship $t_1(\rho_x) \geq \gamma$ is already estabilished from eq. (4.11). To show the strict inequality, we write

$$
\begin{aligned}
\lim_{l_x \to \rho_x} t_1(l_x) &= \rho_x + \ln \left. \frac{\sum_{k=0}^{N_y-1} \frac{(\Delta_1 z)^k}{k!} e^{-\Delta_1 z}}{\sum_{k=0}^{N_y-1} \frac{(\Delta_0 z)^k}{k!} e^{-\Delta_0 z}} \right|_{z = \gamma - \rho_x - \rho_y} \\
&= \rho_x + \gamma - \rho_x - \rho_y + \ln \left. \frac{\sum_{k=0}^{N_y-1} \frac{(\Delta_1 z)^k}{k!}}{\sum_{k=0}^{N_y-1} \frac{(\Delta_0 z)^k}{k!}} \right|_{z = \gamma - \rho_x - \rho_y},
\end{aligned}
$$

where we have used $\Delta_0 = \Delta_1 + 1$. Consider now the argument of the logarithm, which is in the form $\sum_k a_k / \sum_k b_k$, where $a_k$, $b_k$ are positive numbers. By exploiting the fact that $\sum_k a_k / \sum_k b_k > c$ provided that $a_k/b_k > c$, $\forall k$, and observing that $a_k/b_k = (\Delta_1/\Delta_0)^k$ with $\Delta_1/\Delta_0 < 1$, we get $a_k/b_k > (\Delta_1/\Delta_0)^{N_y}$, for all $k < N_y$. Consequently, the argument of the logarithm is strictly greater than $(\Delta_1/\Delta_0)^{N_y}$, namely the logarithm is strictly greater than $\rho_y = N_y \ln(\Delta_1/\Delta_0)$, and the desired result follows.

Finally, for the left limit of $t_0(l_x)$ we have

$$
\lim_{l_x \to \rho_x} t_0(l_x) = \rho_x + \ln \left. \frac{1 - \sum_{k=0}^{N_y-1} \frac{(\Delta_1 z)^k}{k!} e^{-\Delta_1 z}}{1 - \sum_{k=0}^{N_y-1} \frac{(\Delta_0 z)^k}{k!} e^{-\Delta_0 z}} \right|_{z = \gamma - \rho_x - \rho_y}.
$$

The argument of the natural logarithm is the ratio of the $L_y(\boldsymbol{Y})$'s cdfs under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively, evaluated in $\gamma - \rho_x - \rho_y$. This ratio can be made arbitrarily small as the scale parameter $\Delta_1$ can be made arbitrarily smaller than $\Delta_0$ by a suitable choice of $\lambda_0$ and $\lambda_1$. Consequently, there is no finite lower bound for $t_0(l_x)$.

## A.2  Proof of Proposition 4.4.2

Suppose that $\gamma_{\mathcal{B}} > \gamma$. In view of (4.21), the pdf of $L_x$ is

$$
f_{L_x}(l_x; \mathcal{H}_1) = \frac{\Delta_1^{N_x}(l_x - \rho_x)^{N_x-1}}{(N_x - 1)!} e^{-\Delta_1(l_x - \rho_x)} u(l_x - \rho_x),
$$

From Proposition 1, we know that $t_1(l_x)$ is strictly increasing, such that the condition $t_1(l_x) < \gamma_{\mathcal{B}}$ maps into $l_x < s_1$, with $s_1$

defined in (4.14). From (4.16), we therefore get $P_{d,\mathcal{B}} = P_d - \int_{\rho_x}^{s_1} P_{dy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_1) dl_x$. Now, assume first $s_1 \leq \gamma - \rho_y$. From (4.24) and using the above expression

$$P_{d,\mathcal{B}} = P_d - \frac{\Delta_1^{N_x} e^{-\Delta_1(\gamma - \rho_x - \rho_y)}}{(N_x - 1)!}$$
$$\times \sum_{k=0}^{N_y-1} \frac{\Delta_1^k}{k!} \int_{\rho_x}^{s_1} (\gamma - \rho_y - l_x)^k (l_x - \rho_x)^{N_x - 1} dl_x.$$

With the change of variable $z = l_x - \rho_x$, and using the binomial theorem,

$$P_{d,\mathcal{B}} = P_d - \frac{\Delta_1^{N_x} e^{-\Delta_1(\gamma - \rho_x - \rho_y)}}{(N_x - 1)!}$$
$$\times \sum_{k=0}^{N_y-1} \frac{\Delta_1^k}{k!} \sum_{i=0}^{k} \binom{k}{i} b^{k-i} (-1)^i \int_0^{s_1 - \rho_x} z^{i + N_x - 1} dz,$$

where $b = \gamma - \rho_x - \rho_y$. The last integral can be computed in closed form yielding the desired result $P_{d,\mathcal{B}} = P_d - I(\Delta_1; s_1)$, where $I(\Delta; s)$ is defined in (4.26).

