

Chapter 5

Statistical Decision Theory

Contents

| | |
|--|------|
| Stochastic Processes | v |
| Statistical Estimation Theory | vii |
| Linear Filtering | ix |
| Spectral Estimation | xi |
| Statistical Decision Theory | xiii |
| 5.1 Introduction to Decision Theory | 1 |
| 5.1.1 Hypotheses-based problems | 1 |
| 5.1.2 Modeling uncertainty | 2 |
| 5.2 Performance metrics | 3 |
| 5.2.1 Probability of error | 4 |
| 5.2.2 Receiver Operating Characteristic (ROC) | 5 |
| 5.2.3 Risk | 6 |
| 5.3 Detector design | 8 |
| 5.3.1 Maximum likelihood and maximum <i>a posteriori</i> detectors | 9 |
| 5.3.2 Bayesian decision-making: the minimum risk detector | 11 |
| 5.3.3 Non-Bayesian detectors | 13 |
| 5.4 Gaussian models | 14 |
| 5.5 Problems | 19 |
| A Transformations of random variables | 23 |
| A.1 Change of Random Variable | 23 |
| A.1.1 Some usual r.v. changes | 26 |
| B Introductory examples | 29 |
| B.1 Some introductory examples | 30 |
| B.1.1 Example 1: Binary detection with no observations | 30 |
| B.1.2 Example 2: Binary decision with observations | 33 |
| B.1.3 Example 3: Working the solution from the likelihoods | 36 |

5.1 Introduction to Decision Theory

In this section we provide a formal presentation of decision theory. Appendix B provides some introductory examples.

Decision theory (also named **detection theory** or **hypothesis testing**) is a mathematical framework used to make optimal choices under conditions of uncertainty. It employs models and statistical analysis to evaluate and compare the outcomes of different decisions, aiming to identify the most advantageous option based on established criteria. This field utilizes concepts from statistics and economics to aid in strategic planning and risk management by calculating the probabilities and impacts of potential scenarios. Decision theory focuses on maximizing benefits and minimizing costs or risks, providing a structured approach to rational decision-making. Through precise quantitative methods, it guides individuals and organizations in policy formulation and decision implementation.

5.1.1 Hypotheses-based problems

In this course, we will only cover a particular class of detection or classification problems to which we will refer as *hypotheses-based problems*. The goal is to infer the correct hypothesis, which cannot be directly observed, from a set of measurements or observations. Thus, we consider a scenario with M hypotheses, and denote the random variable that identifies the hypothesis as H . This is depicted in Fig. 5.1, where $H \in \{0, 1, \dots, M-1\}$. We also assume that we have access to an observation vector \mathbf{x} , which can be considered as the realization of a random variable \mathbf{X} lying in the observation space \mathcal{X} . We assume also that there is a certain statistical relationship between H and \mathbf{X} . Otherwise, i.e., if H and \mathbf{X} were independent, it would make no sense to use \mathbf{x} to make an informed inference about the value of H .

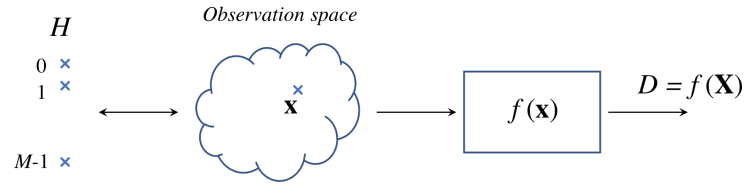


Fig. 5.1 Diagram block of hypothesis testing problems.

In this context, a detector is a function of \mathbf{x} that outputs a value d in the range $\{0, 1, \dots, M-1\}$, i.e., a guess on the value of the hypothesis that is unknown beforehand. Depending on the application scenario, the detector receives another names, like decision-maker or classifier. In this chapter we will take these terms as synonymous.

We should make a few considerations about the functions $f(\mathbf{x})$ that we admit as valid detectors in this course:

- We consider that $d = f(\mathbf{x})$ is a deterministic function. This implies that if the same vector is presented several times, the function will output the same value each time. Note that,

even though $f(\cdot)$ is deterministic, its output can be modeled as a random variable since the input is the random vector \mathbf{X} .

- The function is surjective, that is, every input \mathbf{x} generates one and only one output, but several inputs could generate the same output. Hence, the function divides the observation space into M non-overlapping regions, \mathcal{X}_d , $d = 0, 1, \dots, M-1$, i.e., one region per hypotheses. The boundaries between regions are known as *decision boundaries*.

Example 5.1 The detector $f(x) = u(x^2 - 1)$, where $u(\cdot)$ is the step function, is defined for any x on the real line, and is characterized by the following decision regions:

$$\begin{aligned}\mathcal{X}_0 &= \{x \in \mathbb{R} | x^2 - 1 < 0\} = (-1, 1), \\ \mathcal{X}_1 &= \{x \in \mathbb{R} | x^2 - 1 \geq 0\} = (-\infty, -1] \cup [1, \infty).\end{aligned}$$

where we have assumed $u(0) = 1$. In this example, the regions are connected and non-empty.

Example 5.2 The detector $f(\mathbf{x}) = \arg \min_i y_i(\mathbf{x})$ defined over $\mathcal{X} = [0, 1]^2$, with

$$\begin{aligned}y_0(\mathbf{x}) &= \|\mathbf{x}\|^2, \\ y_1(\mathbf{x}) &= x_1 - x_0 + 1, \\ y_2(\mathbf{x}) &= x_0 - x_1 + 1,\end{aligned}$$

is characterized by the decision regions depicted in Fig. 5.2.

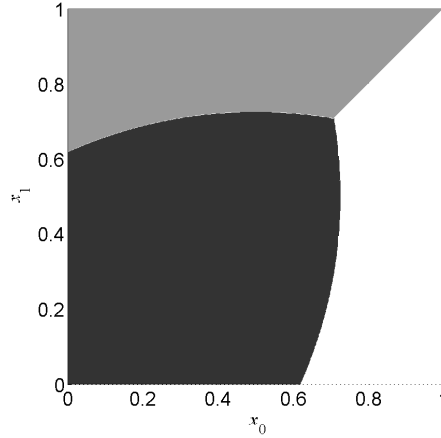


Fig. 5.2 Decision regions for the detector given in Example 5.2: \mathcal{X}_0 (black), \mathcal{X}_1 (grey), and \mathcal{X}_2 (white).

5.1.2 Modeling uncertainty

We review now the main distributions that will be employed in detection problems:

- *A priori* probability distribution of the hypotheses: This is a discrete distribution that quantifies the probability of each hypothesis independently of the observations. If we did not have access to any observations, our design would have to rely entirely on these probabilities, as it was the case in Section B.1.1,

$$P_H(h), \quad \text{for } h = 0, 1, \dots, M-1.$$

- Likelihoods of the hypotheses: This represents the probability of the observations given the hypothesis. Note that, even though we refer to these distribution as the likelihoods of the hypotheses, what we actually have is a collection of distributions over the random variable X (unidimensional case) or \mathbf{X} (multidimensional case), one for each hypothesis,

$$p_{\mathbf{X}|H}(\mathbf{x}|h) \quad \text{for } \mathbf{x} \in \mathcal{X} \text{ and } h = 0, 1, \dots, M-1,$$

where we have assumed a multidimensional case with continuous observations. Note that random variable \mathbf{X} may lie in different regions depending on the hypothesis.

