

# Chapter 2

## Background

### 2.1 Introduction

This chapter provides background information for both robust and distributed detection, and this underpins the development of theory presented in subsequent chapters. Robust detection is discussed under two sub-sections: minimax hypothesis testing in Sect. 2.2.1, which provides a basis for Chap. 7, and robust hypothesis testing in Sect. 2.2.2, which underpins Chaps. 3–5. In Sect. 2.3, decentralized detection is introduced and this material provides the background for Chaps. 6–7. In this section some fundamental results from an optimal design procedure are noted. The conclusions of this chapter are detailed in Sect. 2.4.

### 2.2 Robust Detection

Robust detection refers to the detection of events with a guaranteed level of detection performance despite the uncertainties imposed on the nominal statistical model. There are mainly two sources of uncertainties in hypothesis testing. The first source is the a priori probabilities (priors) of the hypotheses,  $P(\mathcal{H}_0)$  and  $P(\mathcal{H}_1)$ , whereas the second source is the probability distributions under each hypothesis. The rationale behind both types of uncertainties is that in reality neither the priors nor the probability distributions can be known exactly. In the following two sections, minimax hypothesis testing and robust hypothesis testing, which provide minimax solutions for the aforementioned two sources of uncertainties, are introduced.

### 2.2.1 Minimax Hypothesis Testing

Bayesian formulation of hypothesis testing assumes that the a priori probabilities of each hypothesis,  $P(\mathcal{H}_0)$  and  $P(\mathcal{H}_1)$ , are known. Depending on the application, this assumption may or may not hold. Digital communication is an example, where this assumption is expected to hold with high probability, because the transmitted bits, zeros and ones, are often equally likely. On the other hand, there are a vast number of applications, for example, radar or cognitive radio, where this assumption does not hold. For such applications, either the priors can be guessed or they can be determined such that the Bayes' risk, e.g. the error probability, is minimized for the least favorable priors. This latter strategy of determining the priors and accordingly the decision rule is called minimax detection and has the advantage of guaranteeing a minimum level of test performance independent of the actual values of priors. In the sequel, minimax hypothesis testing will be explained with the following example.

Let  $(\Omega, \mathcal{A})$  be a measurable space with two distinct probability measures  $P_0$  and  $P_1$  defined on it. Furthermore, let  $p_0$  and  $p_1$  be the probability density functions of  $P_0$  and  $P_1$ , respectively. Consider the following binary hypothesis testing problem:

$$\begin{aligned}\mathcal{H}_0 : Y &\sim P_0 \\ \mathcal{H}_1 : Y &\sim P_1\end{aligned}\tag{2.1}$$

where  $Y : \Omega \mapsto \Omega$  is a random variable (r.v.) which is distributed as  $P_j$  when  $\mathcal{H}_j$  is true,  $j \in \{0, 1\}$ . Without loss of generality,  $\Omega$  can be any interval of real numbers. To decide for the hypothesis  $\mathcal{H}_j$ , given an observation  $y \in \Omega$ , let  $\phi : \Omega \mapsto \{0, 1\}$  be a decision rule, which separates  $\Omega$  into two non-overlapping sets. Furthermore, let  $\pi_0 = P(\mathcal{H}_0)$  and  $\pi_1 = P(\mathcal{H}_1)$  be the a priori probabilities, and  $C_{ij}$ ,  $i, j \in \{0, 1\}$  be the costs of making a decision  $i$  when hypothesis  $j$  is true. Given the a priori probabilities and the costs, the Bayesian risk is defined as

$$R(\phi, \pi_0) = \sum_{i,j} \pi_j C_{ij} P[\phi(Y) = i | \mathcal{H}_j].$$

The overall error probability is then obtained from the Bayesian risk by setting the costs of detection probabilities to zero,  $C_{00} = C_{11} = 0$  and the costs of error probabilities to one  $C_{10} = C_{01} = 1$ , i.e.

$$P_E(\phi, \pi_0) = \pi_0 P_F(\phi, P_0) + (1 - \pi_0) P_M(\phi, P_1),\tag{2.2}$$

where  $P_F(\phi, P_0) = P[\phi = 1 | \mathcal{H}_0]$  is the false alarm probability and  $P_M(\phi, P_1) = P[\phi = 0 | \mathcal{H}_1]$  is the miss detection probability. As the hypotheses are simple, and the costs are assumed given, it follows that  $R$  and  $P_E$  are not explicit functions of  $P_0$ ,  $P_1$  and  $C_{ij}$ .