Let us now switch to the case $s_1 > \gamma - \rho_y$. From (4.24), we have $P_{dy}(\gamma - l_x) = 1$, such that

$$P_{d,\mathcal{B}} = P_d - \int_{\rho_x}^{\infty} P_{dy}(\gamma - l_x) f_{L_x}(l_x; \mathcal{H}_1) dl_x$$
$$+ \int_{s_1}^{\infty} f_{L_x}(l_x; \mathcal{H}_1) dl_x = \Pr\{\rho_x + \Gamma(N_x, \Delta_1) > s_1\},$$

where the first integral equals $P_d$ by definition. This completes the proof of eq. (4.27). The proof of eq. (4.28) is straightforwardly obtained once that $\Delta_1$ is replaced by $\Delta_0$. The proof of eqs. (4.29) are obtained by reasoning in a similar fashion, and are omitted for reasons of space.

# A.3   Proof of Proposition 4.4.3

Let us introduce the function $r_1(z) = g(z)/Q(z) - z$, where $g(\cdot)$ is the pdf of a standard Gaussian random variable. The above function is clearly always positive on the negative axis, while, from a well-known property, $Q(z) < g(z)/z$, $\forall z \geq 0$. Thus $r_1(z) > 0$ $\forall z$. Now, let us consider $t_1(l_x)$ given in (4.33), and write $d$ for $d_y$. Setting $x = \gamma - l_x$, yields

$$h(x) = t_1(\gamma - x) = \ln \frac{Q\left(\frac{x-d}{\sqrt{2\,d}}\right)}{Q\left(\frac{x+d}{\sqrt{2\,d}}\right)} + \gamma - x,$$

and we now prove that $h(x)$ is strictly decreasing. Taking the first derivative of $h(x)$, we get

$$h'(x) = \frac{g\left(\frac{x+d}{\sqrt{2\,d}}\right)}{\sqrt{2\,d}\,Q\left(\frac{x+d}{\sqrt{2\,d}}\right)} - \frac{g\left(\frac{x-d}{\sqrt{2\,d}}\right)}{\sqrt{2\,d}\,Q\left(\frac{x-d}{\sqrt{2\,d}}\right)} - 1.$$

Setting $z = (x + d)/\sqrt{2\,d}$, the above can be rewritten, with a slight abuse of notation, as $h'(z) = [r_1(z) - r_1(z - \sqrt{2\,d})]/\sqrt{2\,d}$. Proving that $h'(z) < 0$ will be thus implied by proving that $r_1(z)$ is strictly decreasing. To this end, we compute the derivative: $r_1'(z) = r_2/Q^2(z)$, where $r_2(z) = g^2(z) - g(z)Q(z)z - Q^2(z)$, and we must show that this latter is negative. Since $\lim_{z \to -\infty} r_2(z) = -1$ and $\lim_{z \to +\infty} r_2(z) = 0$, invoking continuity, it is enough to show that $r_2(z)$ is strictly increasing. We have $r_2'(z) = g(z)r_3(z)$, where $r_3(z) = Q(z)(1 + z^2) - z\,g(z)$. Now $r_3(z) > 0$ if $r_3(z)$ is strictly decreasing, because $\lim_{z \to -\infty} r_3(z) = +\infty$ and $\lim_{z \to +\infty} r_3(z) = 0$. Computing the derivative yields $r_3'(z) = -2\,Q(z)r_1(z)$ which is indeed negative, since $r_1(z)$ is positive. This completes the proof of the strict monotonicity for $t_1(l_x)$; that of $t_0(l_x)$ is simply obtained by similar derivations. As to the limiting behavior of $t_1(l_x)$, it is obvious that $\lim_{l_x \to +\infty} t_1(l_x) = +\infty$, while a simple application of l'Hôpital's rule gives $\lim_{l_x \to -\infty} t_1(l_x) = \gamma$. By symmetry, the results for $t_0(l_x)$ also follow.