- *A posteriori* distribution of the hypotheses: This distribution provides information about the probabilities of the hypothesis, but conditioning them on each possible value of the observation vector

$$P_{H|\mathbf{X}}(h|\mathbf{x}), \quad \text{for } h = 0, 1, \dots, M-1.$$

Since designing a detector consists in deciding what should be the decision for each value of the observation vector, and this distribution expresses directly what are the probabilities of the hypothesis conditioned on every \mathbf{x} , *a posteriori* probabilities play a fundamental role for the statistical design of detectors.

A priori and *a posteriori* probabilities are related by Bayes' Theorem, which states

$$P_{H|\mathbf{X}}(h|\mathbf{x}) = \frac{p_{\mathbf{X}|H}(\mathbf{x}|h)P_H(h)}{p_{\mathbf{X}}(\mathbf{x})}.$$

Bayes' Theorem shows how observing \mathbf{x} modifies the information about the probabilities of the different hypotheses. Without them, we could only use $P_H(h)$ to make decisions. However, once the observation vector comes into play, a more accurate estimation of these probabilities can be achieved via $P_{H|\mathbf{X}}(h|\mathbf{x})$, and these probabilities can be used to obtain a more informed decision. Note also that if we know both the *a priori* probabilities of the hypothesis and their likelihoods, the joint distribution of \mathbf{X} and H can be calculated. This joint distribution is the most complete characterization of the random variables, and from it any other probability function can be calculated as well.

In the following, we consider two different kinds of problems involving M -ary hypothesis testing problems:

- Analysis of detectors: Here, the detector is given, and the objective is to analyze its performance with respect to certain performance metrics.
- Detector design: The goal is to build a function $f(\mathbf{x})$ to optimize a desired performance metric.

5.2 Performance metrics

The first problem that we consider is the evaluation of the performance of a given detector. In this section, we review different metrics that can be used to assess performance. In all

cases, we consider first the multiple hypothesis test scenario, and afterwards we specialize it to the binary case.

5.2.1 Probability of error

The probability of error is the probability of a wrong decision, i.e., the output of the statistic is not equal to the actual hypothesis. Under a frequentist approach, this probability can be interpreted as the average number of experiments in which an incorrect decision is taken, when the number of experiments tends to infinity. However, since we are assuming that the statistical characterization of the problem is available through the different probability distributions that we just reviewed, the probability of error can be calculated in closed-form as:

$$\begin{aligned}
 P_e &= P(D \neq H) = 1 - P(D = H) \\
 &= 1 - \sum_{h=0}^{M-1} P(D = h, H = h) \\
 &= 1 - \sum_{h=0}^{M-1} P(D = h | H = h) P_H(h) \\
 &= 1 - \sum_{h=0}^{M-1} P_H(h) \int_{\mathcal{X}_h} p_{\mathbf{X}|H}(\mathbf{x}|h) d\mathbf{x},
 \end{aligned}$$

where we have exploited that the probability of error is one minus the probability of correct decision. This is, in most cases, more convenient since the number of combinations where D and H are equal is (much) smaller than the number of combinations where they differ. Moreover, the last line of the previous expression follows from

$$P(D = h | H = h) = P(\mathbf{x} \in \mathcal{X}_d | H = h) = \int_{\mathcal{X}_h} p_{\mathbf{X}|H}(\mathbf{x}|h) d\mathbf{x},$$

which states that, conditioned on $H = h$, the probability of $D = h$ is precisely the integral of the likelihood of that hypothesis in the region where the given detector decides in favor of hypothesis h , i.e., the region \mathcal{X}_h .

Finally, note that it is also possible to compute the probability of error for a particular observation vector \mathbf{x} . If \mathbf{x} belongs to \mathcal{X}_d , the associated probability of error would be

$$P(H \neq d | \mathbf{x}) = 1 - P(H = d | \mathbf{x}) = 1 - P_{H|\mathbf{X}}(d | \mathbf{x}) = \sum_{\substack{l=0 \\ l \neq d}}^{M-1} P_{H|\mathbf{X}}(l | \mathbf{x}) \quad (5.1)$$

In other words, the probability of error at a particular $\mathbf{x} \in \mathcal{X}_d$ is the sum of the *a posteriori* probabilities of hypothesis different from d conditioned on this particular observation. For instance, imagine that in a three-hypothesis testing problem for a given \mathbf{x}_o a detector selects hypothesis 0. Then, the probability of error for \mathbf{x}_o is the sum of the probabilities of hypo-

esis 1 and 2 conditioned on $\mathbf{X} = \mathbf{x}_o$, i.e., the sum of *a posteriori* probabilities $P_{H|\mathbf{X}}(1|\mathbf{x}_o)$ and $P_{H|\mathbf{X}}(2|\mathbf{x}_o)$.

5.2.1.1 Binary case: P_e , P_{FA} , P_M and P_D

For the binary case, contrary to the multiple hypotheses test, computing the probability of error involves as many terms as the probability of a correct decision since

$$\begin{aligned} P_e &= P(D = 0, H = 1) + P(D = 1, H = 0) \\ &= P(D = 0|H = 1)P_H(1) + P(D = 1|H = 0)P_H(0). \end{aligned}$$

In the expression above we find two terms that are normally referred to as the *probability of false alarm* (also known as probability of Type I error or significance level) and the *probability of missing* (or probability of Type II error):

$$\begin{aligned} P_{FA} &= P(D = 1|H = 0) = \int_{\mathcal{X}_1} p_{\mathbf{X}|H}(\mathbf{x}|0)d\mathbf{x}, \\ P_M &= P(D = 0|H = 1) = \int_{\mathcal{X}_0} p_{\mathbf{X}|H}(\mathbf{x}|1)d\mathbf{x}. \end{aligned}$$

Similarly, the probability of detection (power or sensitivity) is defined as

$$P_D = P(D = 1|H = 1) = 1 - P_M,$$

and

$$P(D = 0|H = 0) = 1 - P_{FA},$$

is the specificity. Using these definitions, the probability of error can now be expressed more compactly as

$$P_e = P_M P_H(1) + P_{FA} P_H(0).$$

Interestingly, for the computation of P_{FA} and P_M , only likelihoods are required. However, in order to compute the overall probability of error, we also need to know the *a priori* probabilities of the hypothesis.