The error probability in (2.2) can be optimized based on two different assumptions, one corresponding to the case when the a priori probabilities can be assigned and the

other when they cannot be assigned. For the former case, let the a priori probabilities be equal, i.e.  $\pi_0 = \pi_1 = 1/2$ . This corresponds to the average error probability and is common for several applications. For this case, the optimum decision rule  $\phi_0$  is a unique solution to

$$\min_{\phi} \frac{1}{2} (P_F(\phi, P_0) + P_M(\phi, P_1)). \quad (2.3)$$

For the latter case, the minimax decision rule  $\phi_r$ , corresponding to the a priori probability  $\pi_{0r}$ , is obtained by solving

$$\max_{\pi_0} \min_{\phi} P_E(\phi, \pi_0) = \min_{\phi} \max_{\pi_0} P_E(\phi, \pi_0). \quad (2.4)$$

The equality in (2.4) (which is  $\leq$  in general) is established, e.g. in [Lev08, pp. 39–42] by considering the properties of  $\pi_0$  and  $\phi$ . This result proves the existence of a saddle value condition

$$P_E(\phi_r, \pi_0) \leq P_E(\phi_r, \pi_{0r}) \leq P_E(\phi, \pi_{0r}),$$

which indicates a guaranteed power of the test despite the uncertainty imposed by  $\pi_0$ . The price paid for this property is an almost surely reduced level of performance compared to the case when  $\pi_0$  is known. Changing the order of minimization or maximization does not play any role in (2.4), hence, performing the maximization first leads to  $P_M(\phi, P_1) = P_F(\phi, P_0)$  and the reformulation of the problem as follows:

$$\phi_r = \arg \min_{\phi} P_M(\phi, P_1) \quad \text{s.t.} \quad P_M(\phi, P_1) = P_F(\phi, P_0). \quad (2.5)$$

A nice property of (2.5) is that as long as  $P_M = P_F$  is satisfied for some decision rule  $\phi$ , not necessarily for  $\phi_r$ , and  $P_E < 1/2$ , the error probability is bounded from above with respect to the variations on  $\pi_0$ , see (2.2) for  $P_M = P_F$ .

*Example 2.2.1* Consider the binary hypothesis testing problem

$$\mathbf{Y} \sim \begin{cases} P_0 = \mathcal{N}(\bar{\mu}_0 \mathbf{1}, \sigma^2 \mathbf{I}), & \text{under } \mathcal{H}_0 \\ P_1 = \mathcal{N}(\bar{\mu}_1 \mathbf{1}, \sigma^2 \mathbf{I}), & \text{under } \mathcal{H}_1 \end{cases}$$

where the random variable  $\mathbf{Y} = (Y_1, \dots, Y_n)$  is distributed as multivariate Gaussian  $\mathcal{N}$  with mean vector  $\bar{\mu}_j \mathbf{1}$  under hypothesis  $j$  and covariance matrix  $\sigma^2 \mathbf{I}$ , where  $\mathbf{1}$  is the vector of ones and  $\mathbf{I}$  is the identity matrix of dimension  $n \times n$ . Assume that the random variables  $Y_i$  are independent and identically distributed (i.i.d.) and  $\bar{\mu}_1 > \bar{\mu}_0$ . Then,

$$l(\mathbf{Y}) = \frac{p_1}{p_0}(\mathbf{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} t \implies T(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n Y_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} t',$$

where  $l$  is called the likelihood ratio function,  $T$  is the test statistic, and  $t, t' \in \mathbb{R}$  are thresholds with a certain bijective mapping  $t \mapsto t'$ . The distribution of the test

statistic is again Gaussian distributed

$$T(\mathbf{Y}) \sim \begin{cases} \mathcal{N}(\bar{\mu}_0, \sigma^2/n), & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\bar{\mu}_1, \sigma^2/n), & \text{under } \mathcal{H}_1 \end{cases}$$