# Appendix B

## B.1  The optimal VUB

Let us consider the notional scheme in Fig. 5.4. At the sensor $n$, the optimal detection statistic can be written as

$$T_n(x_1, \ldots, x_n) = T_n(\boldsymbol{x}, y) = L_x(\boldsymbol{x}) + L_y(y), \qquad \text{(B.1)}$$

because the observations are i.i.d.; moreover, we put $y = x_n$ and $\boldsymbol{x} = x_1, \ldots, x_{n-1}$. We denote the optimal decision for sensor $n$ as $b_n(\boldsymbol{x}, y)$. At node number $n-1$, the detection statistic depends on the observations $\boldsymbol{x} = x_1, \ldots, x_{n-1}$ and the decision $b_n(\boldsymbol{x}, y)$, that is

$$T_{n-1}(\boldsymbol{x}, b_n(\boldsymbol{x}, y)), \qquad \text{(B.2)}$$

where $\boldsymbol{x}$ and $y$ are independent observations. Accordingly, by exploiting the well-known results on the unlucky broker problem [31] showed in Chapter 4, it is straightforward to obtain the optimal decision strategy. We know that if we require the false alarm probability to be creasing, we have to retain all the decisions $b_n$ in favor of $\mathcal{H}_0$. All the decisions in favor of $\mathcal{H}_1$ are double-checked by comparing e statistic $T_{n-1}(\boldsymbol{x}, b_n(\boldsymbol{x}, y))$ to an appropriate threshold. The optimal decision at sensor $n-1$ is $b_{n-1}(\boldsymbol{x}, y)$.

It is interesting now to figure out what happens at stage $n-2$, where the detection statistic is

$$T_{n-2}(\hat{\boldsymbol{x}}, b_{n-1}(\boldsymbol{x}, b_n(\boldsymbol{x}, y))) = T_{n-2}(\hat{\boldsymbol{x}}, b_{n-1}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})), \qquad \text{(B.3)}$$

where we put $\hat{\boldsymbol{x}} = x_1, \ldots, x_{n-2}$ and $\hat{\boldsymbol{y}} = x_{n-1}, b_n(\boldsymbol{x}, y)$. The statistic in (B.3) is of the same type of that (B.2),but for the observations $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ are correlated. In Section 4.6 we have shown that also in this se, if we want a decrease of the false alarm rate, we have to retain all the decisions $b_{n-1} = 0$. When $b_{n-1} = 1$, the optimal strategy consists of comparing the statistic $T_{n-2}$ to an appropriate threshold. The reasoning in the above can be applied to every stage of the system. Accordingly, if we want a decrease of the false alarm probability along the chain, we have to retain all the decisions in favor of the null hypothesis, while in presence of a decision in favor of $\mathcal{H}_1$ it is necessary to compare a very complicated statistic of the current data and the previous decision to an opportune threshold.

# Bibliography

[1] F. L. Lewis, "Wireless Sensor Networks," *Smart Environments: Technologies, Protocols, and Applications*, ch. 4 ed. D.J. Cook and S.K. Das, John Wiley, New York, 2004

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no.8 pp. 104–112, 2002.

[3] J. Yick, B. Mukherjee, D. Ghosal, "Wireless Sensor Network Survey," Computer Networks, vol. 52, pp. 2292–2330, 2008.

[4] R. R. Tenney and J. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, pp. 501–510, Jul. 1981.

[5] Z. Chair and P. K. Varshney, "Optimal Data Fusion in multiple sensor detection system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 98–101, Jan. 1986.

[6] J. N. Tsitsiklis, "Decentralized Detection," *Advances in Statistical Signal Processing, Signal Detection* , vol. 2, pp. 297–344, 1993.

[7] J. N. Tsitsiklis, "Decentralized Detection by a large number of sensors," *Mathematics of Control, Signal, and Systems*, pp. 167–182, 1988.

[8] S. C. A. Thomopoulos, R . Viswanathan, and D. K. Bougoulias, "Optimal decision fusion in multiple sensor systems,"

*IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 644–653, Sep. 1987.

[9] S. C. A. Thomopoulos, R . Viswanathan, and D. K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep. 1989.

[10] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 24–30, Jan. 1987.

[11] F. A. Sadjadi, "Hypotheses testing in a distributed environment ," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 134–137, Mar. 1986.

[12] K. L. Boettcher and R. R. Tenney, "Distributed Decision-making with Constrained Decision-Makers ," *IEEE Trans. Systems, Man, Cybernetics*, vol. 16, No.6. 1986.