5.2.2 Receiver Operating Characteristic (ROC)

We also introduce here an important concept for the analysis of binary hypothesis tests: the receiver operating characteristic (ROC) curve. The ROC curve plots the probability of false alarm, P_{FA} , against the probability of detection, P_D for different values of some parameter. Figure 5.3 shows the ROC curves of two different detectors, Detector 1 and Detector 2. As can be seen in this figure, the performance of Detector 2 is clearly better than that of Detector 1, since for each P_{FA} , the P_D of Detector 2 is equal or larger than that of Detector 1. Moreover, both detectors perform better than a random decision whose ROC curve is also shown in the figure. One final comment is in order. For almost all detectors it is not

possible to increase the probability of detection without increasing the probability of false alarm.

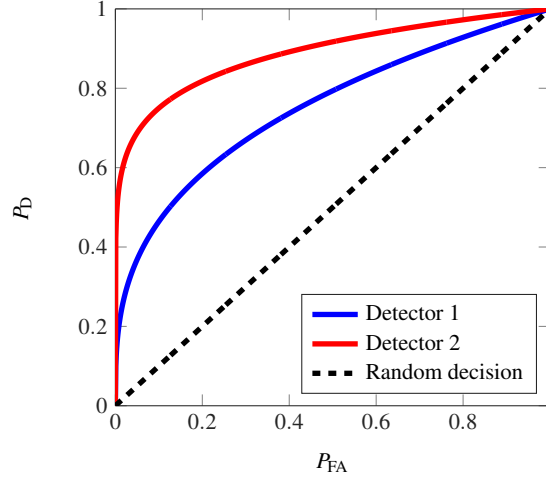


Fig. 5.3 ROC curves for two different detectors.

5.2.3 Risk

In scenarios where the consequences of each type of error are different, the probability of error is not an adequate performance measure. Imagine, for instance, a detector that discriminates whether there are or not suspicious tumor masses in a medical image. Such detector is used as a pre-diagnosis system, so that patients that can have a tumor are then explored with more accurate (but also invasive) techniques. In this case, there is a clear asymmetry between both kinds of errors: The incorrect decision that tumor masses are present would result in an unnecessary biopsy and inconvenience for the patient, but the opposite error could delay the diagnosis until a time when the process is irreversible.

To assign a penalty to different kinds of errors, we can define a cost function

$$c_{DH}, \quad D, H = 0, \dots, M-1.$$

Such function will take as many values as combinations of decisions and hypotheses, in such a way that each particular value c_{dh} is the cost of deciding $D = d$ when hypothesis $H = h$ is the true one. As already pointed out, we assume deterministic costs in this course, in the sense that the cost for each particular d and h is fixed. However, since the cost is a function of the random variables D and H , it is a random variable. Given a detector $D = \phi(\mathbf{X})$, we define the **risk** R_ϕ as the expected value of the cost,

$$R_\phi = \mathbb{E}\{c_{DH}\} = \sum_{h=0}^{M-1} \sum_{d=0}^{M-1} c_{dh} P_H(h) P_{D|H}(d|h). \quad (5.2)$$

Since $\phi(\mathbf{x}) = d$ when the observation belongs to the decision region of $D = d$, that is, $\mathbf{x} \in \mathcal{X}_d$, the conditional probabilities $P_{D|H}(d|h)$ can be calculated as

$$P_{D|H}(d|h) = P\{X \in \mathcal{X}_d | H = h\} = \int_{\mathcal{X}_d} p_{\mathbf{X}|H}(\mathbf{x}|h) d\mathbf{x} \quad (5.3)$$

therefore

$$R_\phi = \sum_{h=0}^{M-1} P_H(h) \sum_{d=0}^{M-1} c_{dh} \int_{\mathbf{x} \in \mathcal{X}_d} p_{\mathbf{X}|H}(\mathbf{x}|h) d\mathbf{x}. \quad (5.4)$$

Example 5.3 Consider a multiclass decision problem with three hypotheses whose likelihoods are:

$$\begin{aligned} p_{X|H}(x|0) &= 1 & 0 < x < 1 \\ p_{X|H}(x|1) &= 2(1-x) & 0 < x < 1 \\ p_{X|H}(x|2) &= 2x & 0 < x < 1 \end{aligned}$$

knowing that the prior probabilities of the hypotheses are: $P_H(0) = 0.4$ and $P_H(1) = P_H(2) = 0.3$, and the cost policy is given by $c_{hh} = 0$, $h = 0, 1, 2$ and $c_{hd} = 1$, $h \neq d$. Obtain the risk of the decision-maker:

$$\phi(x) = \begin{cases} 1, & x < 0.5 \\ 2, & x > 0.5 \end{cases}$$

Applying the expression (5.2) to this problem we have:

$$\begin{aligned} R_\phi &= c_{10}P_H(0)P_{D|H}(1|0) + c_{20}P_H(0)P_{D|H}(2|0) + c_{01}P_H(1)P_{D|H}(0|1) \\ &\quad + c_{21}P_H(1)P_{D|H}(2|1) + c_{02}P_H(2)P_{D|H}(0|2) + c_{12}P_H(2)P_{D|H}(1|2) \end{aligned}$$

where the terms $P_{D|H}(d|h)$ can be calculated using (5.3)

$$\begin{aligned} P_{D|H}(0|1) &= P_{D|H}(0|2) = 0 \\ P_{D|H}(1|0) &= \int_{\mathcal{X}_1} p_{X|H}(x|0) dx = \int_0^{0.5} 1 dx = 0.5 \\ P_{D|H}(2|0) &= \int_{\mathcal{X}_2} p_{X|H}(x|0) dx = \int_{0.5}^1 1 dx = 0.5 \\ P_{D|H}(1|2) &= \int_{\mathcal{X}_1} p_{X|H}(x|2) dx = \int_0^{0.5} 2x dx = 0.25 \\ P_{D|H}(2|1) &= \int_{\mathcal{X}_2} p_{X|H}(x|1) dx = \int_{0.5}^1 2(1-x) dx = 0.25 \end{aligned}$$

and substituting these, we arrive at

$$R_\phi = 0.4 \cdot 0.5 + 0.4 \cdot 0.5 + 0.3 \cdot 0.25 + 0.3 \cdot 0.25 = 0.55$$

Finally, we define the *conditional risk* as the expected cost conditioned on a given value of \mathbf{x} , $\mathbb{E}\{c_{dH}|\mathbf{x}\}$. Taking into account that, for a given \mathbf{x} and a given detector, the decision value is fixed, it is only required to take expectations with respect to such hypothesis. The conditional risk for a given observation $\mathbf{x} \in \mathcal{X}_d$ is given by

$$\mathbb{E}\{c_{dH}|\mathbf{x}\} = \sum_{h=0}^{M-1} c_{dh}P_{H|X}(h|\mathbf{x}). \quad (5.5)$$