False alarm and miss detection probabilities are then

$$P_F(\phi(t'), P_0) = P[T(\mathbf{Y}) > t' | \mathcal{H}_0] = 1 - F\left(\frac{t' - \bar{\mu}_0}{\sigma/\sqrt{n}}\right) \quad (2.6)$$

and

$$P_M(\phi(t'), P_1) = P[T(\mathbf{Y}) \leq t' | \mathcal{H}_1] = F\left(\frac{t' - \bar{\mu}_1}{\sigma/\sqrt{n}}\right) \quad (2.7)$$

where  $F$  is the standard Gaussian cumulative distribution function (c.d.f.) and  $\phi$  is the threshold test. The threshold  $t'$  can be singled out from (2.6) as

$$t' = \frac{\sigma}{\sqrt{n}} F^{-1}(1 - P_F) + \bar{\mu}_0 \quad (2.8)$$

where  $(\cdot)^{-1}$  stands for the generalized inverse function. Let  $d = \bar{\mu}_1 - \bar{\mu}_0$ . Then, plugging (2.8) into (2.7) results in

$$P_M(P_F, n, \sigma, d) = F\left(F^{-1}(1 - P_F) - \frac{d}{\sigma/\sqrt{n}}\right). \quad (2.9)$$

Hence, the error probability (2.2) becomes

$$P_E(\pi_0, P_F; n, \sigma, d) = \pi_0 P_F + (1 - \pi_0) P_M(P_F, n, \sigma, d).$$

For every  $P_F$ , the corresponding  $P_M$  given by (2.9) is minimum and the minimum of  $P_M$  is achievable without randomization as long as  $P[l(Y) = t | \mathcal{H}_j] = 0$  for all  $t \in \mathbb{R}_{\geq 0}$  and  $j \in \{0, 1\}$ , see, e.g. [Tsi93]. This follows from the Neyman-Pearson lemma and suggests that for every fixed  $\pi_0$ , minimization over  $\phi$  in (2.4) can be performed over  $(P_F, P_M)$  in (2.9). Accordingly,

$$\frac{\partial P_E(\pi_0, P_F; n, \sigma, d)}{\partial P_F} = 0 \implies P_F^*(\pi_0; \cdot) = F\left(\frac{-d^2 n + 2\sigma^2 \ln\left(\frac{1-\pi_0}{\pi_0}\right)}{2d\sqrt{n}\sigma}\right)$$

and thus,

$$P_E^*(\pi_0; n, \sigma, d) = \pi_0 P_F^*(\pi_0; \cdot) + (1 - \pi_0) F \left( F^{-1} (1 - P_F^*(\pi_0; \cdot)) - \frac{\sqrt{nd}}{\sigma} \right). \quad (2.10)$$

Similarly, the maximization step yields

$$\frac{\partial P_E^*(\pi_0; n, \sigma, d)}{\partial \pi_0} = 0 \implies \pi_0^* = 1/2, \quad \forall n, \sigma, d.$$

Inserting  $\pi_0^*$  back into (2.10) gives

$$P_E^{**}(n, \sigma, d) = \frac{1}{2} \left( P_F^*(1/2; \cdot) + F \left( F^{-1} (1 - P_F^*(1/2; \cdot)) - \frac{\sqrt{nd}}{\sigma} \right) \right). \quad (2.11)$$

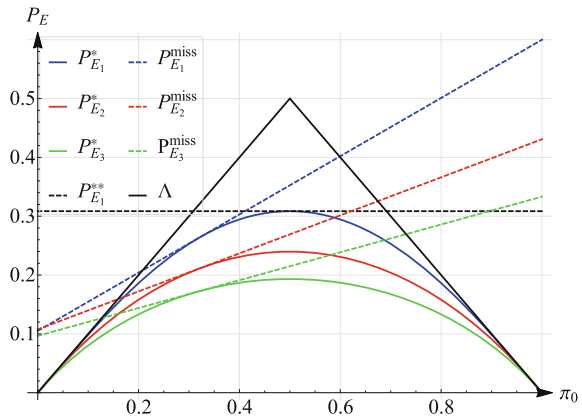
The minimax solution (2.11) can be compared to two cases one of which corresponds to the known a priori probability and the other to the mismatch case, where error minimizing decision rule is chosen for some probably wrong a priori probability  $\pi_0^{\text{miss}}$ . In the latter case, the error probability