[13] L. K. Ekchian and R. R. Tenney, "Detection networks," in *21th IEEE Conference on Decision and Control*, Orlando, 1982, pp. 686–691.

[14] J. D. Papastavrou, "Distributed detection with selective communications," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 1986.

[15] R. Viswanathan, S. Thomopoulus, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 366–376 510, Jul. 1988.

[16] J. Pothiawala, "Analysis of a two-sensor tandem distributed detection network," Master thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1989.

[17] L. K. Ekchian, "Optimal design of distributed detection networks," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 1982.

[18] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, Jul. 1992.

[19] P. E. Swaszek, "On the Performance of Serial Networks in Distributed Detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, Jan. 1993.

[20] T. M. Cover, "Hypothesis testing with finite statistics," *Annals of Mathematical Statistics.*, vol. 40, pp. 828–835, 1969.

[21] M. E. Hellmann, T. M. Cover, "Learning with finite memory," *Annals of Mathematical Statistics.*, vol. 41, pp. 765–782, 1970.

[22] C. T. Mullis, R. A. Roberts, "Memory limitation and multistage decision processes," *IEEE Trans. Sys. Sc. and Cyb.*, vol. 4, pp. 307–316, Sept. 1968.

[23] W. P. Tay, J. N. Tsitsiklis, M. Z. Win "On the Subexponential Decay of Detection Error Probabilities in Long Tandems," *IEEE Trans. Inf. Th.*, vol. 54, no. 10, pp. 4767–4771, Oct. 2008

[24] W. P. Tay, "Decentralized detection in resource-limited sensor network architectures," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 2008.

[25] Z. B. Tang, K. R. Pattipati, D. L. Kleinman, "Optimization of Detection Networks: Part I: Tandem Structures," *IEEE Trans. Systems, Man., and Cybernetics* , vol. 21, no. 5, pp. 1044–1059, Sep/Oct. 1991.

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

[27] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II-Advanced Topics," *Proceeding of the IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[28] S. Kar, R. Tandon, H. V. Poor, and S. Cui "Distributed detection in Noisy Sensor Networks," *Proceeding of the IEEE International Symposium on Information Theory*, pp. 2782–2786, 2011.

[29] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Trans. Sig. Proc.*, vol. 58, pp. 814–825, Feb. 2010.

[30] R. Olfati-Saber, A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, pp. 215–233, Jan. 2007.

[31] S. Marano, V. Matta, and F. Mazzarella, "Refining decisions after losing data: The unlucky broker problem," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 1980–1990, Apr. 2010.

[32] S. Marano, V. Matta, and F. Mazzarella, "The Bayesian Unlucky Broker," *Proc. of 18-th European Signal Processing Conference (EUSIPCO)*, 2010 , pp. 154–158.

[33] C. Park and S. R. Paranjape, "Probabilities of Wiener paths crossing differentiable curves," *Pacific Journal of Mathematics*, vol. 53, pp. 579–583, 1974.

[34] C. Park and F. J. Schuurmann, "Evaluation of Barrier-Crossing Probabilities of Wiener Paths," *J. Appl. Prob*, vol. 13, pp. 267–275, 1976.

[35] P. K. Varshney, *Distributed Detection and Data Fusion.* New York: Springer-Verlag, 1996.

[36] H. V. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 1988.

[37] H. L. Van Trees, *Detection Estimation and Modulation Theory*, Vol I. New York: Wiley, 1968.

[38] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Vol. II: Detection Theory Prentice Hall, 1998.

[39] C. W. Hesltrom, *Elements of Signal Detection and Estimation* Prentice Hall, 1995.

[40] S. A. Kassam, *Signal Detection in Non-Gaussian Noise.* Springer-Verlag, 1987.

[41] L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis.* Addison-Wesley Publishing Company, 1991.

[42] T. Cover and J. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, 1991.

[43] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications.* New York: Springer-Verlag, 1998.

[44] S. Kullback, *Information Theory and Statistics.* New York: Dover, 1968.

[45] R. Blahut, *Principles and Practice of Information Theory.* Addison-Wesley, 1987.

[46] P. Billingsley, *Convergence of Probability Measures*, 2nd ed. New York: Wiley-Interscience, 1999.

[47] R. G. Gallager, *Discrete Stochastic Processes.* Kluwer Academic Publisher, 1996.

[48] A. Gut, *Stopped Random Walks*, 2nd ed. New York, USA: Springer, 2009.

[49] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[50] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1.* New York: John Wiley & Sons, 1994.

[51] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* New York: Dover, 1970.