Example 5.4 Continuing with Example 5.3, the conditional risk of each decision can be calculated as:

$$\mathbb{E}\{c(d, H)|x\} = c_{d0}P_{H|X}(0|x) + c_{d1}P_{H|X}(1|x) + c_{d2}P_{H|X}(2|x)$$

where the posterior distributions can be obtained by applying Bayes' Theorem:

$$\begin{aligned}
P_{H|X}(0|x) &= \frac{p_{X|H}(x|0)P_H(0)}{\sum_{h=0}^2 p_{X|H}(x|h)P_H(h)} = \frac{1 \cdot 0.4}{1 \cdot 0.4 + 2(1-x) \cdot 0.3 + 2x \cdot 0.3} = 0.4 \\
P_{H|X}(1|x) &= \frac{p_{X|H}(x|1)P_H(1)}{\sum_{h=0}^2 p_{X|H}(x|h)P_H(h)} = \frac{2(1-x) \cdot 0.3}{1} = 0.6(1-x) \\
P_{H|X}(2|x) &= \frac{p_{X|H}(x|2)P_H(2)}{\sum_{h=0}^2 p_{X|H}(x|h)P_H(h)} = \frac{2x \cdot 0.3}{1} = 0.6x
\end{aligned}$$

This leads to:

- if $d = 0$:
$$\mathbb{E}\{c(0, H)|x\} = c_{00}P_{H|X}(0|x) + c_{01}P_{H|X}(1|x) + c_{02}P_{H|X}(2|x) = 0 \cdot 0.4 + 1 \cdot 0.6(1-x) + 1 \cdot 0.6x = 0.6$$
- if $d = 1$:
$$\mathbb{E}\{c(1, H)|x\} = c_{10}P_{H|X}(0|x) + c_{11}P_{H|X}(1|x) + c_{12}P_{H|X}(2|x) = 1 \cdot 0.4 + 0 \cdot 0.6(1-x) + 1 \cdot 0.6x = 0.4 + 0.6x$$
- if $d = 2$:
$$\mathbb{E}\{c(2, H)|x\} = c_{20}P_{H|X}(0|x) + c_{21}P_{H|X}(1|x) + c_{22}P_{H|X}(2|x) = 1 \cdot 0.4 + 1 \cdot 0.6(1-x) + 0 \cdot 0.6x = 1 - 0.6x$$

5.2.3.1 Binary case: risk

For the binary case, a simpler expression can be obtained in terms of P_{FA} , P_{M} , and P_{D} as follows

$$\begin{aligned}
R_\phi &= c_{00}P(D=0, H=0) + c_{01}P(D=0, H=1) \\
&= c_{00}P(D=0|H=0)P_H(0) + c_{01}P_{\text{M}}P_H(1) + c_{10}P_{\text{FA}}P_H(0) + c_{11}P_{\text{D}}P_H(1). \\
&= c_{00}(1 - P_{\text{FA}})P_H(0) + c_{01}P_{\text{M}}P_H(1) + c_{10}P_{\text{FA}}P_H(0) + c_{11}(1 - P_{\text{M}})P_H(1). \\
&= (c_{00}P_H(0) + c_{11}P_H(1)) + (c_{01} - c_{11})P_{\text{M}}P_H(1) + (c_{10} - c_{00})P_{\text{FA}}P_H(0) \quad (5.6)
\end{aligned}$$

The previous expression shows that the risk of a decision-maker is the sum of three components:

- $(c_{00}P_H(0) + c_{11}P_H(1))$ is the minimum risk of the ideal decision-maker, the one with $P_{\text{M}} = 0$ and $P_{\text{FA}} = 0$ who succeeds with probability 1.
- $(c_{01} - c_{11})P_H(1)P_{\text{M}}$ is the increase in risk caused by miss errors.
- $(c_{10} - c_{00})P_H(0)P_{\text{FA}}$ is the increase in risk caused by false alarms.

Note that the ideal decision-maker is, in general, unachievable, because if the likelihoods of the hypotheses overlap, it is not possible to avoid errors. The optimal decision-maker will be the one who finds a good compromise between miss errors and false alarms, such that the risk in (5.6) is minimized.

5.3 Detector design

Once we have studied different ways of analyzing the performance of a given detector, we turn our attention to the problem of designing detectors that maximize one of these performance metrics.

5.3.1 Maximum likelihood and maximum *a posteriori* detectors

A first possibility would be to rely directly on the maximization of the available probability density functions:

- The detector that maximizes the likelihood is known as the *maximum likelihood* (ML) detector:

$$d_{ML} = \arg \max_h p_{\mathbf{X}|H}(\mathbf{x}|h).$$

- The detector that selects the hypothesis with maximum *a posteriori* probability is known as the maximum *a posteriori* (MAP) detector:

$$d_{MAP} = \arg \max_h P_{H|\mathbf{X}}(h|\mathbf{x}).$$

These detectors proceed as follows. Designing a detector is equivalent to specifying a unique decision for each possible value of the observation vector \mathbf{x} . Then, the ML and MAP strategies are based on evaluating either the likelihoods or the *a posteriori* probabilities for each \mathbf{x} in the observation space, and select, for each \mathbf{x} , the hypothesis that maximizes $p_{\mathbf{X}|H}(\mathbf{x}|h)$ (ML) or $P_{H|\mathbf{X}}(h|\mathbf{x})$ (MAP).

Finally, there are two properties that are worth considering with respect to these detectors:

1. When the *a priori* probabilities of the hypothesis are the same, i.e., $P_H(h) = 1/M$, the ML and MAP detectors are identical. This can be shown from the Bayes' Theorem, since in this case

$$d_{MAP} = \arg \max_h P_{H|\mathbf{X}}(h|\mathbf{x}) = \arg \max_h \frac{p_{\mathbf{X}|H}(\mathbf{x}|h)P_H(h)}{p_{\mathbf{X}}(\mathbf{x})} = \arg \max_h p_{\mathbf{X}|H}(\mathbf{x}|h) = d_{ML}.$$

2. The MAP detector minimizes the probability of error. Note that according to (5.1) the probability of error for a given \mathbf{x} can be expressed as

$$P(D \neq H|\mathbf{x}) = 1 - P_{H|\mathbf{X}}(h|\mathbf{x}).$$

Since the MAP detector selects for every \mathbf{x} the hypothesis that maximizes $P_{H|\mathbf{X}}(h|\mathbf{x})$, it therefore minimizes the probability of error for each vector of the observation space. Thus, as the probability of error is minimized for each point of the observation space, it is also minimized overall. That is,

$$P(D \neq H) = \int_{\mathcal{X}} P(D \neq H|\mathbf{x})p_{\mathbf{X}}(\mathbf{x})d\mathbf{x},$$

and we can check that the value of the integral (the probability of error) is minimized if, for each \mathbf{x} , the decisions minimize $P(D \neq H|\mathbf{x})$, i.e., the MAP detector.