$$P_E^{\text{miss}}(\pi_0; \pi_0^{\text{miss}}, \cdot) = \pi_0 P_F^*(\pi_0^{\text{miss}}; \cdot) + (1 - \pi_0) F \left( F^{-1} (1 - P_F^*(\pi_0^{\text{miss}}; \cdot)) - \frac{\sqrt{nd}}{\sigma} \right)$$

is linear in  $\pi_0$ , because the other terms are fixed. Let the parameters be chosen as  $d = 1$ ,  $\sigma = 1$ ,  $n \in \{1, 2, 3\}$  and  $\pi_0^{\text{miss}} \approx 0.32$  which corresponds to solving  $P_E^{\text{miss}}(1; \pi_0^{\text{miss}}, 1, 1, 1) = 0.6$ . For this setup, Fig. 2.1 illustrates the minimum error probability curves, mismatch lines, the minimax solution and the theoretical bound for the minimum error probability, which is

$$\Lambda(\pi_0) = \pi_0 \mathbf{1}_{[0, 1/2)}(\pi_0) + (1 - \pi_0) \mathbf{1}_{[1/2, 1]}(\pi_0)$$

**Fig. 2.1** Minimum error probability curves  $P_{E_n}^*$ , mismatch lines  $P_{E_n}^{\text{miss}}$ , for  $n \in \{1, 2, 3\}$  the minimax solution  $P_{E_n}^{**}$  (only for  $n = 1$ ) and the theoretical bound  $\Lambda$



where  $\mathbf{1}_{\{\cdot\}}(\cdot)$  is the indicator function. The role of  $n$ ,  $\sigma$ , and  $d$  are similar. Let us consider  $n = 1$ . A metric for a comparison is the area under the curve (AUC). If a priori probabilities are known, the area under the minimum error probability curve  $P_E^*$  for  $n = 1$  (denoted by  $P_{E_1}^*$  in Fig. 2.1) equals  $\approx 0.20$ . The AUC for the minimax test, i.e. AUC of  $P_{E_1}^{**}$  is  $\approx 0.31$  and for the mismatch line,  $P_{E_1}^{\text{miss}}$ , the AUC is  $\approx 0.35$ . Minimax decision rule not only bounds the error probability but also minimizes the average loss for the case when the a priori probability is unknown. The average loss compared to the known a priori probability case is  $\approx 0.11$ , which is the price paid to achieve robustness.

### 2.2.2 Robust Hypothesis Testing

Robust hypothesis testing is similar to minimax hypothesis testing, in the sense that both schemes provide a desired property on the objective function, e.g. bounded error probability. They also have differences; for instance, the error probability which is constant for all  $\pi_0$  for minimax hypothesis testing, varies and is usually less than expected for robust hypothesis testing. The main objective of robust hypothesis testing is to design a decision maker, which preserves a certain detection performance irrespective of the variations on the nominal distributions. As it may be obvious, the first step of such a design is to build the proximity sets of nominal distributions, which account for the secondary physical effects that go unmodeled. The next step deals with finding a robust decision maker for the given uncertainty sets. In the sequel, the fundamentals of robust hypothesis testing will be introduced with an example following the same order as mentioned above.

