

AN INTRODUCTION TO ESTIMATION AND DETECTION THEORY¹

Armand M. Makowski² and Prakash L.K. Narayan³

¹©1994-2016 by Armand M. Makowski and Prakash L.K. Narayan

²Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742. E-mail: armand@isr.umd.edu. Phone: (301) 405-6844

³Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742. E-mail: prakash@isr.umd.edu. Phone: (301) 405-3661

Notation and conventions

Throughout, we use \mathbb{R} to denote the set of all real numbers, or equivalently, the real line $(-\infty, \infty)$. The set of all non-negative real numbers is denoted by \mathbb{R}_+ . The set of $\{0, 1, \dots\}$ of all non-negative integers is denoted \mathbb{N} . The notation \mathbb{N}_0 will be used for the set $\{1, 2, \dots\}$ of all positive integers.

With p a positive integer, let \mathbb{R}^p denote the p^{th} cartesian product of \mathbb{R} . An element \mathbf{x} in \mathbb{R}^p , whose p components are denoted x_1, \dots, x_p , is always interpreted as a column vector $(x_1, \dots, x_p)'$ (with $'$ denoting transpose).

With scalar m and $\sigma^2 > 0$, the rv X is said to be a Gaussian rv with mean m and variance σ^2 , written $X \sim N(m, \sigma^2)$, if its cumulative probability distribution function is given by

$$\mathbb{P}[X \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\xi-m}{\sigma}\right)^2} d\xi, \quad x \in \mathbb{R}.$$

The zero mean unit variance Gaussian rv is often referred to as a standard Gaussian rv; its probability density function $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}, \tag{1}$$

and its cumulative probability distribution function is then

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt, \quad x \in \mathbb{R}. \tag{2}$$

Obviously, if $X \sim N(m, \sigma^2)$ and $Z \sim N(0, 1)$, then X and $m + \sigma Z$ have the same distribution.

In the context of digital communications, it is customary to use the Q -function $Q : \mathbb{R}_+ \rightarrow [0, 1]$ given by

$$Q(x) = 1 - \Phi(x) = \int_x^\infty \varphi(t) dt, \quad x \geq 0.$$

Part I
Detection Theory

Chapter 1

Simple hypothesis testing

In the statistical hypothesis testing problem, a decision has to be made as to which of several hypotheses (or states of nature) is the correct one. The states of nature are encoded in a rv H and a decision has to be made on the basis of an observation Y which is statistically related to H .

1.1 Motivating examples

Control process

A simple communication example

Testing means

1.2 The probabilistic model

The *binary* hypothesis testing problem is the simplest version of this problem in that nature can be in either of two states, say $H = 0$ or $H = 1$ for sake of concreteness. If F_0 and F_1 are probability distribution functions on \mathbb{R}^k , this situation is summarized by

$$\begin{aligned} H_1 : \mathbf{Y} &\sim F_1 \\ H_0 : \mathbf{Y} &\sim F_0. \end{aligned} \tag{1.1}$$

The hypothesis H_0 is called the *null hypothesis* and hypothesis H_1 is referred to as the *non-null hypothesis* or the *alternative*.

Probabilistically, the symbolic statement (1.1) is understood as follows: On some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, consider the rvs H and \mathbf{Y} defined on Ω which take values in $\{0, 1\}$ and \mathbb{R}^k , respectively. The probability distribution functions F_0 and F_1 then have the interpretation as conditional probability distribution of \mathbf{Y} given $H = 0$ and $H = 1$, respectively, namely

$$F_h(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = h], \quad \begin{array}{l} \mathbf{y} \in \mathbb{R}^k, \\ h = 0, 1. \end{array}$$

The probability distribution of the rv H is specified by p in $[0, 1]$ with

$$p = \mathbb{P}[H = 1] = 1 - \mathbb{P}[H = 0].$$

We refer to the pmf $(1-p, p)$ on $\{0, 1\}$, or just to p , as the *prior*. This construction is always possible.

Since

$$\begin{aligned} \mathbb{P}[\mathbf{Y} \leq \mathbf{y}, H = h] &= \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = h] \mathbb{P}[H = h] \\ &= \begin{cases} (1-p)F_0(\mathbf{y}) & \text{if } h = 0, \mathbf{y} \in \mathbb{R}^k \\ pF_1(\mathbf{y}) & \text{if } h = 1, \mathbf{y} \in \mathbb{R}^k, \end{cases} \end{aligned} \quad (1.2)$$

then the law of total probability shows that

$$\begin{aligned} \mathbb{P}[\mathbf{Y} \leq \mathbf{y}] &= \sum_{h=0}^1 \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = h] \mathbb{P}[H = h] \\ &= pF_1(\mathbf{y}) + (1-p)F_0(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \quad (1.3)$$

Thus, the conditional probability distributions of the observations given the hypothesis *and* the probability distribution of H completely specify the *joint* distribution of the rvs H and \mathbf{Y} .

During the discussion, several assumptions will be enforced on the probability distributions F_0 and F_1 . The assumptions that will be most often encountered are denoted by **(A.1)** and **(A.2)** for sake of convenience. They are stated and discussed in some details below.

Condition **(A.1)** corresponds to the assumption of usual absolute continuity that was used in Part II. The probability distributions F_0 and F_1 on \mathbb{R}^k satisfy

Condition **(A.1)** if they are both *absolutely continuous* with respect to some distribution F on \mathbb{R}^k – In general F may not be a probability distribution. Condition **(A.1)** is equivalent to saying that there exist Borel mappings $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ such that

$$F_h(\mathbf{y}) = \int_{-\infty}^{\mathbf{y}} f_h(\boldsymbol{\eta}) dF(\boldsymbol{\eta}), \quad \begin{array}{l} \mathbf{y} \in \mathbb{R}^k, \\ h = 0, 1. \end{array} \quad (1.4)$$

In some basic sense, this condition is hardly constraining since we can always take F to be the average of the two probability distributions F_0 and F_1 . i.e.,

$$F(\mathbf{y}) \equiv \frac{1}{2}F_0(\mathbf{y}) + \frac{1}{2}F_1(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \quad (1.5)$$

in which case F is also a probability distribution. This choice for F is usually not operationally convenient and therefore discarded. However, the most often encountered situations arise when F is either Lebesgue measure on \mathbb{R}^k or a counting measure on some countable subset of \mathbb{R}^k , in which case F is not a probability distribution.

When F is Lebesgue measure on \mathbb{R}^k , the Borel mappings $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ are just the probability density functions induced by F_0 and F_1 in the usual sense. When F is counting measure on a countable subset $S \subseteq \mathbb{R}^k$, then the Borel mappings $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ are best thought as *probability mass functions* (pdfs) $\mathbf{f}_0 = \{f_0(\mathbf{y}), \mathbf{y} \in S\}$ and $\mathbf{f}_1 = \{f_1(\mathbf{y}), \mathbf{y} \in S\}$, i.e.,

$$0 \leq f_h(\mathbf{y}) \leq 1, \quad \begin{array}{l} \mathbf{y} \in S, \\ h = 0, 1. \end{array}$$

and

$$\sum_{\mathbf{y} \in S} f_h(\mathbf{y}) = 1, \quad h = 0, 1.$$

The condition (1.4) now takes the form

$$\mathbb{P}[\mathbf{Y} \in B | H = h] = \sum_{\boldsymbol{\eta} \in S \cap B} f_h(\boldsymbol{\eta}), \quad \begin{array}{l} B \in \mathcal{B}(\mathbb{R}^k) \\ h = 0, 1. \end{array}$$

The second assumption of interest here is Condition **(A.2)** which asserts that the probability distribution F_1 is *absolutely continuous* with respect to the probability distribution F_0 . Under Condition **(A.1)**, with the notation introduced earlier, this is equivalent to requiring

$$f_0(\mathbf{y}) = 0 \quad \text{implies} \quad f_1(\mathbf{y}) = 0. \quad (1.6)$$

1.3 Admissible tests

An *admissible* decision rule (or test) is any *Borel* mapping $d : \mathbb{R}^k \rightarrow \{0, 1\}$. The collection of all admissible rules is denoted by \mathcal{D} . The measurability requirement entering the definition of admissibility is imposed to guarantee that the mapping $d(\mathbf{Y}) : \Omega \rightarrow \{0, 1\} : \omega \rightarrow d(Y(\omega))$ is indeed a rv, i. e., $[\omega \in \Omega : d(Y(\omega)) = h]$ is an event in \mathcal{F} for all $h = 0, 1$.

It is plain that every test d in \mathcal{D} is completely specified by the *Borel* subset $C(d)$ defined by

$$C(d) \equiv \{\mathbf{y} \in \mathbb{R}^k : d(\mathbf{y}) = 0\}. \quad (1.7)$$

Conversely, any Borel measurable subset C of \mathbb{R}^k uniquely determines an admissible rule d_C in \mathcal{D} through

$$d_C(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \notin C \\ 0 & \text{if } \mathbf{y} \in C. \end{cases}$$

We note that $C(d_C) = C$ as expected.

Any admissible rule d in \mathcal{D} induces *two* types of error: Upon observing \mathbf{Y} , either $H = 0$ is true and $d(\mathbf{Y}) = 1$ or $H = 1$ is true and $d(\mathbf{Y}) = 0$. These two possibilities are the so-called errors of the *first* and *second* type associated with the decision rule d ; they are quantified by

$$\alpha(d) \equiv \mathbb{P}[d(\mathbf{Y}) = 1 | H = 0] \quad (1.8)$$

and

$$\beta(d) \equiv \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1], \quad (1.9)$$

respectively. The quantity $\alpha(d)$ is sometimes called the *size* of the test d . In radar parlance, these probabilities are referred to as probabilities of *false alarm* and *miss*, respectively, with alternate notation

$$P_F(d) \equiv \mathbb{P}[d(\mathbf{Y}) = 1 | H = 0] \quad (1.10)$$

and

$$P_M(d) \equiv \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1]. \quad (1.11)$$

Throughout we shall use this terminology. Sometimes, it is convenient to consider the so-called probability of *detection* given by

$$P_D(d) \equiv \mathbb{P}[d(\mathbf{Y}) = 1 | H = 1] = 1 - P_M(d). \quad (1.12)$$

1.4 The Bayesian formulation

The Bayesian formulation assumes *knowledge* of the conditional distributions F_1 and F_0 , and of the prior distribution p of the rv H .

The Bayesian optimization problem The cost incurred for making decisions is quantified by the mapping $C : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$ with the interpretation that

$$C(h, d) = \left(\begin{array}{l} \text{Cost incurred for deciding } d \\ \text{when } H = h \end{array} \right), \quad d, h = 0, 1.$$

As the sample ω in Ω is realized, the observation $Y(\omega)$ is recorded and the use of the admissible rule d in \mathcal{D} incurs a cost $C(H(\omega), d(\mathbf{Y}(\omega)))$. Although it is tempting to seek to minimize this quantity, this is not possible. Indeed, the rv \mathbf{Y} is observed, whence $d(\mathbf{Y})$ is known once the test d has been specified, but the state of nature H is *not* directly observable. Consequently, the value of the cost $C(H, d(\mathbf{Y}))$ is not available. To remedy to this difficulty, we introduce the *expected cost function* $J : \mathcal{D} \rightarrow \mathbb{R}$ given by

$$J(d) \equiv \mathbb{E}[C(H, d(\mathbf{Y}))], \quad d \in \mathcal{D}.$$

The *Bayesian Problem* \mathcal{P}_B is the minimization problem

$$\mathcal{P}_B : \quad \text{Minimize } J(d) \text{ over } d \text{ in } \mathcal{D}.$$

This amounts to finding an admissible test d^* in \mathcal{D} such that

$$J(d^*) \leq J(d), \quad d \in \mathcal{D}. \quad (1.13)$$

Any admissible test d^* which satisfies (1.13) is called a Bayesian test, and the value

$$J(d^*) = \inf_{d \in \mathcal{D}} J(d) = \min_{d \in \mathcal{D}} J(d) \quad (1.14)$$

is known as the *Bayesian cost*.

Representation results for the Bayesian cost J The solution to the Bayesian problem \mathcal{P}_B is developed with the help of an auxiliary result concerning the form of the Bayesian cost. This representation result will be useful in several places and is given here for sake of easy reference.

Fix d in \mathcal{D} . Recall that the rvs H and $d(\mathbf{Y})$ are $\{0, 1\}$ -valued rvs, and that the events $[d(\mathbf{Y}) = H]$ and $[d(\mathbf{Y}) \neq H]$ form a partition of Ω , i. e.,

$$\mathbf{1} [d(\mathbf{Y}) = H] + \mathbf{1} [d(\mathbf{Y}) \neq H] = \mathbf{1} [\Omega] = 1.$$

It readily follows that

$$\begin{aligned} C(H, d(\mathbf{Y})) &= \mathbf{1} [d(\mathbf{Y}) = H] C(H, H) + \mathbf{1} [d(\mathbf{Y}) \neq H] C(H, 1 - H) \\ &= (1 - \mathbf{1} [d(\mathbf{Y}) \neq H]) C(H, H) + \mathbf{1} [d(\mathbf{Y}) \neq H] C(H, 1 - H) \\ &= C(H, H) + (C(H, 1 - H) - C(H, H)) \mathbf{1} [d(\mathbf{Y}) \neq H] \\ &= C(H, H) + \mathbf{1} [d(\mathbf{Y}) \neq H] \Gamma_H \end{aligned} \quad (1.15)$$

as we introduce the relative costs Γ_0 and Γ_1 given by

$$\Gamma_h \equiv C(h, 1 - h) - C(h, h), \quad h = 0, 1.$$

Taking expectations on both sides of (1.15) we find

$$J(d) = \mathbb{E} [C(H, H)] + \mathbb{E} [\mathbf{1} [d(\mathbf{Y}) \neq H] \Gamma_H].$$

This last relation points to the auxiliary expected cost function $\hat{J} : \mathcal{D} \rightarrow \mathbb{R}$ defined by

$$\hat{J}(d) = \mathbb{E} [\mathbf{1} [d(\mathbf{Y}) \neq H] \Gamma_H], \quad d \in \mathcal{D} \quad (1.16)$$

so that

$$J(d) = \mathbb{E} [C(H, H)] + \hat{J}(d), \quad d \in \mathcal{D}. \quad (1.17)$$

The law of total probabilities gives

$$\begin{aligned} \hat{J}(d) &= \mathbb{E} [\Gamma_0 \mathbf{1} [d(\mathbf{Y}) \neq 0] \mathbf{1} [H = 0] + \Gamma_1 \mathbf{1} [d(\mathbf{Y}) \neq 1] \mathbf{1} [H = 1]] \\ &= \Gamma_0(1 - p) \cdot \mathbb{P} [d(\mathbf{Y}) \neq 0 | H = 0] + \Gamma_1 p \cdot \mathbb{P} [d(\mathbf{Y}) \neq 1 | H = 1] \\ &= \Gamma_0(1 - p) \cdot \mathbb{P} [d(\mathbf{Y}) = 1 | H = 0] + \Gamma_1 p \cdot \mathbb{P} [d(\mathbf{Y}) = 0 | H = 1], \end{aligned} \quad (1.18)$$

and direct substitution easily yields the following expressions.

Lemma 1.4.1 *For any admissible rule d in \mathcal{D} , the relations*

$$\hat{J}(d) = \Gamma_0(1 - p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d) \quad (1.19)$$

and

$$J(d) = \mathbb{E} [C(H, H)] + \Gamma_0(1 - p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d) \quad (1.20)$$

hold.

It is plain from these expressions that the Bayesian cost under a decision rule is completely determined by its probabilities of false alarm and of miss. We have

$$\begin{aligned} \widehat{J}(d) &= \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1] \\ &\quad - \Gamma_0(1-p) \cdot \mathbb{P}[d(\mathbf{Y}) = 0 | H = 0], \quad d \in \mathcal{D}. \end{aligned} \quad (1.21)$$

as an immediate consequence of (1.18).

1.5 Solving the Bayesian problem \mathcal{P}_B

It follows from (1.17) that solving \mathcal{P}_B is *equivalent* to solving the auxiliary problem $\widehat{\mathcal{P}}_B$ where

$$\widehat{\mathcal{P}}_B : \quad \text{Minimize } \widehat{J}(d) \text{ over } d \text{ in } \mathcal{D}.$$

To solve this auxiliary problem $\widehat{\mathcal{P}}_B$, it will be necessary to assume that the probability distributions F_0 and F_1 satisfy the absolute continuity condition **(A1)** given earlier, namely that there exists a single distribution F on \mathbb{R}^k with respect to which both F_0 and F_1 are absolutely continuous. For any test d in \mathcal{D} , we get

$$\begin{aligned} \mathbb{P}[d(\mathbf{Y}) = 0 | H = h] &= \int_{C(d)} dF_h(\mathbf{y}) \\ &= \int_{C(d)} f_h(\mathbf{y}) dF(\mathbf{y}), \quad h = 0, 1 \end{aligned} \quad (1.22)$$

with $C(d)$ defined at (1.7). It is now easy to see from (1.21) that

$$\widehat{J}(d) = \Gamma_0(1-p) + \int_{C(d)} h(\mathbf{y}) dF(\mathbf{y}) \quad (1.23)$$

where the mapping $h : \mathbb{R}^k \rightarrow \mathbb{R}$ is given by

$$h(\mathbf{y}) := \Gamma_1 p \cdot f_1(\mathbf{y}) - \Gamma_0(1-p) \cdot f_0(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \quad (1.24)$$

Theorem 1.5.1 *Assume the absolute continuity condition **(A.1)** to hold. Define the Borel set C^* by*

$$C^* \equiv \{y \in \mathbb{R}^k : h(\mathbf{y}) < 0\}$$

with $h : \mathbb{R}^k \rightarrow \mathbb{R}$ given by (1.24). The decision rule $d^ : \mathbb{R}^k \rightarrow \{0, 1\}$ given by*

$$d^*(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \notin C^* \\ 0 & \text{if } \mathbf{y} \in C^* \end{cases} \quad (1.25)$$

is admissible and solves the Problem $\widehat{\mathcal{P}}_B$, hence solves the Bayesian Problem \mathcal{P}_B .

Proof. The set C^* is a Borel subset of \mathbb{R}^k due to the fact that the functions $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ are themselves Borel measurable. The test d^* is therefore an admissible decision rule in \mathcal{D} since $C(d^*) = C^*$. We now show that d^* satisfies

$$\widehat{J}(d^*) \leq \widehat{J}(d), \quad d \in \mathcal{D}. \quad (1.26)$$

Indeed, for every test d in \mathcal{D} , we see from (1.23) that

$$\widehat{J}(d) = \Gamma_0(1-p) + \int_{C(d) \setminus C^*} h(\mathbf{y}) dF(\mathbf{y}) + \int_{C(d) \cap C^*} h(\mathbf{y}) dF(\mathbf{y})$$

and

$$\widehat{J}(d^*) = \Gamma_0(1-p) + \int_{C^* \setminus C(d)} h(\mathbf{y}) dF(\mathbf{y}) + \int_{C(d) \cap C^*} h(\mathbf{y}) dF(\mathbf{y}).$$

Therefore,

$$\widehat{J}(d) - \widehat{J}(d^*) = \int_{C(d) \setminus C^*} h(\mathbf{y}) dF(\mathbf{y}) + \int_{C^* \setminus C(d)} (-h(\mathbf{y})) dF(\mathbf{y}) \geq 0$$

since

$$\int_{C(d) \setminus C^*} h(\mathbf{y}) dF(\mathbf{y}) \geq 0 \quad \text{and} \quad \int_{C^* \setminus C(d)} h(\mathbf{y}) dF(\mathbf{y}) \leq 0$$

by the very definition of C^* . The problem $\widehat{\mathcal{P}}_B$ is therefore solved by the test d^* defined at (1.25). \blacksquare

The solution to the Bayesian problem is *not* unique: It should be plain that C^* could be replaced by

$$C^{**} \equiv \{y \in \mathbb{R}^k : h(\mathbf{y}) \leq 0\}$$

(with corresponding test d^{**}) without affecting the conclusion of optimality since

$$\int_{\{\mathbf{y} \in \mathbb{R}^k : h(\mathbf{y})=0\}} h(\mathbf{y}) dF(\mathbf{y}) = 0.$$

While it is true that $J(d^*) = J(d^{**})$, it is not necessarily the case that either of the equalities $P_F(d^*) = P_F(d^{**})$ or $P_M(d^*) = P_M(d^{**})$ holds.

1.6 Likelihood ratio tests

Assume that $0 < p < 1$ to avoid trivial situations, and that the relative costs satisfy the conditions

$$\Gamma_h > 0, \quad h = 0, 1,$$

i.e., the cost of making an incorrect decision is greater than the cost of making a correct decision. This is of course a most reasonable assumption which always holds in applications. Under this condition, the Bayesian decision rule d^* given in Theorem 1.5.1 takes the equivalent form

$$d^*(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < \frac{\Gamma_0(1-p)}{\Gamma_1 p} f_0(\mathbf{y}). \quad (1.27)$$

This suggests introducing the following class of admissible tests $\{d_\eta, \eta \geq 0\}$ where for each $\eta \geq 0$, the mapping $d_\eta : \mathbb{R}^k \rightarrow \{0, 1\}$ is defined by

$$d_\eta(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < \eta f_0(\mathbf{y}). \quad (1.28)$$

The Bayesian test d^* described in Section 1.4 is such a test d_η with

$$\eta \equiv \frac{\Gamma_0(1-p)}{\Gamma_1 p}.$$

It is plain from the definition (1.28) (with $\eta = 0$) that d_0 is simply the test that always selects the non-null hypothesis $H = 1$. On the other hand, substituting $\eta = \infty$ in (1.28) will be problematic at observation points where $f_0(\mathbf{y}) = 0$. However, by *convention* we shall interpret d_∞ as the test that always selects the null hypothesis $H = 0$ – More on that later in Section 1.20.

Such tests take an even simpler form under the additional Condition **(A.2)** as will be seen shortly. First, we note that (1.28) can be rewritten as

$$d_\eta(\mathbf{y}) = 0 \quad \text{if} \quad \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} < \eta \quad \text{whenever} \quad f_0(\mathbf{y}) > 0.$$

Taking our cue from this last statement, we define the *likelihood ratio* as any Borel mapping $L : \mathbb{R}^k \rightarrow \mathbb{R}$ of the form

$$L(\mathbf{y}) \equiv \begin{cases} \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})} & \text{if } f_0(\mathbf{y}) > 0 \\ \Lambda(\mathbf{y}) & \text{otherwise} \end{cases} \quad (1.29)$$

for some arbitrary Borel mapping $\Lambda : \mathbb{R}^k \rightarrow \mathbb{R}_+$. Different choices of this arbitrary non-negative function produce different versions of the likelihood ratio function.