5.3.1.1 Binary case: ML and MAP detectors

The expressions of the ML and MAP detectors become fairly simple for the binary case:

- Maximum likelihood detector:

$$\begin{array}{c} D = 1 \\ p_{\mathbf{X}|H}(\mathbf{x}|1) \geq p_{\mathbf{X}|H}(\mathbf{x}|0), \\ D = 0 \end{array}$$

which can be expressed as a *likelihood ratio test* (LRT)

$$\frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\geq}} 1,$$

where we have taken into account that the likelihoods are non-negative. Sometimes, it will be more convenient to work with the *log-likelihood ratio test* (LLRT)

$$\log \left[\frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} \right] = \log p_{\mathbf{X}|H}(\mathbf{x}|1) - \log p_{\mathbf{X}|H}(\mathbf{x}|0) \underset{D=0}{\overset{D=1}{\geq}} 0, \quad (5.7)$$

which can be done because the logarithm is a monotonically increasing function.

- Maximum *a posteriori* detector:

$$\begin{array}{c} D = 1 \\ p_{H|\mathbf{X}}(1|\mathbf{x}) \geq p_{H|\mathbf{X}}(0|\mathbf{x}), \\ D = 0 \end{array}$$

which can also be expressed as a LRT as

$$\frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\geq}} \frac{P_H(0)}{P_H(1)}. \quad (5.8)$$

As in the general case with M hypothesis, the MAP detector minimizes the probability of error and the ML and MAP detectors are the same if $P_H(0) = P_H(1) = 0.5$. Moreover, we can see that both detectors can be expressed as a LRT

$$\frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\geq}} \eta, \quad (5.9)$$

where η is a threshold. When this threshold is 1, the LRT is the ML detector and for $\eta = P_H(0)/P_H(1)$, the LRT becomes the MAP detector, that is, minimum P_e detector. Hence, we get two different points in the ROC curve. Actually, sweeping the value of the threshold generates the complete ROC curves in Figure 5.3.¹

¹ This actually applies to all detectors that can be written as $\phi(\mathbf{x}) \underset{D=0}{\overset{D=1}{\geq}} \eta$. That is, comparing a function of the observations with a threshold achieves a given (P_{FA}, P_D) point in the ROC curve. These detectors are known as threshold detectors.

5.3.2 Bayesian decision-making: the minimum risk detector

As we have already studied, sometimes it makes more sense to measure the performance of a detector in terms of the expected cost. Therefore, it is important to tackle the problem of designing a detector that is optimum with respect to the expected cost.

Remember that the expected cost of a detector deciding d for an observation \mathbf{x} is given by Equation (5.5), which we reproduce here for convenience:

$$\mathbb{E}\{c_{dH}|\mathbf{x}\} = \sum_{h=0}^{M-1} c_{dh} P_{H|\mathbf{X}}(h|\mathbf{x}). \quad (5.10)$$

Minimizing the expected cost over the whole observation space requires that decisions for each observation minimize the conditional expected cost. That is, for each \mathbf{x} the above expression should be minimized, and the expression of the minimum mean cost detector can be stated as follows:

$$d^* = \arg \min_d \sum_{h=0}^{M-1} c_{dh} P_{H|\mathbf{X}}(h|\mathbf{x}).$$

Hence, when designing the detector, we need to evaluate the cost of the different decisions for each observation vector, and select the decision for which the expected cost is minimized.

Example 5.5 Continuing with Example 5.3, the conditional risk of each decision can be calculated as:

$$\mathbb{E}\{c(d, H)|x\} = c_{d0}P_{H|X}(0|x) + c_{d1}P_{H|X}(1|x) + c_{d2}P_{H|X}(2|x)$$

where the posterior distributions can be obtained by applying Bayes' Theorem:

$$\begin{aligned} P_{H|X}(0|x) &= \frac{P_{X|H}(x|0)P_H(0)}{\sum_{h=0}^2 P_{X|H}(x|h)P_H(h)} = \frac{1 \cdot 0.4}{1 \cdot 0.4 + 2(1-x) \cdot 0.3 + 2x \cdot 0.3} = 0.4 \\ P_{H|X}(1|x) &= \frac{P_{X|H}(x|1)P_H(1)}{\sum_{h=0}^2 P_{X|H}(x|h)P_H(h)} = \frac{2(1-x) \cdot 0.3}{1} = 0.6(1-x) \\ P_{H|X}(2|x) &= \frac{P_{X|H}(x|2)P_H(2)}{\sum_{h=0}^2 P_{X|H}(x|h)P_H(h)} = \frac{2x \cdot 0.3}{1} = 0.6x \end{aligned}$$

This leads to:

- if $d = 0$:

$$\mathbb{E}\{c(0, H)|x\} = c_{00}P_{H|X}(0|x) + c_{01}P_{H|X}(1|x) + c_{02}P_{H|X}(2|x) = 0 \cdot 0.4 + 1 \cdot 0.6(1-x) + 1 \cdot 0.6x = 0.6$$
- if $d = 1$:

$$\mathbb{E}\{c(1, H)|x\} = c_{10}P_{H|X}(0|x) + c_{11}P_{H|X}(1|x) + c_{12}P_{H|X}(2|x) = 1 \cdot 0.4 + 0 \cdot 0.6(1-x) + 1 \cdot 0.6x = 0.4 + 0.6x$$
- if $d = 2$:

$$\mathbb{E}\{c(2, H)|x\} = c_{20}P_{H|X}(0|x) + c_{21}P_{H|X}(1|x) + c_{22}P_{H|X}(2|x) = 1 \cdot 0.4 + 1 \cdot 0.6(1-x) + 0 \cdot 0.6x = 1 - 0.6x$$

It is interesting to point out that when the cost function penalizes equally all kinds of errors, i.e.,

$$c_{dh} = \begin{cases} 0, & d = h \\ c, & d \neq h \end{cases}$$

the detector with minimum expected cost becomes the MAP one. This is easily proved by replacing these costs into the expression for the minimum expected cost detector

$$\begin{aligned}
 d^* &= \arg \min_d \sum_{h=0}^{M-1} c_{dh} P_{H|X}(h|\mathbf{x}) \\
 &= \arg \min_d c \sum_{h \neq d} P_{H|X}(h|\mathbf{x}) \\
 &= \arg \min_d 1 - P_{H|X}(d|\mathbf{x}) \\
 &= \arg \max_d P_{H|X}(d|\mathbf{x}) \\
 &= d_{MAP}.
 \end{aligned} \tag{5.11}$$

5.3.2.1 Binary case: Minimum risk detector

In the binary case, we can also express the optimum detector with respect to a cost function as a LRT. Let us start by particularizing (5.10) for $d = 0$ and $d = 1$, and then follow the criterion of deciding in favor of the minimum cost, i.e.,

$$\begin{array}{c} D = 1 \\ \mathbb{E} \{c_{0H} | \mathbf{x}\} \geq \mathbb{E} \{c_{1H} | \mathbf{x}\}. \\ D = 0 \end{array}$$