Let  $F_0$  and  $F_1$  be the nominal probability measures, and  $G_0$  and  $G_1$  be the actual probability measures all defined on  $(\Omega, \mathcal{A})$  and have the density functions  $f_0, f_1, g_0$ , and  $g_1$  respectively, with respect to a dominating measure  $\mu$ , i.e.  $F_0, F_1, G_0, G_1 \ll \mu$ . Assume that  $F_0$  and  $F_1$  are distinct, i.e.  $F_0 \neq F_1$  at least on some measurable set  $A \in \mathcal{A}$ . Here, and in the following sections every probability measure e.g.  $G[\cdot]$  will be associated with its distribution function  $G(\cdot)$  i.e.  $G(y) = G[Y \leq y]$ . Consider the binary composite hypotheses test

$$\mathcal{H}_0 : Y \sim G_0$$

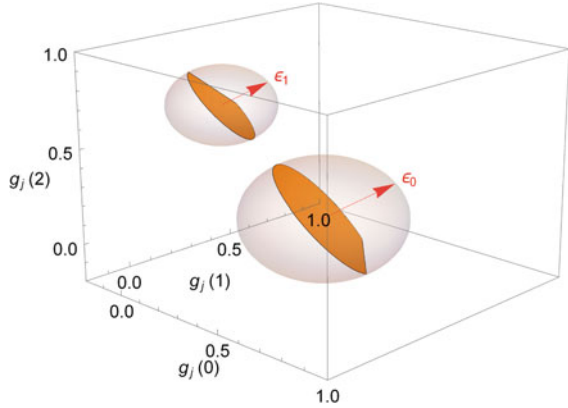
$$\mathcal{H}_1 : Y \sim G_1$$

where each measure  $G_j$  belongs to the closed ball

$$\mathcal{G}_j = \{G_j : D(G_j, F_j) \leq \epsilon_j\}, \quad j \in \{0, 1\}$$

with respect to a distance  $D$ . In other words, every distribution  $G_j$ , which is at least  $\epsilon_j$  close to the nominal distribution  $F_j$  is a member of the uncertainty class  $\mathcal{G}_j$ . Clearly, this model extends the simple hypothesis testing scheme given by (2.1), where the

**Fig. 2.2** Uncertainty classes  $\mathcal{G}_0$  and  $\mathcal{G}_1$  described in Example 2.2.2 before and after the constraints are applied



random variable  $Y$  potentially follows uncountably many probability measures  $G_j$  belonging to the set  $\mathcal{G}_j$ .

*Example 2.2.2* Let  $Y : \{0, 1, 2\} \rightarrow \{0, 1, 2\}$  be a discrete random variable and  $D$  be the Euclidean distance. Assume that the nominal densities correspond to two points,  $f_0 = (0.5, 0.3, 0.2)$  and  $f_1 = (0.1, 0.2, 0.7)$ , in three dimensional Euclidean space, and the robustness parameters are chosen as  $\epsilon_0 = 0.3$  and  $\epsilon_1 = 0.2$ . For this setup, Fig. 2.2 illustrates the two spheres, each corresponding to a set of functions (here points),  $g_0$  and  $g_1$ , before the constraints  $\sum_{i=0}^2 g_0(i) = 1$ ,  $\sum_{i=0}^2 g_1(i) = 1$  are applied, whereas the planes inside the spheres correspond to the densities after the constraints are applied, i.e. all  $g_0$  and  $g_1$  such that  $G_0 \in \mathcal{G}_0$  and  $G_1 \in \mathcal{G}_1$ .

After determining the uncertainty sets, the aim is to find a decision rule (function) which satisfies certain properties. For every observation  $y$ , the decision rule  $\phi$  is almost surely either 0 or 1, therefore it is non-randomized, and is an element of the subset of the set of all decision rules  $\Delta = \mathcal{B}(\Omega, [0, 1])$ , where  $\mathcal{B}$  denotes the set of all bounded functions. A randomized decision rule, on the other hand, generalizes  $\phi$ , and can be defined as the pair  $(U, \delta)$ , where  $U(Y)$  is a point-wise Bernoulli random variable with success probability

$$\delta(y) = P[U(Y) = 1 | Y = y],$$

i.e. for every  $y$ , the decision  $u(y)$  is 1 with probability  $\delta(y)$ . For any possible choice of  $\delta \in \Delta$ ,  $G_0 \in \mathcal{G}_0$  and  $G_1 \in \mathcal{G}_1$ , the false alarm and the miss detection probabilities are defined as

$$P_F(\delta, g_0) = \int_{\Omega} \delta g_0 d\mu$$

and

$$P_M(\delta, g_1) = \int_{\Omega} (1 - \delta) g_1 d\mu.$$

Then, the overall error probability can be given as

$$P_E(\delta, g_0, g_1) = P(\mathcal{H}_0)P_F(\delta, g_0) + P(\mathcal{H}_1)P_M(\delta, g_1).$$

It is well known that  $P_E$  is minimized if the decision rule is the likelihood ratio test, which can be represented by the decision rule

$$\delta(y) = \begin{cases} 0, & l(y) < \rho \\ \kappa(y), & l(y) = \rho \\ 1, & l(y) > \rho \end{cases}$$

where  $\rho = P(\mathcal{H}_0)/P(\mathcal{H}_1)$  is the threshold,  $l(y) := f_1/f_0(y)$  is the likelihood ratio of the observation  $y$ , and  $\kappa : \Omega \rightarrow [0, 1]$ . It can be seen that  $P_E$  is a functional of  $(g_0, g_1) \in \mathcal{G}_0 \times \mathcal{G}_1$  and  $\delta \in \Delta$ , and the condition of bounded error probability amounts to finding a triple  $(\hat{\delta}, \hat{g}_0, \hat{g}_1) \in \Delta \times (\mathcal{G}_0 \times \mathcal{G}_1)$  such that

$$P_E(\delta, \hat{g}_0, \hat{g}_1) \geq P_E(\hat{\delta}, \hat{g}_0, \hat{g}_1) \geq P_E(\hat{\delta}, g_0, g_1). \quad (2.12)$$

The condition in (2.12) is called the saddle value condition, and it says that when the pair  $(\hat{g}_0, \hat{g}_1)$  is known, the decision rule  $\hat{\delta}$  must be the best (error minimizing) decision rule among all  $\delta \in \Delta$  due to the first inequality and when  $\hat{\delta}$  is known, the pair  $(\hat{g}_0, \hat{g}_1)$  must be the worst (error maximizing) density functions due to the second inequality.

*Example 2.2.3* Visualization of a saddle value is possible only in three or lower dimensions. For this purpose, assume that  $P_E$  is a function of  $g \in [0, 1]$  and  $\delta \in [0, 1]$ , e.g.

$$P_E(\delta, g) = 1/2 + 1/50 \left( (-5 + 10\delta)^2 - (-5 + 10g)^2 \right).$$

Then, the saddle value condition ((2.12) with  $g := (g_0, g_1)$  and  $\hat{g} := (\hat{g}_0, \hat{g}_1)$ ) holds and the saddle value (red point) is shown in Fig. 2.3.

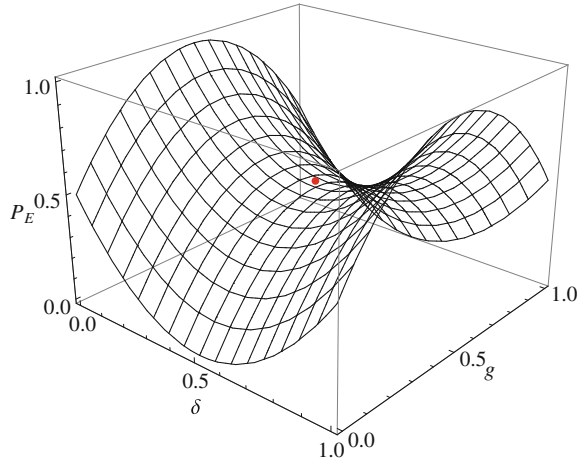
A formal definition of the existence of a saddle value was first given by Von Neuman [Neu28]. Its restatement for the robust hypothesis testing problem is

$$\max_{(g_0, g_1) \in \mathcal{G}_0 \times \mathcal{G}_1} \min_{\delta \in \Delta} P_E(\delta, g_0, g_1) = \min_{\delta \in \Delta} \max_{(g_0, g_1) \in \mathcal{G}_0 \times \mathcal{G}_1} P_E(\delta, g_0, g_1)$$

where the objective function  $P_E$  is required to be bilinear in all its arguments. Shiftman generalizes this condition: it is enough that  $P_E$  is convex in  $\delta$  and concave in  $g_0$  and  $g_1$  [Shi49]. Additionally, the sets  $\Delta$  and  $\mathcal{G}_0 \times \mathcal{G}_1$  must be convex and compact. It can be seen in Fig. 2.3 that,  $P_E$  is convex in  $\delta$  and concave in  $g$ . This property guarantees the existence of a saddle value for this example, and in general for subsets of finite dimensional vector spaces, if they are convex, bounded, and include all end points, i.e. if they are closed. The compactness argument is a straightforward application of Heine–Borel theorem [Rud76, Theorem 2.41].

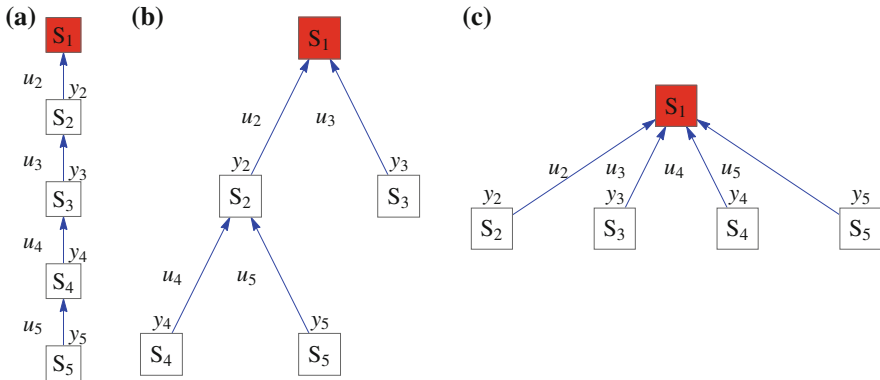


**Fig. 2.3** An example of a saddle value for two variable error function  $P_E$ , defined by Example 2.2.3



## 2.3 Decentralized Detection

There are two major reasons, among others, to build a sensor network: the first is to increase the overall detection performance and the second is to infer the state of the hypothesis under the test in different geographical locations. A designer first determines the topology of the sensor network, which may be parallel, tandem, tree, or an arbitrary topology of interest, cf. Sect. 1.1. Examples of three basic network topologies are illustrated in Fig. 2.4. For all sensor networks, each sensor  $S_i$  makes an observation  $y_i$ , and gives a decision  $u_i$ . The decisions are shared with the sensors



**Fig. 2.4** Decentralized detection networks with three different topologies; **a** Tandem topology, **b** Tree topology, and **c** Parallel topology

in connection. The final decision is given by the fusion center marked with  $S_1$  in red color.

The design of a decentralized detection network comprises finding suitable, and ideally optimum in some sense, decision and fusion rules. The decision rules can be deterministic, for example threshold rules, or (independently) randomized, as introduced in the previous sections. In the presence of multiple sensors, the decision rules can be extended to dependently randomized decision rules, allowing every decision  $u_i$  to be possibly dependent on every other decision  $u_j$ . An example is a random convex combination of two decisions with a third decision, i.e.  $u_4 = u_3u_1 + (1 - u_3)u_2$ . The same idea applies to the randomization of fusion rules as well, e.g. [Var96, pp. 87–89] for the randomization of AND and OR fusion rules.

An optimum choice of decision and fusion rules requires either a Bayesian or Neyman-Pearson formulation of the problem. Considering the Bayesian formulation and without further assumptions on the observations, the problem of determining the optimal decentralized detection strategy is shown to be NP-complete [TA85] for the parallel network topology. Optimum decision rules are also known not to be restricted to threshold tests [Tsi93]. Assume that there are two hypotheses: the null hypothesis and the alternative. Then, a person by person optimum (PBPO) solution to the general problem requires  $K + 2^K$  coupled non-linear equations to be solved [Var96], where  $K$  is the total number of sensors in the network. The problem significantly simplifies if the random variables, each associated with the observations of one sensor, are independent under each hypothesis. In this case, the number of equations to be solved for the PBPO solution does not decrease, but the computations become simpler. There are  $2^{2^K}$  fusion rules in general and with independence condition, the optimum fusion rules are monotone and the number of monotone fusion rules are much less than  $2^{2^K}$  [Var96, p. 65]. A drastic reduction in the number of computations occurs, if the local sensor decisions are restricted to be identical. Note that identical sensor decisions are not always optimum, see counterexamples in [CK92], but they often result in little or no loss of performance [Var96] and they are asymptotically optimum [Tsi88]. The sufficient statistic for the fusion center, which also reduces the total number of fusion rules to  $K$ , is then  $\sum_{i=1}^K u_i$ .

In Fig. 2.4, the fusion centers are assumed only to fuse the received information, not to make an observation of their own. The existence of a fusion center in the sensor network makes the sensor network prone to single link failure, i.e. a failure of the fusion center can jeopardize the whole network, and increases the complexity of an optimum design. On the other hand, the detection performance of a sensor network with a fusion center can outperform the performance of a sensor network without a fusion center with a significant margin [Var96]. Note that if there is no fusion center in the network, the fusion is established among sensors via exchange of information in an iterative way, e.g. via belief propagation [AVSA04].

Another classification of networks is made by centralization or decentralization. In centralized schemes sensors forward their observations directly to the fusion center. In decentralized schemes the sensors exchange only a summary information of their observations. The performance loss due to decentralization is usually negligible,

especially if multi-bit information is allowed, i.e. if the sensors quantize their observations to more than a single bit [TVB87, SZ11, Var96].

## 2.4 Conclusions

In this chapter, fundamentals of robust detection and distributed detection have been introduced. The theory and notations introduced in this chapter will be used in the following chapters with a brief reminder for the convenience of the reader. The observation space  $(\Omega, \mathcal{A})$  and the set of functions  $\mathcal{G}_0$  and  $\mathcal{G}_1$  have not been given explicitly and the concepts are treated at a basic level. Throughout the book  $\Omega = \mathbb{R}$  will be considered whenever needed, although the results are valid for vector valued observations as well. Hence, infinite dimensional function spaces will be of interest. The random variables will be denoted with capital letters and the sets will be denoted with calligraphic letters. For mathematical analysis, sensors will be called as decision makers. Classes and sets will be used synonymously. The necessary theory will be introduced, whenever needed.

## References

- [AVSA04] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron, Distributed Bayesian hypothesis testing in sensor networks, *Proceedings of the American Control Conference 2004*, vol. 6, pp. 5369–5374 2004.
- [CK92] M. Cherikh and P. B. Kantor, “Counterexamples in distributed detection,” *IEEE Transactions on Information Theory*, vol. 38, no. 1, pp. 162–165, 1992.
- [Lev08] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [Neu28] J. V. Neumann, “Zur theorie der gesellschaftsspiele,” *Mathematische Annalen*, vol. 100, pp. 295–320, 1928.
- [Rud76] W. Rudin, *Principles of Mathematical Analysis*, ser. International series in pure and applied mathematics. Paris: McGraw-Hill, 1976.
- [Shi49] M. Shiftman, “On the equality min max=max min, and the theory of games,” *RAND Report*, no. RM-243, 1949.
- [SZ11] F. Suratman and A. Zoubir, “Collaborative spectrum sensing in cognitive radio using hard decision combining with quality information,” in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, June 2011, pp. 377–380.
- [TVB87] S. Thomopoulos, R. Viswanathan, and D. Bougoulas, “Optimal decision fusion in multiple sensor systems,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-23, no. 5, pp. 644–653, Sept 1987.
- [Tsi88] J. N. Tsitsiklis, “Decentralized detection by a large numb. of sensors,” *Mathematics of Control, Signals, and Systems*, pp. 167–182, 1988.
- [Tsi93] J. N. Tsitsiklis, “Decentralized detection,” in *In Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.
- [TA85] J. Tsitsiklis and M. Athans, “On the complexity of decentralized decision making and detection problems,” *Automatic Control, IEEE Transactions on*, vol. 30, no. 5, pp. 440–446, May 1985.
- [Var96] P. K. Varshney, *Distributed detection and data fusion*, 1st ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1996.

Robust and Distributed Hypothesis Testing

Gül, G.

2017, XXI, 141 p. 42 illus., 40 illus. in color., Hardcover

ISBN: 978-3-319-49285-8