Given a version of the likelihood ratio function in (1.29), we define the *likelihood ratio* test with *threshold* $\eta \geq 0$ to be the admissible decision rule $Lrt_\eta : \mathbb{R}^k \rightarrow \{0, 1\}$ given by

$$Lrt_\eta(\mathbf{y}) \equiv \begin{cases} 1 & \text{if } L(\mathbf{y}) \geq \eta \\ 0 & \text{if } L(\mathbf{y}) < \eta. \end{cases} \quad (1.30)$$

With

$$B_h = \{\mathbf{y} \in \mathbb{R}^k : f_h(\mathbf{y}) = 0\}, \quad h = 0, 1,$$

we note that

$$\mathbb{P}[f_0(\mathbf{Y}) = 0 | H = h] = \int_{B_0} f_h(\mathbf{y}) dF(\mathbf{y}), \quad h = 0, 1. \quad (1.31)$$

Under **(A.2)**, as the inclusion $B_0 \subseteq B_1$ holds, we conclude that

$$\mathbb{P}[f_0(\mathbf{Y}) = 0 | H = h] = 0, \quad h = 0, 1.$$

For any value η of the threshold it is plain that the tests d_η and Lrt_η coincide on the set $\{\mathbf{y} \in \mathbb{R}^k : f_0(\mathbf{y}) > 0\}$ (while possibly disagreeing on the complement B_0). Thus, for each $h = 0, 1$, we find that

$$\begin{aligned} & \mathbb{P}[d_\eta(\mathbf{Y}) = 0 | H = h] \\ &= \mathbb{P}[d_\eta(\mathbf{Y}) = 0, f_0(\mathbf{Y}) > 0 | H = h] + \mathbb{P}[d_\eta(\mathbf{Y}) = 0, f_0(\mathbf{Y}) = 0 | H = h] \\ &= \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 0, f_0(\mathbf{Y}) > 0 | H = h] \\ &= \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 0, f_0(\mathbf{Y}) > 0 | H = h] + \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 0, f_0(\mathbf{Y}) = 0 | H = h] \\ &= \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 0 | H = h]. \end{aligned}$$

This discussion leads to the following fact.

Lemma 1.6.1 *Assume the absolute continuity conditions **(A.1)**–**(A.2)** to hold. For each $\eta \geq 0$, the tests d_η and Lrt_η are equivalent in that they have identical performance with $P_M(d_\eta) = P_M(Lrt_\eta)$ and $P_F(d_\eta) = P_F(Lrt_\eta)$.*

It follows from (1.20) that $J(d_\eta) = J(Lrt_\eta)$ regardless of the cost function $C : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$. The same argument also shows that any two versions of the likelihood ratio function will generate likelihood ratio tests which are equivalent.

Equipped with Lemma 1.6.1 we can now restate Theorem 1.5.1.

Theorem 1.6.1 *Assume the absolute continuity conditions (A.1)–(A.2) to hold. Whenever $\Gamma_h > 0$ for $h = 0, 1$, the Bayesian decision rule d^* identified in Theorem 2.1 is equivalent to the likelihood ratio test Lrt_{η^*} where*

$$\eta^* \equiv \frac{\Gamma_0(1-p)}{\Gamma_1 p} = \frac{C(0,1) - C(0,0)}{C(1,0) - C(1,1)} \cdot \frac{1-p}{p}.$$

1.7 The probability of error criterion

A special case of great interest is obtained when the cost function C takes the form

$$C(h, d) = \mathbf{1}[h \neq d], \quad h, d = 0, 1.$$

The corresponding expected cost then reduces to the probability of making an incorrect decision, namely the *probability of error*, and is given by

$$P_E(d) \equiv \mathbb{P}[d(\mathbf{Y}) \neq H], \quad d \in \mathcal{D}.$$

We check that

$$\Gamma_h = C(h, 1-h) - C(h, h) = 1, \quad h = 0, 1,$$

and the relation (1.20) becomes

$$\begin{aligned} P_E(d) &= (1-p) \cdot P_F(d) + p \cdot P_M(d) \\ &= p + (1-p) \cdot P_F(d) - p \cdot P_D(d), \quad d \in \mathcal{D}. \end{aligned} \quad (1.32)$$

For the probability of error criterion, the threshold η^* appearing in Theorem 1.6.1 has the simpler form

$$\eta^* = \frac{1-p}{p}.$$

The optimal decision rule d^* , as described at (1.27), can now be rewritten as

$$d^*(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < \frac{1-p}{p} f_0(\mathbf{y}). \quad (1.33)$$

The ML test In the uniform prior case, i.e., $p = \frac{1}{2}$, the Bayesian test (1.33) becomes

$$d^*(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < f_0(\mathbf{y}). \quad (1.34)$$

In other words, the optimal decision is to select that hypothesis whose likelihood is largest given the observation \mathbf{y} . We refer to this strategy as the *Maximum Likelihood* (ML) test.

The MAP computer Finally, (1.33) can also be rewritten as

$$d^*(\mathbf{y}) = 0 \quad \text{iff} \quad \mathbb{P}[H = 1 | \mathbf{Y} = \mathbf{y}] < \mathbb{P}[H = 0 | \mathbf{Y} = \mathbf{y}] \quad (1.35)$$

since for each \mathbf{y} in \mathbb{R}^k , we have

$$\mathbb{P}[H = 1 | \mathbf{Y} = \mathbf{y}] = \frac{pf_1(\mathbf{y})}{pf_1(\mathbf{y}) + (1-p)f_0(\mathbf{y})}$$

and

$$\mathbb{P}[H = 0 | \mathbf{Y} = \mathbf{y}] = \frac{(1-p)f_0(\mathbf{y})}{pf_1(\mathbf{y}) + (1-p)f_0(\mathbf{y})}$$

by Bayes' Theorem. For each $h = 0, 1$, the conditional probability $\mathbb{P}[H = h | \mathbf{Y} = \mathbf{y}]$ is known as the *posterior* probability that $H = h$ occurs given the observation \mathbf{y} . Put differently, the optimal test (1.35) compares these posterior probabilities given the observation \mathbf{y} , and selects the hypothesis with the largest posterior probability, hence the terminology *Maximum A Posteriori* (MAP) computer.

1.8 The Gaussian case

Assume that the observation rv \mathbf{Y} is conditionally Gaussian given H , i.e.,

$$\begin{aligned} H_1 : \quad \mathbf{Y} &\sim N(\mathbf{m}_1, \mathbf{R}_1) \\ H_0 : \quad \mathbf{Y} &\sim N(\mathbf{m}_0, \mathbf{R}_0) \end{aligned}$$

where \mathbf{m}_1 and \mathbf{m}_0 are elements in \mathbb{R}^k , and the $k \times k$ symmetric matrices \mathbf{R}_1 and \mathbf{R}_0 are *positive definite* (thus *invertible*). Throughout the pairs $(\mathbf{m}_0, \mathbf{R}_0)$ and $(\mathbf{m}_1, \mathbf{R}_1)$ are distinct so that the probability density functions $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ are distinct since

$$f_h(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^k \det \mathbf{R}_h}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m}_h)' \mathbf{R}_h^{-1} (\mathbf{y}-\mathbf{m}_h)}, \quad \mathbf{y} \in \mathbb{R}^k, \quad h = 0, 1.$$

Both conditions (A.1) and (A.2) obviously hold, and for each $\eta > 0$, the test d_η and Lrt_η coincide.

The likelihood ratio and the likelihood ratio tests For this example, the likelihood ratio function is given by

$$L(\mathbf{y}) = \sqrt{\frac{\det(\mathbf{R}_0)}{\det(\mathbf{R}_1)}} \cdot e^{\frac{1}{2}Q(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^k$$

where we have used the notation

$$Q(\mathbf{y}) = (\mathbf{y} - \mathbf{m}_0)' \mathbf{R}_0^{-1} (\mathbf{y} - \mathbf{m}_0) - (\mathbf{y} - \mathbf{m}_1)' \mathbf{R}_1^{-1} (\mathbf{y} - \mathbf{m}_1).$$

Fix $\eta > 0$. By direct substitution, we conclude that

$$Lrt_\eta(\mathbf{y}) = 0 \quad \text{iff} \quad e^{\frac{1}{2}Q(\mathbf{y})} < \sqrt{\eta^2 \cdot \frac{\det \mathbf{R}_1}{\det \mathbf{R}_0}},$$

and a simple logarithmic transformation yields

$$Lrt_\eta(\mathbf{y}) = 0 \quad \text{iff} \quad Q(\mathbf{y}) < \log \left(\eta^2 \frac{\det \mathbf{R}_1}{\det \mathbf{R}_0} \right).$$

The equal covariance case If the covariances are identical under both hypotheses, i.e.,

$$\mathbf{R}_0 = \mathbf{R}_1 \equiv \mathbf{R},$$

with $\mathbf{m}_1 \neq \mathbf{m}_0$, then

$$\begin{aligned} Q(\mathbf{y}) &= (\mathbf{y} - \mathbf{m}_0)' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{m}_0) - (\mathbf{y} - \mathbf{m}_1)' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{m}_1) \\ &= 2\mathbf{y}' \mathbf{R}^{-1} (\mathbf{m}_1 - \mathbf{m}_0) - (\mathbf{m}_1' \mathbf{R}^{-1} \mathbf{m}_1 - \mathbf{m}_0' \mathbf{R}^{-1} \mathbf{m}_0). \end{aligned} \quad (1.36)$$

The form of Lrt_η simplifies even further to read

$$Lrt_\eta(\mathbf{y}) = 0 \quad \text{iff} \quad \mathbf{y}' \mathbf{R}^{-1} \Delta \mathbf{m} < \tau(\eta)$$

where we have set

$$\Delta \mathbf{m} \equiv \mathbf{m}_1 - \mathbf{m}_0 \quad (1.37)$$

and

$$\tau(\eta) \equiv \frac{1}{2} (\mathbf{m}_1' \mathbf{R}^{-1} \mathbf{m}_1 - \mathbf{m}_0' \mathbf{R}^{-1} \mathbf{m}_0) + \log \eta. \quad (1.38)$$

Evaluating probabilities We will now evaluate the probabilities of false alarm and miss under Lrt_η . It is plain that

$$\begin{aligned} P_F(Lrt_\eta) &= \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 1 | H = 0] \\ &= \mathbb{P}[L(\mathbf{Y}) \geq \eta | H = 0] \\ &= \mathbb{P}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} \geq \tau(\eta) | H = 0] \end{aligned} \quad (1.39)$$

and

$$\begin{aligned} P_M(Lrt_\eta) &= \mathbb{P}[Lrt_\eta(\mathbf{Y}) = 0 | H = 1] \\ &= \mathbb{P}[L(\mathbf{Y}) < \eta | H = 1] \\ &= \mathbb{P}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} < \tau(\eta) | H = 1] \\ &= 1 - \mathbb{P}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} \geq \tau(\eta) | H = 1]. \end{aligned} \quad (1.40)$$

To carry out the calculations further, recall that for each $h = 0, 1$, given $H = h$, the rv \mathbf{Y} is conditionally Gaussian with mean vector \mathbf{m}_h and covariance matrix \mathbf{R} . Therefore, the scalar rv $\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m}$ is also conditionally Gaussian with mean and variance given by

$$\mathbb{E}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} | H = h] = \mathbf{m}'_h \mathbf{R}^{-1}\Delta\mathbf{m}$$

and

$$\begin{aligned} \text{Var}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} | H = h] &= (\mathbf{R}^{-1}\Delta\mathbf{m})' \text{Cov}[\mathbf{Y} | H = h] (\mathbf{R}^{-1}\Delta\mathbf{m}) \\ &= (\mathbf{R}^{-1}\Delta\mathbf{m})' \mathbf{R} (\mathbf{R}^{-1}\Delta\mathbf{m}) \\ &= \Delta\mathbf{m}' \mathbf{R}^{-1}\Delta\mathbf{m}, \end{aligned} \quad (1.41)$$

respectively. In obtaining this last relation we have used the fact that

$$\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} = (\mathbf{R}^{-1}\Delta\mathbf{m})'\mathbf{Y}.$$

Consequently, for all $h = 0, 1$,

$$\begin{aligned} &\mathbb{P}[\mathbf{Y}'\mathbf{R}^{-1}\Delta\mathbf{m} \geq \tau(\eta) | H = h] \\ &= \mathbb{P}\left[\mathbf{m}'_h \mathbf{R}^{-1}\Delta\mathbf{m} + \sqrt{\Delta\mathbf{m}' \mathbf{R}^{-1}\Delta\mathbf{m}} \cdot Z \geq \tau(\eta)\right] \\ &= \mathbb{P}\left[Z \geq \frac{\tau(\eta) - \mathbf{m}'_h \mathbf{R}^{-1}\Delta\mathbf{m}}{\sqrt{\Delta\mathbf{m}' \mathbf{R}^{-1}\Delta\mathbf{m}}}\right] \end{aligned} \quad (1.42)$$

where $Z \sim N(0, 1)$.

For the sake of convenience, pose

$$d^2 \equiv \Delta \mathbf{m}' \mathbf{R}^{-1} \Delta \mathbf{m}, \quad (1.43)$$

and note that

$$\tau(\eta) - \mathbf{m}'_h \mathbf{R}^{-1} \Delta \mathbf{m} = \begin{cases} \log \eta - \frac{1}{2} d^2 & \text{if } h = 1 \\ \log \eta + \frac{1}{2} d^2 & \text{if } h = 0. \end{cases}$$

It is now clear that

$$P_F(Lrt_\eta) = 1 - \Phi \left(\frac{\log \eta + \frac{1}{2} d^2}{d} \right)$$

and

$$P_M(Lrt_\eta) = \Phi \left(\frac{\log \eta - \frac{1}{2} d^2}{d} \right).$$

We finally obtain

$$P_D(Lrt_\eta) = 1 - \Phi \left(\frac{\log \eta - \frac{1}{2} d^2}{d} \right).$$

The ML test The ML test corresponds to $\eta = 1$, in which case these expressions become

$$P_F(d_{\text{ML}}) = 1 - \Phi \left(\frac{d}{2} \right) = Q \left(\frac{d}{2} \right)$$

and

$$P_M(d_{\text{ML}}) = \Phi \left(-\frac{d}{2} \right) = Q \left(\frac{d}{2} \right).$$

Therefore,

$$P_E(d_{\text{ML}}) = (1 - p)P_F(d_{\text{ML}}) + pP_M(d_{\text{ML}}) = Q \left(\frac{d}{2} \right)$$

regardless of the prior p .

1.9 The Bernoulli case

Consider now the binary hypothesis testing problem

$$\begin{aligned} H_1 &: Y \sim \text{Ber}(a_1) \\ H_0 &: Y \sim \text{Ber}(a_0) \end{aligned}$$

with $a_1 < a_0$ in $(0, 1)$. The case $a_0 < a_1$ is left as an exercise. Thus,

$$\mathbb{P}[Y = 1|H = h] = a_h = 1 - \mathbb{P}[Y = 0|H = h], \quad h = 0, 1$$

and Conditions **(A.1)** and **(A.2)** obviously hold with respect to counting measure F on $\{0, 1\}$. The likelihood rate function is given by

$$L(y) \left(\frac{1 - a_1}{1 - a_0} \right)^{1-y} \left(\frac{a_1}{a_0} \right)^y, \quad y \in \mathbb{R}.$$

For each $\eta > 0$, the test d_η takes the following form

$$\begin{aligned} d_\eta(y) = 0 & \quad \text{iff} \quad \left(\frac{1 - a_1}{1 - a_0} \right)^{1-y} \cdot \left(\frac{a_1}{a_0} \right)^y < \eta \\ & \quad \text{iff} \quad \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} \right)^y < \eta \frac{1 - a_0}{1 - a_1} \end{aligned} \quad (1.44)$$

with $y = 0, 1$. Therefore,

$$\begin{aligned} P_F(d_\eta) &= \mathbb{P}[d_\eta(Y) = 1|H = 0] \\ &= \mathbb{P} \left[\left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} \right)^Y \geq \eta \frac{1 - a_0}{1 - a_1} \middle| H = 0 \right] \\ &= \mathbb{P} \left[Y = 1, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} \right)^Y \geq \eta \frac{1 - a_0}{1 - a_1} \middle| H = 0 \right] \\ &\quad + \mathbb{P} \left[Y = 0, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} \right)^Y \geq \eta \frac{1 - a_0}{1 - a_1} \middle| H = 0 \right] \\ &= a_0 \mathbf{1} \left[\eta \frac{1 - a_0}{1 - a_1} \leq \frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} \right] + (1 - a_0) \mathbf{1} \left[\eta \frac{1 - a_0}{1 - a_1} \leq 1 \right] \\ &= a_0 \mathbf{1} \left[\eta \leq \frac{a_1}{a_0} \right] + (1 - a_0) \mathbf{1} \left[\eta \frac{1 - a_0}{1 - a_1} \leq 1 \right] \\ &= a_0 \mathbf{1} \left[\eta \leq \frac{a_1}{a_0} \right] + (1 - a_0) \mathbf{1} \left[\eta \leq \frac{1 - a_1}{1 - a_0} \right] \end{aligned} \quad (1.45)$$

Similarly, we get

$$\begin{aligned}
P_M(d_\eta) &= \mathbb{P}[d_\eta(Y) = 0 | H = 1] \\
&= \mathbb{P}\left[\left(\frac{1-a_0}{1-a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \frac{1-a_0}{1-a_1} \middle| H = 1\right] \\
&= \mathbb{P}\left[Y = 1, \left(\frac{1-a_0}{1-a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \frac{1-a_0}{1-a_1} \middle| H = 1\right] \\
&\quad + \mathbb{P}\left[Y = 0, \left(\frac{1-a_0}{1-a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \frac{1-a_0}{1-a_1} \middle| H = 1\right] \\
&= a_1 \mathbf{1}\left[\frac{1-a_0}{1-a_1} \cdot \frac{a_1}{a_0} < \eta \frac{1-a_0}{1-a_1}\right] + (1-a_1) \mathbf{1}\left[1 < \eta \frac{1-a_0}{1-a_1}\right] \\
&= a_1 \mathbf{1}\left[\frac{a_1}{a_0} < \eta\right] + (1-a_1) \mathbf{1}\left[1 < \eta \frac{1-a_0}{1-a_1}\right] \\
&= a_1 \mathbf{1}\left[\frac{a_1}{a_0} < \eta\right] + (1-a_1) \mathbf{1}\left[\frac{1-a_1}{1-a_0} < \eta\right]. \tag{1.46}
\end{aligned}$$

1.10 Additional examples

We now present several examples where Conditions **(A.1)** or **(A.2)** fail. In all cases we assume $\Gamma_0 > 0$ and $\Gamma_1 > 0$.

An example where absolute continuity (A.2) fails Here, the observation is the scalar rv Y with

$$f_0(y) = \begin{cases} 1 - |y| & \text{if } |y| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_1(y) = \begin{cases} \frac{1}{3} & \text{if } -1 \leq y \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

Condition **(A.1)** holds (with Lebesgue measure) but the absolute continuity condition **(A.2)** is clearly not satisfied. However, simple substitution reveals that

$$h(y) = \Gamma_1 p \cdot f_1(y) - \Gamma_0 (1-p) \cdot f_0(y)$$

$$= \begin{cases} 0 & \text{if } y < -1 \\ \frac{1}{3}\Gamma_1 p - \Gamma_0(1-p)(1-|y|) & \text{if } |y| \leq 1 \\ \frac{1}{3}\Gamma_1 p & \text{if } 1 < y \leq 2 \\ 0 & \text{if } 2 < y. \end{cases} \quad (1.47)$$

The Bayesian test d^* is simply

$$d^*(y) = 0 \quad \text{iff} \quad |y| < 1 - \frac{\frac{1}{3}\Gamma_1 p}{\Gamma_0(1-p)}.$$

Another example where absolute continuity (A.2) fails Here, the observation is again the scalar rv Y with

$$f_0(y) = \begin{cases} 1 - |y| & \text{if } |y| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_1(y) = \begin{cases} \frac{1}{3} & \text{if } 0 \leq y \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

Condition **(A.1)** holds (with Lebesgue measure) but **(A.2)** fails. Simple substitution reveals that

$$\begin{aligned} h(y) &= \Gamma_1 p \cdot f_1(y) - \Gamma_0(1-p) \cdot f_0(y) \\ &= \begin{cases} 0 & \text{if } y < -1 \\ -\Gamma_0(1-p)(1+y) & \text{if } -1 \leq y \leq 0 \\ \frac{1}{3}\Gamma_1 p - \Gamma_0(1-p)(1-y) & \text{if } 0 < y \leq 1 \\ \frac{1}{3}\Gamma_1 p & \text{if } 1 < y \leq 3 \\ 0 & \text{if } 3 < y, \end{cases} \end{aligned} \quad (1.48)$$

and it is straightforward to check that the Bayesian test d^* is simply

$$d^*(y) = 0 \quad \text{iff} \quad \begin{array}{c} -1 < y \leq 0 \\ \text{or} \\ 0 < y \leq 1, y < 1 - \frac{\frac{1}{3}\Gamma_1 p}{\Gamma_0(1-p)}. \end{array}$$

Equivalently, d^* can be described as

$$d^*(y) = 0 \quad \text{iff} \quad y \in \left(-1, \left(1 - \frac{\Gamma_1 p}{3\Gamma_0(1-p)} \right)^+ \right).$$

An example where both (A.1) and (A.2) fail Consider the binary hypothesis testing problem

$$\begin{aligned} H_1 &: \mathbf{Y} \sim F_1 \\ H_0 &: \mathbf{Y} \sim F_0. \end{aligned}$$

where F_0 is a discrete distribution uniform on $\{0, 1\}$, and F_1 is uniform on the interval $(0, 1)$. Thus, F_1 admits a probability density function $f_1 : \mathbb{R} \rightarrow \mathbb{R}_+$ with respect to Lebesgue measure given by

$$f_1(y) = \begin{cases} 1 & \text{if } y \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbb{P}[Y = 0|H = 0] = \mathbb{P}[Y = 1|H = 0] = \frac{1}{2}.$$

For each test d in \mathcal{D} we recall that we have

$$\begin{aligned} & \widehat{J}(d) \\ = & \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}[d(Y) = 0|H = 1] - \Gamma_0(1-p) \cdot \mathbb{P}[d(Y) = 0|H = 0] \end{aligned}$$

with

$$\mathbb{P}[d(Y) = 0|H = 0] = \begin{cases} \frac{1}{2} & \text{if } 0 \in C(d), 1 \notin C(d) \\ \frac{1}{2} & \text{if } 1 \in C(d), 0 \notin C(d) \\ 1 & \text{if } 0 \in C(d), 1 \in C(d) \\ 0 & \text{if } 0 \notin C(d), 1 \notin C(d) \end{cases}$$

and

$$\mathbb{P}[d(Y) = 0|H = 1] = \int_{C(d)} f_1(y) dy = |C(d) \cap [0, 1]|.$$

Adding or deleting a finite number of points from $C(d)$ will *not* affect the value of $\mathbb{P}[d(Y) = 0|H = 1]$, but it will change the value of $\mathbb{P}[d(Y) = 0|H = 0]$. Therefore, with $C(d)$ given, modify it, if needed, by adding both points 0 and 1.