Now, using the definition of expectation, the criterion becomes

$$\begin{array}{c} D = 1 \\ c_{00}P_{H|X}(0|\mathbf{x}) + c_{01}P_{H|X}(1|\mathbf{x}) \geq c_{10}P_{H|X}(0|\mathbf{x}) + c_{11}P_{H|X}(1|\mathbf{x}), \\ D = 0 \end{array}$$

which after some algebra can be rewritten as

$$\begin{array}{c} D = 1 \\ \frac{P_{H|X}(1|\mathbf{x})}{P_{H|X}(0|\mathbf{x})} \geq \frac{c_{10} - c_{00}}{c_{01} - c_{11}}. \\ D = 0 \end{array}$$

Finally, using Bayes' Theorem, we may rewrite the *a posteriori* probabilities in terms of the likelihoods and the *a priori* probabilities, which finally yields

$$\begin{array}{c} D = 1 \\ \frac{P_{X|H}(\mathbf{x}|1)}{P_{X|H}(\mathbf{x}|0)} \geq \frac{c_{10} - c_{00}}{c_{01} - c_{11}} \frac{P_H(0)}{P_H(1)}, \\ D = 0 \end{array}$$

and corresponds to yet another point of the ROC curve of the LRT.

5.3.3 Non-Bayesian detectors

Non-Bayesian detectors are those that do not ground on a probability model for the hypothesis. Their design depends on the likelihood functions only. This is the case, for instance, of the ML detector. Other non-Bayesian detectors, in the binary case can be expressed as the LRT in (5.9) for different values of η . The Neyman-Pearson detector is a classical example.

5.3.3.1 Binary case: Neyman-Pearson detector

The Neyman-Pearson (NP) detector is a well known detector for binary problems, which maximizes the probability of detection while it provides a bound on the probability of false alarm. Before proceeding with the derivation, let us recall the definitions of probability of false alarm and detection

$$P_{\text{FA}} = \int_{\mathcal{X}_1} p_{\mathbf{X}|H}(\mathbf{x}|0) d\mathbf{x},$$

$$P_{\text{D}} = \int_{\mathcal{X}_1} p_{\mathbf{X}|H}(\mathbf{x}|1) d\mathbf{x}.$$

Now, the NP detector can be derived as the solution to

$$\text{maximize } P_{\text{D}}, \quad \text{subject to } P_{\text{FA}} = \alpha,$$

which is an optimization problem with constraints. The solution to this kind of problems is obtained from the Lagrangian, which is given by

$$\begin{aligned} \mathcal{L}(\mathcal{X}_1, \eta) &= P_{\text{D}} - \eta(P_{\text{FA}} - \alpha) \\ &= \int_{\mathcal{X}_1} p_{\mathbf{X}|H}(\mathbf{x}|1) d\mathbf{x} - \eta \left(\int_{\mathcal{X}_1} p_{\mathbf{X}|H}(\mathbf{x}|0) d\mathbf{x} - \alpha \right) \\ &= \int_{\mathcal{X}_1} (p_{\mathbf{X}|H}(\mathbf{x}|1) - \eta p_{\mathbf{X}|H}(\mathbf{x}|0)) d\mathbf{x} + \eta \alpha. \end{aligned}$$

Note, that the optimization variable is the region where we decide $d = 1$. Next, we need to maximize the Lagrangian, and therefore the P_{D} , which is achieved by maximizing the above integral. To do so, and taken into account that an integral may be seen as a sum, we need to design \mathcal{X}_1 such that the integrand is positive, i.e.

$$\mathcal{X}_1 = \{\mathbf{x} | p_{\mathbf{X}|H}(\mathbf{x}|1) - \eta p_{\mathbf{X}|H}(\mathbf{x}|0) \geq 0\} \Rightarrow \frac{p_{\mathbf{X}|H}(\mathbf{x}|1)}{p_{\mathbf{X}|H}(\mathbf{x}|0)} \underset{D=0}{\overset{D=1}{\geq}} \eta,$$

and η is selected to achieve the desired probability of false alarm.

5.3.3.2 Minimax classifiers

Minimax classifiers are designed in such a way that their error probability is independent on the prior probabilities of the hypothesis. For binary decision problems, they are given by the LRT such that P_{FA} and P_M are the same

$$P_{FA} = P_M \quad (5.12)$$

5.4 Gaussian models

In this section, we will derive the likelihood ratio test for Gaussian observations under several assumptions. Then, depending on the threshold, we would obtain the different detectors: NP, ML, MAP, and minimum cost.

Before proceeding, we introduce the multivariate real Gaussian probability density function (PDF), which is given by

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \mathbf{V}^{-1} (\mathbf{x} - \mathbf{m}) \right),$$

where \mathbf{x} is an N -dimensional vector, \mathbf{m} is the mean vector, and \mathbf{V} is the cross-covariance matrix. Then, under hypothesis $h = 0$, the likelihood is

$$P_{\mathbf{X}|H}(\mathbf{x}|0) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}_0|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_0)^T \mathbf{V}_0^{-1} (\mathbf{x} - \mathbf{m}_0) \right),$$

whereas it is

$$P_{\mathbf{X}|H}(\mathbf{x}|1) = \frac{1}{(2\pi)^{N/2} |\mathbf{V}_1|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mathbf{m}_1)^T \mathbf{V}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \right),$$

under hypothesis $h = 1$. For this hypothesis test, the LLRT in (5.7) becomes

$$\begin{aligned} -\frac{1}{2} \log |\mathbf{V}_1| - \frac{1}{2} (\mathbf{x} - \mathbf{m}_1)^T \mathbf{V}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \\ + \frac{1}{2} \log |\mathbf{V}_0| + \frac{1}{2} (\mathbf{x} - \mathbf{m}_0)^T \mathbf{V}_0^{-1} (\mathbf{x} - \mathbf{m}_0) \underset{D=0}{\overset{D=1}{\geq}} \log(\eta) \end{aligned}$$

or, equivalently,

$$(\mathbf{x} - \mathbf{m}_0)^T \mathbf{V}_0^{-1} (\mathbf{x} - \mathbf{m}_0) - (\mathbf{x} - \mathbf{m}_1)^T \mathbf{V}_1^{-1} (\mathbf{x} - \mathbf{m}_1) \underset{D=0}{\overset{D=1}{\geq}} \mu \quad (5.13)$$

where

$$\mu = 2 \log(\eta) + \log |\mathbf{V}_1| - \log |\mathbf{V}_0|,$$

with η being a threshold selected according to the performance criterion.

After a careful look at (5.13), it can be shown that the optimal detector in the Gaussian case is given by a second-order polynomial function. Hence, the decision boundaries² are quadratic surfaces. For instance, for 2D problems ($N = 2$), these boundaries are hyperbolas, parabolas, ellipses or straight lines.

In the following sections, we consider a few particular cases, and we conclude this section with two examples.

Example 5.6 Figure 5.4 shows the decision boundaries for the ML detector ($\eta = 1$ in (5.13)), for a detection problem with 2D Gaussian observations with the following means and cross-covariance matrices:

$$\mathbf{m}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{V}_0 = \begin{pmatrix} 1.2 & 0.43 \\ 0.43 & 1.75 \end{pmatrix},$$

and

$$\mathbf{m}_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \mathbf{V}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.$$

In this figure, the gray color gradient represents the value of the likelihoods $P_{\mathbf{X}|H}(\mathbf{x}|0)$ and $P_{\mathbf{X}|H}(\mathbf{x}|1)$, where darker colors denote larger values. Moreover, the white curves are the iso-probability lines and the black curve is the decision boundary, which in this case is a hyperbola (the symmetric part is not shown in this figure).

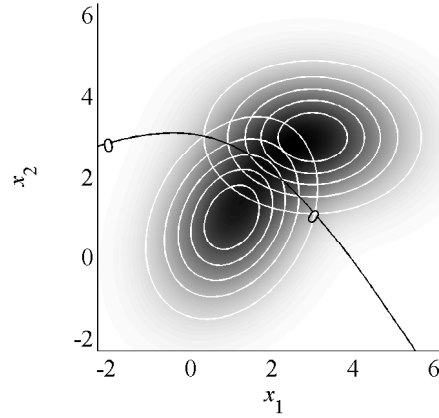


Fig. 5.4 Hyperbolic decision boundary of the ML detector and likelihoods for a Gaussian detection problem with 2D observations.

Example 5.7 Figure 5.5 shows an equivalent figure to that of the previous example, but for a problem with the following means and cross-covariance matrices:

$$\mathbf{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{V}_0 = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.7 \end{pmatrix},$$

and

$$\mathbf{m}_1 = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix}, \mathbf{V}_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.2 \end{pmatrix}.$$

In this case, the decision boundary is an ellipse.

² We obtain the decision boundaries for the equality in (5.13).

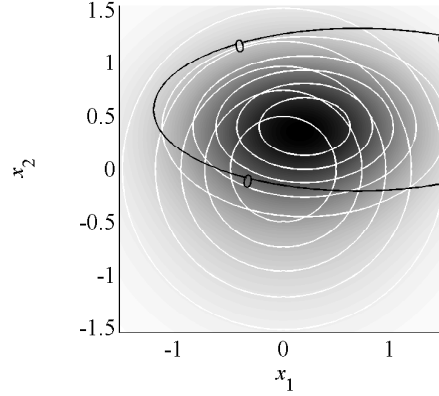


Fig. 5.5 Elliptic decision boundary of the ML detectors and likelihoods for a Gaussian detection problem with 2D observations.

5.4.0.1 Identical cross-covariance matrices

This section considers the case of $\mathbf{V}_1 = \mathbf{V}_0 = \mathbf{V}$. Then, the LLRT becomes

$$\begin{array}{l} D = 1 \\ (\mathbf{x} - \mathbf{m}_0)^T \mathbf{V}^{-1} (\mathbf{x} - \mathbf{m}_0) - (\mathbf{x} - \mathbf{m}_1)^T \mathbf{V}^{-1} (\mathbf{x} - \mathbf{m}_1) \geq \mu. \\ D = 0 \end{array}$$

Now, expanding the quadratic forms, the above expression simplifies to

$$\begin{array}{l} D = 1 \\ (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{V}^{-1} \mathbf{x} \geq \tilde{\mu}, \\ D = 0 \end{array} \quad (5.14)$$

where $\tilde{\mu} = \mu/2 + \mathbf{m}_1^T \mathbf{V}^{-1} \mathbf{m}_1/2 - \mathbf{m}_0^T \mathbf{V}^{-1} \mathbf{m}_0/2$. In this particular case, the LLRT in (5.14) is a linear function of the observation vector \mathbf{x} .

Example 5.8 Figure 5.6 shows three decision boundaries for an example with

$$\mathbf{m}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{V}_0 = \begin{pmatrix} 0.44 & 0.32 \\ 0.32 & 0.81 \end{pmatrix}$$

and

$$\mathbf{m}_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \mathbf{V}_1 = \begin{pmatrix} 0.44 & 0.32 \\ 0.32 & 0.81 \end{pmatrix}.$$

The label of each decision boundary is $\log(\eta)$. Then, $\log(\eta) = 0$ corresponds to the ML detector.

Example 5.9 (Matched filter) In this example, we derive one of the most well-known detectors, the matched filter (MF). The MF is the LLRT to the detection of a known signal contaminated by zero-mean Gaussian noise. Concretely, under hypothesis $h = 0$, the observations are given by noise only:

$$x[n] = w[n], \quad n = 0, \dots, N-1,$$

and under hypothesis $h = 1$, the observations are

$$x[n] = s[n] + w[n], \quad n = 0, \dots, N-1,$$

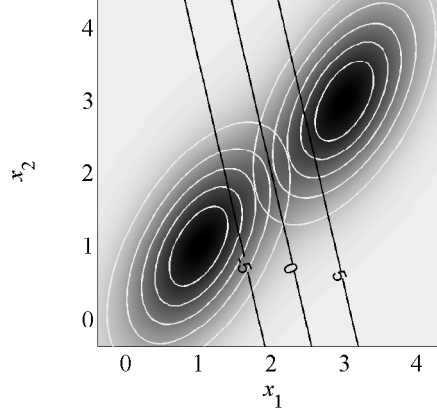


Fig. 5.6 Decision boundaries of the LLRT and likelihoods for a Gaussian detection problem with 2D observations and identical covariance matrices.

where $s[n]$ is a known signal and $w[n]$ is additive white Gaussian noise with zero mean and variance σ^2 , i.e., $w[n] \sim \mathcal{N}(0, \sigma^2)$. To use the LLRT already derived in this section, we must first define the vector

$$\mathbf{x} = (x[0] \ x[1] \ \dots \ x[N-1])^T = \mathbf{s} + \mathbf{w},$$

with $\mathbf{s} = (s[0] \ s[1] \ \dots \ s[N-1])^T$ and $\mathbf{w} = (w[0] \ w[1] \ \dots \ w[N-1])^T$, and obtain the distributions of \mathbf{x} under both hypothesis. Under hypothesis $h = 0$, the observation vector \mathbf{x} collects samples of a Gaussian process, which makes it also Gaussian. Hence, only the mean and cross-covariance matrices are required:

$$\mathbf{m}_0 = \mathbb{E}\{\mathbf{x}|0\} = \mathbb{E}\{\mathbf{w}\} = (\mathbb{E}\{w[0]\} \ \mathbb{E}\{w[1]\} \ \dots \ \mathbb{E}\{w[N-1]\})^T = \mathbf{0},$$

and

$$\begin{aligned} \mathbf{V}_0 &= \mathbb{E}\{(\mathbf{x} - \mathbf{m}_0)(\mathbf{x} - \mathbf{m}_0)^T | 0\} = \mathbb{E}\{\mathbf{w}\mathbf{w}^T\} \\ &= \mathbb{E}\left\{ \begin{pmatrix} w[0] & w[1] & \dots & w[N-1] \end{pmatrix}^T \begin{pmatrix} w[0] & w[1] & \dots & w[N-1] \end{pmatrix} \right\} \\ &= \begin{pmatrix} \mathbb{E}\{w^2[0]\} & \mathbb{E}\{w[0]w[1]\} & \dots & \mathbb{E}\{w[0]w[N-1]\} \\ \mathbb{E}\{w[1]w[0]\} & \mathbb{E}\{w^2[1]\} & \dots & \mathbb{E}\{w[1]w[N-1]\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{w[N-1]w[0]\} & \mathbb{E}\{w[N-1]w[1]\} & \dots & \mathbb{E}\{w^2[N-1]\} \end{pmatrix}. \end{aligned}$$