If C' denotes this Borel subset of \mathbb{R} , then $C' = C(d) \cup \{0, 1\}$; if d' denotes the corresponding test, then $C(d') = C'$. Obviously

$$\mathbb{P}[d(Y) = 0|H = 1] = \mathbb{P}[d'(Y) = 0|H = 1] = |C(d') \cap [0, 1]|$$

since $|C(d') \cap [0, 1]| = |C(d) \cap [0, 1]|$, while

$$\mathbb{P}[d(Y) = 0|H = 0] \leq \mathbb{P}[d'(Y) = 0|H = 0] = 1.$$

We can now conclude that

$$\begin{aligned} & \widehat{J}(d) \\ &= \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}[d(Y) = 0|H = 1] - \Gamma_0(1-p) \cdot \mathbb{P}[d(Y) = 0|H = 0] \\ &\geq \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}[d'(Y) = 0|H = 1] - \Gamma_0(1-p) \cdot \mathbb{P}[d'(Y) = 0|H = 0] \\ &= \Gamma_0(1-p) + \Gamma_1 p \cdot |C(d') \cap [0, 1]| - \Gamma_0(1-p) \\ &= \Gamma_1 p \cdot |C(d') \cap [0, 1]| \geq 0. \end{aligned} \tag{1.49}$$

Consider the test $d^* : \mathbb{R} \rightarrow \{0, 1\}$ given by

$$d^*(y) = 0 \quad \text{iff} \quad y \in \{0, 1\}.$$

The arguments leading to (1.49) also show that

$$\widehat{J}(d^*) = \Gamma_1 p \cdot |C(d^*) \cap [0, 1]| = 0,$$

and this shows that d^* is the Bayesian decision rule.

1.11 Randomized tests

As we shall see shortly, a solution cannot always be found to the minimax and Neyman–Pearson formulations of the hypothesis testing problem if the search is restricted to the class of decision rules \mathcal{D} as was done for the Bayesian set-up. In some very real sense this class of tests \mathcal{D} is not large enough to yield a solution, and we enlarge it by considering the class of *randomized* tests or decision rules.

A randomized test δ is a Borel mapping $\delta : \mathbb{R}^k \rightarrow [0, 1]$ with the following interpretation as conditional probability: Having observed $\mathbf{Y} = \mathbf{y}$, it is decided that the state of nature is 1 (resp. 0) with probability $\delta(\mathbf{y})$ (resp. $1 - \delta(\mathbf{y})$). The collection of all randomized tests will be denoted by \mathcal{D}^* . Obviously, any test d in \mathcal{D} can be mechanized as a randomized test, say $\delta_d : \mathbb{R}^k \rightarrow [0, 1]$, given by

$$\delta_d(\mathbf{y}) \equiv d(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

A test in \mathcal{D} is often referred to as a *pure* strategy.

Probabilistic construction A natural question then arises as to how such randomization mechanisms can be incorporated into the probabilistic framework introduced earlier in Section 1.2:

The model data is unchanged as we are given two probability distributions F_0 and F_1 on \mathbb{R}^k and a prior p in $[0, 1]$. We still consider a sample space Ω equipped with a σ -field of events \mathcal{F} , and on it we now define the three rvs H , \mathbf{Y} and D which take values in $\{0, 1\}$, \mathbb{R}^k and $\{0, 1\}$, respectively. The rvs H and \mathbf{Y} have the same interpretation as before, as state of nature and observation, respectively, while the rv D now encodes the decision to be taken on the basis of the observation \mathbf{Y} .

With each decision rule δ in \mathcal{D}^* we associate a probability measure \mathbb{P}_δ on \mathcal{F} such that the following constraints are satisfied: As before, this time under \mathbb{P}_δ , we still have

$$\mathbb{P}_\delta [\mathbf{Y} \leq \mathbf{y} | H = h] = F_h(\mathbf{y}), \quad \begin{array}{l} \mathbf{y} \in \mathbb{R}^k, \\ h = 0, 1 \end{array}$$

and

$$p = \mathbb{P}_\delta [H = 1] = 1 - \mathbb{P}_\delta [H = 0].$$

Moreover, for $h = 0, 1$ and \mathbf{y} in \mathbb{R}^k , we require that

$$\mathbb{P}_\delta [D = d | H = h, \mathbf{Y} = \mathbf{y}] = \begin{cases} 1 - \delta(\mathbf{y}) & \text{if } d = 0 \\ \delta(\mathbf{y}) & \text{if } d = 1. \end{cases} \quad (1.50)$$

The joint probability distribution of the rvs H , D and \mathbf{Y} (under \mathbb{P}_δ) can now be completely specified: With $h, d = 0, 1$ and a Borel subset B of \mathbb{R}^k , a preconditioning argument gives

$$\begin{aligned} & \mathbb{P}_\delta [H = h, D = d, \mathbf{Y} \in B] \\ &= \mathbb{E}_\delta [\mathbf{1} [H = h, \mathbf{Y} \in B] \mathbb{P}_\delta [D = d | H, \mathbf{Y}]] \\ &= \mathbb{E}_\delta [\mathbf{1} [H = h, \mathbf{Y} \in B] (d\delta(\mathbf{Y}) + (1-d)(1-\delta(\mathbf{Y})))] \\ &= \mathbb{P}_\delta [H = h] \cdot \int_B (d\delta(\mathbf{y}) + (1-d)(1-\delta(\mathbf{y}))) dF_h(\mathbf{y}) \\ &= \begin{cases} \mathbb{P}_\delta [H = h] \cdot \int_B (1-\delta(\mathbf{y})) dF_h(\mathbf{y}) & \text{if } d = 0 \\ \mathbb{P}_\delta [H = h] \cdot \int_B \delta(\mathbf{y}) dF_h(\mathbf{y}) & \text{if } d = 1. \end{cases} \end{aligned} \quad (1.51)$$

An alternate framework The class \mathcal{D}^* of randomized strategies gives rise to a *collection* of probability triples, namely

$$\{(\Omega, \mathcal{F}, \mathbb{P}_\delta), \delta \in \mathcal{D}^*\}.$$

It is however possible to provide an equivalent probabilistic framework using a *single* probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. To see how this can be done, imagine that the original probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ is sufficiently rich that there exists a rv $U : \Omega \rightarrow [0, 1]$ which is uniformly distributed and independent of the pair of rvs H and \mathbf{Y} . This amounts to

$$\mathbb{P}[U \leq t, H = h, \mathbf{Y} \leq \mathbf{y}] = \mathbb{P}[U \leq t] \mathbb{P}[H = h, \mathbf{Y} \leq \mathbf{y}], \quad \begin{array}{l} t \in \mathbb{R} \\ h = 0, 1, \\ \mathbf{y} \in \mathbb{R}^k \end{array}$$

with

$$\mathbb{P}[U \leq t] = \begin{cases} 0 & \text{if } t \leq 0 \\ \min(t, 1) & \text{if } t \geq 0. \end{cases}$$

Now, for each decision rule δ in \mathcal{D}^* , define the $\{0, 1\}$ -valued rv D_δ given by

$$D_\delta = \mathbf{1}[U \leq \delta(\mathbf{Y})].$$

Note that

$$\begin{aligned} \mathbb{P}[D_\delta = 1 | H = h, \mathbf{Y} = \mathbf{y}] &= \mathbb{E}[\mathbf{1}[U \leq \delta(\mathbf{Y})] | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{E}[\mathbf{1}[U \leq \delta(\mathbf{y})] | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{P}[U \leq \delta(\mathbf{y})] \\ &= \delta(\mathbf{y}) \end{aligned} \tag{1.52}$$

while

$$\begin{aligned} \mathbb{P}[D_\delta = 0 | H = h, \mathbf{Y} = \mathbf{y}] &= \mathbb{E}[\mathbf{1}[U > \delta(\mathbf{Y})] | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{E}[\mathbf{1}[U > \delta(\mathbf{y})] | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{P}[U > \delta(\mathbf{y})] \\ &= 1 - \delta(\mathbf{y}) \end{aligned} \tag{1.53}$$

under the enforced independence assumption. Therefore, the conditional distribution of D_δ (under \mathbb{P}) given H and \mathbf{Y} coincides with the conditional distribution of D (under \mathbb{P}_δ) given H and \mathbf{Y} , and the two formalisms are probabilistically equivalent.

Evaluating error probabilities Consider a randomized test δ in \mathcal{D}^* . In analogy with (1.10) and (1.11), we evaluate the probabilities of false alarm and miss under δ as

$$P_F(\delta) \equiv \mathbb{P}_\delta [D = 1 | H = 0] \quad (1.54)$$

and

$$P_M(\delta) \equiv \mathbb{P}_\delta [D = 0 | H = 1]. \quad (1.55)$$

It is also convenient to consider the so-called probability of *detection* given by

$$P_D(\delta) \equiv \mathbb{P}_\delta [D = 1 | H = 1] = 1 - P_M(\delta). \quad (1.56)$$

Because

$$\mathbb{P}_\delta [D = h | H] = \mathbb{E}_\delta [\mathbb{P}_\delta [D = h | H, \mathbf{Y}] | H], \quad h = 0, 1$$

we readily get that

$$P_F(\delta) = \int_{\mathbb{R}^k} \delta(\mathbf{y}) dF_0(\mathbf{y}) \quad (1.57)$$

and

$$P_M(\delta) = \int_{\mathbb{R}^k} (1 - \delta(\mathbf{y})) dF_1(\mathbf{y}). \quad (1.58)$$

1.12 The Bayesian problem revisited

Assuming the cost function $C : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$ introduced in Section 1.4, we define the expected cost function $J^* : \mathcal{D}^* \rightarrow \mathbb{R}$ given by

$$J^*(\delta) = \mathbb{E}_\delta [C(H, D)], \quad \delta \in \mathcal{D}^*.$$

When considering randomized decision rules, the original Bayesian Problem \mathcal{P}_B is now reformulated as the minimization problem

$$\mathcal{P}_B^* : \quad \text{Minimize } J^*(\delta) \text{ over } \delta \text{ in } \mathcal{D}^*.$$

This amounts to finding an admissible test δ^* in \mathcal{D}^* such that

$$J^*(\delta^*) \leq J^*(\delta), \quad \delta \in \mathcal{D}^*. \quad (1.59)$$

Any admissible test δ^* which satisfies (1.59) is called a randomized Bayesian test, and the value

$$J^*(\delta^*) = \inf_{\delta \in \mathcal{D}^*} J^*(\delta) = \min_{\delta \in \mathcal{D}^*} J^*(\delta) \quad (1.60)$$

is sometimes referred to as the randomized *Bayesian cost*.

Obviously, since $\mathcal{D} \subset \mathcal{D}^*$ with

$$J^*(\delta_d) = J(d), \quad d \in \mathcal{D},$$

it is plain that

$$\inf_{\delta \in \mathcal{D}^*} J^*(\delta) \leq \inf_{d \in \mathcal{D}} J(d).$$

While in principle this last inequality could be strict, we now show that it is not so and that the Bayesian problem is not affected by considering the larger set of randomized decision rules.

Theorem 1.12.1 *Under the absolute continuity condition (A.1), it holds that*

$$\inf_{\delta \in \mathcal{D}^*} J^*(\delta) = \inf_{d \in \mathcal{D}} J(d). \quad (1.61)$$

Proof. Pick an arbitrary test δ in \mathcal{D}^* , A simple preconditioning argument shows that

$$\begin{aligned} J^*(\delta) &= \mathbb{E}_\delta [C(H, D)] \\ &= \mathbb{E}_\delta [\mathbb{E}_\delta [C(H, D) | H, \mathbf{Y}]] \\ &= \mathbb{E}_\delta [C(H, 1)\mathbb{P}_\delta [D = 1 | H, \mathbf{Y}] + C(H, 0)\mathbb{P}_\delta [D = 0 | H, \mathbf{Y}]] \\ &= \mathbb{E}_\delta [C(H, 1) \cdot \delta(\mathbf{Y}) + C(H, 0) \cdot (1 - \delta(\mathbf{Y}))] \\ &= \mathbb{E}_\delta [C(H, 0)] + \mathbb{E}_\delta [(C(H, 1) - C(H, 0)) \cdot \delta(\mathbf{Y})] \end{aligned} \quad (1.62)$$

with

$$\begin{aligned} &\mathbb{E}_\delta [(C(H, 1) - C(H, 0)) \cdot \delta(\mathbf{Y})] \\ &= \mathbb{E}_\delta [(C(H, 1) - C(H, 0)) \cdot \mathbb{E}_\delta [\delta(\mathbf{Y}) | H]] \\ &= (C(1, 1) - C(1, 0)) \mathbb{E}_\delta [\delta(\mathbf{Y}) | H = 1] \mathbb{P}_\delta [H = 1] \\ &\quad + (C(0, 1) - C(0, 0)) \mathbb{E}_\delta [\delta(\mathbf{Y}) | H = 0] \mathbb{P}_\delta [H = 0] \\ &= -\Gamma_1 p \cdot \mathbb{E}_\delta [\delta(\mathbf{Y}) | H = 1] + \Gamma_1 (1 - p) \cdot \mathbb{E}_\delta [\delta(\mathbf{Y}) | H = 0]. \end{aligned} \quad (1.63)$$

Using the absolute continuity condition (A.1) we can now write

$$\mathbb{E}_\delta [\delta(\mathbf{Y}) | H = h] = \int_{\mathbb{R}^k} \delta(\mathbf{y}) dF_h(\mathbf{y}) = \int_{\mathbb{R}^k} \delta(\mathbf{y}) f_h(\mathbf{y}) dF(\mathbf{y}), \quad h = 0, 1$$

so that

$$\begin{aligned}
J^*(\delta) &= -\Gamma_1 p \cdot \int_{\mathbb{R}^k} \delta(\mathbf{y}) f_1(\mathbf{y}) dF(\mathbf{y}) + \Gamma_0(1-p) \cdot \int_{\mathbb{R}^k} \delta(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}) \\
&= \int_{\mathbb{R}^k} (-\Gamma_1 p f_1(\mathbf{y}) + \Gamma_0(1-p) f_0(\mathbf{y})) \delta(\mathbf{y}) dF(\mathbf{y}) \\
&= - \int_{\mathbb{R}^k} h(\mathbf{y}) \delta(\mathbf{y}) dF(\mathbf{y})
\end{aligned} \tag{1.64}$$

where the mapping $h : \mathbb{R}^k \rightarrow \mathbb{R}$ is given by (1.24)

From Theorem 1.5.1 recall that the Bayesian rule which solves Problem \mathcal{P}_B^* is the test $d^* : \mathbb{R}^k \rightarrow \{0, 1\}$ in \mathcal{D} given by (1.25). Note that d^* can also be interpreted as the randomized rule $\delta^* : \mathbb{R}^k \rightarrow [0, 1]$ given by

$$\delta^*(\mathbf{y}) = \begin{cases} 0 & \text{if } h(\mathbf{y}) < 0 \\ 1 & \text{if } h(\mathbf{y}) \geq 0. \end{cases}$$

Equivalently, this can be written as

$$\delta^*(\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{y} \in C^* \\ 1 & \text{if } \mathbf{y} \notin C^*. \end{cases}$$

The desired result will be established if we show that

$$J^*(d^*) \leq J^*(\delta), \quad \delta \in \mathcal{D}^*.$$

The approach is reminiscent of the one used in the proof of Theorem 1.5.1: For an arbitrary δ in \mathcal{D}^* , earlier calculations show that

$$\begin{aligned}
J^*(\delta) - J^*(d^*) &= - \int_{\mathbb{R}^k} h(\mathbf{y}) \delta(\mathbf{y}) dF(\mathbf{y}) + \int_{\mathbb{R}^k} h(\mathbf{y}) \delta^*(\mathbf{y}) dF(\mathbf{y}) \\
&= \int_{\mathbb{R}^k} h(\mathbf{y}) (\delta^*(\mathbf{y}) - \delta(\mathbf{y})) dF(\mathbf{y}) \\
&= \int_{C^*} (-h(\mathbf{y})) \delta(\mathbf{y}) dF(\mathbf{y}) + \int_{\mathbb{R}^k \setminus C^*} (1 - \delta(\mathbf{y})) h(\mathbf{y}) dF(\mathbf{y}) \\
&\geq 0
\end{aligned} \tag{1.65}$$

since

$$\int_{C^*} (-h(\mathbf{y})) dF(\mathbf{y}) \geq 0 \quad \text{and} \quad \int_{\mathbb{R}^k \setminus C^*} (1 - \delta(\mathbf{y})) h(\mathbf{y}) dF(\mathbf{y}) \geq 0$$

by the very definition of the set C^* and of the mapping $h : \mathbb{R}^k \rightarrow \mathbb{R}$. ■

1.13 Randomizing between two pure decision rules

Consider two pure strategies d_1 and d_2 in \mathcal{D} . With a in $(0, 1)$, we introduce a randomized policy δ_a in \mathcal{D}^* which first selects the pure strategy d_1 (resp. d_2) with probability a (resp. $1 - a$), and then uses the pure policy that was selected. Formally, $\delta_a : \mathbb{R}^k \rightarrow [0, 1]$ is given by

$$\delta(\mathbf{y}) = ad_1(\mathbf{y}) + (1 - a)d_2(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

One very concrete way to implement such a randomized policy on the original triple $(\Omega, \mathcal{F}, \mathbb{P})$ is as follows: Consider the original probabilistic framework introduced in Section 1.2 and assume it to be sufficiently rich to carry an extra \mathbb{R} -valued rv V independent of the rvs H and \mathbf{Y} (under \mathbb{P}), and uniformly distributed on the interval $[0, 1]$. Define the $\{0, 1\}$ -valued rv B_a given by

$$B_a = \mathbf{1}[V \leq a].$$

It is plain that the rv B_a is independent of the rvs H and \mathbf{Y} (under \mathbb{P}), with

$$\mathbb{P}[B_a = 1] = a = 1 - \mathbb{P}[B_a = 0].$$

Define the decision rv D_a given by

$$D_a = B_a d_1(\mathbf{Y}) + (1 - B_a) d_2(\mathbf{Y}).$$

It is easy to check that

$$\begin{aligned} & \mathbb{P}[D_a = 1 | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{P}[B_a d_1(\mathbf{Y}) + (1 - B_a) d_2(\mathbf{Y}) = 1 | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{P}[B_a d_1(\mathbf{y}) + (1 - B_a) d_2(\mathbf{y}) = 1 | H = h, \mathbf{Y} = \mathbf{y}] \\ &= \mathbb{P}[B_a = 1, d_1(\mathbf{y}) = 1 | H = h, \mathbf{Y} = \mathbf{y}] + \mathbb{P}[B_a = 0, d_2(\mathbf{y}) = 1 | H = h, \mathbf{Y} = \mathbf{y}] \\ &= d_1(\mathbf{y}) \mathbb{P}[B_a = 1 | H = h, \mathbf{Y} = \mathbf{y}] + d_2(\mathbf{y}) \mathbb{P}[B_a = 0 | H = h, \mathbf{Y} = \mathbf{y}] \\ &= d_1(\mathbf{y}) \mathbb{P}[B_a = 1] + d_2(\mathbf{y}) \mathbb{P}[B_a = 0] \\ &= ad_1(\mathbf{y}) + (1 - a)d_2(\mathbf{y}), \quad \begin{array}{l} \mathbf{y} \in \mathbb{R}^k, \\ h = 0, 1 \end{array} \end{aligned} \tag{1.66}$$

as desired.

Applying the expressions (1.57) and (1.58) with the randomized test δ_a we get

$$\begin{aligned}
 P_F(\delta_a) &= \int_{\mathbb{R}^k} \delta_a(\mathbf{y}) dF_0(\mathbf{y}) \\
 &= \int_{\mathbb{R}^k} (ad_1(\mathbf{y}) + (1-a)d_2(\mathbf{y})) dF_0(\mathbf{y}) \\
 &= a \int_{\mathbb{R}^k} d_1(\mathbf{y}) dF_0(\mathbf{y}) + (1-a) \int_{\mathbb{R}^k} d_2(\mathbf{y}) dF_0(\mathbf{y}) \\
 &= aP_F(d_1) + (1-a)P_F(d_2).
 \end{aligned} \tag{1.67}$$

Similarly we find that

$$\begin{aligned}
 P_M(\delta_a) &= \int_{\mathbb{R}^k} (1 - \delta_a(\mathbf{y})) dF_1(\mathbf{y}) \\
 &= \int_{\mathbb{R}^k} (1 - ad_1(\mathbf{y}) - (1-a)d_2(\mathbf{y})) dF_1(\mathbf{y}) \\
 &= a \int_{\mathbb{R}^k} (1 - d_1(\mathbf{y})) dF_1(\mathbf{y}) + (1-a) \int_{\mathbb{R}^k} (1 - d_2(\mathbf{y})) dF_1(\mathbf{y}) \\
 &= aP_M(d_1) + (1-a)P_M(d_2).
 \end{aligned} \tag{1.68}$$

1.14 The minimax formulation

The Bayesian formulation *implicitly* assumes knowledge of the prior distribution of the hypothesis rv H . In many situations, this assumption cannot be adequately justified, and the Bayesian formulation has to be abandoned for the so-called *minimax criterion*.

The basic idea For each p in $[0, 1]$, let $J_p(d)$ denote the expected cost associated with the admissible decision rule d in \mathcal{D} when the prior on H is p , i.e.,

$$J_p(d) \equiv \mathbb{E}_p [C(H, d(H))]$$

where $\mathbb{E}_p [\cdot]$ denotes expectation with prior p . The Bayesian problem now reads

$$\mathcal{P}_{p,B} : \quad \text{Minimize } J_p(d) \text{ over } d \text{ in } \mathcal{D}.$$

As shown earlier, under the mild Conditions **(A.1)** and **(A.2)** (assumed from now on), $\mathcal{P}_{p,B}$ has a solution which is now denoted by $d^*(p)$ to indicate its dependence on the prior p . Clearly, any such solution satisfies

$$J_p(d^*(p)) \leq J_p(d), \quad d \in \mathcal{D}. \quad (1.69)$$

Let the corresponding Bayesian cost be denoted by

$$V(p) \equiv \min_{d \in \mathcal{D}} J_p(d) \quad (1.70)$$

It is plain that

$$V(p) = J_p(d^*(p)). \quad (1.71)$$

Since the exact value of the prior p is not available, a reasonable way to proceed consists in using the Bayesian test for that value of p which yields the *largest* Bayesian cost (1.70): Thus, with the notation introduced, let p_m in $[0, 1]$ such that

$$V(p_m) = \max_{p \in [0,1]} V(p), \quad (1.72)$$

and use the Bayesian rule $d^*(p_m)$ – The existence of p_m is guaranteed by the fact that the mapping $V : [0, 1] \rightarrow \mathbb{R}$ is continuous on the closed bounded interval $[0, 1]$ by Lemma 1.15.1, hence achieves its maximum value on $[0, 1]$.

The test $d^*(p_m)$, hereafter denoted d_m^* , is known as the *minimax* decision rule; it tries to *compensate* for the uncertainty in the modeling assumptions, namely, that the exact value of p is not known. We refer to the cost value

$$V(p_m) = J_{p_m}(d^*(p_m))$$

as the *minimax* cost; it will be denoted by V_m .

That there is a performance cost to pay for this uncertainty is not surprising in the least: Indeed, let p_{True} be the true (but *unknown*) value of the prior p . If $p_m = p_{\text{True}}$, then d_m^* coincides with the Bayesian test $d^*(p_{\text{True}})$ and this choice is optimal. On the other hand, if $p_m \neq p_{\text{True}}$, then

$$J_{p_{\text{True}}}(d^*(p_{\text{True}})) \leq J_{p_{\text{True}}}(d^*(p_m)) = J_{p_{\text{True}}}(d_m^*)$$

with a *strict* inequality in most cases, i.e., as expected, a better performance level could be reached if the value p_{True} were known.

The Minimax Theorem

Theorem 1.14.1 *Under the absolute continuity condition (A.1), the minimax cost V_m can be characterized by*

$$V_m = \min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right) = \max_{p \in [0,1]} \left(\min_{d \in \mathcal{D}} J_p(d) \right). \quad (1.73)$$

Proof. Combining (1.69) and (1.71) we get

$$V(p) \leq J_p(d), \quad \begin{array}{l} d \in \mathcal{D} \\ p \in [0, 1]. \end{array}$$

Therefore, for each d in \mathcal{D} it holds that

$$\max_{p \in [0,1]} V(p) \leq \max_{p \in [0,1]} J_p(d),$$

whence

$$\min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} V(p) \right) \leq \min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right).$$

This is equivalent to

$$\max_{p \in [0,1]} V(p) \leq \min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right),$$

an inequality that also reads as

$$\max_{p \in [0,1]} \left(\min_{d \in \mathcal{D}} J_p(d) \right) \leq \min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right).$$

The equality (1.73) will be established if we show that

$$\min_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right) \leq \max_{p \in [0,1]} \left(\min_{d \in \mathcal{D}} J_p(d) \right). \quad (1.74)$$

■

1.15 The minimax equation

The main issue in constructing the minimax rule d_m^* consists in determining the value of p_m such that (1.72) holds; its characterization is achieved through the *Minimax Equation* discussed below. In view of (1.71) this requirement also reads

$$J_{p_m}(d^*(p_m)) = \max_{p \in [0,1]} V(p). \quad (1.75)$$

An auxiliary concavity result The following technical fact will be useful in the derivation of the minimax equation.

Lemma 1.15.1 *The mapping $V : [0, 1] \rightarrow \mathbb{R}$ is concave and continuous on the closed interval $[0, 1]$ with boundary values $V(0) = C(0, 0)$ and $V(1) = C(1, 1)$.*

For easy reference, recall that for each test d in \mathcal{D} , the expression

$$\begin{aligned} J_p(d) &= pC(1, 1) + (1 - p)C(0, 0) \\ &\quad + \Gamma_0(1 - p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d) \end{aligned} \quad (1.76)$$

holds with p in $[0, 1]$.

Proof. With $p = 0$ in (1.76) we get

$$J_0(d) = C(0, 0) + \Gamma_0 P_F(d), \quad d \in \mathcal{D}.$$

Therefore,

$$V(0) = C(0, 0) + \Gamma_0 \cdot \inf_{d \in \mathcal{D}} P_F(d)$$

upon using $\Gamma_0 > 0$. However, $P_F(d_\infty) = 0$, so that $\inf_{d \in \mathcal{D}} P_F(d) = 0$, whence $V(0) = C(0, 0)$. A similar argument shows that $V(1) = C(1, 1)$.

The probabilities $P_F(d)$ and $P_M(d)$ appearing in (1.76) do *not* depend on p , but rather on F_0 , F_1 and d . Thus, the mapping $p \rightarrow J_p(d)$ is affine, hence concave in p . As a result, the mapping $V : [0, 1] \rightarrow \mathbb{R}$ is concave on the closed interval $[0, 1]$, being the infimum of the family $\{J_p(d), d \in \mathcal{D}\}$ of concave functions.

Because a concave function defined on an open interval is necessarily continuous on that open interval, the mapping $V : [0, 1] \rightarrow \mathbb{R}$ is continuous on $(0, 1)$. Therefore, it remains only to show that this mapping is also continuous at the boundary points $p = 0$ and $p = 1$. We discuss only the case $p = 0$; the case $p = 1$ can be handled *mutatis mutandis* and is left to the interested reader as an exercise.

In view of (1.76) continuity of the mapping $V : [0, 1] \rightarrow \mathbb{R}$ at $p = 0$ is equivalent to

$$\lim_{p \rightarrow 0} \left(\inf_{d \in \mathcal{D}} (\Gamma_0(1-p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d)) \right) = 0 \quad (1.77)$$

since $V(0) = C(0, 0)$ by the first part of the proof.

To do so, write

$$\Delta(p) = \inf_{d \in \mathcal{D}} (\Gamma_0(1-p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d)), \quad p \in [0, 1].$$

Thus, for any fixed p in $(0, 1)$, we get

$$\begin{aligned} \Delta(p) &= \inf_{d \in \mathcal{D}} (\Gamma_0 \cdot P_F(d) + p (\Gamma_1 \cdot P_M(d) - \Gamma_0 \cdot P_F(d))) \\ &= \inf_{d \in \mathcal{D}} (\Gamma_0 \cdot P_F(d) + p \cdot A(d)) \end{aligned} \quad (1.78)$$

where we have set

$$A(d) = \Gamma_1 \cdot P_M(d) - \Gamma_0 \cdot P_F(d), \quad d \in \mathcal{D}.$$

Next, note the elementary bounds

$$\inf_{d \in \mathcal{D}} (p \cdot A(d)) \leq \Delta(p) \leq \Gamma_0 \cdot P_F(d) + p \cdot A(d), \quad d \in \mathcal{D}. \quad (1.79)$$

It is clear that $\lim_{p \rightarrow 0} \inf_{d \in \mathcal{D}} (p \cdot A(d)) = 0$ since

$$\left| \inf_{d \in \mathcal{D}} (p \cdot A(d)) \right| \leq p \cdot \sup_{d \in \mathcal{D}} |A(d)| \leq p (\Gamma_0 + \Gamma_1).$$

It then follows from the first inequality in (1.79) that $\liminf_{p \rightarrow 0} \Delta(p) \geq 0$. On the other hand, the second inequality in (1.79) yields

$$\limsup_{p \rightarrow 0} \Delta(p) \leq \Gamma_0 \cdot P_F(d), \quad d \in \mathcal{D}.$$

Taking $d = d_\infty$ in this last inequality gives $\limsup_{p \rightarrow 0} \Delta(p) \leq 0$, and combining the last two limiting statements we get the desired conclusion $\lim_{p \rightarrow 0} \Delta(p) = 0$. ■

Figure

The minimax equation Fix p in $[0, 1]$. The mapping $\alpha \rightarrow J_\alpha(d^*(p))$ is affine in the variable α on the interval $[0, 1]$ as we recall that

$$\begin{aligned} J_\alpha(d^*(p)) &= \alpha C(1, 1) + (1 - \alpha)C(0, 0) \\ &\quad + \Gamma_0(1 - \alpha) \cdot P_F(d^*(p)) + \Gamma_1\alpha \cdot P_M(d^*(p)) \end{aligned} \quad (1.80)$$

with α in $[0, 1]$ upon specializing (1.76) to the test $d^*(p)$. Therefore, the graph of the mapping $\alpha \rightarrow J_\alpha(d^*(p))$ is a straight line, whose slope is constant over $[0, 1]$ and given by

$$\frac{d}{d\alpha} J_\alpha(d^*(p)) = C(1, 1) - C(0, 0) + \Gamma_1 \cdot P_M(d^*(p)) - \Gamma_0 \cdot P_F(d^*(p)). \quad (1.81)$$

By its definition, the Bayesian cost satisfies

$$V(\alpha) \leq J_\alpha(d), \quad \begin{array}{l} d \in \mathcal{D} \\ \alpha \in [0, 1] \end{array}$$

with strict inequality for most tests. With $d = d^*(p)$ this inequality becomes an equality when $\alpha = p$, namely

$$V(p) = J_p(d^*(p))$$

while

$$V(\alpha) \leq J_\alpha(d^*(p)), \quad \alpha \in [0, 1].$$

With p in $(0, 1)$, if the concave mapping $\alpha \rightarrow V(\alpha)$ is *differentiable* at $\alpha = p$, then the straight line $\alpha \rightarrow J_\alpha(d^*(p))$ will be tangential to the mapping $\alpha \rightarrow V(\alpha)$ at $\alpha = p$ – This is a consequence of the concavity established in Lemma 1.15.1. Thus,

$$\left. \frac{d}{d\alpha} V(\alpha) \right|_{\alpha=p} = \left. \frac{d}{d\alpha} J_\alpha(d^*(p)) \right|_{\alpha=p}. \quad (1.82)$$

In particular, if p_m is an element of $(0, 1)$ and the mapping $\alpha \rightarrow V(\alpha)$ is differentiable at $\alpha = p_m$, then

$$\left. \frac{d}{d\alpha} V(\alpha) \right|_{\alpha=p_m} = \left. \frac{d}{d\alpha} J_\alpha(d^*(p_m)) \right|_{\alpha=p_m}. \quad (1.83)$$

But p_m being a maximum for the function $\alpha \rightarrow V(\alpha)$, we must have

$$\left. \frac{d}{d\alpha} V(\alpha) \right|_{\alpha=p_m} = 0,$$

whence

$$\left. \frac{d}{d\alpha} J_\alpha(d^*(p_m)) \right|_{\alpha=p_m} = 0.$$

Using (1.81) we conclude that

$$C(1, 1) - C(0, 0) + \Gamma_0 P_F(d^*(p_m)) - \Gamma_1 P_M(d^*(p_m)) = 0. \quad (1.84)$$

This equation characterizing p is called the Minimax Equation.

For the probability of error criterion, the Minimax Equation takes the simpler form

$$P_F(d^*(p_m)) = P_M(d^*(p)).$$

This analysis does not cover the cases when (i) $p_m = 0$, (ii) $p_m = 1$ and (iii) p_m is an element of $(0, 1)$ but the mapping $\alpha \rightarrow V(\alpha)$ is not differentiable at $\alpha = p_m$,

1.16 The minimax formulation – Two examples

The Gaussian case The setting is that of Section 1.8 to which we refer the reader for the notation.

As shown there, for any $\eta > 0$ we have

$$P_F(Lrt_\eta) = 1 - \Phi\left(\frac{\log \eta + \frac{1}{2}d^2}{d}\right)$$

and

$$P_M(Lrt_\eta) = \Phi\left(\frac{\log \eta - \frac{1}{2}d^2}{d}\right)$$

For each p in $[0, 1]$, with

$$\eta(p) = \frac{1-p}{p} \cdot \frac{\Gamma_0}{\Gamma_1}$$

we have $d^*(p) = Lrt_{\eta(p)}$ and the expression (1.76) becomes

$$\begin{aligned} J_p(d^*(p)) &= pC(1, 1) + (1-p)C(0, 0) \\ &\quad + \Gamma_0(1-p) \cdot P_F(d^*(p)) + \Gamma_1 p \cdot P_M(d^*(p)) \\ &= pC(1, 1) + (1-p)C(0, 0) \\ &\quad + \Gamma_0(1-p) \cdot \left(1 - \Phi\left(\frac{\log \eta(p) + \frac{1}{2}d^2}{d}\right)\right) \\ &\quad + \Gamma_1 p \cdot \Phi\left(\frac{\log \eta(p) - \frac{1}{2}d^2}{d}\right). \end{aligned} \quad (1.85)$$

Therefore,

$$\begin{aligned}
 V(p) &= pC(1, 1) + (1 - p)C(0, 0) \\
 &\quad + \Gamma_0(1 - p) \cdot \left(1 - \Phi \left(\frac{\log \eta(p) + \frac{1}{2}d^2}{d} \right) \right) \\
 &\quad + \Gamma_1 p \cdot \Phi \left(\frac{\log \eta(p) - \frac{1}{2}d^2}{d} \right). \tag{1.86}
 \end{aligned}$$

The boundary cases $p = 0$ and $p = 1$ are easily recovered upon formally substituting these values in the expression above. The Minimax Equation takes the form

$$\begin{aligned}
 &C(1, 1) - C(0, 0) \\
 &= \Gamma_1 \Phi \left(\frac{\log \eta(p_m) - \frac{1}{2}d^2}{d} \right) - \Gamma_0 \left(1 - \Phi \left(\frac{\log \eta(p_m) + \frac{1}{2}d^2}{d} \right) \right).
 \end{aligned}$$

For the probability of error case, simplifications occur. The last expression becomes

$$\begin{aligned}
 V(p) &= (1 - p) \cdot \left(1 - \Phi \left(\frac{\frac{1-p}{p} + \frac{1}{2}d^2}{d} \right) \right) \\
 &\quad + p \cdot \Phi \left(\frac{\log \frac{1-p}{p} - \frac{1}{2}d^2}{d} \right) \tag{1.87}
 \end{aligned}$$

and the Minimax Equation reduces to

$$\Phi \left(\frac{\log \eta(p_m) - \frac{1}{2}d^2}{d} \right) + \Phi \left(\frac{\log \eta(p_m) + \frac{1}{2}d^2}{d} \right) = 1.$$

It is easy to see that this requires $\log \eta(p_m) = 0$ so that $p_m = \frac{1}{2}$ (indeed in $(0, 1)$), an intuitively satisfying conclusion!

The Bernoulli case The setting is that of Section 1.9 to which we refer the reader for the notation. We discuss only the case $a_1 < a_0$, and leave the case $a_0 < a_1$ as an exercise for the interested reader.

Note that the condition $a_1 < a_0$ is equivalent to $1 < \frac{1-a_1}{1-a_0}$, so that the expressions (1.45) and (1.46) for the probabilities $P_F(d_\eta)$ and $P_M(d_\eta)$, respectively, are

piecewise constant functions of η with different constant values on the intervals $(0, \frac{a_1}{a_0}]$, $(\frac{a_1}{a_0}, \frac{1-a_1}{1-a_0}]$ and $(\frac{1-a_1}{1-a_0}, \infty)$: Direct inspection of the expression (1.45) yields

$$P_F(d_\eta) = \begin{cases} 1 & \text{if } 0 < \eta \leq \frac{a_1}{a_0} \\ 1 - a_0 & \text{if } \frac{a_1}{a_0} < \eta \leq \frac{1-a_1}{1-a_0} \\ 0 & \text{if } \frac{1-a_1}{1-a_0} < \eta. \end{cases} \quad (1.88)$$

Similarly, using (1.46) we find

$$P_M(d_\eta) = \begin{cases} 0 & \text{if } 0 < \eta \leq \frac{a_1}{a_0} \\ a_1 & \text{if } \frac{a_1}{a_0} < \eta \leq \frac{1-a_1}{1-a_0} \\ 1 & \text{if } \frac{1-a_1}{1-a_0} < \eta. \end{cases} \quad (1.89)$$

Thus, for each p in $[0, 1]$, we see from (1.76) that the cost $J_p(d_\eta)$ takes a different value on each of the intervals $(0, \frac{a_1}{a_0}]$, $(\frac{a_1}{a_0}, \frac{1-a_1}{1-a_0}]$ and $(\frac{1-a_1}{1-a_0}, \infty)$: Specifically, we have:

On $(0, \frac{a_1}{a_0}]$,

$$\begin{aligned} J_p(d_\eta) &= pC(1, 1) + (1-p)C(0, 0) + \Gamma_0(1-p) \\ &= pC(1, 1) + (1-p)C(0, 1); \end{aligned} \quad (1.90)$$

On $(\frac{a_1}{a_0}, \frac{1-a_1}{1-a_0}]$,

$$\begin{aligned} J_p(d_\eta) &= pC(1, 1) + (1-p)C(0, 0) + \Gamma_0(1-p) \cdot (1-a_0) + \Gamma_1 p \cdot a_1 \\ &= pC(1, 1) + (1-p)C(0, 1) + \Gamma_1 p \cdot a_1 - \Gamma_0(1-p) \cdot a_0 \\ &= p(C(1, 1) + \Gamma_1 a_1) + (1-p)(C(0, 1) - \Gamma_0 a_0); \end{aligned} \quad (1.91)$$

On $(\frac{1-a_1}{1-a_0}, \infty)$,

$$\begin{aligned} J_p(d_\eta) &= pC(1, 1) + (1-p)C(0, 0) + \Gamma_1 p \\ &= pC(1, 0) + (1-p)C(0, 0). \end{aligned} \quad (1.92)$$

Recall that

$$V(p) = J_p(d_{\eta(p)}) \text{ with } \eta(p) = \frac{\Gamma_0(1-p)}{\Gamma_1 p}, \quad 0 < p \leq 1.$$

As the mapping $p : (0, 1] \rightarrow \mathbb{R}_+ : p \rightarrow \eta(p)$ is strictly decreasing, the equations

$$\eta(p) = \frac{1 - a_1}{1 - a_0}, \quad 0 < p \leq 1$$

and

$$\eta(p) = \frac{a_1}{a_0}, \quad 0 < p \leq 1$$

have each a unique solution in $(0, 1)$. Their solutions, denoted p_- and p_+ , respectively, are given by

$$p_- = \frac{\Gamma_0(1 - a_0)}{\Gamma_1(1 - a_1) + \Gamma_0(1 - a_0)}$$

and

$$p_+ = \frac{\Gamma_0 a_0}{\Gamma_1 a_1 + \Gamma_0 a_0}.$$

As expected $p_- < \frac{1}{2} < p_+$.

Earlier expressions can now be used

$$V(p) = \begin{cases} pC(1, 0) + (1 - p)C(0, 0) & \text{if } p \in (0, p_-] \\ p(C(1, 1) + \Gamma_1 a_1) + (1 - p)(C(0, 1) - \Gamma_0 a_0) & \text{if } p \in (p_-, p_+] \\ pC(1, 1) + (1 - p)C(0, 1) & \text{if } p \in (p_+, 1). \end{cases}$$

It is plain that the function $V : [0, 1] \rightarrow \mathbb{R}$ is piecewise linear with three distinct segments, namely $(0, p_-]$, $(p_-, p_+]$ and $(p_+, 1]$. There are two kinks at $p = p_-$ and $p = p_+$, respectively. That the function is concave can be seen by computing the left and right-derivatives at these points. However, the function $V : [0, 1] \rightarrow \mathbb{R}$ is differentiable everywhere except at these kinks. However the maximum occurs at one of these points so that $p_m \in \{p_-, p_+\}$.

For the probability of error criterion, we find that

$$V(p) = \begin{cases} p & \text{if } p \in (0, p_-] \\ pa_1 + (1 - p)(1 - a_0) & \text{if } p \in (p_-, p_+] \\ 1 - p & \text{if } p \in (p_+, 1) \end{cases} \quad (1.93)$$

with

$$p_- = \frac{1 - a_0}{(1 - a_1) + (1 - a_0)}$$

and

$$p_+ = \frac{a_0}{a_1 + a_0}.$$

It is trivial to check that $V(p_{\pm-}) = V(p_{\pm+})$, establishing continuity at the kink points. Note that here $p_m = p_+$ so that the minimax cost is given by

$$V_m = 1 - p_+ = \frac{a_1}{a_1 + a_0}.$$

1.17 The Neyman-Pearson formulation

In many situations, not only is the prior probability p not available but it is quite difficult to make meaningful cost assignments. This is typically the case in radar applications – After all, what is the real cost of failing to detect an incoming missile? While it is tempting to seek to minimize *both* the probabilities of miss and false alarm, these are (usually) conflicting objectives and a *constrained* optimization problem is considered instead. For $0 \leq \alpha \leq 1$, consider the constrained optimization problem NP_α where

$$\text{NP}_\alpha : \quad \text{Maximize } P_D(d) \text{ over } d \text{ in } \mathcal{D}_\alpha$$

where \mathcal{D}_α is the collection of admissible tests in \mathcal{D} of size at most α , i. e.,

$$\mathcal{D}_\alpha = \{d \in \mathcal{D} : P_F(d) \leq \alpha\}.$$

Solving NP_α amounts to finding a test $d_{\text{NP}}(\alpha)$ in \mathcal{D}_α with the property that

$$P_D(d) \leq P_D(d_{\text{NP}}(\alpha)), \quad d \in \mathcal{D}_\alpha.$$

Such a test $d_{\text{NP}}(\alpha)$, when it exists, is called a *Neyman–Pearson* test of size α . Following the accepted terminology, its *power* $\beta(\alpha)$ is given by

$$\beta(\alpha) := \sup_{d \in \mathcal{D}_\alpha} P_D(d)$$

The Lagrangian argument When reformulated as

$$\text{NP}_\alpha : \quad \text{Minimize } P_M(d) \text{ over } d \text{ in } \mathcal{D}_\alpha,$$

the constrained optimization problem NP_α can be solved by the following Lagrangian arguments: First, for each $\lambda \geq 0$ consider the Lagrangian functional $J_\lambda : \mathcal{D} \rightarrow \mathbb{R}$ given by

$$J_\lambda(d) = P_M(d) + \lambda(P_F(d) - \alpha), \quad d \in \mathcal{D}.$$

The *Lagrangian* problem LP_λ is then defined as the *unconstrained* minimization problem

$$\text{LP}_\lambda : \quad \text{Minimize } J_\lambda(d) \text{ over } d \text{ in } \mathcal{D}$$

Its solution is readily obtained as follows: For any test d in \mathcal{D} , we observe that

$$\begin{aligned} J_\lambda(d) &= \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1] + \lambda(\mathbb{P}[d(\mathbf{Y}) = 1 | H = 0] - \alpha) \\ &= \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1] + \lambda(1 - \mathbb{P}[d(\mathbf{Y}) = 0 | H = 0] - \alpha) \\ &= \lambda(1 - \alpha) + \mathbb{P}[d(\mathbf{Y}) = 0 | H = 1] - \lambda\mathbb{P}[d(\mathbf{Y}) = 0 | H = 0] \\ &= \lambda(1 - \alpha) + \int_{C(d)} h_\lambda(\mathbf{y}) dF(\mathbf{y}) \end{aligned} \quad (1.94)$$

with $h_\lambda : \mathbb{R}^k \rightarrow \mathbb{R}$ given by

$$h_\lambda(\mathbf{y}) = f_1(\mathbf{y}) - \lambda f_0(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

By the arguments leading to Theorem 1.5.1, the Lagrangian problem LP_λ is seen to be solved by the test d_λ given by

$$d_\lambda(\mathbf{y}) = 0 \quad \text{iff} \quad h_\lambda(\mathbf{y}) < 0, \quad (1.95)$$

or equivalently,

$$d_\lambda(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < \lambda f_0(\mathbf{y}). \quad (1.96)$$

The next step consists in finding some value $\lambda(\alpha) > 0$ of the Lagrangian multiplier such that the test $d_{\lambda(\alpha)}$ meets the constraint, i.e.,

$$P_F(d_{\lambda(\alpha)}) = \alpha. \quad (1.97)$$

If such value $\lambda(\alpha)$ exists, then we get

$$J_{\lambda(\alpha)}(d_{\lambda(\alpha)}) \leq J_{\lambda(\alpha)}(d), \quad d \in \mathcal{D},$$

or equivalently,

$$P_M(d_{\lambda(\alpha)}) \leq P_M(d) + \lambda(\alpha)(P_F(d) - \alpha), \quad d \in \mathcal{D}.$$

Consequently, for every test d in \mathcal{D}_α (and not merely in \mathcal{D}), we conclude that

$$P_M(d_{\lambda(\alpha)}) \leq P_M(d)$$

since then $P_M(d) \leq \alpha$. The test $d_{\lambda(\alpha)}$ being in \mathcal{D}_α , it is clear that $d_{\lambda(\alpha)}$ solves NP_α . In other words, $d_{\text{NP}}(\alpha)$ can be taken to be $d_{\lambda(\alpha)}$.

Meeting the constraint (1.97) The Lagrangian argument hinges upon the possibility of finding a value $\lambda(\alpha)$ of the Lagrange multiplier such that $P_F(d_{\lambda(\alpha)}) = \alpha$. Unfortunately, this may not be always possible, unless additional assumptions are imposed. To see this, note that for every $\lambda \geq 0$, it holds that

$$\begin{aligned} P_F(d_\lambda) &= \mathbb{P}[d_\lambda(\mathbf{Y}) = 1 | H = 0] \\ &= \mathbb{P}[f_1(\mathbf{Y}) \geq \lambda f_0(\mathbf{Y}) | H = 0]. \end{aligned} \quad (1.98)$$

The mapping $\mathbb{R}_+ \rightarrow [0, 1] : \lambda \rightarrow P_F(d_\lambda)$ is clearly *monotone non-increasing* with boundary values $P_F(d_0) = 1$ and $\lim_{\lambda \uparrow \infty} P_F(d_\lambda) = P_F(d_\infty) = 0$. However, the constraint $P_F(d_\lambda) = \alpha$ may fail to hold for some α in $[0, 1]$ because the set of values $\{P_F(d_\lambda), \lambda \geq 0\}$ need not contain α . This will occur if the mapping $\lambda \rightarrow P_F(d_\lambda)$ is not continuous at some point, say $\lambda^* > 0$, with

$$\lim_{\lambda \uparrow \lambda^*} P_F(d_\lambda) < \alpha < \lim_{\lambda \downarrow \lambda^*} P_F(d_\lambda).$$

Later in this section we illustrate this situation on the simple example of deciding between the two hypotheses on the basis of a Poisson rv.

Although randomized policies are introduced to solve this difficulty, there are however situations where this can be avoided because each one of the problems NP_α has a solution within the set of non-randomized policies \mathcal{D} . One such situation occurs when F is Lebesgue measure on \mathbb{R}^k and the absolute continuity condition **(A.2)** holds.

Lemma 1.17.1 *Assume F to be Lebesgue measure on \mathbb{R}^k and that the absolute continuity condition **(A.2)** holds. Then the mapping $\mathbb{R}_+ \rightarrow [0, 1] : \lambda \rightarrow P_F(d_\lambda)$ is continuous.*

The proof of Lemma 1.17.1 is given Section 1.22. Continuity, together with the monotonicity property discussed earlier, implies $\{P_F(d_\lambda), \lambda \geq 0\} = [0, 1]$, and the requirement can always be met.

1.18 The Neyman-Pearson Lemma

The discussion of Section 1.17 suggests the need to consider an extended version of the Neyman-Pearson formulation. First for each α in $[0, 1]$ let \mathcal{D}_α^* denote the collection of all randomized tests in \mathcal{D}^* of size at most α , i. e.,

$$\mathcal{D}_\alpha^* = \{\delta \in \mathcal{D}^* : P_F(\delta) \leq \alpha\}.$$

Now consider the following constrained optimization problem NP_α^* where

$$\text{NP}_\alpha^* : \quad \text{Maximize } P_D(\delta) \text{ over } \delta \text{ in } \mathcal{D}_\alpha^*$$

Solving NP_α^* amounts to finding a test $\delta_{\text{NP}}(\alpha)$ in \mathcal{D}_α^* with the property that

$$P_D(\delta) \leq P_D(\delta_{\text{NP}}(\alpha)), \quad \delta \in \mathcal{D}_\alpha^*.$$

Such a test $\delta_{\text{NP}}(\alpha)$, when it exists, is also called a *Neyman-Pearson* test of size α .

Its existence and characterization are consequences of two technical facts, namely Lemma 1.18.1 and Lemma 1.18.2 below, which are collectively referred as the Neyman-Pearson Lemma. Optimality is handled first.

Optimality With $\eta > 0$ and Borel mapping $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ (to be selected shortly), define the randomized test $\delta^* : \mathbb{R}^k \rightarrow [0, 1]$ in \mathcal{D}^* given by

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \text{if } \eta f_0(\mathbf{y}) < f_1(\mathbf{y}) \\ \gamma(\mathbf{y}) & \text{if } f_1(\mathbf{y}) = \eta f_0(\mathbf{y}) \\ 0 & \text{if } f_1(\mathbf{y}) < \eta f_0(\mathbf{y}). \end{cases} \quad (1.99)$$

Lemma 1.18.1 For any test $\delta : \mathbb{R}^k \rightarrow [0, 1]$ in \mathcal{D}^* , the inequality

$$P_D(\delta^*) - P_D(\delta) \geq \eta(P_F(\delta^*) - P_F(\delta)) \quad (1.100)$$

holds where the randomized test $\delta^* : \mathbb{R}^k \rightarrow [0, 1]$ in \mathcal{D}^* given by (1.99).

Proof. Let $\delta : \mathbb{R}^k \rightarrow [0, 1]$ be an arbitrary test in \mathcal{D}^* . Recall that

$$P_F(\delta) = \int_{\mathbb{R}^k} \delta(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}) \quad \text{and} \quad P_F(\delta^*) = \int_{\mathbb{R}^k} \delta^*(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}),$$

while

$$P_D(\delta) = \int_{\mathbb{R}^k} \delta(\mathbf{y}) f_1(\mathbf{y}) dF(\mathbf{y}) \quad \text{and} \quad P_D(\delta^*) = \int_{\mathbb{R}^k} \delta^*(\mathbf{y}) f_1(\mathbf{y}) dF(\mathbf{y}).$$

Now,

$$\begin{aligned} & \int_{\mathbb{R}^k} (\delta^*(\mathbf{y}) - \delta(\mathbf{y})) (f_1(\mathbf{y}) - \eta f_0(\mathbf{y})) dF(\mathbf{y}) \\ &= \int_{\mathbb{R}^k} (\delta^*(\mathbf{y}) - \delta(\mathbf{y})) f_1(\mathbf{y}) dF(\mathbf{y}) - \eta \int_{\mathbb{R}^k} (\delta^*(\mathbf{y}) - \delta(\mathbf{y})) f_0(\mathbf{y}) dF(\mathbf{y}) \\ &= P_D(\delta^*) - P_D(\delta) - \eta (P_F(\delta^*) - P_F(\delta)). \end{aligned} \quad (1.101)$$

But direct inspection shows that it always the case that

$$(\delta^*(\mathbf{y}) - \delta(\mathbf{y})) (f_1(\mathbf{y}) - \eta f_0(\mathbf{y})) \geq 0, \quad \mathbf{y} \in \mathbb{R}^k. \quad (1.102)$$

Therefore,

$$P_D(\delta^*) - P_D(\delta) - \eta (P_F(\delta^*) - P_F(\delta)) \geq 0 \quad (1.103)$$

and the desired conclusion (1.100) follows. \blacksquare

Fix α in $(0, 1)$. If we select $\eta > 0$ and $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ so that

$$P_F(\delta^*) = \alpha, \quad (1.104)$$

then the inequality (1.100) implies

$$P_D(\delta^*) - P_D(\delta) \geq \eta (\alpha - P_F(\delta)), \quad \delta \in \mathcal{D}^*. \quad (1.105)$$

Therefore, for any test $\delta : \mathbb{R}^k \rightarrow [0, 1]$ in \mathcal{D}_α^* , we get

$$P_D(\delta^*) - P_D(\delta) \geq \eta (\alpha - P_F(\delta)) \geq 0 \quad (1.106)$$

since $P_F(\delta) \leq \alpha$. In other words,

$$P_D(\delta) \leq P_D(\delta^*), \quad \delta \in \mathcal{D}_\alpha^*$$

and the test δ^* solves the constrained problem NP_α^* .

Meeting the constraint (1.104) We now show that the parameter $\eta > 0$ and the Borel mapping $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ can be selected so that the test δ^* indeed satisfies (1.104).

Lemma 1.18.2 *For every α in $(0, 1]$ it is always possible to select $\eta > 0$ and a Borel mapping $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ so that (1.104) holds.*

Proof. Note that

$$\begin{aligned} & P_F(\delta^*) \\ &= \int_{\mathbb{R}^k} \delta^*(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}) \\ &= \int_{\{\mathbf{y} \in \mathbb{R}^k : f_1(\mathbf{y}) = \eta f_0(\mathbf{y})\}} \gamma(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}) + \int_{\{\mathbf{y} \in \mathbb{R}^k : f_1(\mathbf{y}) > \eta f_0(\mathbf{y})\}} f_0(\mathbf{y}) dF(\mathbf{y}) \\ &= \int_{\{\mathbf{y} \in \mathbb{R}^k : f_1(\mathbf{y}) = \eta f_0(\mathbf{y})\}} \gamma(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}) + \mathbb{P}[f_1(\mathbf{Y}) > \eta f_0(\mathbf{Y}) | H = 0]. \end{aligned}$$

Therefore, as we seek to satisfy (1.104) we need to select $\eta > 0$ and a Borel mapping $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ such that

$$\alpha - \mathbb{P}[f_1(\mathbf{Y}) > \eta f_0(\mathbf{Y}) | H = 0] = \int_{\{\mathbf{y} \in \mathbb{R}^k : f_1(\mathbf{y}) = \eta f_0(\mathbf{y})\}} \gamma(\mathbf{y}) f_0(\mathbf{y}) dF(\mathbf{y}).$$

The form of this last relation suggests introducing the quantity $\eta(\alpha)$ defined by

$$\eta(\alpha) = \inf \{ \eta > 0 : \mathbb{P}[f_1(\mathbf{Y}) > \eta f_0(\mathbf{Y}) | H = 0] < \alpha \}.$$

The definition of $\eta(\alpha)$ is well posed since $\eta \rightarrow \mathbb{P}[f_1(\mathbf{Y}) > \eta f_0(\mathbf{Y}) | H = 0]$ is non-increasing on $(0, \infty)$.

Two cases are possible: If

$$\mathbb{P}[f_1(\mathbf{Y}) > \eta(\alpha) f_0(\mathbf{Y}) | H = 0] < \alpha,$$

then take $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ to be constant, say

$$\gamma(\mathbf{y}) = \gamma(\alpha), \quad \mathbf{y} \in \mathbb{R}^k.$$

This requires that $\gamma(\alpha)$ be selected so that

$$\begin{aligned} \alpha - \mathbb{P}[f_1(\mathbf{Y}) > \eta(\alpha)f_0(\mathbf{Y})|H = 0] \\ = \gamma(\alpha) \int_{\{\mathbf{y} \in \mathbb{R}^k: f_1(\mathbf{y}) = \eta(\alpha)f_0(\mathbf{y})\}} f_0(\mathbf{y})dF(\mathbf{y}). \end{aligned} \quad (1.107)$$

More compactly, we find that $\gamma(\alpha)$ is given by

$$\gamma(\alpha) = \frac{\alpha - \mathbb{P}[f_1(\mathbf{Y}) > \eta(\alpha)f_0(\mathbf{Y})|H = 0]}{\mathbb{P}[f_1(\mathbf{Y}) = \eta(\alpha)f_0(\mathbf{Y})|H = 0]}. \quad (1.108)$$

If

$$\mathbb{P}[f_1(\mathbf{Y}) > \eta(\alpha)f_0(\mathbf{Y})|H = 0] = \alpha,$$

then we must select the mapping $\gamma : \mathbb{R}^k \rightarrow [0, 1]$ so that

$$\int_{\{\mathbf{y} \in \mathbb{R}^k: f_1(\mathbf{y}) = \eta f_0(\mathbf{y})\}} \gamma(\mathbf{y})f_0(\mathbf{y})dF(\mathbf{y}) = 0.$$

Just take the constant mapping given by

$$\gamma(\mathbf{y}) = 0, \quad \mathbf{y} \in \mathbb{R}^k.$$

■

1.19 Examples

The Gaussian case Consider again the situation discussed in Section 1.10 where the observation rv \mathbf{Y} is conditionally Gaussian given H , i.e.,

$$\begin{aligned} H_1 : \mathbf{Y} &\sim N(\mathbf{m}_1, \mathbf{R}) \\ H_0 : \mathbf{Y} &\sim N(\mathbf{m}_0, \mathbf{R}) \end{aligned}$$

where \mathbf{m}_1 and \mathbf{m}_0 are distinct elements in \mathbb{R}^k , and the $k \times k$ symmetric matrix \mathbf{R} is positive definite (thus invertible). From the discussion given in Section 1.10, it follows for each $\lambda > 0$ the test d_λ takes the form

$$d_\lambda(\mathbf{y}) = 0 \quad \text{iff} \quad \mathbf{y}'\mathbf{R}^{-1}\Delta\mathbf{m} > \phi(\lambda)$$

with Δm and $\phi(\lambda)$ given by (1.37) and (1.38), respectively. We also have

$$P_F(d_\lambda) = 1 - \Phi\left(\frac{\log \lambda + \frac{1}{2}d^2}{d}\right).$$

where d^2 is given by (1.43) – It is plain that the function $\lambda \rightarrow P_F(d_\lambda)$ is continuous on \mathbb{R}_+ . Given α in the unit interval $(0, 1)$, the value $\lambda(\alpha)$ is *uniquely* determined through the relation

$$1 - \alpha = \Phi\left(\frac{\log \lambda + \frac{1}{2}d^2}{d}\right).$$

This is equivalent to

$$\lambda(\alpha) = e^{d \cdot x_1 - \alpha - \frac{1}{2}d^2}.$$

where for t in $(0, 1)$, let x_t denote the only solution to the equation

$$\Phi(x) = t, \quad x \in \mathbb{R}.$$

Discontinuity with Bernoulli rvs The setting is that of Section 1.9 to which we refer the reader for the notation. We discuss only the case $a_1 < a_0$, and leave the case $a_0 < a_1$ as an exercise for the interested reader. In Section 1.16 we have shown that

$$P_F(d_\lambda) = \begin{cases} 1 & \text{if } 0 < \lambda \leq \frac{a_1}{a_0} \\ 1 - a_0 & \text{if } \frac{a_1}{a_0} < \lambda \leq \frac{1-a_1}{1-a_0} \\ 0 & \text{if } \frac{1-a_1}{1-a_0} < \lambda \end{cases} \quad (1.109)$$

as λ ranges over $(0, \infty)$. Note that $\lambda \rightarrow P_F(d_\lambda)$ is left-continuous.

Discontinuity with Poisson rvs With $\mathcal{P}(m)$ denoting the Poisson pmf on \mathbb{N} with parameter $m > 0$, consider the following simple binary hypothesis testing problem

$$\begin{aligned} H_1 : & Y \sim \mathcal{P}(m_1) \\ H_0 : & Y \sim \mathcal{P}(m_0) \end{aligned}$$

where $m_1 \neq m_0$ in $(0, \infty)$. Thus,

$$\mathbb{P}[Y = k | H = h] = \frac{(m_h)^k}{k!} e^{-m_h}, \quad \begin{array}{l} h = 0, 1 \\ k = 0, 1, \dots \end{array}$$

In this example, we take F to be the counting measure on \mathbb{N} , and for every $\lambda \geq 0$, the definition of d_λ reduces to

$$\begin{aligned} d_\lambda(k) = 0 & \text{ iff } \frac{(m_1)^k}{k!} e^{-m_1} < \lambda \frac{(m_0)^k}{k!} e^{-m_0} \\ & \text{ iff } \left(\frac{m_1}{m_0} \right)^k < \lambda e^{-(m_0-m_1)} \end{aligned} \quad (1.110)$$

with $k = 0, 1, \dots$

If $m_0 < m_1$, then

$$\begin{aligned} d_\lambda(k) = 0 & \text{ iff } \frac{(m_1)^k}{k!} e^{-m_1} < \lambda \frac{(m_0)^k}{k!} e^{-m_0} \\ & \text{ iff } \left(\frac{m_1}{m_0} \right)^k < \lambda e^{-(m_0-m_1)} \\ & \text{ iff } k < \eta(\lambda) \end{aligned} \quad (1.111)$$

with $k = 0, 1, \dots$, where

$$\eta(\lambda) = \frac{\log \lambda e^{-(m_0-m_1)}}{\log \left(\frac{m_1}{m_0} \right)}.$$

It follows that

$$\begin{aligned} P_F(d_\lambda) &= \mathbb{P}[d_\lambda(Y) = 1 | H = 0] \\ &= \mathbb{P}[Y \geq \eta(\lambda) | H = 0] \\ &= \sum_{k=0: \eta(\lambda) \leq k}^{\infty} \frac{(m_0)^k}{k!} e^{-m_0}. \end{aligned} \quad (1.112)$$

In this last expression only the integer ceiling $\lceil \eta(\lambda) \rceil$ of $\eta(\lambda)$ matters, where $\lceil \eta(\lambda) \rceil = \inf \{k \in \mathbb{N} : \eta(\lambda) \leq k\}$, whence

$$P_F(d_\lambda) = \sum_{k=\lceil \eta(\lambda) \rceil}^{\infty} \frac{(m_0)^k}{k!} e^{-m_0}.$$

As a result, the mapping $\lambda \rightarrow P_F(d_\lambda)$ is easily seen to be a *left-continuous piecewise constant* mapping with

$$P_F(d_\lambda) = P_F(d_{\lambda_n}), \quad \begin{aligned} &\lambda_n < \lambda \leq \lambda_{n+1} \\ &n = 0, 1, \dots \end{aligned}$$

where $\{\lambda_n, n = 1, 2, \dots\}$ is a strictly monotone increasing sequence determined by the relation

$$n = \frac{\log \lambda_n e^{-(m_0 - m_1)}}{\log \left(\frac{m_1}{m_0} \right)}. \quad n = 1, 2, \dots$$

or equivalently,

$$\lambda_n = \left(\frac{m_1}{m_0} \right)^n e^{-(m_1 - m_0)}, \quad n = 1, 2, \dots$$

It is now plain that whenever α is chosen in $[0, 1]$ such that

for some integer $n = 0, 1, \dots$ then the requirement that $P_F(d_{\lambda(\alpha)}) = \alpha$ cannot be met. This difficulty is circumvented by enlarging \mathcal{D} with *randomized* policies; see Section 1.11.

1.20 The receiver operating characteristic (ROC)

A careful inspection of the solutions to the three formulations discussed so far shows that under mild assumptions, the tests of interest all take the form

$$d_\eta(\mathbf{y}) = 0 \quad \text{iff} \quad f_1(\mathbf{y}) < \eta f_0(\mathbf{y}) \quad (1.113)$$

for some $\eta > 0$. It is only the value of the threshold η that varies with the problem formulation. With the notation used earlier, we have

In the Bayesian formulation,

$$\eta_B = \frac{\Gamma_0(1 - p)}{\Gamma_1 p}$$

In the minimax formulation,

$$\eta_m = \frac{\Gamma_0(1 - p_m)}{\Gamma_1 p_m}$$

with p_m such that

$$V(p_m) = \max(V(p) : p \in [0, 1]).$$

When p is a point in $(0, 1)$ at which $V : [0, 1] \rightarrow \mathbb{R}$ is differentiable, then p_m can be characterized through the Minimax Equation.

In the Neyman–Pearson formulation,

$$\eta_{\text{NP}}(\alpha) = \lambda(\alpha).$$

with $\lambda(\alpha)$ satisfying the constraint (1.97).

In view of this, it seems natural to analyze in some details the performance of the tests (1.113). This is done by considering how their probabilities of detection and of false alarm, namely $P_F(d_\eta)$ and $P_D(d_\eta)$, vary in relation to each other as η ranges from $\eta = 0$ to $\eta = +\infty$. This is best understood by plotting the graph (Γ) of the detection probability against the probability of false alarm. Such a graph is analogous to a *phase portrait* for two-dimensional non-linear ODEs, and is called a *receiver operating characteristic (ROC) curve*. Its *parametric* representation is given by

$$\mathbb{R}_+ \rightarrow [0, 1] \times [0, 1] : \eta \rightarrow (P_F(d_\eta), (P_D(d_\eta))),$$

whence

$$(\Gamma) : \quad \{(P_F(d_\eta), (P_D(d_\eta))), \eta \geq 0\}.$$

This graph is *completely* determined by the probability distributions F_0 and F_1 of the observation rv \mathbf{Y} under the two hypotheses (through the densities f_0 and f_1 with respect to the underlying distribution F) and not by cost assignments or the prior probabilities. A typical ROC curve is drawn below.

Geometric properties The following geometric properties of the ROC curve are useful for using it.

Theorem 1.20.1 *Assume that the absolute continuity conditions (A.1) and (A.2) hold.*

(i): *Both mappings $\mathbb{R}_+ \rightarrow [0, 1] : \eta \rightarrow P_F(d_\eta)$ and $\mathbb{R}_+ \rightarrow [0, 1] : \eta \rightarrow P_D(d_\eta)$ are monotone non-increasing, with $P_F(d_0) = 1$ and $P_D(d_0) = 1$. Moreover,*

$$\lim_{\eta \uparrow \infty} P_F(d_\eta) = P_F(d_\infty) = 0$$

and

$$\lim_{\eta \uparrow \infty} P_D(d_\eta) = P_D(d_\infty) = 0.$$

(ii): If the right-derivative of $\eta \rightarrow P_F(d_\eta)$ exists at $\eta = \lambda$ for some $\lambda \geq 0$, then the right-derivative of $\eta \rightarrow P_D(d_\eta)$ exists at $\eta = \lambda$, and the relation

$$\frac{d^+}{d\eta} P_D(d_\eta) \Big|_{\eta=\lambda} = \lambda \cdot \frac{d^+}{d\eta} P_F(d_\eta) \Big|_{\eta=\lambda} \quad (1.114)$$

holds.

(iii): If the left-derivative of $\eta \rightarrow P_F(d_\eta)$ exists at $\eta = \lambda$ for some $\lambda > 0$, then the left-derivative of $\eta \rightarrow P_D(d_\eta)$ exists at $\eta = \lambda$, and the relation

$$\frac{d^-}{d\eta} P_D(d_\eta) \Big|_{\eta=\lambda} = \lambda \cdot \frac{d^-}{d\eta} P_F(d_\eta) \Big|_{\eta=\lambda} \quad (1.115)$$

holds.

The proof of Theorem 1.20.1 is available in Section 1.22. It follows from this last result that whenever the mapping $\eta \rightarrow P_F(d_\eta)$ is *strictly* decreasing, then the mapping $\eta \rightarrow P_D(d_\eta)$ is also *strictly* decreasing, whence the curve (Γ) can be represented as the graph of a function $\Gamma : [0, 1] \rightarrow [0, 1] : P_F \rightarrow P_D = \Gamma(P_F)$, namely

$$P_D(d_\eta) = \Gamma(P_F(d_\eta)), \quad \eta \geq 0. \quad (1.116)$$

In such circumstance, Theorem 1.20.1 yields the the following information concerning this mapping.

Corollary 1.20.1 *Assume the absolute continuity conditions (A.1)-(A.2) to hold. Whenever the mapping $\mathbb{R}_+ \rightarrow [0, 1] : \eta \rightarrow P_F(d_\eta)$ is differentiable and strictly decreasing, so is the mapping $\mathbb{R}_+ \rightarrow [0, 1] : \eta \rightarrow P_D(d_\eta)$. In that case, the mapping $\Gamma : [0, 1] \rightarrow [0, 1]$ is differentiable, strictly increasing and concave with*

$$\frac{d\Gamma}{dP_F}(P_F(d_\eta)) = \eta, \quad \eta \geq 0. \quad (1.117)$$

Proof. By the very definition of the function Γ , the identity (1.116) must hold. Now Theorem 1.20.1 and the Implicit Function Theorem yield the differentiability of the mapping $\Gamma : [0, 1] \rightarrow [0, 1]$. Differentiating both sides of (1.116) with respect of η we find

$$\begin{aligned} \frac{d}{d\eta} P_D(d_\eta) &= \frac{d}{d\eta} \Gamma(P_F(d_\eta)) \\ &= \frac{d\Gamma}{dP_F}(P_F(d_\eta)) \cdot \frac{d}{d\eta} P_F(d_\eta) \end{aligned} \quad (1.118)$$

as we use the Chain Rule. But Theorem 1.20.1 implies also that

$$\frac{d}{d\eta} P_D(d_\eta) = \eta \frac{d}{d\eta} P_F(d_\eta).$$

Combining these facts we conclude that

$$\eta \frac{d}{d\eta} P_F(d_\eta) = \frac{d\Gamma}{dP_F}(P_F(d_\eta)) \cdot \frac{d}{d\eta} P_F(d_\eta).$$

The mapping $\mathbb{R}_+ \rightarrow [0, 1] : \eta \rightarrow P_F(d_\eta)$ being assumed differentiable and strictly decreasing, we have $\frac{d}{d\eta} P_F(d_\eta) < 0$. Dividing by $\frac{d}{d\eta} P_F(d_\eta)$ we get (1.117). The other properties follow readily. ■

1.21 Operating the ROC and some examples

How to use the ROC These results are most useful for operationally using the ROC curve:

For the Neyman–Pearson test of size α , consider the point on the ROC curve with abscissa α . It is determined by the threshold value $\eta(\alpha)$ with the property that

$$P_F(d_{\eta(\alpha)}) = \alpha.$$

and $d_{NP}(\alpha)$ is simply $d_{\eta(\alpha)}$. Note that $\eta(\alpha)$ is the *slope* of the tangent to the ROC curve at the point with abscissa α and the power $\beta(\alpha)$ of the test is simply the ordinate of that point.

For the Bayesian problem, η is determined by the cost assignment and the prior distribution of the rv H . The values of $P_D(d_\eta)$ and $P_F(d_\eta)$ can be easily determined by finding the point on the ROC where the tangent has slope η .

The Minimax Equation takes the form

$$C(1, 1) - C(0, 0) + \Gamma_1 P_M(d_\eta) - \Gamma_0 P_F(d_\eta) = 0,$$

or equivalently

$$C(1, 1) - C(0, 0) + \Gamma_1 = \Gamma_1 P_D(d_\eta) + \Gamma_0 P_F(d_\eta).$$

This shows that the minimax rule d_m^* is obtained as follows. Consider the straight line (L) in the (P_F, P_D) -plane with equation

$$(L) \quad C(1, 1) - C(0, 0) + \Gamma_1 = \Gamma_1 P_D + \Gamma_0 P_F.$$

Let (P_F^*, P_D^*) be the point of intersection of the straight line (L) with the ROC curve (Γ) , and let η^* be the corresponding threshold value, i.e., $P_F^* = P_F(d_{\eta^*})$ and $P_D^* = P_D(d_{\eta^*})$. It is now clear that $d_m^* = d_{\eta^*}$.

Building on material developed earlier we now discuss the ROC in the Gaussian and bernoulli cases, respectively.

The Gaussian case The setting is that of Section 1.8 to which we refer the reader for the notation. As shown there, for any $\eta > 0$ we have

$$P_F(d_\eta) = 1 - \Phi\left(\frac{\log \eta + \frac{1}{2}d^2}{d}\right)$$

and

$$P_M(d_\eta) = \Phi\left(\frac{\log \eta - \frac{1}{2}d^2}{d}\right)$$

so that

$$P_D(d_\eta) = 1 - \Phi\left(\frac{\log \eta - \frac{1}{2}d^2}{d}\right).$$

To find the ROC curve, note that

$$d\Phi^{-1}(1 - P_F(d_\eta)) - \frac{d^2}{2} = \log \eta,$$

while we must have

$$d\Phi^{-1}(1 - P_D(d_\eta)) + \frac{d^2}{2} = \log \eta,$$

whence

$$d\Phi^{-1}(1 - P_F(d_\eta)) - \frac{d^2}{2} = d\Phi^{-1}(1 - P_D(d_\eta)) + \frac{d^2}{2}.$$

It follows that

$$\Phi^{-1}(1 - P_D(d_\eta)) = \Phi^{-1}(1 - P_F(d_\eta)) - d$$

so that

$$1 - P_D(d_\eta) = \Phi\left(\Phi^{-1}(1 - P_F(d_\eta)) - d\right)$$

This shows that here the mapping $\Gamma : [0, 1] \rightarrow [0, 1]$ is well defined and given by

$$P_D = 1 - \Phi\left(\Phi^{-1}(1 - P_F) - d\right), \quad P_F \in [0, 1].$$

The Bernoulli case The setting is that of Section 1.9 to which we refer the reader for the notation. We discuss only the case $a_1 < a_0$, and leave the case $a_0 < a_1$ as an exercise for the interested reader. As shown in Section 1.16 we have

$$P_F(d_\eta) = \begin{cases} 1 & \text{if } 0 < \eta \leq \frac{a_1}{a_0} \\ 1 - a_0 & \text{if } \frac{a_1}{a_0} < \eta \leq \frac{1-a_1}{1-a_0} \\ 0 & \text{if } \frac{1-a_1}{1-a_0} < \eta \end{cases} \quad (1.119)$$

and

$$P_D(d_\eta) = \begin{cases} 1 & \text{if } 0 < \eta \leq \frac{a_1}{a_0} \\ 1 - a_1 & \text{if } \frac{a_1}{a_0} < \eta \leq \frac{1-a_1}{1-a_0} \\ 0 & \text{if } \frac{1-a_1}{1-a_0} < \eta. \end{cases} \quad (1.120)$$

Therefore,

$$\{(P_F(d_\eta), P_D(d_\eta)), \eta \geq 0\} = \{(0, 0), (1 - a_0, 1 - a_1), (1, 1)\}.$$

Strictly speaking, in this case the ROC is not a “curve” as it spans only three points. However, the points are on two segments $[(0, 0), (1 - a_0, 1 - a_1)]$ and $[(1 - a_0, 1 - a_1), (1, 1)]$ are achievable through randomization.

1.22 Proofs

Preliminaries We start with some facts that prove useful in discussing both Lemma 1.17.1 and Theorem 1.20.1.

Fix $\lambda > 0$, and define the set

$$R(\lambda) \equiv \{\mathbf{y} \in \mathbb{R}^k : f_1(\mathbf{y}) \geq \lambda f_0(\mathbf{y})\}.$$

Noting that

$$d_\lambda(\mathbf{y}) = 1 \quad \text{iff} \quad \mathbf{y} \in R(\lambda),$$

it is plain that

$$\begin{aligned}
P_F(d_\lambda) &= \mathbb{P}[d_\lambda(\mathbf{y}) = 1 | H = 0] \\
&= \mathbb{P}[\mathbf{Y} \in R(\lambda) | H = 0] \\
&= \int_{R(\lambda)} f_0(\mathbf{y}) dF(\mathbf{y})
\end{aligned} \tag{1.121}$$

and

$$\begin{aligned}
P_D(d_\lambda) &= \mathbb{P}[d_\lambda(\mathbf{y}) = 1 | H = 1] \\
&= \int_{R(\lambda)} f_1(\mathbf{y}) dF(\mathbf{y}).
\end{aligned} \tag{1.122}$$

For each $\Delta\lambda > 0$, easy algebra now leads to

$$\begin{aligned}
P_F(d_{\lambda+\Delta\lambda}) - P_F(d_\lambda) &= \int_{R(\lambda+\Delta\lambda)} f_0(\mathbf{y}) dF(\mathbf{y}) - \int_{R(\lambda)} f_0(\mathbf{y}) dF(\mathbf{y}) \\
&= - \int_{R_+(\lambda; \Delta\lambda)} f_0(\mathbf{y}) dF(\mathbf{y})
\end{aligned} \tag{1.123}$$

where

$$R_+(\lambda; \Delta\lambda) \equiv \{\mathbf{y} \in \mathbb{R}^k : \lambda f_0(\mathbf{y}) \leq f_1(\mathbf{y}) < (\lambda + \Delta\lambda) f_0(\mathbf{y})\}.$$

Similarly, we have

$$\begin{aligned}
P_F(d_{\lambda-\Delta\lambda}) - P_F(d_\lambda) &= \int_{R(\lambda-\Delta\lambda)} f_0(\mathbf{y}) dF(\mathbf{y}) - \int_{R(\lambda)} f_0(\mathbf{y}) dF(\mathbf{y}) \\
&= \int_{R_-(\lambda; \Delta\lambda)} f_0(\mathbf{y}) dF(\mathbf{y})
\end{aligned} \tag{1.124}$$

where

$$R_-(\lambda; \Delta\lambda) \equiv \{\mathbf{y} \in \mathbb{R}^k : (\lambda - \Delta\lambda) f_0(\mathbf{y}) \leq f_1(\mathbf{y}) < \lambda f_0(\mathbf{y})\}.$$

A proof of Lemma 1.17.1 Fix $\lambda > 0$ and $\Delta\lambda > 0$. When F is Lebesgue measure on \mathbb{R}^k , these relations become

$$P_F(d_{\lambda \pm \Delta\lambda}) - P_F(d_\lambda) = \mp \int_{R_\pm(\lambda; \Delta\lambda)} f_0(\mathbf{y}) d\mathbf{y}$$

$$\begin{aligned}
P_F(d_{\lambda+\Delta\lambda}) - P_F(d_\lambda) &= - \int_{R_+(\lambda; \Delta\lambda)} f_0(\mathbf{y}) d\mathbf{y} \\
&= - \int_{\mathbb{R}} \mathbf{1}[\mathbf{y} \in R_+(\lambda; \Delta\lambda)] f_0(\mathbf{y}) d\mathbf{y} \quad (1.125)
\end{aligned}$$

and

$$\begin{aligned}
P_F(d_{\lambda-\Delta\lambda}) - P_F(d_\lambda) &= \int_{R_-(\lambda; \Delta\lambda)} f_0(\mathbf{y}) d\mathbf{y} \\
&= \int_{\mathbb{R}} \mathbf{1}[\mathbf{y} \in R_-(\lambda; \Delta\lambda)] f_0(\mathbf{y}) d\mathbf{y}. \quad (1.126)
\end{aligned}$$

Because

$$\cap_{\Delta\lambda>0} R_+(\lambda; \Delta\lambda) = \{\mathbf{y} \in \mathbb{R}^k : \lambda f_0(\mathbf{y})\}$$

and

$$\cap_{\Delta\lambda>0} R_-(\lambda; \Delta\lambda) = \{\mathbf{y} \in \mathbb{R}^k : \lambda f_0(\mathbf{y})\}$$

we note that

$$\lim_{\Delta\lambda \downarrow 0} \mathbf{1}[\mathbf{y} \in R_+(\lambda; \Delta\lambda)] = \mathbf{1}[\mathbf{y} \in R^*(\lambda)], \quad \mathbf{y} \in \mathbb{R}^k$$

and

$$\lim_{\Delta\lambda \downarrow 0} \mathbf{1}[\mathbf{y} \in R_-(\lambda; \Delta\lambda)] = \mathbf{1}[\mathbf{y} \in R^*(\lambda)], \quad \mathbf{y} \in \mathbb{R}^k$$

A proof of Theorem 1.20.1 The first part of the theorem readily follows from the monotonicity of the sets $\{R(\eta), \eta \geq 0\}$, namely

$$R(\eta_2) \subseteq R(\eta_1), \quad 0 \leq \eta_1 < \eta_2.$$

Next, fix $\eta \geq 0$. With $\Delta\eta > 0$, it is easy to see that The very definition of $R(\eta; \Delta\eta)$ implies the inequalities

$$\eta \int_{R(\eta; \Delta\eta)} f_0(\mathbf{y}) dF(\mathbf{y}) \leq \int_{R(\eta; \Delta\eta)} f_1(\mathbf{y}) dF(\mathbf{y})$$

and

$$\int_{R(\eta; \Delta\eta)} f_1(\mathbf{y}) dF(\mathbf{y}) \leq (\eta + \Delta\eta) \int_{R(\eta; \Delta\eta)} f_0(\mathbf{y}) dF(\mathbf{y}).$$

It then follows that

$$(\eta + \Delta\eta) \cdot \frac{P_F(d_{\eta+\Delta\eta}) - P_F(d_\eta)}{\Delta\eta} \leq \frac{P_D(d_{\eta+\Delta\eta}) - P_D(d_\eta)}{\Delta\eta}$$

and

$$\frac{P_D(d_{\eta+\Delta\eta}) - P_D(d_\eta)}{\Delta\eta} \leq \eta \cdot \frac{P_F(d_{\eta+\Delta\eta}) - P_F(d_\eta)}{\Delta\eta}.$$

If the right-derivative of $\eta \rightarrow P_F(d_\eta)$ exists, then

$$\frac{d^+}{d\eta} P_F(d_\eta) = \lim_{\Delta\eta \downarrow 0} \frac{P_F(d_{\eta+\Delta\eta}) - P_F(d_\eta)}{\Delta\eta}$$

and an easy sandwich argument shows that the limit

$$\lim_{\Delta\eta \downarrow 0} \frac{P_D(d_{\eta+\Delta\eta}) - P_D(d_\eta)}{\Delta\eta}$$

also exists. Therefore, the right-derivative of $\eta \rightarrow P_D(d_\eta)$ also exists and is given by

$$\frac{d^+}{d\eta} P_D(d_\eta) = \eta \cdot \frac{d^+}{d\eta} P_F(d_\eta).$$

■

1.23 Exercises

1.24 References

Chapter 2

The M -ary hypothesis testing problem

As we shall see in this chapter and in the next one, the simple binary hypothesis testing problem of Chapter 1 admits several important generalizations. The more general version discussed here assumes that there are more than two hypotheses; it is of particular relevance to the design of optimal receivers in digital modulation. In this formulation, nature assumes M distinct states, labeled $0, 1, \dots, M-1$, and is now encoded in a rv H which take values in the discrete set $\{0, 1, \dots, M-1\}$. A decision has to be made as to which of these M hypotheses is the correct one on the basis of an observation Y which is statistically related to H .

2.1 Motivating examples

Digital communications

Manufacturing

2.2 The probabilistic model

To formulate the M -ary hypothesis testing problem we proceed very much as in Chapter 1: With positive integer $M \geq 2$, we are given M distinct probability distribution functions F_0, \dots, F_{M-1} on \mathbb{R}^k , and a pmf $\mathbf{p} = (p_0, \dots, p_{M-1})$ on

$\{0, 1, \dots, M-1\}$ with

$$0 \leq p_m \leq 1, \quad m = 0, 1, \dots, M-1 \quad \text{and} \quad \sum_{m=0}^{M-1} p_m = 1.$$

The situation is summarized by

$$H_m : \mathbf{Y} \sim F_m, \quad m = 0, 1, \dots, M-1. \quad (2.1)$$

We construct a sample space Ω equipped with a σ -field of events \mathcal{F} , and rvs H and \mathbf{Y} defined on it and taking values in $\{0, 1, \dots, M-1\}$ and \mathbb{R}^k , respectively. Now the probability distribution functions F_0, \dots, F_{M-1} have the interpretation that

$$F_m(\mathbf{y}) = \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = m], \quad \mathbf{y} \in \mathbb{R}^k, \quad m = 0, 1, \dots, M-1.$$

The probability distribution of the rv H is specified by the pmf $\mathbf{p} = (p_0, \dots, p_{M-1})$ with

$$p_m = \mathbb{P}[H = m], \quad m = 0, 1, \dots, M-1.$$

Again, the conditional probability distributions of the observations given the hypothesis *and* the probability distribution of H completely specify the *joint* distribution of the rvs H and \mathbf{Y} : Indeed, for each $m = 0, 1, \dots, M-1$,

$$\begin{aligned} \mathbb{P}[\mathbf{Y} \leq \mathbf{y}, H = m] &= \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = m] \mathbb{P}[H = m] \\ &= p_m F_m(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \quad (2.2)$$

The *unconditional* probability distribution function of the rv \mathbf{Y} is easily determined to be

$$\mathbb{P}[\mathbf{Y} \leq \mathbf{y}] = \sum_{m=0}^{M-1} \mathbb{P}[\mathbf{Y} \leq \mathbf{y}, H = m], \quad \mathbf{y} \in \mathbb{R}^k$$

by the law of total probabilities. For future reference, we set

$$\begin{aligned} G(\mathbf{y}) &\equiv \mathbb{P}[\mathbf{Y} \leq \mathbf{y}] \\ &= \sum_{m=0}^{M-1} p_m F_m(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \quad (2.3)$$

During the discussion, several assumptions will be enforced on the probability distributions F_0, \dots, F_{M-1} . The most common assumption is denoted by **(A.3)**

for sake of convenience; it parallels Condition **(A.1)** made in the binary case: The probability distributions F_0, \dots, F_{M-1} on \mathbb{R}^k satisfy condition **(A.3)** if they are *absolutely continuous* with respect to some distribution F on \mathbb{R}^k . This is equivalent to saying that for each $m = 0, 1, \dots, M-1$, there exists a Borel mapping $f_m : \mathbb{R}^k \rightarrow \mathbb{R}_+$ such that

$$F_m(\mathbf{y}) = \int_{-\infty}^{\mathbf{y}} f_m(\boldsymbol{\eta}) dF(\boldsymbol{\eta}), \quad \mathbf{y} \in \mathbb{R}^k, \quad m = 0, 1, \dots, M-1. \quad (2.4)$$

We refer to these Borel mappings as probability density functions with respect to F .

This condition is hardly constraining since we can always take F to be the average of the M distributions F_0, \dots, F_{M-1} . i.e.,

$$F(\mathbf{y}) \equiv \frac{1}{M} \left(\sum_{m=0}^{M-1} F_m(\mathbf{y}) \right), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.5)$$

However, in most applications F is either Lebesgue measure on \mathbb{R}^k or a counting measure on some countable subset of \mathbb{R}^k .

Under Condition **(A.3)**, it is plain that the unconditional probability distribution function of the rv \mathbf{Y} is automatically absolutely continuous with respect to the distribution F on \mathbb{R}^k . Indeed it follows from (2.4) and (2.3) that

$$\begin{aligned} G(\mathbf{y}) &= \sum_{m=0}^{M-1} p_m \int_{-\infty}^{\mathbf{y}} f_m(\boldsymbol{\eta}) dF(\boldsymbol{\eta}) \\ &= \int_{-\infty}^{\mathbf{y}} \left(\sum_{m=0}^{M-1} p_m f_m(\boldsymbol{\eta}) \right) dF(\boldsymbol{\eta}) \\ &= \int_{-\infty}^{\mathbf{y}} g(\boldsymbol{\eta}) dF(\boldsymbol{\eta}). \quad \mathbf{y} \in \mathbb{R}^k \end{aligned} \quad (2.6)$$

with Borel mapping $g : \mathbb{R}^k \rightarrow \mathbb{R}_+$ given by

$$g(\mathbf{y}) = \sum_{m=0}^{M-1} p_m f_m(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k. \quad (2.7)$$

In other words, the unconditional probability distribution function of \mathbf{Y} admits $g : \mathbb{R}^k \rightarrow \mathbb{R}_+$ as probability density function with respect to F .

2.3 Admissible tests

An *admissible* decision rule or test is any *Borel* mapping $d : \mathbb{R}^k \rightarrow \{0, 1, \dots, M-1\}$. The collection of all such admissible rules is still denoted by \mathcal{D} . Again the measurability requirement is imposed to guarantee that the mapping $d(Y) : \Omega \rightarrow \{0, 1, \dots, M-1\} : \omega \rightarrow d(Y(\omega))$ is indeed a rv, i.e., $\{\omega \in \Omega : d(Y(\omega)) = m\}$ is an event in \mathcal{F} for all $m = 0, 1, \dots, M-1$. For every rule d in \mathcal{D} , we introduce the sets

$$C_m(d) \equiv \{\mathbf{y} \in \mathbb{R}^k : d(\mathbf{y}) = m\}, \quad m = 0, 1, \dots, M-1.$$

The collection of M sets $C_0(d), \dots, C_{M-1}(d)$ forms an M -ary Borel partition of \mathbb{R}^k in the following sense:

Definition 2.3.1 *The collection of M subsets c if the following three conditions are satisfied:*

- (i) *For each $m = 0, 1, \dots, M-1$, the set C_m is a Borel subset of \mathbb{R}^k ;*
- (ii) *We have*

$$C_m \cap C_k = \emptyset, \quad \begin{array}{l} m \neq k \\ m, k = 0, 1, \dots, M-1 \end{array}$$

and

- (iii) *The condition*

$$\cup_{m=0}^{M-1} C_m = \mathbb{R}^k$$

holds.

The collection of all M -ary Borel partitions of \mathbb{R}^k is denoted $\mathcal{P}_M(\mathbb{R}^k)$. Note that any M -ary Borel measurable partition C_0, \dots, C_{M-1} in $\mathcal{P}_M(\mathbb{R}^k)$ determines an admissible test $d_{C_1, \dots, C_{M-1}}$ in \mathcal{D} through

$$d_{C_1, \dots, C_{M-1}}(\mathbf{y}) = m \quad \text{if } \mathbf{y} \in C_m.$$

Therefore, there is a one-to-one correspondence between \mathcal{D} and $\mathcal{P}_M(\mathbb{R}^k)$.

2.4 The Bayesian formulation

The probabilistic model The Bayesian formulation assumes *knowledge* of the prior distribution $\mathbf{p} = (p_0, \dots, p_{M-1})$ of the rv H , and of the conditional distributions F_0, \dots, F_{M-1} of the rv \mathbf{Y} given H .

The optimization problem A cost is incurred for making decisions. This is captured through the mapping $C : \{0, 1, \dots, M - 1\} \times \{0, 1, \dots, M - 1\} \rightarrow \mathbb{R}$ with the interpretation that

$$C(m, \ell) = \left(\begin{array}{l} \text{Cost incurred for deciding } \ell \\ \text{when } H = m \end{array} \right), \quad \ell, m = 0, 1, \dots, M - 1.$$

Using the admissible rule d in \mathcal{D} incurs a cost $C(H, d(\mathbf{Y}))$, but as for the binary hypothesis testing problem, the value of the cost $C(H, d(\mathbf{Y}))$ is not available, and attention focuses on the *expected cost* $J : \mathcal{D} \rightarrow \mathbb{R}$ given by

$$J(d) \equiv \mathbb{E}[C(H, d(\mathbf{Y}))], \quad d \in \mathcal{D}.$$

Here as well, the *Bayesian Problem* \mathcal{P}_B is the minimization problem

$$\mathcal{P}_B : \quad \text{Minimize } J(d) \text{ over } d \text{ in } \mathcal{D}.$$

Its solution is any test d^* in \mathcal{D} such that

$$J(d^*) \leq J(d), \quad d \in \mathcal{D}. \quad (2.8)$$

Any test d^* in \mathcal{D} which satisfies (2.8) is called a *Bayesian test*, and the value

$$J(d^*) = \inf_{d \in \mathcal{D}} J(d) = \min_{d \in \mathcal{D}} J(d) \quad (2.9)$$

is known as the *Bayesian cost*.

Solving the Bayesian problem The solution of \mathcal{P}_B parallels that given in Section 1.5 for the binary case but with some important differences: Fix a test d in \mathcal{D} . The decomposition

$$\mathbf{1}[d(\mathbf{Y}) = H] + \mathbf{1}[d(\mathbf{Y}) \neq H] = 1$$

holds so that

$$\begin{aligned} C(H, d(\mathbf{Y})) &= \mathbf{1}[d(\mathbf{Y}) = H] C(H, H) + \mathbf{1}[d(\mathbf{Y}) \neq H] C(H, d(\mathbf{Y})) \\ &= C(H, H) + (C(H, d(\mathbf{Y})) - C(H, H)) \mathbf{1}[d(\mathbf{Y}) \neq H]. \end{aligned} \quad (2.10)$$

Defining the auxiliary expected cost function $\hat{J} : \mathcal{D} \rightarrow \mathbb{R}$ to be

$$\hat{J}(d) = \mathbb{E}[(C(H, d(\mathbf{Y})) - C(H, H)) \mathbf{1}[d(\mathbf{Y}) \neq H]], \quad d \in \mathcal{D} \quad (2.11)$$

we again conclude that

$$J(d) = E[C(H, H)] + \widehat{J}(d), \quad d \in \mathcal{D}.$$

Therefore, solving \mathcal{P}_B is equivalent to solving the auxiliary problem $\widehat{\mathcal{P}}_B$ where

$$\widehat{\mathcal{P}}_B : \quad \text{Minimize } \widehat{J}(d) \text{ over } d \text{ in } \mathcal{D}.$$

Preparatory computations Fix d in \mathcal{D} . Note that

$$\begin{aligned} \widehat{J}(d) &= \mathbb{E} [(C(H, d(\mathbf{Y})) - C(H, H)) \mathbf{1} [d(\mathbf{Y}) \neq H]] \\ &= \sum_{m=0}^{M-1} \left(\sum_{\ell=0, \ell \neq m}^{M-1} \mathbb{E} [(C(m, \ell) - C(m, m)) \mathbf{1} [H = m, d(\mathbf{Y}) = \ell]] \right) \\ &= \sum_{m=0}^{M-1} \left(\sum_{\ell=0, \ell \neq m}^{M-1} (C(m, \ell) - C(m, m)) \mathbb{P} [H = m, d(\mathbf{Y}) = \ell] \right) \\ &= \sum_{m=0}^{M-1} \left(\sum_{\ell=0, \ell \neq m}^{M-1} (C(m, \ell) - C(m, m)) \mathbb{P} [d(\mathbf{Y}) = \ell | H = m] p_m \right) \\ &= \sum_{m=0}^{M-1} \left(\sum_{\ell=0, \ell \neq m}^{M-1} p_m (C(m, \ell) - C(m, m)) \int_{C_\ell(d)} f_m(\mathbf{y}) dF(\mathbf{y}) \right). \end{aligned}$$

Interchanging the summations we get

$$\begin{aligned} \widehat{J}(d) &= \sum_{\ell=0}^{M-1} \left(\sum_{m=0, m \neq \ell}^{M-1} p_m (C(m, \ell) - C(m, m)) \int_{C_\ell(d)} f_m(\mathbf{y}) dF(\mathbf{y}) \right) \\ &= \sum_{\ell=0}^{M-1} \int_{C_\ell(d)} \left(\sum_{m=0, m \neq \ell}^{M-1} p_m (C(m, \ell) - C(m, m)) f_m(\mathbf{y}) \right) dF(\mathbf{y}) \\ &= \sum_{\ell=0}^{M-1} \int_{C_\ell(d)} h_\ell(\mathbf{y}) dF(\mathbf{y}) \tag{2.12} \end{aligned}$$

where for each $\ell = 0, 1, \dots, M-1$, the mapping $h_\ell : \mathbb{R}^k \rightarrow \mathbb{R}$ is given by

$$h_\ell(\mathbf{y}) = \sum_{m=0, m \neq \ell}^{M-1} p_m (C(m, \ell) - C(m, m)) f_m(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

Solving $\widehat{\mathcal{P}}_B$ Now define the mapping $h : \mathbb{R}^k \rightarrow \mathbb{R}$ given by

$$h(\mathbf{y}) = \min_{\ell=0,1,\dots,M-1} h_\ell(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

The following facts

$$h(\mathbf{y}) = h_{d^*}(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.13)$$

and

$$h(\mathbf{y}) \leq h_\ell(\mathbf{y}), \quad \ell = 0, 1, \dots, M-1, \quad \mathbf{y} \in \mathbb{R}^k \quad (2.14)$$

are simple consequences of the definitions.

Theorem 2.4.1 *Assume the absolute continuity condition (A.1) to hold. The test $d^* : \mathbb{R}^k \rightarrow \{0, 1, \dots, M-1\}$ given by*

$$d^*(\mathbf{y}) = \arg \min (\ell = 0, \dots, M-1 : h_\ell(\mathbf{y})), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.15)$$

(with a lexicographic tiebreaker in the event of ties) is admissible and solves Problem $\widehat{\mathcal{P}}_B$, hence solves the Bayesian Problem \mathcal{P}_B .

Proof. Pick an arbitrary test in \mathcal{D} . Using the expression (2.9) we get

$$\begin{aligned} & \widehat{J}(d) - \widehat{J}(d^*) \\ &= \sum_{\ell=0}^{M-1} \int_{C_\ell(d)} h_\ell(\mathbf{y}) dF(\mathbf{y}) - \sum_{\ell=0}^{M-1} \int_{C_\ell(d^*)} h_\ell(\mathbf{y}) dF(\mathbf{y}) \\ &\geq \sum_{\ell=0}^{M-1} \int_{C_\ell(d)} h(\mathbf{y}) dF(\mathbf{y}) - \sum_{\ell=0}^{M-1} \int_{C_\ell(d^*)} h_\ell(\mathbf{y}) dF(\mathbf{y}) \\ &= \sum_{\ell=0}^{M-1} \int_{C_\ell(d)} h(\mathbf{y}) dF(\mathbf{y}) - \sum_{\ell=0}^{M-1} \int_{C_\ell(d^*)} h(\mathbf{y}) dF(\mathbf{y}) \\ &= \int_{\mathbb{R}^k} h(\mathbf{y}) dF(\mathbf{y}) - \int_{\mathbb{R}^k} h(\mathbf{y}) dF(\mathbf{y}) \\ &= 0 \end{aligned} \quad (2.16)$$

We made use of the fact that the two collections $C_0(d), \dots, C_{M-1}(d)$ and $C_0(d^*), \dots, C_{M-1}(d^*)$ of Borel subsets of \mathbb{R}^k each form a Borel partition of \mathbb{R}^k . In particular, this implies that

$$\sum_{\ell=0}^{M-1} \int_{C_\ell(d)} h(\mathbf{y}) dF(\mathbf{y}) = \int_{\mathbb{R}^k} h(\mathbf{y}) dF(\mathbf{y})$$

and

$$\sum_{\ell=0}^{M-1} \int_{C_\ell(d^*)} h(\mathbf{y}) dF(\mathbf{y}) = \int_{\mathbb{R}^k} h(\mathbf{y}) dF(\mathbf{y}).$$

We conclude

$$\widehat{J}(d) - \widehat{J}(d^*) \geq 0, \quad d \in \mathcal{D}$$

whence

$$J(d) - J(d^*) \geq 0, \quad d \in \mathcal{D},$$

and the optimality of d^* is now established. ■

2.5 The probability of error criterion

When C takes the form

$$C(m, k) = \mathbf{1}[\ell \neq m], \quad m, \ell = 0, 1, \dots, M-1,$$

the expected cost reduces to the *probability of error*

$$P_E(d) = \mathbb{P}[d(\mathbf{Y}) \neq H], \quad d \in \mathcal{D}.$$

The Bayesian test $d^* : \mathbb{R}^k \rightarrow \{0, 1, \dots, M-1\}$ given by (2.15) now takes the following form: For each $\ell = 0, \dots, M-1$, the mapping $h_\ell : \mathbb{R}^k \rightarrow \mathbb{R}$ is now given by

$$h_\ell(\mathbf{y}) = \sum_{m=0, m \neq \ell}^{M-1} p_m f_m(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k.$$

But the probability distribution function $G : \mathbb{R}^k \rightarrow [0, 1]$ of the observation rv \mathbf{Y} is given by (2.3), and under condition **(A.1)** it has probability density function $g : \mathbb{R}^k \rightarrow \mathbb{R}_+$ given by (2.7). Therefore, for each $\ell = 0, \dots, M-1$, we have

$$h_\ell(\mathbf{y}) = g(\mathbf{y}) - p_\ell f_\ell(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^k$$

and the test $d^* : \mathbb{R}^k \rightarrow \{0, 1, \dots, M-1\}$ can be rewritten more compactly as

$$d^*(\mathbf{y}) = \arg \max (\ell = 0, \dots, M-1 : p_\ell f_\ell(\mathbf{y})), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.17)$$

with a lexicographic tiebreaker in the event of ties.

The ML test When all the hypotheses are *equally likely*, namely

$$p_0 = \dots = p_{M-1} = \frac{1}{M},$$

then (2.17) becomes

$$d^*(\mathbf{y}) = \arg \max (\ell = 0, \dots, M-1 : f_\ell(\mathbf{y})), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.18)$$

with a lexicographic tiebreaker in the event of ties, so that

$$d^*(\mathbf{y}) = m \quad \text{iff} \quad f_m(\mathbf{y}) = \max (f_\ell(\mathbf{y}), \ell = 0, 1, \dots, M-1) \quad (2.19)$$

with a lexicographic tiebreaker in the event of ties.

The MAP computer Bayes's Theorem gives

$$\mathbb{P}[H = \ell | \mathbf{Y} = \mathbf{y}] = \frac{p_\ell f_\ell(\mathbf{y})}{\sum_{m=0}^{M-1} p_m f_m(\mathbf{y})}, \quad \begin{array}{l} \ell = 0, 1, \dots, M-1, \\ \mathbf{y} \in \mathbb{R}^k. \end{array}$$

This allows rewriting the test $d^* : \mathbb{R}^k \rightarrow \{0, 1, \dots, M-1\}$ in the more compact form

$$d^*(\mathbf{y}) = \arg \max (\ell = 0, \dots, M-1 : \mathbb{P}[H = \ell | \mathbf{Y} = \mathbf{y}]), \quad \mathbf{y} \in \mathbb{R}^k \quad (2.20)$$

with a lexicographic tiebreaker in the event of ties. As with the binary case, we refer to this rule as the *Maximum A Posteriori* computer.

2.6 The Gaussian case

Consider the case where the rv \mathbf{Y} is Gaussian under each hypothesis, namely

$$H_m : \mathbf{Y} \sim N(\mathbf{a}_m, \mathbf{R}_m). \quad m = 0, 1, \dots, M-1. \quad (2.21)$$

where $\mathbf{a}_0, \dots, \mathbf{a}_{M-1}$ are elements in \mathbb{R}^k , and the $k \times k$ symmetric matrices $\mathbf{R}_0, \dots, \mathbf{R}_{M-1}$ are *positive definite* (thus *invertible*). Condition **(A.3)** holds with respect to Lebesgue measure.

Throughout the M pairs $(\mathbf{a}_0, \mathbf{R}_0), \dots, (\mathbf{a}_{M-1}, \mathbf{R}_{M-1})$ are assumed to be distinct so that the probability density function $f_0, \dots, f_{M-1} : \mathbb{R}^k \rightarrow \mathbb{R}_+$ are distinct. Indeed, for each $m = 0, \dots, M-1$, the probability density function $f_m : \mathbb{R}^k \rightarrow \mathbb{R}_+$ is given by

$$f_m(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^k \det \mathbf{R}_m}} e^{-\frac{1}{2} Q_m(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^k$$

with

$$Q_m(\mathbf{y}) = (\mathbf{y} - \mathbf{a}_m)' \mathbf{R}_m^{-1} (\mathbf{y} - \mathbf{a}_m), \quad \mathbf{y} \in \mathbb{R}^k.$$

Thus, for each $\ell = 0, \dots, M - 1$, we have

$$\begin{aligned} h_\ell(\mathbf{y}) &= \sum_{m=0, m \neq \ell}^{M-1} p_m (C(m, \ell) - C(m, m)) \frac{1}{\sqrt{(2\pi)^k \det \mathbf{R}_m}} e^{-\frac{1}{2} Q_m(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^k \end{aligned} \quad (2.22)$$

For the probability of error criterion, for each $\ell = 0, \dots, M - 1$, this last expression becomes

$$\begin{aligned} h_\ell(\mathbf{y}) &= \sum_{m=0, m \neq \ell}^{M-1} \frac{p_m}{\sqrt{(2\pi)^k \det \mathbf{R}_m}} e^{-\frac{1}{2} Q_m(\mathbf{y})} \\ &= g(\mathbf{y}) - \frac{p_\ell}{\sqrt{(2\pi)^k \det \mathbf{R}_\ell}} e^{-\frac{1}{2} Q_\ell(\mathbf{y})}, \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \quad (2.23)$$

The equal covariance case When

$$\mathbf{R}_0 = \dots = \mathbf{R}_{M-1} \equiv \mathbf{R},$$

additional simplifications occur: For each $\ell = 0, \dots, M - 1$, we find

$$\begin{aligned} h_\ell(\mathbf{y}) &= \frac{1}{\sqrt{(2\pi)^k \det \mathbf{R}}} \sum_{m=0, m \neq \ell}^{M-1} \frac{p_m (C(m, \ell) - C(m, m))}{\sqrt{(2\pi)^k \det \mathbf{R}}} e^{-\frac{1}{2} Q(\mathbf{y} - \mathbf{a}_m)}, \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned}$$

with

$$Q(\mathbf{y}) = \mathbf{y}' \mathbf{R}^{-1} \mathbf{y}, \quad \mathbf{y} \in \mathbb{R}^k.$$

For the probability of error criterion, for each $\ell = 0, \dots, M - 1$, we find

$$\begin{aligned} h_\ell(\mathbf{y}) &= \sum_{m=0, m \neq \ell}^{M-1} \frac{p_m}{\sqrt{(2\pi)^k \det \mathbf{R}}} e^{-\frac{1}{2} Q(\mathbf{y} - \mathbf{a}_m)} \\ &= g(\mathbf{y}) - \frac{p_\ell}{\sqrt{(2\pi)^k \det \mathbf{R}}} e^{-\frac{1}{2} Q(\mathbf{y} - \mathbf{a}_\ell)}, \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \quad (2.24)$$

Writing

$$d(p; \mathbf{y}) = \log p - \frac{1}{2}Q(\mathbf{y}), \quad \begin{array}{l} p \in (0, 1) \\ \mathbf{y} \in \mathbb{R}^k, \end{array}$$

the Bayesian test $d^* : \mathbb{R}^k \rightarrow \{0, \dots, M-1\}$ now reduces to

$$d^*(\mathbf{y}) = m \quad \text{iff} \quad \begin{array}{l} d(p_m; \mathbf{y} - \mathbf{a}_m) \\ = \max (d(p_\ell; \mathbf{y} - \mathbf{a}_\ell), \ell = 0, 1, \dots, M-1) \end{array}$$

with a lexicographic tiebreaker in the event of ties.

Chapter 3

Composite hypothesis testing problems

3.1 A motivating example

Consider the following problem of deciding between two hypotheses. Under the null hypothesis H_0 , the observation \mathbf{Y} is an \mathbb{R}^k -valued rv which is normally distributed with mean vector \mathbf{m} and covariance matrix \mathbf{R} which are both *known*. Under the alternative hypothesis H_1 , the \mathbb{R}^k -valued rv \mathbf{Y} is normally distributed with mean vector $\boldsymbol{\theta}$ and covariance matrix \mathbf{R} where $\boldsymbol{\theta} \neq \mathbf{m}$ is only known to lie in a subset Θ_1 of \mathbb{R}^k , and is otherwise unspecified. We assume that \mathbf{m} is *not* an element of Θ_1 .

This problem of testing for the binary hypothesis H_0 vs. H_1 can also be interpreted as one of deciding between the hypotheses

$$\begin{aligned} H_1 &: \{H_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_1\} \\ H_0 &: \mathbf{Y} \sim N(\mathbf{m}, \mathbf{R}). \end{aligned} \tag{3.1}$$

where

$$H_{\boldsymbol{\theta}} : \mathbf{Y} \sim N(\boldsymbol{\theta}, \mathbf{R}), \quad \begin{array}{l} \boldsymbol{\theta} \neq \mathbf{m}, \\ \boldsymbol{\theta} \in \Theta_1 \end{array}$$

Thus, the alternative hypothesis can be viewed as a *composite* hypothesis $\{H_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_1\}$. We emphasize that we seek to decide between H_0 and H_1 , or equivalently, between H_0 and $\{H_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_1\}$; the precise value of $\boldsymbol{\theta}$ is not thought

3.2 The probabilistic model

More generally consider two non-empty Borel subsets Θ_0 and Θ_1 of \mathbb{R}^p for some positive integer p . Assume that

$$\Theta_0 \cap \Theta_1 = \emptyset.$$

We set

$$\Theta = \Theta_0 \cup \Theta_1.$$

so that the pair Θ_0 and Θ_1 forms a Borel partition of Θ .

Given is a family of probability distributions $\{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ on \mathbb{R}^k . For mathematical reasons, it is required that the mapping $\mathbb{R}^k \times \Theta \rightarrow [0, 1] : (\mathbf{y}, \boldsymbol{\theta}) \rightarrow F_{\boldsymbol{\theta}}(\mathbf{y})$ is Borel measurable.

We are given a measurable space (Ω, \mathcal{F}) . The state of nature is modeled by means of a rv $\boldsymbol{\vartheta} : \Omega \rightarrow \Theta$ defined on (Ω, \mathcal{F}) . The observation is given by an \mathbb{R}^k -valued rv $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^k$ defined on the same measurable space (Ω, \mathcal{F}) . with the interpretation that

$$\mathbb{P}[\mathbf{Y} \leq \mathbf{y} | \boldsymbol{\vartheta} = \boldsymbol{\theta}] = F_{\boldsymbol{\theta}}(\mathbf{y}), \quad \begin{array}{l} \mathbf{y} \in \mathbb{R}^k, \\ \boldsymbol{\theta} \in \Theta. \end{array} \quad (3.2)$$

The state of nature and the corresponding observation are denoted

$$H_{\boldsymbol{\theta}} : \mathbf{Y} \sim F_{\boldsymbol{\theta}} \quad (3.3)$$

with $\boldsymbol{\theta}$ ranging in Θ .

The *composite* binary hypothesis testing problem is then the problem of deciding between the two composite hypotheses $H_0 = \{H_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_0\}$ and $H_1 = \{H_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_1\}$ on the basis of the observation \mathbf{Y} .

If either Θ_0 or Θ_1 is reduced to a single element, the corresponding hypothesis is termed *simple*. Obviously the problem of simple binary hypothesis testing discussed in Chapter 1 obtains when each of the sets Θ_0 and Θ_1 contains exactly one element.

The composite binary hypothesis testing problem is concisely denoted by

$$\begin{array}{l} H_1 : \mathbf{Y} \sim F_{\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \in \Theta_1 \\ H_0 : \mathbf{Y} \sim F_{\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \in \Theta_0 \end{array} \quad (3.4)$$

In a manner reminiscent of the parameter estimation problem of Chapter ?? there are two possible cases, depending on whether or not $\boldsymbol{\theta}$ is modeled as a rv; these are the Bayesian and non-Bayesian cases, respectively.

3.3 The Bayesian case

Here we assume that there exists a Θ -valued rv ϑ defined on the measurable space (Ω, \mathcal{F}) , and let $K : \mathbb{R}^p \rightarrow [0, 1]$ denote its probability distribution.

Write

$$H = \mathbf{1}[\vartheta \in \Theta_1].$$

Note that

$$\begin{aligned} p &= \mathbb{P}[H = 1] \\ &= \mathbb{P}[\vartheta \in \Theta_1] \\ &= \int_{\Theta_1} dK(t). \end{aligned} \tag{3.5}$$

In a similar manner we conclude that

$$\begin{aligned} 1 - p &= \mathbb{P}[H = 0] \\ &= \mathbb{P}[\vartheta \in \Theta_0] \\ &= \int_{\Theta_0} dK(t). \end{aligned} \tag{3.6}$$

For each $h = 0, 1$, we find that

$$\begin{aligned} \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = h] &= \frac{\mathbb{P}[\mathbf{Y} \leq \mathbf{y}, H = h]}{\mathbb{P}[H = h]} \\ &= \frac{\mathbb{P}[\mathbf{Y} \leq \mathbf{y}, \vartheta \in \Theta_h]}{\mathbb{P}[\vartheta \in \Theta_h]} \\ &= \frac{\mathbb{E}[\mathbb{P}[\mathbf{Y} \leq \mathbf{y} | \vartheta] \mathbf{1}[\vartheta \in \Theta_h]]}{\mathbb{P}[\vartheta \in \Theta_h]} \\ &= \frac{\mathbb{E}[F_{\vartheta}(\mathbf{y}) \mathbf{1}[\vartheta \in \Theta_h]]}{\mathbb{P}[\vartheta \in \Theta_h]} \\ &= \frac{\int_{\Theta_h} F_t(\mathbf{y}) dK(t)}{\int_{\Theta_h} dK(t)}, \quad \mathbf{y} \in \mathbb{R}^k. \end{aligned} \tag{3.7}$$

Condition **(A.4)** parallels Condition **(A.2)** given in the binary case and Condition **(A.3)** given in the M -ary case. It states that for each θ in Θ , the probability distribution on \mathbb{R}^k is absolutely continuous with respect to some distribution F on

\mathbb{R}^k . This is equivalent to saying that there exists a Borel mapping $f_\theta : \mathbb{R}^k \rightarrow \mathbb{R}_+$ such that

$$F_\theta(\mathbf{y}) = \int_{-\infty}^{\mathbf{y}} f_\theta(\boldsymbol{\eta}) dF(\boldsymbol{\eta}), \quad \mathbf{y} \in \mathbb{R}^k. \quad (3.8)$$

For mathematical reasons we will require that the mapping

$$\mathbb{R}^k \times \Theta \rightarrow \mathbb{R}_+ : (\mathbf{y}, \theta) \rightarrow f_\theta(\mathbf{y})$$

is Borel measurable

Fix $h = 0, 1$ and \mathbf{y} in \mathbb{R}^k . Under Assumption **(A.4)** we note that

$$\begin{aligned} F_h(\mathbf{y}) &\equiv \mathbb{P}[\mathbf{Y} \leq \mathbf{y} | H = h] \\ &= \frac{\int_{\Theta_h} F_t(\mathbf{y}) dK(t)}{\int_{\Theta_h} dK(t)} \end{aligned} \quad (3.9)$$

with

$$\begin{aligned} \int_{\Theta_h} F_t(\mathbf{y}) dK(t) &= \int_{\Theta_h} \left(\int_{-\infty}^{\mathbf{y}} f_t(\boldsymbol{\eta}) dF(\boldsymbol{\eta}) \right) dK(t) \\ &= \int_{-\infty}^{\mathbf{y}} \left(\int_{\Theta_h} f_t(\boldsymbol{\eta}) dK(t) \right) dF(\boldsymbol{\eta}) \end{aligned} \quad (3.10)$$

by Tonelli's Theorem. Therefore,

$$F_h(\mathbf{y}) = \int_{-\infty}^{\mathbf{y}} \left(\frac{\int_{\Theta_h} f_t(\boldsymbol{\eta}) dK(t)}{\int_{\Theta_h} dK(t)} \right) dF(\boldsymbol{\eta}). \quad (3.11)$$

This shows that Condition **(A.1)** holds for F_0 and F_1 with respect to F with probability density functions $f_0, f_1 : \mathbb{R}^k \rightarrow \mathbb{R}_+$ given by

$$f_h(\mathbf{y}) = \frac{\int_{\Theta_h} f_t(\mathbf{y}) dK(t)}{\int_{\Theta_h} dK(t)}, \quad h = 0, 1, \quad \mathbf{y} \in \mathbb{R}^k.$$

In this setting the tests $\{d_\eta, \eta \geq 0\}$ now take the form

$$\begin{aligned} d_\eta(\mathbf{y}) = 0 &\quad \text{iff} \quad f_1(\mathbf{y}) < \eta f_0(\mathbf{y}) \\ &\quad \text{iff} \quad \int_{\Theta_1} f_t(\mathbf{y}) dK(t) < \eta \int_{\Theta_0} f_t(\mathbf{y}) dK(t). \end{aligned} \quad (3.12)$$

The cost incurred for making decisions is quantified by the mapping $C : \Theta \times \{0, 1\} \rightarrow \mathbb{R}$ with the interpretation that

$$C(\boldsymbol{\theta}, d) = \begin{pmatrix} \text{Cost incurred for deciding } d \\ \text{when } \boldsymbol{\vartheta} = \boldsymbol{\theta} \end{pmatrix}, \quad \begin{array}{l} \boldsymbol{\theta} \in \Theta \\ d = 0, 1 \end{array}$$

With any admissible test $d : \mathbb{R}^k \rightarrow \{0, 1\}$, we define the expected cost

$$J(d) = \mathbb{E} [C(\boldsymbol{\vartheta}, d(\mathbf{Y}))]$$

Note that

$$\begin{aligned} J(d) &= \\ &= \end{aligned} \tag{3.13}$$

As before, the *Bayesian Problem* \mathcal{P}_B is the minimization problem

$$\mathcal{P}_B : \quad \text{Minimize } J(d) \text{ over } d \text{ in } \mathcal{D}.$$

This amounts to finding an admissible test d^* in \mathcal{D} such that

$$J(d^*) \leq J(d), \quad d \in \mathcal{D}. \tag{3.14}$$

Any admissible test d^* which satisfies (3.14) is called a *Bayesian test*, and the value

$$J(d^*) = \inf_{d \in \mathcal{D}} J(d) = \min_{d \in \mathcal{D}} J(d) \tag{3.15}$$

is known as the *Bayesian cost*.

3.4 The non-Bayesian case

In this formulation θ is an unknown parameter, and a Neyman-Pearson approach seems warranted. However, since now composite hypotheses are involved, earlier definitions given in Section 1.17 need to be modified.

Consider a test $d : \mathbb{R}^k \rightarrow \{0, 1\}$ in \mathcal{D} that tests H_0 against the alternative H_1 . We define its *size* to be the quantity

$$\alpha(d) \equiv \sup_{\boldsymbol{\theta} \in \Theta_0} (\mathbb{P}_{\boldsymbol{\theta}} [d(\mathbf{Y}) = 0]). \tag{3.16}$$

Note that with θ in Θ_0 , the probability $\mathbb{P}_\theta [d(\mathbf{Y}) = 0]$ can be interpreted as the probability of false alarm under the test d given that the hypothesis H_θ is indeed correct. With α in $[0, 1]$, let \mathcal{D}_α denote the collection of all tests in \mathcal{D} whose size is no greater than α , namely,

$$\mathcal{D}_\alpha \equiv \{d \in \mathcal{D} : \alpha(d) \leq \alpha\}. \quad (3.17)$$

The Neyman-Pearson formulation now takes the following form: For α in $[0, 1]$, find a test $d_{\text{NP}}(\alpha)$ in \mathcal{D}_α such that

$$\mathbb{P}_\theta [d(\mathbf{Y}) = 1] \leq \mathbb{P}_\theta [d_{\text{NP}}(\alpha)(\mathbf{Y}) = 1], \quad \begin{array}{l} \theta \in \Theta_1, \\ d \in \mathcal{D}_\alpha. \end{array} \quad (3.18)$$

Such a test $d_{\text{NP}}(\alpha)$, when it exists, is called a *Uniformly Most Powerful (UMP)* test of size α .

3.5 The non-Bayesian case – UMP tests

As will become shortly apparent, a UMP test may not always exist. To get an idea of the possible subtleties, consider the situation where the hypothesis H_0 and H_1 are simple and composite, respectively, say $\Theta_0 = \{\theta_0\}$ and $\theta_0 \notin \Theta_1$. In this setting the definition of \mathcal{D}_α now reduces to

$$\mathcal{D}_\alpha \equiv \{d \in \mathcal{D} : \mathbb{P}_{\theta_0} [d(\mathbf{Y}) = 0] \leq \alpha\}.$$

Fix θ_1 in Θ_1 . Recall that the Neyman-Pearson approach to the simple hypothesis testing problem

$$\begin{array}{l} H_1 : \mathbf{Y} \sim F_{\theta_1} \\ H_0 : \mathbf{Y} \sim F_{\theta_0}. \end{array} \quad (3.19)$$

leads to the test $d_{\text{NP}}(\alpha; \theta_1)$ in \mathcal{D}_α characterized by

$$\mathbb{P}_{\theta_0} [d_{\text{NP}}(\alpha; \theta_1)(\mathbf{Y}) = 1] = \alpha$$

and

$$\mathbb{P}_{\theta_1} [d(\mathbf{Y}) = 1] \leq \mathbb{P}_{\theta_1} [d_{\text{NP}}(\alpha; \theta_1)(\mathbf{Y}) = 1], \quad d \in \mathcal{D}_\alpha.$$

However, there is *a priori* no reason to expect that for any $\tilde{\theta}_1 \neq \theta_1$ in Θ_1 , the Neyman-Pearson test $d_{\text{NP}}(\alpha; \tilde{\theta}_1)$ also solves the problem

$$\text{Maximize } \mathbb{P}_{\tilde{\theta}_1} [d(\mathbf{Y}) = 1] \text{ over } d \text{ in } \mathcal{D}_\alpha$$

In fact this will happen if the tests $d_{\text{NP}}(\alpha; \theta_1)$ and $d_{\text{NP}}(\alpha; \tilde{\theta}_1)$ satisfy

$$\mathbb{P}_{\theta_1} [d_{\text{NP}}(\alpha; \theta_1)(\mathbf{Y}) = 1] = \mathbb{P}_{\theta_1} [d_{\text{NP}}(\alpha; \tilde{\theta}_1)(\mathbf{Y}) = 1]$$

and

$$\mathbb{P}_{\tilde{\theta}_1} [d_{\text{NP}}(\alpha; \theta_1)(\mathbf{Y}) = 1] = \mathbb{P}_{\tilde{\theta}_1} [d_{\text{NP}}(\alpha; \tilde{\theta}_1)(\mathbf{Y}) = 1]$$

More generally, a test $d^* : \mathbb{R}^k \rightarrow \{0, 1\}$ in \mathcal{D}_α is UMP of size α if

$$\alpha(d^*) = \alpha$$

and

$$\mathbb{P}_\theta [d^*(\mathbf{Y}) = 1] = \mathbb{P}_\theta [d_{\text{NP}}(\alpha; \theta_1)(\mathbf{Y}) = 1], \quad \theta \in \Theta_1.$$

This is clearly a non-trivial restriction on the problem. It is therefore not surprising that UMP tests do not always exist. This is discussed on an example later on in Section 3.6.

Theorem 3.5.1 *Let H_0 be a simple hypothesis and let H_1 be a composite hypothesis. A UMP test of size α exists if and only if the Neyman-Pearson tests $\{d_{\text{NP}}(\alpha; \text{myvetheta}), \theta \in \Theta_1\}$ can be completely specified without knowledge of the parameter θ , i.e., the Borels subsets*

$$C(d_{\text{NP}}(\alpha; \theta)) = \{\mathbf{y} \in \mathbb{R}^k : d_{\text{NP}}(\alpha; \theta)(\mathbf{y}) = 0\}, \quad \theta \in \Theta_1$$

do not depend on θ as it ranges over Θ_1 .

Proof. If a UMP test of size α exists, say d_α^* , then

$$\mathbb{P}_{\theta_0} [d_\alpha^*(\mathbf{Y}) = 1] = \alpha$$

and

$$\mathbb{P}_\theta [d(\mathbf{Y}) = 1] \leq \mathbb{P}_\theta [d_\alpha^*(\mathbf{Y}) = 1], \quad \begin{array}{l} \theta \in \Theta_1, \\ d \in \mathcal{D}_\alpha. \end{array}$$

Conversely, if the set $C(d_{\text{NP}}(\alpha; \theta))$ is independent of θ in Θ_1 , set

$$C_\alpha^* \equiv \{\mathbf{y} \in \mathbb{R}^k : \}$$

■

3.6 The non-Bayesian case – An example

We illustrate the developments of Section with an example. With θ arbitrary in \mathbb{R} ,

$$H_\theta : Y \sim N(\theta, 1)$$

so that F_θ admits the density $f_\theta : \mathbb{R} \rightarrow \mathbb{R}_+$ given by

$$f_\theta(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}, \quad y \in \mathbb{R}.$$

With distinct θ_0 and θ_1 in \mathbb{R} , consider the Neyman-Pearson formulation for the binary hypothesis problem

$$\begin{aligned} H_1 : Y &\sim N(\theta_1, 1) \\ H_0 : Y &\sim N(\theta_0, 1) \end{aligned} \quad (3.20)$$

As shown in Section 1.19 this problem has a complete solution: Fix $\lambda > 0$. The test $d_\lambda : \mathbb{R} \rightarrow \{0, 1\}$ takes the form

$$\begin{aligned} d_\lambda(y) = 0 & \quad \text{iff} \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta_1)^2} < \lambda \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta_0)^2} \\ & \quad \text{iff} \quad (y - \theta_0)^2 < 2 \log \lambda + (y - \theta_1)^2 \\ & \quad \text{iff} \quad 2y(\theta_1 - \theta_0) < 2 \log \lambda + \theta_1^2 - \theta_0^2 \end{aligned} \quad (3.21)$$

Two cases arise: If $\theta_0 < \theta_1$, then

$$d_\lambda(y) = 0 \quad \text{iff} \quad y < \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 + \theta_0}{2} \quad (3.22)$$

Then, by standard arguments we get

$$\begin{aligned} \mathbb{P}_{\theta_0} [d_\lambda(\mathbf{Y}) = 1] &= \mathbb{P}_{\theta_0} \left[Y \geq \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 + \theta_0}{2} \right] \\ &= \mathbb{P}_{\theta_0} \left[Y - \theta_0 \geq \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} \right] \\ &= 1 - \Phi \left(\frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} \right). \end{aligned} \quad (3.23)$$

If $\theta_1 < \theta_0$, then

$$d_\lambda(y) = 0 \quad \text{iff} \quad y > \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 + \theta_0}{2}. \quad (3.24)$$

This time we find

$$\begin{aligned}
\mathbb{P}_{\theta_0} [d_\lambda(\mathbf{Y}) = 1] &= \mathbb{P}_{\theta_0} \left[Y \leq \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 + \theta_0}{2} \right] \\
&= \mathbb{P}_{\theta_0} \left[Y - \theta_0 \leq \frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} \right] \\
&= \Phi \left(\frac{\log \lambda}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} \right). \tag{3.25}
\end{aligned}$$

Fix α in $(0, 1)$. The Neyman-Pearson test of size α for testing H_{θ_1} vs. H_{θ_0} is the test $d_{\lambda(\theta_1, \theta_0; \alpha)}$ where $\lambda(\theta_1, \theta_0; \alpha)$ is that value of $\lambda > 0$ determined by $\mathbb{P}_{\theta_0} [d_\lambda(\mathbf{Y}) = 1] = \alpha$.

If $\theta_0 < \theta_1$, then

$$\frac{\log \lambda(\theta_1, \theta_0; \alpha)}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} = \Phi^{-1}(1 - \alpha)$$

and the test $d_{\lambda(\theta_1, \theta_0; \alpha)}$ is given by

$$d_{\lambda(\theta_1, \theta_0; \alpha)}(y) = 0 \quad \text{iff} \quad y < \theta_0 + \Phi^{-1}(1 - \alpha).$$

If $\theta_1 < \theta_0$, then

$$\frac{\log \lambda(\theta_1, \theta_0; \alpha)}{\theta_1 - \theta_0} + \frac{\theta_1 - \theta_0}{2} = \Phi^{-1}(\alpha)$$

and the test $d_{\lambda(\theta_1, \theta_0; \alpha)}$ is given by

$$d_{\lambda(\theta_1, \theta_0; \alpha)}(y) = 0 \quad \text{iff} \quad y > \theta_0 + \Phi^{-1}(\alpha).$$

We now consider four different situations, each associated with different sets Θ_0 and Θ_1 .

Case I: $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = (\theta_0, \infty)$

Case II: $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = (-\infty, \theta_0)$

Case III: $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \mathbb{R} - \{\theta_0\}$

Case IV: $\Theta_0 = (-\infty, \theta_0]$ and $\Theta_1 = (\theta_0, \infty)$ Fix σ arbitrary in Θ_0 , so $\sigma \leq \theta_0$. Consider now the composite hypothesis testing problem The composite binary hypothesis testing problem is concisely denoted by

$$\begin{aligned} H_1 : & Y \sim N(\theta, 1), \quad \theta \in \Theta_1 \\ H_0 : & Y \sim N(\sigma, 1) \end{aligned} \quad (3.26)$$

For this problem we have seen that a UMP test of size α exists; it was discussed in Case I, and is identified as the test $d^*(\alpha; \sigma) : \mathbb{R}^k \rightarrow \{0, 1\}$ given by

$$d^*(\alpha; \sigma)(y) = 0 \quad \text{iff} \quad y < \sigma + \Phi^{-1}(1 - \alpha)$$

Thus, for every θ in Θ_1 , it holds that

$$\begin{aligned} \mathbb{P}_\theta [d^*(\alpha; \sigma)(Y) = 1] &= \mathbb{P}_\theta [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] \\ &\geq \mathbb{P}_\theta [d(Y) = 1] \end{aligned} \quad (3.27)$$

for every test d such that

$$\mathbb{P}_\sigma [d(Y) = 1] \leq \alpha.$$

Recall that

$$\begin{aligned} \mathcal{D}_\alpha &= \left\{ d \in \mathcal{D} : \sup_{\sigma \in \Theta_0} (\mathbb{P}_\sigma [d(Y) = 1]) \leq \alpha \right\} \\ &= \left\{ d \in \mathcal{D} : \sup_{\sigma \leq \theta_0} (\mathbb{P}_\sigma [d(Y) = 1]) \leq \alpha \right\}. \end{aligned} \quad (3.28)$$

For any d in \mathcal{D}_α , the following statement holds: For every θ in Θ_1 , we have

$$\mathbb{P}_\theta [d(Y) = 1] \leq \mathbb{P}_\theta [Y \geq \sigma + \Phi^{-1}(1 - \alpha)], \quad \sigma \leq \theta_0$$

Thus, for every $\sigma \leq \theta_0$, the rule

$$d^*(\alpha; \sigma)(y) = 0 \quad \text{iff} \quad y < \sigma + \Phi^{-1}(1 - \alpha)$$

is a candidate for the UMP test. Next we show that only when $\sigma = \theta_0$, does this rule belong to \mathcal{D}_α , in which case it is the UMP test.

With $\sigma < \theta_0$, we consider the quantity

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)], \quad \sigma' \leq \theta_0$$

Lemma 3.6.1 *We have*

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] \begin{cases} < \alpha & \text{if } \sigma' < \sigma < \theta_0 \\ = \alpha & \text{if } \sigma' = \sigma \\ > \alpha & \text{if } \sigma < \sigma' \leq \theta_0. \end{cases} \quad (3.29)$$

Proof. We consider the three cases separately. Note that for arbitrary σ and σ' in \mathbb{R} we always have

$$\begin{aligned} & \mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] \\ &= \mathbb{P}_{\sigma'} [Y \geq \sigma' + (\sigma - \sigma') + \Phi^{-1}(1 - \alpha)] \\ &= \mathbb{P}_{\sigma'} [Y - \sigma' \geq (\sigma - \sigma') + \Phi^{-1}(1 - \alpha)] \\ &= 1 - \Phi((\sigma - \sigma') + \Phi^{-1}(1 - \alpha)). \end{aligned} \quad (3.30)$$

Also recall that

$$\Phi(\Phi^{-1}(1 - \alpha)) = 1 - \alpha.$$

If $\sigma' < \sigma < \theta_0$, then

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] < 1 - \Phi(\Phi^{-1}(1 - \alpha))$$

by monotonicity since $\Phi(\Phi^{-1}(1 - \alpha)) < \Phi((\sigma - \sigma') + \Phi^{-1}(1 - \alpha))$. Thus,

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] < \alpha.$$

Next, if $\sigma' = \sigma$, then

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] = 1 - \Phi(\Phi^{-1}(1 - \alpha)) = \alpha.$$

Finally, if $\sigma < \sigma' \leq \theta_0$, then

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] > 1 - \Phi(\Phi^{-1}(1 - \alpha))$$

by monotonicity since $\Phi(\Phi^{-1}(1 - \alpha)) > \Phi((\sigma - \sigma') + \Phi^{-1}(1 - \alpha))$. This time we get

$$\mathbb{P}_{\sigma'} [Y \geq \sigma + \Phi^{-1}(1 - \alpha)] > \alpha.$$

This establishes the desired result. ■

In generalizing theorem to both hypotheses being composite

There exists θ_* in Θ_0 with d_{NP} for θ_* vs $\theta \in \Theta_1$ which is independent of θ in Θ_1 .

Part II
Estimation Theory

Part III
APENDICES

Chapter 4

Analysis

4.1 Convex functions

Let I denote an interval in \mathbb{R} . A function $g : I \rightarrow \mathbb{R}$ is said to be *convex* if for every x_0 and x_1 in I , it holds that

$$g((1 - \lambda)x_0 + \lambda x_1) \leq (1 - \lambda)g(x_0) + \lambda g(x_1), \quad \lambda \in [0, 1]. \quad (4.1)$$

A function $g : I \rightarrow \mathbb{R}$ is said to be *concave* if the function $-g$ is convex. Here are some well-known facts concerning convex functions.

With $x < y < z$ in I , write

$$y = (1 - \lambda)x + \lambda z$$

with

$$\lambda = \frac{y - x}{z - x}.$$

It is plain that λ is an element of $(0, 1)$. Therefore,

$$\begin{aligned} g(y) &\leq (1 - \lambda)g(x) + \lambda g(z) \\ &= \frac{z - y}{z - x}g(x) + \frac{y - x}{z - x}g(z) \end{aligned} \quad (4.2)$$

so that

$$\begin{aligned} g(y) - g(x) &\leq \frac{z - y}{z - x}g(x) + \frac{y - x}{z - x}g(z) - g(x) \\ &= \left(\frac{y - x}{z - x} \right) (g(z) - g(x)). \end{aligned} \quad (4.3)$$

It follows that

$$\frac{g(y) - g(x)}{y - x} \leq \frac{g(z) - g(x)}{z - x}.$$

Fact 4.1.1 *If the mapping $g : I \rightarrow \mathbb{R}$ is convex on an interval I , then it is also continuous on the interior of I .*

Proof. ■

Fact 4.1.2 *If the mapping $g : I \rightarrow \mathbb{R}$ is convex function on some interval I , then its left and right-derivatives always exist at every point in I (whenever appropriate).*

Proof. The existence of the right-derivative at x in I is a simple consequence of the monotonicity

$$\frac{g(y) - g(x)}{y - x} \leq \frac{g(z) - g(x)}{z - x}, \quad \begin{array}{l} x < y < z \\ x, y, z \in I \end{array}$$
■