The cross-covariance matrix \mathbf{V}_0 can be simplified taking into account that the noise is white, i.e., $\mathbb{E}\{w[n]w[n-m]\} = \sigma^2 \delta[m]$, which yields

$$\mathbf{V}_0 = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}.$$

Similarly, under hypothesis $h = 1$, the observations are Gaussian with mean

$$\mathbf{m}_1 = \mathbb{E}\{\mathbf{x}|1\} = \mathbb{E}\{\mathbf{s} + \mathbf{w}\} = \mathbb{E}\{\mathbf{s}\} + \mathbb{E}\{\mathbf{w}\} = \mathbf{s},$$

since \mathbf{s} is deterministic, and cross-covariance matrix

$$\mathbf{V}_1 = \mathbb{E}\{(\mathbf{x} - \mathbf{m}_1)(\mathbf{x} - \mathbf{m}_1)^T | 1\} = \mathbb{E}\{(\mathbf{s} + \mathbf{w} - \mathbf{s})(\mathbf{s} + \mathbf{w} - \mathbf{s})^T\} = \mathbb{E}\{\mathbf{w}\mathbf{w}^T\} = \sigma^2 \mathbf{I}.$$

Hence, the detection problem is that of Gaussian observations with identical covariance matrices, for which the LLRT is

$$(\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{V}^{-1} \mathbf{x} = \frac{1}{\sigma^2} \mathbf{s}^T \mathbf{x} \underset{D=0}{\overset{D=1}{\geq}} \tilde{\mu} \Rightarrow \underbrace{\sum_{n=0}^{N-1} s[n]x[n]}_{MF} \underset{D=0}{\overset{D=1}{\geq}} \sigma^2 \tilde{\mu}.$$

Alternatively, and the motivation for the term matched filter, is because the above detector can be rewritten as a filtering of the signal $x[n]$ with the filter $h[n] = s[N-1-n]$, followed by sampling every N samples. Finally, we also would like to point out that the matched filter is a filter that maximizes the signal-to-noise ratio.

5.4.0.2 Zero means

We consider now that $\mathbf{m}_0 = \mathbf{m}_1 = \mathbf{0}$, which yields

$$\mathbf{x}^T (\mathbf{V}_0^{-1} - \mathbf{V}_1^{-1}) \mathbf{x} \underset{D=0}{\overset{D=1}{\geq}} \mu.$$

Example 5.10 Figure 5.7 shows the ML decision boundary for 2D Gaussian observations with

$$\mathbf{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{V}_0 = \begin{pmatrix} 0.62 & -0.22 \\ -0.22 & 0.37 \end{pmatrix},$$

and

$$\mathbf{m}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{V}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

The region \mathcal{R}_0 is given by the interior of the ellipse. Moreover, since the variance of the observations in every direction is larger under hypothesis $h = 1$, points further away from the origin should be assigned $d = 1$.

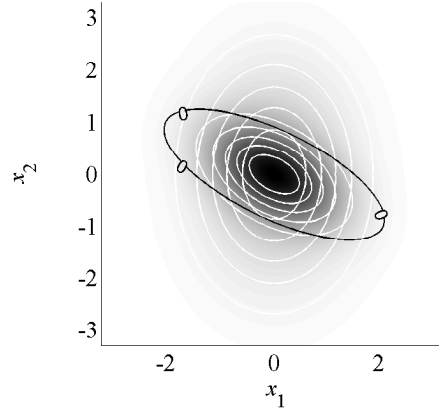


Fig. 5.7 Elliptic decision boundary for a 2D Gaussian problem with zero means.

Example 5.11 Figure 5.8 shows the ML decision boundary for 2D Gaussian observations with

$$\mathbf{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{V}_0 = \begin{pmatrix} 0.33 & 0.39 \\ 0.39 & 0.77 \end{pmatrix}$$

and

$$\mathbf{m}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{v}_1 = \begin{pmatrix} 0.39 & -0.19 \\ -0.19 & 0.16 \end{pmatrix}.$$

In this example, the variance under hypothesis $h = 1$ is larger only along dimension 1, whereas it is smaller along dimension 2. Hence, as a consequence, the boundary is a hyperbola.

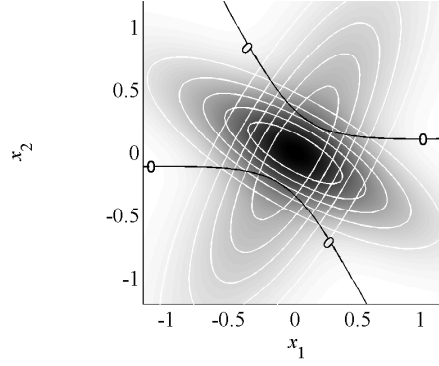


Fig. 5.8 Hyperbolic decision boundary for a 2D Gaussian problem with zero means.

5.5 Problems

5.1 Consider the decision problem with three hypotheses given by the observation $\mathbf{x} = (x_1, x_2) \in [0, 1]^2$ and likelihoods:

$$p_{\mathbf{X}|H}(\mathbf{x}|0) = 2(1 - x_1) \quad (5.15)$$

$$p_{\mathbf{X}|H}(\mathbf{x}|1) = 2x_1 \quad (5.16)$$

$$p_{\mathbf{X}|H}(\mathbf{x}|2) = 2x_2 \quad (5.17)$$

- a) Determine the ML (Maximum Likelihood) decider
- b) Represent the decision regions.

5.2 Consider the decision problem given by the observation $x \in [0, 1]$, likelihoods:

$$p_{X|H}(x|0) = 2(1 - x) \quad (5.18)$$

$$p_{X|H}(x|1) = 1 \quad (5.19)$$

and prior probability $P_H(1) = 1/4$.

- a) Determine the ML decider.
- b) Determine the MAP (Maximum A Posteriori) decider.
- c) Given that $c_{01} = 2$, $c_{10} = 1$, $c_{11} = c_{00} = 0$, determine the decider of minimum average cost.
- d) Consider a threshold decider over x in the form: