AN INTRODUCTION TO ESTIMATION AND DETECTION THEORY[1]

Armand M. Makowski [2] and Prakash L.K. Narayan [3]

[2]Department 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

[3]Department 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 $\{0, 1, \ldots\}$ of all non-negative integers is denoted $\mathbb{N}$. The notation $\mathbb{N}_0$ will be used for the set $\{1, 2, \ldots\}$ of all positive integers.

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

Part I

DETECTION THEORY

Chapter 1

Simple binary hypothesis testing

A decision has to be made as to which of two 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 machine produces circuit boards. It is either fully functioning ($H = 1$) or worn out ($H = 0$). Checking the state of the machine is not feasible as it would require that the production be stopped, incurring a loss of revenue for the manufacturer if the machine were indeed shown to be fully functionally. Instead, a batch of circuits is collected and tested for a number of performance parameters, say $Y_1, \ldots, Y_k$. It is known that

A simple communication example

Testing means

1.2 The probabilistic model

These examples can be cast as *binary* hypothesis testing problems: Nature is in either of two states, say $H = 0$ or $H = 1$ for sake of concreteness, and the observations are organized into an $\mathbb{R}^k$-valued rv $\boldsymbol{Y}$. We assume given two probability distribution functions $F_0, F_1 : \mathbb{R}^k \to [0, 1]$ on $\mathbb{R}^k$; they will act as conditional

probability distribution of $\boldsymbol{Y}$ given $H = 0$ and $H = 1$, respectively. This situation is summarized by

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

In the statistical literature 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: Given some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ (whose existence is discussed shortly), consider rvs $H : \Omega \to \{0, 1\}$ and $\boldsymbol{Y} : \Omega \to \mathbb{R}^k$. The probability distribution functions F_0 and F_1 being interpreted as conditional probability distribution of $\boldsymbol{Y}$ given $H = 0$ and $H = 1$, respectively, we must have

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

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

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

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

Because

$$
\begin{aligned}
\mathbb{P}\left[\boldsymbol{Y} \leq \boldsymbol{y}, H = h\right] \;=\;& \mathbb{P}\left[\boldsymbol{Y} \leq \boldsymbol{y} | H = h\right] \mathbb{P}\left[H = h\right] \\
\;=\;& \begin{cases} (1 - p)F_0(\boldsymbol{y}) & \text{if } h = 0, \boldsymbol{y} \in \mathbb{R}^k \\[2mm] pF_1(\boldsymbol{y}) & \text{if } h = 1, \boldsymbol{y} \in \mathbb{R}^k, \end{cases}
\end{aligned}
\tag{1.2}
$$

the law of total probability shows that

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

In other words, 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 $\boldsymbol{Y}$.

1.3 A construction

The existence of the model described in Section 1.2 can be guaranteed through the following construction: Take $\Omega = \{0, 1\} \times \mathbb{R}^k$ with generic element $\omega = (h, \boldsymbol{y})$ with $h = 0, 1$ and $\boldsymbol{y}$ an arbitrary element of $\mathbb{R}^k$. We endow Ω with the σ-field $\mathcal{F}$ given by

$$\mathcal{F} = \sigma\left(\mathcal{P}(\{0, 1\}) \times \mathcal{B}(\mathbb{R}^k)\right)$$

where $\mathcal{P}(\{0, 1\})$ is the power set of $\{0, 1\}$, and $\mathcal{B}(\mathbb{R}^k)$ is the Borel σ-field on $\mathbb{R}^k$.

We define the mappings $H : \Omega \to \mathbb{R}$ and $\boldsymbol{Y} : \Omega \to \mathbb{R}^k$ by

$$H(\omega) = h \quad \text{and} \quad \boldsymbol{Y}(\omega) = \boldsymbol{y}, \quad \omega = (h, \boldsymbol{y}) \in \Omega.$$

Both projection mappings are Borel measurable, and therefore define rvs.

If $\mathbb{P}$ is any probability measure on the σ-field $\mathcal{F}$, then by *construction* of the rvs H and $\boldsymbol{Y}$ just given, the joint probability distribution of the pair $(H, \boldsymbol{Y})$ is necessarily given by

$$
\begin{aligned}
\mathbb{P}\left[H = h, \boldsymbol{Y} \leq \boldsymbol{y}\right] &= \mathbb{P}\left[\{\omega \in \Omega : H(\omega) = h, \boldsymbol{Y}(\omega) \leq \boldsymbol{y}\}\right] \\
&= \mathbb{P}\left[\{h\} \times (-\infty, \boldsymbol{y}]\right], \quad \begin{aligned} h &= 0, 1 \\ \boldsymbol{y} &\in \mathbb{R}^k \end{aligned}
\end{aligned}
\tag{1.4}
$$

since

$$\{\omega \in \Omega : H(\omega) = h, \boldsymbol{Y}(\omega) \leq \boldsymbol{y}\} = \{h\} \times (-\infty, \boldsymbol{y}].$$

On the way to identify *a* probability $\mathbb{P}$ on $\mathcal{F}$ under which the joint probability distribution of the pair $(H, \boldsymbol{Y})$ satisfies (1.2), we readily conclude from (1.4) that $\mathbb{P}$ is necessarily determined on certain rectangles, namely

$$
\mathbb{P}\left[\{h\} \times (-\infty, \boldsymbol{y}]\right] = \begin{cases} (1 - p)F_0(\boldsymbol{y}) & \text{if } h = 0 \\[2mm] pF_1(\boldsymbol{y}) & \text{if } h = 1 \end{cases}
\tag{1.5}
$$

for every $\boldsymbol{y}$ in $\mathbb{R}^k$. At this point we recall the following fact from Measure Theory: Any probability measure on the σ-field $\mathcal{F}$ carried by the product space $\{0, 1\} \times \mathbb{R}^k$ is uniquely determined on the entire σ-field $\mathcal{F}$ by its values on the rectangle sets of the form

$$\{h\} \times (-\infty, \boldsymbol{y}], \quad \begin{aligned} h &= 0, 1 \\ \boldsymbol{y} &\in \mathbb{R}^k. \end{aligned}$$

Therefore, by virtue of (1.5) there exists a *unique* probability measure $\mathbb{P}$ on $\mathcal{F}$ such that (1.4) holds. More generally, it is also the case that

$$\mathbb{P}\left[\{h\} \times (-\infty, \boldsymbol{y}]\right] = \begin{cases} (1-p)\int_B dF_0(\boldsymbol{y}) & \text{if } h = 0 \\ \\ p\int_B dF_1(\boldsymbol{y}) & \text{if } h = 1 \end{cases} \tag{1.6}$$

for every Borel set B in $\mathbb{R}^k$ as a result of the fact that

$$\sigma\left((-\infty, \boldsymbol{y}], \ \boldsymbol{y} \in \mathbb{R}^k\right) = \mathcal{B}(\mathbb{R}^k).$$

Finally, under this probability measure $\mathbb{P}$ it is plain (1.5) immediately implies

$$\mathbb{P}\left[H = h\right] = \mathbb{P}\left[\{h\}\right] = \begin{cases} (1-p) & \text{if } h = 0 \\ \\ p & \text{if } h = 1 \end{cases} \tag{1.7}$$

and

$$\begin{aligned} \mathbb{P}\left[\boldsymbol{Y} \leq \boldsymbol{y} | H = h\right] &= \frac{\mathbb{P}\left[H = h, \boldsymbol{Y} \leq \boldsymbol{y}\right]}{\mathbb{P}\left[H = h\right]} \\ &= F_h(\boldsymbol{y}) \end{aligned} \tag{1.8}$$

for every $\boldsymbol{y}$ in $\mathbb{R}^k$, as required.

1.4   Basic assumptions

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)**: The probability distributions F_0 and F_1 on $\mathbb{R}^k$ 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 $f_0, f_1 : \mathbb{R}^k \to \mathbb{R}_+$ such that

$$F_h(\boldsymbol{y}) = \int_{-\infty}^{\boldsymbol{y}} f_h(\boldsymbol{\eta}) dF(\boldsymbol{\eta}), \qquad \begin{matrix} \boldsymbol{y} \in \mathbb{R}^k, \\ h = 0, 1. \end{matrix} \tag{1.9}$$

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.,

$$F(\boldsymbol{y}) \equiv \frac{1}{2}F_0(\boldsymbol{y}) + \frac{1}{2}F_1(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^k. \tag{1.10}$$

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 $f_0, f_1 : \mathbb{R}^k \to \mathbb{R}_+$ 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 $f_0, f_1 : \mathbb{R}^k \to \mathbb{R}_+$ are best thought as *probability mass functions* (pdfs) $\boldsymbol{f}_0 = \{f_0(\boldsymbol{y}), \ \boldsymbol{y} \in S\}$ and $\boldsymbol{f}_1 = \{f_1(\boldsymbol{y}), \ \boldsymbol{y} \in S\}$, i.e.,

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

and

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

The condition (1.9) now takes the form

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

Condition **(A.2)**: 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

$$f_0(\boldsymbol{y}) = 0 \quad \text{implies} \quad f_1(\boldsymbol{y}) = 0. \tag{1.11}$$

1.5 Admissible tests

Decisions as to which state of nature occurred are taken on the basis of observations; this is formalized through the following definition.

An *admissible* decision rule (or test) is any *Borel* mapping $d : \mathbb{R}^k \to \{0, 1\}$. 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 $d(\boldsymbol{Y}) : \Omega \to \{0, 1\} : \omega \to d(Y(\omega))$ is indeed a rv, i. e., $[\omega \in \Omega : d(Y(\omega)) = h]$ is an event in $\mathcal{F}$ for all $h = 0, 1$. The need for this technical condition will become apparent in subsequent chapters.

The next fact will prove useful in some of the discussion

Lemma 1.5.1 *The set $\mathcal{D}$ of admissible decision rules is in one-to-one correspondence with $\mathcal{B}(\mathbb{R}^k)$.*

Proof. By definition of admissibility every test d in $\mathcal{D}$ is completely specified by the *Borel* subset $C(d)$ defined by

$$C(d) \equiv \{\boldsymbol{y} \in \mathbb{R}^k : d(\boldsymbol{y}) = 0\}. \tag{1.12}$$

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

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

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

Any admissible rule d in $\mathcal{D}$ induces *two* types of error: Upon observing $\boldsymbol{Y}$, either $H = 0$ is true and $d(\boldsymbol{Y}) = 1$ *or* $H = 1$ is true and $d(\boldsymbol{Y}) = 0$.

These two possibilities are the so–called errors of the *first* and *second* type associated with the decision rule d; they are quantified by

$$\alpha(d) \equiv \mathbb{P}\left[d(\boldsymbol{Y}) = 1 | H = 0\right] \tag{1.13}$$

and

$$\beta(d) \equiv \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right], \tag{1.14}$$

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

$$P_{\mathrm{F}}(d) \equiv \mathbb{P}\left[d(\boldsymbol{Y}) = 1 | H = 0\right] \qquad (1.15)$$

and

$$P_{\mathrm{M}}(d) \equiv \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right]. \qquad (1.16)$$

Sometimes, it is convenient to consider the so–called probability of *detection* given by

$$P_{\mathrm{D}}(d) \equiv \mathbb{P}\left[d(\boldsymbol{Y}) = 1 | H = 1\right] = 1 - P_{\mathrm{M}}(d). \qquad (1.17)$$

Throughout we shall use this terminology.

1.6 Likelihood ratio tests

In subsequent chapters we consider several formulations of the binary hypothesis problem. In all cases the tests of interest are related to tests in the class of admissible tests $\{d_\eta, \eta \geq 0\}$ which we now introduce.

For each $\eta \geq 0$, test $d_\eta : \mathbb{R}^k \to \{0, 1\}$ is defined by

$$d_\eta(\boldsymbol{y}) = 0 \quad \text{iff} \quad f_1(\boldsymbol{y}) < \eta f_0(\boldsymbol{y}). \qquad (1.18)$$

Its admissibility is a simple consequence of the Borel measurability of the probability density functions $f_0, f_1 : \mathbb{R}^k \to \mathbb{R}_+$.

It is plain from the definition (1.18) (with $\eta = 0$) that d_0 is simply the test that always selects the non-null hypothesis $H = 1$, i.e., $d_0(\boldsymbol{y}) = 1$ for every $\boldsymbol{y}$ in $\mathbb{R}^k$. On the other hand, formally substituting $\eta = \infty$ in (1.18) will be problematic at observation points where $f_0(\boldsymbol{y}) = 0$. However, by *convention* we shall interpret d_∞ as the test that always selects the null hypothesis $H = 0$, i.e., $d_0(\boldsymbol{y}) = 0$ for every $\boldsymbol{y}$ in $\mathbb{R}^k$.

Such tests take an even simpler form under the additional Condition **(A.2)** as will be seen shortly: Note that (1.18) can be rewritten as

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

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

$$
L(\boldsymbol{y}) \equiv
\begin{cases}
\frac{f_1(\boldsymbol{y})}{f_0(\boldsymbol{y})} & \text{if } f_0(\boldsymbol{y}) > 0 \\[2ex]
\Lambda(\boldsymbol{y}) & \text{if } f_0(\boldsymbol{y}) = 0
\end{cases}
\tag{1.19}
$$

for some arbitrary Borel mapping $\Lambda : \mathbb{R}^k \to \mathbb{R}_+$. 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 (1.19), we define the *likelihood ratio* test with *threshold* $\eta \geq 0$ to be the admissible decision rule $Lrt_\eta : \mathbb{R}^k \to \{0, 1\}$ given by

$$
Lrt_\eta(\boldsymbol{y}) \equiv
\begin{cases}
1 & \text{if } L(\boldsymbol{y}) \geq \eta \\[2ex]
0 & \text{if } L(\boldsymbol{y}) < \eta.
\end{cases}
\tag{1.20}
$$

With

$$
B_h = \left\{ \boldsymbol{y} \in \mathbb{R}^k : \ f_h(\boldsymbol{y}) = 0 \right\}, \quad h = 0, 1,
\tag{1.21}
$$

we note that

$$
\mathbb{P}\left[f_0(\boldsymbol{Y}) = 0 | H = h \right] = \int_{B_0} f_h(\boldsymbol{y}) dF(\boldsymbol{y}), \quad h = 0, 1.
\tag{1.22}
$$

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

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

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

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

This discussion leads to the following fact.

Lemma 1.6.1 *Assume Conditions* **(A.1)**–**(A.2)** *to hold. For each* $\eta \geq 0$, *the tests* d_η *and* Lrt_η *are equivalent in the sense that* $P_M(d_\eta) = P_M(Lrt_\eta)$ *and* $P_F(d_\eta) = P_F(Lrt_\eta)$.

1.7 Exercises

1.8 References

Chapter 2

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. Two other formulations, namely the Minimax formulation and the Neyman-Pearson formulation, will be studied in Chapters 4 and 5, respectively.

2.1 The Bayesian optimization problem

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

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

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 $C(H(\omega), d(\boldsymbol{Y}(\omega)))$. Although it is tempting to seek to minimize this quantity, this is not possible. Indeed, the rv $\boldsymbol{Y}$ is observed, whence $d(\boldsymbol{Y})$ 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 $C(H, d(\boldsymbol{Y}))$ is not available. To remedy to this difficulty, we introduce the *expected cost function* $J : \mathcal{D} \to \mathbb{R}$ given by

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

The *Bayesian* Problem $\mathcal{P}_{\mathrm{B}}$ is the minimization problem

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

This amounts to finding an admissible test $d^\star : \mathbb{R}^k \to \{0, 1, \ldots, M - 1\}$ in $\mathcal{D}$ such that

$$J(d^\star) \leq J(d), \quad d \in \mathcal{D}. \tag{2.1}$$

Any admissible test $d^\star$ which satisfies (2.1) is called a Bayesian test, and the value

$$J(d^\star) = \inf_{d \in \mathcal{D}} J(d) = \min_{d \in \mathcal{D}} J(d) \tag{2.2}$$

is known as the *Bayesian* cost.

The solution to the Bayesian problem $\mathcal{P}_{\mathrm{B}}$ 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: Introduce the *relative* costs Γ_0 and Γ_1 given by

$$\Gamma_h \equiv C(h, 1 - h) - C(h, h), \quad h = 0, 1 \tag{2.3}$$

and define the auxiliary expected cost function $\widehat{J} : \mathcal{D} \to \mathbb{R}$ to be

$$\widehat{J}(d) = \mathbb{E}\left[\mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right] \Gamma_H\right], \quad d \in \mathcal{D}. \tag{2.4}$$

Lemma 2.1.1 *For any admissible rule d in $\mathcal{D}$, the relation*

$$J(d) = \mathbb{E}\left[C(H, H)\right] + \widehat{J}(d) \tag{2.5}$$

holds with

$$\widehat{J}(d) = \Gamma_0(1 - p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d). \tag{2.6}$$

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

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

It readily follows that

$$
\begin{aligned}
C(H, d(\boldsymbol{Y})) &= \mathbf{1}\left[d(\boldsymbol{Y}) = H\right] C(H, H) + \mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right] C(H, 1 - H) \\
&= (1 - \mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right]) C(H, H) + \mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right] C(H, 1 - H) \\
&= C(H, H) + (C(H, 1 - H) - C(H, H)) \mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right] \\
&= C(H, H) + \mathbf{1}\left[d(\boldsymbol{Y}) \neq H\right] \Gamma_H \quad\quad\quad\quad (2.7)
\end{aligned}
$$

with the relative costs Γ_0 and Γ_1 given by (2.3). Taking expectations on both sides of (2.7) we obtain (2.5).

The law of total probabilities gives

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

and the desired expression (2.6) is obtained. ∎

The Bayesian cost under a given decision rule is completely determined by its probabilities of false alarm and of miss. We also note that

$$
\begin{aligned}
\widehat{J}(d) &= \Gamma_0(1 - p) + \Gamma_1 p \cdot \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right] \\
&\quad - \Gamma_0(1 - p) \cdot \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 0\right], \quad d \in \mathcal{D} \quad\quad (2.8)
\end{aligned}
$$

as an immediate consequence of (2.6).

Therefore, by Lemma 1.6.1 it follows from (2.5)-(2.6) that $J(d_\eta) = J(Lrt_\eta)$ *regardless* of the cost function $C : \{0, 1\} \times \{0, 1\} \to \mathbb{R}$. The same argument also shows that any two versions of the likelihood ratio function will generate likelihood ratio tests which are equivalent.

2.2 Solving the Bayesian problem $\mathcal{P}_B$

It follows from (2.6) 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 do so, 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

$$
\begin{aligned}
\mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = h\right] &= \int_{C(d)} dF_h(\boldsymbol{y}) \\
&= \int_{C(d)} f_h(\boldsymbol{y}) dF(\boldsymbol{y}), \quad h = 0, 1
\end{aligned}
\tag{2.9}
$$

with $C(d)$ defined at (1.12). It is now easy to see from (2.8) that

$$
\widehat{J}(d) = \Gamma_0(1-p) + \int_{C(d)} h(\boldsymbol{y}) dF(\boldsymbol{y})
\tag{2.10}
$$

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

$$
h(\boldsymbol{y}) \equiv \Gamma_1 p \cdot f_1(\boldsymbol{y}) - \Gamma_0(1-p) \cdot f_0(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^k.
\tag{2.11}
$$

Theorem 2.2.1 *Assume the absolute continuity condition* **(A.1)** *to hold. Define the Borel set $C^\star$ by*

$$
C^\star \equiv \{y \in \mathbb{R}^k : h(\boldsymbol{y}) < 0\}
\tag{2.12}
$$

with $h : \mathbb{R}^k \to \mathbb{R}$ given by (2.11). The decision rule $d^\star : \mathbb{R}^k \to \{0, 1\}$ induced by $C^\star$ is given by

$$
d^\star(\boldsymbol{y}) = \begin{cases} 1 & \text{if } x \notin C^\star \\[2mm] 0 & \text{if } x \in C^\star; \end{cases}
\tag{2.13}
$$

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

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

$$
\widehat{J}(d^\star) \leq \widehat{J}(d), \quad d \in \mathcal{D}.
\tag{2.14}
$$

Indeed, for every test d in $\mathcal{D}$, we see from (2.10) that

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

and

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

Therefore,

$$
\begin{aligned}
J(d) &- J(d^\star) \\
&= \widehat{J}(d) - \widehat{J}(d^\star) \\
&= \int_{C(d) \setminus C^\star} h(\boldsymbol{y}) dF(\boldsymbol{y}) + \int_{C^\star \setminus C(d)} (-h(\boldsymbol{y}))\, dF(\boldsymbol{y}) \geq 0 \qquad (2.15)
\end{aligned}
$$

since

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

by the very definition of $C^\star$. The problems $\widehat{\mathcal{P}}_B$ and $\mathcal{P}_B$ are therefore simultaneously solved by the test $d^\star$ defined at (2.13). ∎

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

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

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

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

While it is true that $J(d^\star) = J(d^{\star\star})$, it is not necessarily the case that the equalities $P_F(d^\star) = P_F(d^{\star\star})$ or $P_M(d^\star) = P_M(d^{\star\star})$ hold.

Implementation using likelihood ratio test Assume that $0 < p < 1$ to avoid trivial situations, and that the relative costs satisfy the conditions

$$\Gamma_h > 0, \quad h = 0, 1, \qquad\qquad (2.16)$$

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 $d^\star$ given in Theorem 2.2.1 takes the equivalent form

$$d^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad f_1(\boldsymbol{y}) < \frac{\Gamma_0(1-p)}{\Gamma_1 p} f_0(\boldsymbol{y}). \tag{2.17}$$

In view of the definition (1.18), the Bayesian test $d^\star$ is indeed a test d_η with η given by

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

Equipped with Lemma 1.6.1 we can now restate Theorem 2.2.1.

Theorem 2.2.2 *Assume Conditions* **(A.1)**–**(A.2)** *to hold. Whenever* $\Gamma_h > 0$ *for* $h = 0, 1$, *the Bayesian decision rule* $d^\star$ *identified in Theorem 2.1 is equivalent to the likelihood ratio test* $Lrt_{\eta^\star}$ *where*

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

2.3   The probability of error criterion

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

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

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

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

We check that

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

and the relations (2.5)-(2.6) yield

$$
\begin{aligned}
P_E(d) &= (1-p) \cdot P_F(d) + p \cdot P_M(d) \\
&= p + (1-p) \cdot P_F(d) - p \cdot P_D(d), \quad d \in \mathcal{D}. \quad (2.18)
\end{aligned}
$$

For the probability of error criterion, the threshold $\eta^\star$ appearing in Theorem 2.2.2 has the simpler form

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

The optimal decision rule $d^\star$, as described at (2.17), can now be rewritten as

$$
d^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad f_1(\boldsymbol{y}) < \frac{1-p}{p} f_0(\boldsymbol{y}). \quad (2.19)
$$

The ML test   In the uniform prior case, i.e., $p = \frac{1}{2}$, the Bayesian test (2.19) becomes

$$
d^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad f_1(\boldsymbol{y}) < f_0(\boldsymbol{y}). \quad (2.20)
$$

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

The MAP computer   Finally, (2.19) can also be rewritten as

$$
d^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad \mathbb{P}\left[H = 1 | \boldsymbol{Y} = \boldsymbol{y}\right] < \mathbb{P}\left[H = 0 | \boldsymbol{Y} = \boldsymbol{y}\right] \quad (2.21)
$$

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

$$
\mathbb{P}\left[H = 1 | \boldsymbol{Y} = \boldsymbol{y}\right] = \frac{p f_1(\boldsymbol{y})}{p f_1(\boldsymbol{y}) + (1-p) f_0(\boldsymbol{y})}
$$

and

$$
\mathbb{P}\left[H = 0 | \boldsymbol{Y} = \boldsymbol{y}\right] = \frac{(1-p) f_0(\boldsymbol{y})}{p f_1(\boldsymbol{y}) + (1-p) f_0(\boldsymbol{y})}
$$

by Bayes' Theorem. For each $h = 0, 1$, the conditional probability $\mathbb{P}\left[H = h | \boldsymbol{Y} = \boldsymbol{y}\right]$ is known as the *posterior* probability that $H = h$ occurs given the observation $\boldsymbol{y}$. Put differently, the optimal test (2.21) compares these posterior probabilities given the observation $\boldsymbol{y}$, and selects the hypothesis with the largest posterior probability, hence the terminology *Maximum A Posteriori* (MAP) computer.

2.4　The Gaussian case

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

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

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

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

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

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

where we have used the notation

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

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

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

and a simple logarithmic transformation yields

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

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

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

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

$$
\begin{aligned}
Q(\boldsymbol{y}) &= (\boldsymbol{y} - \boldsymbol{m}_0)' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{m}_0) - (\boldsymbol{y} - \boldsymbol{m}_1)' \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{m}_1) \\
&= 2 \boldsymbol{y}' \boldsymbol{R}^{-1} (\boldsymbol{m}_1 - \boldsymbol{m}_0) - \left(\boldsymbol{m}_1' \boldsymbol{R}^{-1} \boldsymbol{m}_1 - \boldsymbol{m}_0' \boldsymbol{R}^{-1} \boldsymbol{m}_0 \right). \quad (2.22)
\end{aligned}
$$

The form of Lrt_η simplifies even further to read

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

where we have set

$$\Delta \boldsymbol{m} \equiv \boldsymbol{m}_1 - \boldsymbol{m}_0 \tag{2.23}$$

and

$$\tau(\eta) \equiv \frac{1}{2} \left(\boldsymbol{m}_1' \boldsymbol{R}^{-1} \boldsymbol{m}_1 - \boldsymbol{m}_0' \boldsymbol{R}^{-1} \boldsymbol{m}_0 \right) + \log \eta. \tag{2.24}$$

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

$$
\begin{aligned}
P_F(Lrt_\eta) &= \mathbb{P}\left[Lrt_\eta(\boldsymbol{Y}) = 1 | H = 0 \right] \\
&= \mathbb{P}\left[L(\boldsymbol{Y}) \geq \eta | H = 0 \right] \\
&= \mathbb{P}\left[\boldsymbol{Y}' \boldsymbol{R}^{-1} \Delta \boldsymbol{m} \geq \tau(\eta) \mid H = 0 \right] \quad (2.25)
\end{aligned}
$$

and

$$
\begin{aligned}
P_M(Lrt_\eta) &= \mathbb{P}\left[Lrt_\eta(\boldsymbol{Y}) = 0 | H = 1 \right] \\
&= \mathbb{P}\left[L(\boldsymbol{Y}) < \eta | H = 1 \right] \\
&= \mathbb{P}\left[\boldsymbol{Y}' \boldsymbol{R}^{-1} \Delta \boldsymbol{m} < \tau(\eta) \mid H = 1 \right] \\
&= 1 - \mathbb{P}\left[\boldsymbol{Y}' \boldsymbol{R}^{-1} \Delta \boldsymbol{m} \geq \tau(\eta) \mid H = 1 \right]. \quad (2.26)
\end{aligned}
$$

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

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

and

$$
\begin{aligned}
\mathrm{Var}\left[\boldsymbol{Y}'\boldsymbol{R}^{-1}\Delta\boldsymbol{m}|H=h\right] &= \left(\boldsymbol{R}^{-1}\Delta\boldsymbol{m}\right)'\mathrm{Cov}\left[\boldsymbol{Y}|H=h\right]\left(\boldsymbol{R}^{-1}\Delta\boldsymbol{m}\right)\\
&= \left(\boldsymbol{R}^{-1}\Delta\boldsymbol{m}\right)'\boldsymbol{R}\left(\boldsymbol{R}^{-1}\Delta\boldsymbol{m}\right)\\
&= \Delta\boldsymbol{m}'\boldsymbol{R}^{-1}\Delta\boldsymbol{m},
\end{aligned}
\tag{2.27}
$$

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

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

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

$$
\begin{aligned}
\mathbb{P}&\left[\boldsymbol{Y}'R^{-1}\Delta\boldsymbol{m} \geq \tau(\eta)|H=h\right]\\
&= \mathbb{P}\left[\boldsymbol{m}'_h\boldsymbol{R}^{-1}\Delta\boldsymbol{m} + \sqrt{\Delta\boldsymbol{m}'\boldsymbol{R}^{-1}\Delta\boldsymbol{m}}\cdot Z \geq \tau(\eta)\right]\\
&= \mathbb{P}\left[Z \geq \frac{\tau(\eta) - \boldsymbol{m}'_h\boldsymbol{R}^{-1}\Delta\boldsymbol{m}}{\sqrt{\Delta\boldsymbol{m}'\boldsymbol{R}^{-1}\Delta\boldsymbol{m}}}\right]
\end{aligned}
\tag{2.28}
$$

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

For the sake of convenience, pose

$$
d^2 \equiv \Delta\boldsymbol{m}'\boldsymbol{R}^{-1}\Delta\boldsymbol{m},
\tag{2.29}
$$

and note that

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

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

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

and

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

whence

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

regardless of the prior p.

2.5   The Bernoulli case

Consider the binary hypothesis testing problem

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

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

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

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

$$\begin{aligned} d_\eta(y) = 0 \quad & \text{iff} \quad \left(\frac{1 - a_1}{1 - a_0}\right)^{1-y} \cdot \left(\frac{a_1}{a_0}\right)^{y} < \eta \\ & \text{iff} \quad \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^{y} < \eta \cdot \frac{1 - a_0}{1 - a_1}, \quad y \in \mathbb{R}. \quad (2.30) \end{aligned}$$

Therefore,

$$
\begin{aligned}
P_F(d_\eta) &= \mathbb{P}\left[d_\eta(Y) = 1 | H = 0\right] \\
&= \mathbb{P}\left[\left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y \geq \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 0\right] \\
&= \mathbb{P}\left[Y = 1, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y \geq \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 0\right] \\
&\quad + \mathbb{P}\left[Y = 0, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y \geq \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 0\right] \\
&= a_0 \mathbf{1}\left[\eta \cdot \frac{1 - a_0}{1 - a_1} \leq \frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right] + (1 - a_0)\mathbf{1}\left[\eta \cdot \frac{1 - a_0}{1 - a_1} \leq 1\right] \\
&= a_0 \mathbf{1}\left[\eta \leq \frac{a_1}{a_0}\right] + (1 - a_0)\mathbf{1}\left[\eta \frac{1 - a_0}{1 - a_1} \leq 1\right] \\
&= a_0 \mathbf{1}\left[\eta \leq \frac{a_1}{a_0}\right] + (1 - a_0)\mathbf{1}\left[\eta \leq \frac{1 - a_1}{1 - a_0}\right]. \quad (2.31)
\end{aligned}
$$

Similarly, we get

$$
\begin{aligned}
P_M(d_\eta) &= \mathbb{P}\left[d_\eta(Y) = 0 | H = 1\right] \\
&= \mathbb{P}\left[\left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 1\right] \\
&= \mathbb{P}\left[Y = 1, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 1\right] \\
&\quad + \mathbb{P}\left[Y = 0, \left(\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0}\right)^Y < \eta \cdot \frac{1 - a_0}{1 - a_1} \Big| H = 1\right] \\
&= a_1 \mathbf{1}\left[\frac{1 - a_0}{1 - a_1} \cdot \frac{a_1}{a_0} < \eta \cdot \frac{1 - a_0}{1 - a_1}\right] + (1 - a_1)\mathbf{1}\left[1 < \eta \cdot \frac{1 - a_0}{1 - a_1}\right] \\
&= a_1 \mathbf{1}\left[\frac{a_1}{a_0} < \eta\right] + (1 - a_1)\mathbf{1}\left[1 < \eta \cdot \frac{1 - a_0}{1 - a_1}\right] \\
&= a_1 \mathbf{1}\left[\frac{a_1}{a_0} < \eta\right] + (1 - a_1)\mathbf{1}\left[\frac{1 - a_1}{1 - a_0} < \eta\right]. \quad (2.32)
\end{aligned}
$$

2.6 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 and F_0 admitting probability density functions $f_0, f_1 : \mathbb{R} \to \mathbb{R}_+$ with respect to Lebesgue measure given by

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

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

$$
\begin{aligned} h(y) &= \Gamma_1 p \cdot f_1(y) - \Gamma_0(1-p) \cdot f_0(y) \\ &= \begin{cases} 0 & \text{if } y < -1 \\ \frac{1}{3}\Gamma_1 p - \Gamma_0(1-p)(1-|y|) & \text{if } |y| \leq 1 \\ \frac{1}{3}\Gamma_1 p & \text{if } 1 < y \leq 2 \\ 0 & \text{if } 2 < y. \end{cases} \end{aligned} \tag{2.33}
$$

The Bayesian test $d^\star$ is simply

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

Another example where absolute continuity (A.2) fails   The observation is the scalar rv Y with F_0 and F_0 admitting probability density functions $f_0, f_1 : \mathbb{R} \to \mathbb{R}_+$ with respect to Lebesgue measure given by

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

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

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

$$
= \begin{cases}
0 & \text{if } y < -1 \\[2mm]
-\Gamma_0(1-p)(1+y) & \text{if } -1 \leq y \leq 0 \\[2mm]
\frac{1}{3}\Gamma_1 p - \Gamma_0(1-p)(1-y) & \text{if } 0 < y \leq 1 \\[2mm]
\frac{1}{3}\Gamma_1 p & \text{if } 1 < y \leq 3 \\[2mm]
0 & \text{if } 3 < y,
\end{cases}
\tag{2.34}
$$

and it is straightforward to check that the Bayesian test $d^\star$ is simply

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

Equivalently, $d^\star$ can be described as

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

A final example   Consider the binary hypothesis testing problem

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

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

$$
f_1(y) = \begin{cases}
1 & \text{if } y \in (0, 1) \\[2mm]
0 & \text{otherwise}
\end{cases}
$$

and

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

In this example F cannot be taken to be either the distribution associated with Lebesgue measure on $\mathbb{R}$ or with the counting measure on $\{0, 1\}$. In principle we could use F given by (1.10) but this would yield complicated expressions for the density functions $f_0, f_1 : \mathbb{R} \to \mathbb{R}^+$. Instead of applying Theorem 2.2.1 with that choice, we provide a direct optimization of the auxiliary expected cost function (2.4): For each test d in $\mathcal{D}$ we recall that we have

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

with

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

and

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

Adding or deleting a finite number of points from $C(d)$ will *not* affect the value of $\mathbb{P}\left[d(Y) = 0 | H = 1\right]$, but it may change the value of $\mathbb{P}\left[d(Y) = 0 | H = 0\right]$. 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

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

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

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

We can now conclude that

$$
\begin{aligned}
&\widehat{J}(d) \\
&= \; \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}\left[d(Y) = 0 | H = 1\right] - \Gamma_0(1-p) \cdot \mathbb{P}\left[d(Y) = 0 | H = 0\right] \\
&\geq \; \Gamma_0(1-p) + \Gamma_1 p \cdot \mathbb{P}\left[d'(Y) = 0 | H = 1\right] - \Gamma_0(1-p) \cdot \mathbb{P}\left[d'(Y) = 0 | H = 0\right] \\
&= \; \Gamma_0(1-p) + \Gamma_1 p \cdot |C(d') \cap [0, 1]| - \Gamma_0(1-p) \\
&= \; \Gamma_1 p \cdot |C(d') \cap [0, 1]| \geq 0. \qquad\qquad (2.35)
\end{aligned}
$$

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

$$d^\star(y) = \mathbf{1}\left[\{0, 1\}\right](y), \quad y \in \mathbb{R}.$$

The arguments leading to (2.35) also show that

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

and the test $d^\star$ is therefore a Bayesian decision rule.

2.7   Exercises

2.8   References

Chapter 3

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 done for the Bayesian set–up. In some very real sense this class $\mathcal{D}$ of tests is not always large enough to guarantee a solution; to remedy this difficulty we enlarge $\mathcal{D}$ by considering the class of *randomized* tests or decision rules.

3.1   Randomized tests

We start with a definition.

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

Obviously, any test d in $\mathcal{D}$ can be mechanized as a randomized test, say $\delta_d : \mathbb{R}^k \to [0,1]$, given by

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

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

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, $\boldsymbol{Y}$ and D which take values in $\{0, 1\}$, $\mathbb{R}^k$ and $\{0, 1\}$, respectively. The rvs H and $\boldsymbol{Y}$ 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 $\boldsymbol{Y}$.

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

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

and

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

Therefore, under $\mathbb{P}_\delta$ the probability distribution of the pair $(H, \boldsymbol{Y})$ does not dependent on δ with

$$\mathbb{P}_\delta\left[H = h, \boldsymbol{Y} \leq \boldsymbol{y}\right] = \mathbb{P}_\delta\left[H = h\right] F_h(\boldsymbol{y}), \quad \begin{array}{l} h = 0, 1, \\ \boldsymbol{y} \in \mathbb{R}^k \end{array} \tag{3.1}$$

as expected. In addition, for $h = 0, 1$ and $\boldsymbol{y}$ in $\mathbb{R}^k$, we now require that

$$\mathbb{P}_\delta\left[D = d|H = h, \boldsymbol{Y} = \boldsymbol{y}\right] = \begin{cases} 1 - \delta(\boldsymbol{y}) & \text{if } d = 0 \\ \\ \delta(\boldsymbol{y}) & \text{if } d = 1 \end{cases}$$
$$= d\delta(\boldsymbol{y}) + (1 - d)\left(1 - \delta(\boldsymbol{y})\right). \tag{3.2}$$

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

$$\begin{aligned}
\mathbb{P}_\delta &\left[H = h, D = d, \boldsymbol{Y} \in B\right] \\
&= \mathbb{E}_\delta\left[\mathbf{1}\left[H = h, \boldsymbol{Y} \in B\right]\mathbb{P}_\delta\left[D = d|H, \boldsymbol{Y}\right]\right] \\
&= \mathbb{E}_\delta\left[\mathbf{1}\left[H = h, \boldsymbol{Y} \in B\right]\left(d\delta(\boldsymbol{Y}) + (1 - d)\left(1 - \delta(\boldsymbol{Y})\right)\right)\right] \\
&= \mathbb{P}_\delta\left[H = h\right] \cdot \int_B \left(d\delta(\boldsymbol{y}) + (1 - d)\left(1 - \delta(\boldsymbol{y})\right)\right) dF_h(\boldsymbol{y}) \\
&= \begin{cases} \mathbb{P}_\delta\left[H = h\right] \cdot \int_B (1 - \delta(\boldsymbol{y}))dF_h(\boldsymbol{y}) & \text{if } d = 0 \\ \\ \mathbb{P}_\delta\left[H = h\right] \cdot \int_B \delta(\boldsymbol{y})dF_h(\boldsymbol{y}) & \text{if } d = 1. \end{cases}
\end{aligned} \tag{3.3}$$

3.2 An alternate framework

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

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

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 on it a rv $U : \Omega \to [0, 1]$ which is uniformly distributed on $(0, 1)$, and independent of the pair of rvs H and $\boldsymbol{Y}$, This amounts to

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

with

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

$$\mathbb{P}\left[H = h, \boldsymbol{Y} \leq \boldsymbol{y}\right] = \mathbb{P}\left[H = h\right] F_h(\boldsymbol{y}), \quad \begin{array}{c} h = 0, 1, \\ \boldsymbol{y} \in \mathbb{R}^k \end{array}$$

and

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

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

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

Note that

$$\begin{aligned} \mathbb{P}\left[D_\delta = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] &= \mathbb{E}\left[\mathbf{1}\left[U \leq \delta(\boldsymbol{Y})\right] | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\ &= \mathbb{E}\left[\mathbf{1}\left[U \leq \delta(\boldsymbol{y})\right] | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\ &= \mathbb{P}\left[U \leq \delta(\boldsymbol{y})\right] \\ &= \delta(\boldsymbol{y}) \end{aligned} \tag{3.4}$$

under the enforced independence assumptions. Similarly it follows that

$$\begin{aligned} \mathbb{P}\left[D_\delta = 0 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] &= 1 - \mathbb{P}\left[D_\delta = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\ &= 1 - \delta(\boldsymbol{y}). \end{aligned} \tag{3.5}$$

Therefore, the conditional distribution of D_δ (under $\mathbb{P}$) given H and $\boldsymbol{Y}$ coincides with the conditional distribution of D (under $\mathbb{P}_\delta$) given H and $\boldsymbol{Y}$, and the two formalisms are probabilistically equivalent.

3.3   Evaluating error probabilities

Evaluating error probabilities under randomized tests can be done rather easily: Consider a randomized test δ in $\mathcal{D}^\star$. In analogy with (1.15) and (1.16), we evaluate the probabilities of false alarm and miss under δ as

$$P_F(\delta) \equiv \mathbb{P}_\delta\left[D = 1|H = 0\right] \tag{3.6}$$

and

$$P_M(\delta) \equiv \mathbb{P}_\delta\left[D = 0|H = 1\right]. \tag{3.7}$$

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

$$P_D(\delta) \equiv \mathbb{P}_\delta\left[D = 1|H = 1\right] = 1 - P_M(\delta). \tag{3.8}$$

Because

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

we readily conclude that

$$P_F(\delta) = \int_{\mathbb{R}^k} \delta(\boldsymbol{y})dF_0(\boldsymbol{y}) \tag{3.9}$$

and

$$P_M(\delta) = \int_{\mathbb{R}^k} \left(1 - \delta(\boldsymbol{y})\right)dF_1(\boldsymbol{y}), \tag{3.10}$$

so that

$$P_D(\delta) = \int_{\mathbb{R}^k} \delta(\boldsymbol{y})dF_1(\boldsymbol{y}), \tag{3.11}$$

3.4   The Bayesian problem revisited

Assuming the cost function $C : \{0,1\} \times \{0,1\} \to \mathbb{R}$ introduced in Section 2.1, we define the expected cost function $J^\star : \mathcal{D}^\star \to \mathbb{R}$ given by

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

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

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

This amounts to finding an admissible test $\delta^\star$ in $\mathcal{D}^\star$ such that

$$J^\star(\delta^\star) \leq J^\star(\delta), \quad \delta \in \mathcal{D}^\star. \tag{3.12}$$

Any admissible test $\delta^\star$ which satisfies (3.12) is called a randomized Bayesian test, and the value

$$J^\star(\delta^\star) = \inf_{\delta \in \mathcal{D}^\star} J^\star(\delta)) \tag{3.13}$$

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

Obviously, since $\mathcal{D} \subset \mathcal{D}^\star$ (with a slight abuse of notation) with

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

it is plain that

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

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; the proof is available in Section 3.6.

Theorem 3.4.1 *Under the absolute continuity condition* **(A.1)**, *it holds that*

$$\inf_{\delta \in \mathcal{D}^\star} J^\star(\delta) = \inf_{d \in \mathcal{D}} J(d). \tag{3.14}$$

It follows from Theorem 2.13 that (3.14) is equivalent to

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

where the deterministic test $d^\star : \mathbb{R}^k \to \{0, 1\}$ is given by (2.13).

For easy reference we close with the following analog of Lemma 2.1.1 for randomized tests; the proof is left as an exercise.

Lemma 3.4.1 *For any admissible rule δ in $\mathcal{D}^{\star}$, the relation*

$$J^{\star}(\delta) = \mathbb{E}\left[C(H, H)\right] + \widehat{J}^{\star}(\delta) \tag{3.16}$$

holds with

$$\widehat{J}^{\star}(\delta) = \Gamma_0(1 - p) \cdot P_F(\delta) + \Gamma_1 p \cdot P_M(\delta). \tag{3.17}$$

3.5   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 $\mathcal{D}^{\star}$ 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, this amounts to defining $\delta_a : \mathbb{R}^k \to [0, 1]$ by

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

Applying the expressions (3.9) and (3.10) with the randomized test δ_a we get

$$
\begin{aligned}
P_F(\delta_a) &= \int_{\mathbb{R}^k} \delta_a(\boldsymbol{y})dF_0(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^k} \left(ad_1(\boldsymbol{y}) + (1 - a)d_2(\boldsymbol{y})\right) dF_0(\boldsymbol{y}) \\
&= a \int_{\mathbb{R}^k} d_1(\boldsymbol{y})dF_0(\boldsymbol{y}) + (1 - a) \int_{\mathbb{R}^k} d_2(\boldsymbol{y})dF_0(\boldsymbol{y}) \\
&= aP_F(d_1) + (1 - a)P_F(d_2). \tag{3.18}
\end{aligned}
$$

Similarly we find that

$$
\begin{aligned}
P_M(\delta_a) &= \int_{\mathbb{R}^k} \left(1 - \delta_a(\boldsymbol{y})\right) dF_1(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^k} \left(1 - ad_1(\boldsymbol{y}) - (1 - a)d_2(\boldsymbol{y})\right) dF_1(\boldsymbol{y}) \\
&= a \int_{\mathbb{R}^k} (1 - d_1(\boldsymbol{y}))dF_1(\boldsymbol{y}) + (1 - a) \int_{\mathbb{R}^k} (1 - d_2(\boldsymbol{y}))dF_1(\boldsymbol{y}) \\
&= aP_M(d_1) + (1 - a)P_M(d_2). \tag{3.19}
\end{aligned}
$$

It immediately follows from (3.16) and (3.17) that

$$J_p^\star(\delta_a) = aJ_p(d_1) + (1-a)J_p(d_2). \qquad (3.20)$$

as we use the relations (3.18) and (3.19).

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

$$B_a = \mathbf{1}\left[V \le a\right].$$

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

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

Define the decision rv D_a given by

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

It is easy to check that

$$
\begin{aligned}
&\mathbb{P}\left[D_a = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\
=\ & \mathbb{P}\left[B_a d_1(\boldsymbol{Y}) + (1 - B_a)d_2(\boldsymbol{Y}) = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\
=\ & \mathbb{P}\left[B_a d_1(\boldsymbol{y}) + (1 - B_a)d_2(\boldsymbol{y}) = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\
=\ & \mathbb{P}\left[B_a = 1, d_1(\boldsymbol{y}) = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] + \mathbb{P}\left[B_a = 0, d_2(\boldsymbol{y}) = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\
=\ & d_1(\boldsymbol{y})\mathbb{P}\left[B_a = 1 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] + d_2(\boldsymbol{y})\mathbb{P}\left[B_a = 0 | H = h, \boldsymbol{Y} = \boldsymbol{y}\right] \\
=\ & d_1(\boldsymbol{y})\mathbb{P}\left[B_a = 1\right] + d_2(\boldsymbol{y})\mathbb{P}\left[B_a = 0\right] \\
=\ & ad_1(\boldsymbol{y}) + (1-a)d_2(\boldsymbol{y}), \quad \begin{matrix} \boldsymbol{y} \in \mathbb{R}^k, \\ h = 0, 1 \end{matrix} \qquad (3.21)
\end{aligned}
$$

as desired.

3.6   A proof of Theorem 3.4.1

Pick an arbitrary test δ in $\mathcal{D}^\star$. A simple preconditioning argument shows that

$$J^\star(\delta) \ =\ \mathbb{E}_\delta\left[C(H, D)\right]$$

$$\begin{aligned}
&= \ \mathbb{E}_\delta\left[\mathbb{E}_\delta\left[C(H,D)|H,\boldsymbol{Y}\right]\right]\\
&= \ \mathbb{E}_\delta\left[C(H,1)\mathbb{P}_\delta\left[D=1|H,\boldsymbol{Y}\right]+C(H,0)\mathbb{P}_\delta\left[D=0|H,\boldsymbol{Y}\right]\right]\\
&= \ \mathbb{E}_\delta\left[C(H,1)\cdot\delta(\boldsymbol{Y})+C(H,0)\cdot(1-\delta(\boldsymbol{Y}))\right]\\
&= \ \mathbb{E}_\delta\left[C(H,0)\right]+\mathbb{E}_\delta\left[(C(H,1)-C(H,0))\cdot\delta(\boldsymbol{Y})\right] \qquad (3.22)
\end{aligned}$$

with

$$\begin{aligned}
&\mathbb{E}_\delta\left[(C(H,1)-C(H,0))\cdot\delta(\boldsymbol{Y})\right]\\
&= \ \mathbb{E}_\delta\left[(C(H,1)-C(H,0))\cdot\mathbb{E}_\delta\left[\delta(\boldsymbol{Y})|H\right]\right]\\
&= \ (C(1,1)-C(1,0))\,\mathbb{E}_\delta\left[\delta(\boldsymbol{Y})|H=1\right]\mathbb{P}_\delta\left[H=1\right]\\
&\quad +(C(0,1)-C(0,0))\,\mathbb{E}_\delta\left[\delta(\boldsymbol{Y})|H=0\right]\mathbb{P}_\delta\left[H=0\right]\\
&= \ -\Gamma_1 p\cdot\mathbb{E}_\delta\left[\delta(\boldsymbol{Y})|H=1\right]+\Gamma_0(1-p)\cdot\mathbb{E}_\delta\left[\delta(\boldsymbol{Y})|H=0\right].\quad (3.23)
\end{aligned}$$

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

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

so that

$$\begin{aligned}
&J^\star(\delta)-\mathbb{E}_\delta\left[C(H,0)\right]\\
&= \ -\Gamma_1 p\cdot\int_{\mathbb{R}^k}\delta(\boldsymbol{y})f_1(\boldsymbol{y})dF(\boldsymbol{y})+\Gamma_0(1-p)\cdot\int_{\mathbb{R}^k}\delta(\boldsymbol{y})f_0(\boldsymbol{y})dF(\boldsymbol{y})\\
&= \ \int_{\mathbb{R}^k}\left(-\Gamma_1 pf_1(\boldsymbol{y})+\Gamma_0(1-p)f_0(\boldsymbol{y})\right)\delta(\boldsymbol{y})dF(\boldsymbol{y})\\
&= \ -\int_{\mathbb{R}^k}h(\boldsymbol{y})\delta(\boldsymbol{y})dF(\boldsymbol{y}) \qquad\qquad\qquad\qquad (3.24)
\end{aligned}$$

where the mapping $h:\mathbb{R}^k\to\mathbb{R}$ is given by (2.11). Note that the term $\mathbb{E}_\delta\left[C(H,0)\right]$ does not depend on the randomized test δ being used.

From Theorem 2.2.1 recall that the Bayesian rule which solves Problem $\mathcal{P}_B$ is the test $d^\star:\mathbb{R}^k\to\{0,1\}$ in $\mathcal{D}$ given by (2.13). Note that $d^\star$ can also be interpreted as the randomized rule $\delta^\star:\mathbb{R}^k\to[0,1]$ given by

$$\delta^\star(\boldsymbol{y})=\left\{\begin{array}{ll} 0 & \text{if }h(\boldsymbol{y})<0\\[2mm] 1 & \text{if }h(\boldsymbol{y})\geq 0\end{array}\right.=\left\{\begin{array}{ll} 0 & \text{if }\boldsymbol{y}\in C^\star\\[2mm] 1 & \text{if }\boldsymbol{y}\notin C^\star\end{array}\right.$$

where $C^\star$ is the Borel subset of $\mathbb{R}^k$ given by (2.12).

The desired result will be established if we show that

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

The approach we take is reminiscent of the one used in the proof of Theorem 2.2.1: For an arbitrary δ in $\mathcal{D}^\star$, earlier calculations (3.24) show that

$$
\begin{aligned}
J^\star(\delta) - J^\star(\delta^\star) &= -\int_{\mathbb{R}^k} h(\boldsymbol{y})\delta(\boldsymbol{y})dF(\boldsymbol{y}) + \int_{\mathbb{R}^k} h(\boldsymbol{y})\delta^\star(\boldsymbol{y})dF(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^k} h(\boldsymbol{y}) \left(\delta^\star(\boldsymbol{y}) - \delta(\boldsymbol{y})\right) dF(\boldsymbol{y}) \\
&= \int_{C^\star} (-h(\boldsymbol{y}))\delta(\boldsymbol{y})dF(\boldsymbol{y}) + \int_{\mathbb{R}^k \setminus C^\star} \left(1 - \delta(\boldsymbol{y})\right) h(\boldsymbol{y})dF(\boldsymbol{y}) \\
&\geq 0
\end{aligned}
$$

as desired since

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

by the very definition of the set $C^\star$ and of the mapping $h : \mathbb{R}^k \to \mathbb{R}$. ∎

3.7 Exercises

3.8 References

Chapter 4

The Minimax formulation

The Bayesian formulation *implicitly* assumes knowledge of the prior distribution on the hypothesis rv H. In many situations, this assumption cannot be adequately justified, and the Bayesian formulation has to abandoned for the so–called *Minimax formulation* discussed in this chapter.

4.1 Keeping track of the prior

To facilitate the discussion, we augment the notation introduced in Chapter 1 and Chapter 3 by explicitly indicating the dependence on the prior probability distribution: As before we are given two distinct probability distributions $F_0, F_1 :$ $\mathbb{R}^k \to [0, 1]$ which act as conditional probability distributions for the observation given the state of nature. As in Chapter 1, we can always construct a collection $\{(\Omega, \mathcal{F}, \mathbb{P}_p), \ p \in [0, 1]\}$ of probability triples, and rvs H and $\boldsymbol{Y}$ defined on Ω which take values in $\{0, 1\}$ and $\mathbb{R}^k$, respectively, such that for each p in $[0, 1]$,

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

and

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

One possible construction was given in Section 1.3: Take $\Omega = \{0, 1\} \times \mathbb{R}^k$ with generic element $\omega = (h, \boldsymbol{y})$ with $h = 0, 1$ and $\boldsymbol{y}$ an arbitrary element of $\mathbb{R}^k$. We endow Ω with the σ-field $\mathcal{F}$ given by

$$\mathcal{F} = \sigma\left(\mathcal{P}(\{0, 1\}) \times \mathcal{B}(\mathbb{R}^k)\right)$$

where $\mathcal{P}(\{0, 1\})$ is the power set of $\{0, 1\}$, and $\mathcal{B}(\mathbb{R}^k)$ is the Borel σ-field on $\mathbb{R}^k$. We define the mappings $H : \Omega \to \mathbb{R}$ and $\boldsymbol{Y} : \Omega \to \mathbb{R}^k$ by

$$H(\omega) = h \quad \text{and} \quad \boldsymbol{Y}(\omega) = \boldsymbol{y}, \quad \omega = (h, \boldsymbol{y}) \in \Omega.$$

Both projection mappings are Borel measurable, and therefore define rvs.

As before, it is plain that

$$
\begin{aligned}
\mathbb{P}_p\left[\boldsymbol{Y} \leq y, H = h\right] &= \mathbb{P}_p\left[\boldsymbol{Y} \leq \boldsymbol{y} | H = h\right] \mathbb{P}_p\left[H = h\right] \\
&= \begin{cases} (1-p)F_0(\boldsymbol{y}) & \text{if } h = 0, \boldsymbol{y} \in \mathbb{R}^k \\[2mm] pF_1(\boldsymbol{y}) & \text{if } h = 1, \boldsymbol{y} \in \mathbb{R}^k. \end{cases}
\end{aligned}
\tag{4.1}
$$

Let $\mathbb{E}_p\left[\cdot\right]$ denote expectation under $\mathbb{P}_p$.

When dealing with randomized strategies we further augment the notation $\mathbb{P}_\delta$ to read $\mathbb{P}_{\delta,p}$ when using the randomized strategy δ in $\mathcal{D}^\star$ with prior p; see Section 3.1 for details on the probabilistic framework to be used.. In that case let $\mathbb{E}_{\delta,p}\left[\cdot\right]$ denote expectation under $\mathbb{P}_{\delta,p}$.

4.2   The Bayesian problems

Fix 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.,

$$J_p(d) \equiv \mathbb{E}_p\left[C(H, d(\boldsymbol{Y}))\right].$$

Similarly, let $J_p^\star(\delta)$ denote the expected cost associated under the randomized decision rule δ in $\mathcal{D}^\star$ when the prior on H is p, i.e.,

$$J_p^\star(\delta) \equiv \mathbb{E}_{\delta,p}\left[C(H, D)\right].$$

The Bayesian problems introduced in Chapters 2 and 3 now read

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

and

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

The corresponding Bayesian costs will be denoted by

$$V(p) \equiv \inf_{d \in \mathcal{D}} J_p(d) \tag{4.2}$$

and

$$V^\star(p) \equiv \inf_{\delta \in \mathcal{D}^\star} J_p^\star(\delta). \tag{4.3}$$

As shown in Chapter 2, under Condition **(A.1)**, for each p in $[0, 1]$ the problem $\mathcal{P}_{p,\mathrm{B}}$ has a solution which we denote $d^\star(p)$ to indicate its dependence on the prior p. Clearly, any such solution satisfies

$$J_p(d^\star(p)) \leq J_p(d), \quad d \in \mathcal{D} \tag{4.4}$$

and the equality

$$V(p) = J_p(d^\star(p)) \tag{4.5}$$

holds. Under the same condition, Theorem 3.4.1 further shows that

$$J_p^\star(\delta_{d^\star(p)}) \leq J_p(\delta), \quad \delta \in \mathcal{D}^\star$$

so that

$$V^\star(p) = V(p). \tag{4.6}$$

The following properties of the value function $V : [0, 1] \to \mathbb{R}$ will be useful in the forthcoming discussion. Conditions **(A.1)** and **(A.2)** are not needed for the results to hold.

Lemma 4.2.1 *Assume $\Gamma_h > 0$ for $h = 0, 1$. The value function $V : [0, 1] \to \mathbb{R}$ is concave and continuous on the closed interval $[0, 1]$ with boundary values $V(0) = C(0, 0)$ and $V(1) = C(1, 1)$. Moreover, its right-derivative (resp. left-derivative) exists and is finite on $[0, 1)$ (resp. $(0, 1])$)*

The proof can be omitted in a first reading, and can be found in Section 4.10. For easy reference, recall that for each p in $[0, 1]$ the expressions

$$
\begin{aligned}
J_p(d) &= pC(1, 1) + (1 - p)C(0, 0) \\
&\quad + \Gamma_0(1 - p) \cdot P_F(d) + \Gamma_1 p \cdot P_M(d), \quad d \in \mathcal{D}
\end{aligned}
\tag{4.7}
$$

and

$$
\begin{aligned}
J_p^\star(\delta) &= pC(1, 1) + (1 - p)C(0, 0) \\
&\quad + \Gamma_0(1 - p) \cdot P_F(\delta) + \Gamma_1 p \cdot P_M(\delta), \quad \delta \in \mathcal{D}^\star
\end{aligned}
\tag{4.8}
$$

hold. The relationships were given in Lemma 2.1.1 and Lemma 3.4.1, respectively.

4.3 The minimax formulation

Since the exact value of the prior p is not available, the Bayesian criterion has to be modified. Two different approaches are possible; each in its own way seeks to *compensate* for the uncertainty in the modeling assumptions.

Minmax On possible approach is to introduce a *worst-case* cost associated with the original cost, and then use it as the new criterion to be minimized. With this in mind, define

$$J_{\mathrm{Max}}(d) \equiv \sup_{p \in [0,1]} J_p(d), \quad d \in \mathcal{D}. \tag{4.9}$$

We are then lead to consider the minimization problem

$$\mathcal{P}_{\mathrm{Max}} : \quad \text{Minimize } J_{\mathrm{Max}}(d) \text{ over } d \text{ in } \mathcal{D}.$$

Solving $\mathcal{P}_{\mathrm{Max}}$ amounts to finding an admissible test $d_{\mathrm{m}}^{\star}$ in $\mathcal{D}$ such that

$$J_{\mathrm{Max}}(d_{\mathrm{m}}^{\star}) \leq J_{\mathrm{Max}}(d), \quad d \in \mathcal{D}. \tag{4.10}$$

When it exists, the test $d_{\mathrm{m}}^{\star}$ is known as a *minimax* test.

A priori there is no guarantee that a test in $\mathcal{D}$ exists which satisfies (4.10) (even under Condition (**A.1**)) – It is not clear that a cost $\widetilde{C} : \{0,1\} \times \{0,1\} \to \mathbb{R}$ (likely related to the original cost $C : \{0,1\} \times \{0,1\} \to \mathbb{R}$) and $\tilde{p}$ in $[0,1]$ can be found such that

$$J_{\mathrm{Max}}(d) = \mathbb{E}_{\tilde{p}} \left[\widetilde{C}(H, d(\boldsymbol{Y})) \right], \quad d \in \mathcal{D}.$$

If that were indeed the case, then Theorem 2.2.1 would guarantee the existence of a minimizer.

For technical reasons to become shortly apparent we also introduce the worst-case cost under randomized strategies, namely

$$J_{\mathrm{Max}}^{\star}(\delta) \equiv \sup_{p \in [0,1]} J_p^{\star}(\delta), \quad \delta \in \mathcal{D}^{\star}. \tag{4.11}$$

The minimization problem of interest here is now defined as

$$\mathcal{P}_{\mathrm{Max}}^{\star} : \quad \text{Minimize } J_{\mathrm{Max}}^{\star}(\delta) \text{ over } \delta \text{ in } \mathcal{D}^{\star}.$$

Solving $\mathcal{P}_{\text{Max}}^\star$ amounts to finding a randomized strategy $\delta_{\text{m}}^\star$ in $\mathcal{D}^\star$ such that

$$J_{\text{Max}}^\star(\delta_{\text{m}}^\star) \leq J_{\text{Max}}^\star(\delta), \quad \delta \in \mathcal{D}^\star. \tag{4.12}$$

Again a priori there is no guarantee that there exists a test in $\mathcal{D}^\star$ satisfying (4.12) (even under Condition (**A.1**)). When it exists, the test $\delta_{\text{m}}^\star$ is also known as a *minimax* test.

It is natural to wonder whether the tests $d_{\text{m}}^\star$ and $\delta_{\text{m}}^\star$ exist, (possibly under additional conditions), whether they are different, and if not, whether $J_{\text{Max}}^\star(\delta_m^\star) = J_{\text{Max}}(d_{\text{m}}^\star)$.

Maxmin   Another reasonable way to proceed consists in using the Bayesian test for that value of p which yields the *largest* Bayesian cost (4.2): With the notation introduced earlier, let p_{m} in $[0, 1]$ such that

$$V(p_{\text{m}}) = \max_{p \in [0,1]} V(p), \tag{4.13}$$

and use the Bayesian rule $d^\star(p_{\text{m}})$ – The existence of p_{m} is guaranteed by the fact that the mapping $V : [0, 1] \to \mathbb{R}$ is continuous on the closed bounded interval $[0, 1]$ by Lemma 4.2.1, hence achieves its maximum value on $[0, 1]$.

The value p_{m} satisfying (4.13) is known as the *least favorable* prior. Although the terminology is not standard, we shall refer to $d^\star(p_{\text{m}})$ as a *maximin* test.

4.4 Preliminary facts

In view of the two competing approaches outlined in Section 4.3, several questions arise: (i) How does one characterize the minimax strategy $d_{\text{m}}^\star$ and develop ways find it; (ii) How does one characterize the least-favorable prior p_{m} and develop ways find it; (iii) Is there a simple relationship between the solutions proposed by two approaches, and in particular, whether is $d^\star(p_{\text{m}})$ is a candidate for $d_{\text{m}}^\star$.

To frame the discussion of these issues we start with a couple of preliminary remarks.

The minimax inequalities   As a first step towards understanding how the two approaches may be related to each other, consider the following arguments: From the definitions it always holds that

$$V(p) \leq J_p(d) \leq J_{\mathrm{Max}}(d), \quad \begin{matrix} p \in [0, 1] \\ d \in \mathcal{D}. \end{matrix} \tag{4.14}$$

It is now immediate that

$$V(p) \leq \inf_{d \in \mathcal{D}} J_{\mathrm{Max}}(d), \quad p \in [0, 1]$$

since $V(p)$ does not depend on d, whence

$$\sup_{p \in [0,1]} V(p) \leq \inf_{d \in \mathcal{D}} J_{\mathrm{Max}}(d).$$

This last inequality can be rewritten as the *minimax inequality*

$$\sup_{p \in [0,1]} \left(\inf_{d \in \mathcal{D}} J_p(d) \right) \leq \inf_{d \in \mathcal{D}} \left(\sup_{p \in [0,1]} J_p(d) \right) \tag{4.15}$$

(in pure policies)

If we were to consider randomized strategies, it is also the case that

$$V^{\star}(p) \leq J_p^{\star}(\delta) \leq J_{\mathrm{Max}}^{\star}(\delta), \quad \begin{matrix} p \in [0, 1] \\ \delta \in \mathcal{D}^{\star} \end{matrix} \tag{4.16}$$

and arguments similar to the ones leading to (4.15) yield the minimax inequality

$$\sup_{p \in [0,1]} \left(\inf_{\delta \in \mathcal{D}^{\star}} J_p^{\star}(\delta) \right) \leq \inf_{\delta \in \mathcal{D}^{\star}} \left(\sup_{p \in [0,1]} J_p^{\star}(\delta) \right) \tag{4.17}$$

in randomized strategies.

Toward minimax equalities   As we contrast the inequalities (4.15) and (4.17), it is natural to wonder whether these inequalities ever hold as *equalities*, namely

$$\sup_{p \in [0,1]} \left(\inf_{d \in \mathcal{D}} J_p(d) \right) = \inf_{d \in \mathcal{D}} \left(\sup_{p \in [0,1]} J_p(d) \right) \tag{4.18}$$

and

$$\sup_{p \in [0,1]} \left(\inf_{\delta \in \mathcal{D}^\star} J_p^\star(\delta) \right) = \inf_{\delta \in \mathcal{D}^\star} \left(\sup_{p \in [0,1]} J_p^\star(\delta) \right). \tag{4.19}$$

When this occurs we shall then say that the minimax property holds in pure and randomized policies, respectively.

It is worth pointing out that the equalities

$$\inf_{\delta \in \mathcal{D}^\star} \left(\sup_{p \in [0,1]} J_p^\star(\delta) \right) \leq \inf_{d \in \mathcal{D}} \left(\sup_{p \in [0,1]} J_p(d) \right) \tag{4.20}$$

and

$$\sup_{p \in [0,1]} \left(\inf_{\delta \in \mathcal{D}^\star} J_p^\star(\delta) \right) = \sup_{p \in [0,1]} \left(\inf_{d \in \mathcal{D}} J_p(d) \right) \tag{4.21}$$

always hold; the latter is a rewrite of (4.6) and is a simple consequence of Theorem 3.4.1. As we combine these observations with (4.17) we conclude that

$$\sup_{p \in [0,1]} \left(\inf_{d \in \mathcal{D}} J_p(d) \right) \leq \inf_{\delta \in \mathcal{D}^\star} \left(\sup_{p \in [0,1]} J_p^\star(\delta) \right) \leq \inf_{d \in \mathcal{D}} \left(\sup_{p \in [0,1]} J_p(d) \right). \tag{4.22}$$

Thus, if (4.18) happens to be true, then (4.19) necessarily holds – Put differently, the minimax property in pure policies is more difficult to achieve than the minimax property in randomized strategies. This disparity will become apparent in the discussion of the Minimax Theorem given in Section 4.5, opening the possibility that we may have to resort to randomized tests (at least in some situations) in order to achieve the minimax equality.

The structure of the worst-case costs (4.9) and (4.11) A little more can be said concerning the auxiliary costs (4.9) and (4.11): For each test d in $\mathcal{D}$, we note from (4.7) and (4.8) that

$$\sup_{p \in [0,1]} J_p(d) = \max_{p \in [0,1]} J_p(d) = \max\{J_0(d), J_1(d)\} \tag{4.23}$$

with the supremum achieved at either $p = 0$ or $p = 1$. Also, $J_0(d)$ and $J_1(d)$ can be given probabilistic interpretations as the conditional interpretations

$$J_0(d) = \mathbb{E}_p \left[C(H, d(\boldsymbol{Y})) | H = 0 \right] \tag{4.24}$$

and
$$J_1(d) = \mathbb{E}_p\left[C(H, d(\boldsymbol{Y}))|H = 1\right] \tag{4.25}$$

with p arbitrary in $[0, 1]$. Similarly, for each randomized strategy δ in $\mathcal{D}^\star$, we have

$$\sup_{p \in [0,1]} J_p^\star(\delta) = \max_{p \in [0,1]} J_p^\star(\delta) = \max\{J_0^\star(\delta), J_1^\star(\delta)\} \tag{4.26}$$

with the supremum achieved at either $p = 0$ or $p = 1$ with probabilistic interpretations

$$J_0^\star(\delta) = \mathbb{E}_{\delta,p}\left[C(H, D)|H = 0\right] \tag{4.27}$$

and

$$J_1^\star(\delta) = \mathbb{E}_{\delta,p}\left[C(H, D)|H = 1\right] \tag{4.28}$$

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

4.5   The minimax equality

The main result concerning the minimax formulation for the binary hypothesis testing problem is summarized in the following special case of the Minimax Theorem from Statistical Decision Theory; see [**?**, Thm. 1, p. 82] for a discussion in a more general setting.

Theorem 4.5.1 *Assume $\Gamma_h > 0$ for all $h = 0, 1$. Under Condition (**A.1**), the minimax equality*

$$\sup_{p \in [0,1]} \left(\inf_{\delta \in \mathcal{D}^\star} J_p^\star(\delta)\right) = \inf_{\delta \in \mathcal{D}^\star} \left(\sup_{p \in [0,1]} J_p^\star(\delta)\right) \tag{4.29}$$

holds in randomized strategies.

In Section 4.6 we present an analysis of the minimax equality which exploits the specific structure of the binary hypothesis problem as reflected through the properties of the value function: As pointed out earlier, there always exists p_m in $[0, 1]$ such that (4.13) holds. From the concavity of the value function it follows that the set of maximizers

$$I_\mathrm{m} \equiv \left\{p_\mathrm{m} \in [0, 1] : V(p_\mathrm{m}) = \max_{p \in [0,1]} V(p)\right\}$$

is a closed interval in $[0, 1]$. The set I_m will often be reduced to a singleton, in which case the value function admits a unique (isolated) maximizer. Four situations can occur depending on the location of I_m and on the smoothness of $p \to V(p)$ at the maximum. In each case we establish a minimax equality and identify the minimax strategy. Throughout we still use $d^\star(p_\mathrm{m})$ to denote the Bayesian test for the selected value p_m in I_m, so that

$$V(p_\mathrm{m}) = J_{p_\mathrm{m}}(d^\star(p_\mathrm{m})) = \min_{d \in \mathcal{D}} J_{p_\mathrm{m}}(d). \tag{4.30}$$

From the discussion of Section 4.4 we see that (4.29) will hold if we can establish the reverse inequality to (4.17), namely

$$\inf_{\delta \in \mathcal{D}^\star} \left(\sup_{p \in [0,1]} J_p^\star(\delta) \right) \leq \sup_{p \in [0,1]} \left(\inf_{\delta \in \mathcal{D}^\star} J_p^\star(\delta) \right). \tag{4.31}$$

Recall that (4.31) will automatically hold if we show the stronger inequality

$$\inf_{d \in \mathcal{D}} \left(\sup_{p \in [0,1]} J_p(d) \right) \leq \sup_{p \in [0,1]} \left(\inf_{d \in \mathcal{D}} J_p(d) \right). \tag{4.32}$$

In the first three cases we show in effect that

$$\inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right) \leq \max_{p \in [0,1]} \left(\min_{d \in \mathcal{D}} J_p(d) \right). \tag{4.33}$$

4.6 A proof of Theorem 4.5.1

We start with the boundary cases $p_\mathrm{m} = 0$ and $p_\mathrm{m} = 1$.

Case 1: Assume $p_\mathrm{m} = 0$ – Thus, $\max_{p \in [0,1]} V(p) = V(0) = J_0(d^\star(0))$. By concavity we have $\frac{d^+}{dp} V(p) \Big|_{p=0} \leq 0$ with the mapping $V : [0, 1] \to \mathbb{R}$ being decreasing. But the straight line $p \to J_p(d^\star(0))$ is tangent to the value function $V : [0, 1] \to \mathbb{R}$ at $p = 0$, whence

$$\frac{d^+}{dp} V(p) \Big|_{p=0} = \frac{d}{dp} J_p(d^\star(0)) \Big|_{p=0} \leq 0.$$

The mapping $p \to J_p(d^\star(0))$ being affine, its derivative is therefore constant with

$$\frac{d}{dp} J_p(d^\star(0)) = \frac{d}{dp} J_p(d^\star(0)) \Big|_{p=0} \leq 0, \quad p \in [0, 1]$$

and the mapping $p \to J_p(d^\star(0))$ is also decreasing on $[0,1]$. This leads to

$$J_0(d^\star(0)) = \max_{p \in [0,1]} J_p(d^\star(0)).$$

With this in mind we get

$$
\begin{aligned}
\max_{p \in [0,1]} \left(\min_{d \in \mathcal{D}} J_p(d) \right) &= \max_{p \in [0,1]} V(p) \\
&= V(0) \\
&= J_0(d^\star(0)) \\
&= \max_{p \in [0,1]} J_p(d^\star(0)). \quad\quad (4.34)
\end{aligned}
$$

The desired inequality (4.32) (hence (4.31)) is now immediate from (4.34) as we note that

$$\max_{p \in [0,1]} J_p(d^\star(0)) \geq \inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right).$$

But the minimax equality being now established in pure strategies, we conclude from the discussion that

$$\max_{p \in [0,1]} J_p(d^\star(0)) = \inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right).$$

This shows that $d_{\mathrm{m}}^\star$ can be taken to be $d^\star(0)$.   ■

Case 2: Assume $p_{\mathrm{m}} = 1$ –   The proof is as in Case 1 *mutatis mutandis*, and is left as an exercise. Again, the minimax equality holds in pure strategies and $d_{\mathrm{m}}^\star$ can be taken to be $d^\star(1)$.   ■

We now turn to cases when p_{m} is selected in $(0, 1)$.

Case 3: Assume that p_{m} in an element of $(0, 1)$ and $p \to V(p)$ is differentiable at $p = p_{\mathrm{m}}$ –   It is plain that $\left. \frac{d}{dp} V(p) \right|_{p=p_{\mathrm{m}}} = 0$ since p_{m} is an interior point by assumption. By concavity the mapping $p \to J_p(d^\star(p_{\mathrm{m}}))$ is tangent to the value function $V : [0, 1] \to \mathbb{R}$ at $p = p_{\mathrm{m}}$, whence

$$\left. \frac{d}{dp} V(p) \right|_{p=p_{\mathrm{m}}} = \left. \frac{d}{dp} J_p(d^\star(p_{\mathrm{m}})) \right|_{p=p_{\mathrm{m}}} = 0.$$

The mapping $p \to J_p(d^\star(p_{\mathrm{m}}))$ being affine, its derivative is constant and given by

$$\frac{d}{dp} J_p(d^\star(p_{\mathrm{m}})) = \frac{d}{dp} J_p(d^\star(p_{\mathrm{m}}))\Big|_{p=0} = 0, \quad p \in [0,1].$$

Therefore, the mapping $p \to J_p(d^\star(p_{\mathrm{m}}))$ is constant on $[0,1]$, and the equality $J_0(d^\star(p_{\mathrm{m}})) = J_1(d^\star(p_{\mathrm{m}}))$ holds. It follows from the first equality in (4.30) that

$$V(p_{\mathrm{m}}) = J_p(d^\star(p_{\mathrm{m}})) = \max_{p \in [0,1]} J_p(d^\star(p_{\mathrm{m}})), \quad p \in [0,1]. \tag{4.35}$$

On the other hand, it is plain that

$$
\begin{aligned}
\inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right) &\leq \max_{p \in [0,1]} J_p(d^\star(p_{\mathrm{m}})) \\
&= J_{p_{\mathrm{m}}}(d^\star(p_{\mathrm{m}})) \\
&= \min_{d \in \mathcal{D}} J_{p_{\mathrm{m}}}(d) \\
&\leq \inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right) \tag{4.36}
\end{aligned}
$$

as we use the second equality in (4.35) with $p = p_{\mathrm{m}}$, and then apply the second equality in (4.30). The inequality (4.32) (hence (4.31)) is now a straightforward consequence of (4.36).

Leveraging the fact that the minimax equality is now known to hold in pure strategies, we conclude from the discussion that

$$\max_{p \in [0,1]} J_p(d^\star(p_{\mathrm{m}})) = \inf_{d \in \mathcal{D}} \left(\max_{p \in [0,1]} J_p(d) \right),$$

and $d_{\mathrm{m}}^\star$ can therefore be taken to be $d^\star(p_{\mathrm{m}})$. ∎

Case 4: Assume that $I_{\mathrm{m}} = \{p_{\mathrm{m}}\} \subseteq (0,1)$ but $p \to V(p)$ is not differentiable at $p = p_{\mathrm{m}}$ – Under such assumptions we must have

$$a_+ \equiv \frac{d^+}{dp} V(p)\Big|_{p=p_{\mathrm{m}}} < \frac{d^-}{dp} V(p)\Big|_{p=p_{\mathrm{m}}} \equiv a_-$$

by concavity with either $a_+ < 0 \leq a_-$ or $a_+ \leq 0 < a_-$. We continue the discussion under the assumption $a_+ < 0 \leq a_-$; the case $a_+ \leq 0 < a_-$ proceeds along similar lines, and is therefore omitted.

Recall that $p \to V(p)$ is defined as the envelope of a family of affine functions. Thus, under the non-differentiability assumption at $p = p_{\mathrm{m}}$, concavity guarantees that there exist two pure strategies, say $d_-, d_+ : \mathbb{R}^k \to \{0, 1\}$, such that $V(p_{\mathrm{m}}) = J_{p_{\mathrm{m}}}(d_-)$ and $V(p_{\mathrm{m}}) = J_{p_{\mathrm{m}}}(d_+)$ (because p_{m} is a maximum) while the straight lines $p \to J_p(d_-)$ and $p \to J_p(d_+)$ are both tangent to the value function at $p = p_{\mathrm{m}}$ – These two strategies are distinct. Hence, as discussed in earlier cases, the function $p \to J_p(d_-)$ (resp. $p \to J_p(d_+)$) is an affine function with constant derivative $a_- \geq 0$ (resp. $a_+ < 0$), hence non-decreasing (resp. decreasing). It follows that $J_0(d_-) \leq J_1(d_-)$ and $J_1(d_+) < J_0(d_+)$.

Next we introduce randomized policies $\{\delta_a, \ a \in [0, 1]\}$ obtained by randomizing two pure strategies d_- and d_+. Thus, with each a in $[0, 1]$ consider the randomized policy $\delta_a : \mathbb{R}^k \to [0, 1]$ given by

$$\delta_a = a d_+ + (1 - a) d_-.$$

The relation (3.20) discussed in Section 3.5 applies, yielding

$$J_p^\star(\delta_a) = a J_p(d_+) + (1 - a) J_p(d_-), \quad p \in [0, 1].$$

By construction we also note that

$$V(p_{\mathrm{m}}) = J_{p_{\mathrm{m}}}^\star(\delta_a), \quad a \in [0, 1]. \tag{4.37}$$

If a suitable of a, we were to have $p \to J_p^\star(\delta_a)$ *constant* over $[0, 1]$, then the test δ_a would a performance insensitive to the value of p. This requirement (on a) is equivalent to the equality $J_0^\star(\delta_a) = J_1^\star(\delta_a)$, i.e.,

$$a J_0(d_+) + (1 - a) J_0(d_-) = a J_1(d_+) + (1 - a) J_1(d_-).$$

Thus,

$$a \left((J_0(d_+) - J_1(d_+)) + (J_1(d_-) - J_0(d_-)) \right) = J_1(d_-) - J_0(d_-)$$

and solving for a we get

$$a^\star = \frac{J_1(d_-) - J_0(d_-)}{(J_0(d_+) - J_1(d_+)) + (J_1(d_-) - J_0(d_-))}.$$

It is a simple matter to check that $a^\star$ lies in $[0,)$ since $J_1(d_-) - J_0(d_-) \geq 0$ and $J_0(d_+) - J_1(d_+) > 0$ as discussed earlier.

It is now plain that

$$V(p_{\mathrm{m}}) = J_p^\star(\delta_{a^\star}) = \max_{p\in[0,1]} J_p^\star(\delta_{a^\star}), \quad p \in [0,1]. \tag{4.38}$$

Therefore, as in the discussion for Case 3, we have

$$
\begin{aligned}
\inf_{\delta\in\mathcal{D}^\star}\left(\max_{p\in[0,1]} J_p^\star(\delta)\right) &\leq \max_{p\in[0,1]} J_p^\star(\delta_{a^\star}) \\
&= V(p_{\mathrm{m}}) \\
&= \inf_{\delta\in\mathcal{D}^\star} J_p^\star(\delta) \\
&\leq \max_{p\in[0,1]}\left(\inf_{\delta\in\mathcal{D}^\star} J_p^\star(\delta)\right) \tag{4.39}
\end{aligned}
$$

and the desired inequality (4.31) is established.

The minimax equality now holds in mixed strategies, whence

$$\max_{p\in[0,1]} J_p^\star(\delta_{a^\star}) = \inf_{\delta\in\mathcal{D}^\star}\left(\max_{p\in[0,1]} J_p^\star(\delta)\right)$$

by virtue of (4.39). The minimax strategy is a randomized strategy $\delta_{\mathrm{m}}^\star$ which is identified as $\delta^\star(p_{\mathrm{m}})$. Note that $\delta^\star(p_{\mathrm{m}})$ is also a (randomized) Bayesian policy for the least favorable prior. ∎

We summarize these findings in the following corollary to Theorem 4.5.1.

Corollary 4.6.1 *Assume $\Gamma_h > 0$ for all $h = 0, 1$. Under Condition (**A.1**), the minimax equality (4.29) holds in randomized strategies. Moreover, the minimax strategy always exists and can be interpreted as a (possibly randomized) Bayesian test under the least favorable prior p_{m}.*

4.7 The minimax equation

The discussion of Section 4.5 shows that finding minimax tests passes through the evaluation of the value function $p \to V(p)$ and its maximizing set I_{m}. As

simple examples already suggest in later sections, this evaluation may not always
be straightforward to carry. Moreover, once an expression for the value function
becomes available, finding its maximizers may turn out to be rather cumbersome.
However, this two-step approach can be bypassed when I_{m} contains an interior
point p_{m} at which the value function is differentiable, in which case the minimax
test is given by the Bayesian test $d^{\star}(p_{\mathrm{m}})$. Instead a simple characterization of p_{m}
is achieved through the so-called *Minimax Equation.*

Lemma 4.7.1 *Assume $\Gamma_h > 0$ for all $h = 0, 1$, and that p_{m} is an element of $(0, 1)$
and $p \to V(p)$ is differentiable at $p = p_{\mathrm{m}}$. Under Condition (**A.1**), p_{m} can be
characterized through the Minimax Equation*

$$C(1, 1) - C(0, 0) = \Gamma_0 \cdot P_F(d^{\star}(p_{\mathrm{m}})) - \Gamma_1 \cdot P_M(d^{\star}(p_{\mathrm{m}})). \qquad (4.40)$$

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

$$P_F(d^{\star}(p_{\mathrm{m}})) = P_M(d^{\star}(p_{\mathrm{m}})). \qquad (4.41)$$

Proof. Fix p in $[0, 1]$. Upon specializing (4.7) to the test $d^{\star}(p)$, we get

$$
\begin{aligned}
J_\alpha(d^{\star}(p)) \;=\;& \alpha C(1, 1) + (1 - \alpha)C(0, 0) \\
& + \Gamma_0(1 - \alpha) \cdot P_F(d^{\star}(p)) + \Gamma_1 \alpha \cdot P_M(d^{\star}(p)) \qquad (4.42)
\end{aligned}
$$

with α in $[0, 1]$ and the mapping $\alpha \to J_\alpha(d^{\star}(p))$ is therefore affine in the variable
α on the interval $[0, 1]$ Therefore, the graph of the mapping $\alpha \to J_\alpha(d^{\star}(p))$ is a
straight line; its slope is given by

$$\frac{d}{d\alpha} J_\alpha(d^{\star}(p)) = C(1, 1) - C(0, 0) + \Gamma_1 \cdot P_M(d^{\star}(p)) - \Gamma_0 \cdot P_F(d^{\star}(p)). \quad (4.43)$$

By its definition, the Bayesian cost satisfies

$$V(\alpha) \le J_\alpha(d), \qquad \begin{matrix} d \in \mathcal{D} \\ \alpha \in [0, \end{matrix}$$

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

$$V(p) = J_p(d^{\star}(p))$$

while

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

With p in $(0,1)$, if the concave mapping $\alpha \to V(\alpha)$ is *differentiable* at $\alpha = p$, then the straight line $\alpha \to J_\alpha(d^\star(p))$ will be a tangent to the mapping $\alpha \to V(\alpha)$ at $\alpha = p$ – This is a consequence of the concavity established in Lemma 4.2.1. Thus,

$$\frac{d}{d\alpha}V(\alpha)\Big|_{\alpha=p} = \frac{d}{d\alpha}J_\alpha(d^\star(p))\Big|_{\alpha=p}. \tag{4.44}$$

In particular, if p_{m} is an element of $(0,1)$ and the mapping $\alpha \to V(\alpha)$ is differentiable at $\alpha = p_{\mathrm{m}}$, then

$$\frac{d}{d\alpha}V(\alpha)\Big|_{\alpha=p_{\mathrm{m}}} = \frac{d}{d\alpha}J_\alpha(d^\star(p_{\mathrm{m}}))\Big|_{\alpha=p_{\mathrm{m}}}. \tag{4.45}$$

But the interior point p_{m} being a maximum for the function $\alpha \to V(\alpha)$, we must have

$$\frac{d}{d\alpha}V(\alpha)\Big|_{\alpha=p_{\mathrm{m}}} = 0,$$

whence

$$\frac{d}{d\alpha}J_\alpha(d^\star(p_{\mathrm{m}}))\Big|_{\alpha=p_{\mathrm{m}}} = 0.$$

The equation (4.40) now follows from (4.43). ∎

Obviously this analysis does not cover the cases when (i) $p_{\mathrm{m}} = 0$, (ii) $p_{\mathrm{m}} = 1$ and (iii) p_{m} is an element of $(0,1)$ but the mapping $\alpha \to V(\alpha)$ is not differentiable at $\alpha = p_{\mathrm{m}}$.

4.8 The Gaussian Case

The setting is that of Section 2.4 to which we refer the reader for the notation. As shown there, for every $\eta > 0$ 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 $d^\star(p) = Lrt_{\eta(p)}$ and the expression (4.7) yields

$$
\begin{aligned}
V(p) &= J_p(d^\star(p)) \\
&= pC(1, 1) + (1 - p)C(0, 0) \\
&\quad + \Gamma_0(1 - p) \cdot \left(1 - \Phi \left(\frac{\log \eta(p) + \frac{1}{2}d^2}{d} \right) \right) \\
&\quad + \Gamma_1 p \cdot \Phi \left(\frac{\log \eta(p) - \frac{1}{2}d^2}{d} \right).
\end{aligned}
\tag{4.46}
$$

The boundary cases $p = 0$ is easily recovered upon formally substituting this value in the expression (4.46). The Minimax Equation (4.40) takes the form

$$
\begin{aligned}
&C(1, 1) - C(0, 0) \\
&= \Gamma_1 \Phi \left(\frac{\log \eta(p_\mathrm{m}) - \frac{1}{2}d^2}{d} \right) - \Gamma_0 \left(1 - \Phi \left(\frac{\log \eta(p_\mathrm{m}) + \frac{1}{2}d^2}{d} \right) \right).
\end{aligned}
\tag{4.47}
$$

Probability of error – Simplifications occur since $C(0, 0) = C(1, 1) = 0$ and $\Gamma_0 = \Gamma_1 = 1$: The expression (4.46) becomes

$$
V(p) = (1 - p) \cdot \left(1 - \Phi \left(\frac{\frac{1-p}{p} + \frac{1}{2}d^2}{d} \right) \right) + p \cdot \Phi \left(\frac{\log \frac{1-p}{p} - \frac{1}{2}d^2}{d} \right),
$$

and the Minimax Equation (4.47) reduces to

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

It is easy to see that this requires $\log \eta(p_\mathrm{m}) = 0$ so that $p_\mathrm{m} = \frac{1}{2}$ (indeed in $(0, 1)$), an intuitively satisfying conclusion! Moreover, the minimax test is given by $d_\mathrm{m}^\star = d(\frac{1}{2})$.

4.9 The Bernoulli case

The setting is that of Section 2.5 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 (2.31) and (2.32) 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 (2.31) yields

$$
P_F(d_\eta) \;=\; \begin{cases} 1 & \text{if } 0 < \eta \le \frac{a_1}{a_0} \\[2mm] 1 - a_0 & \text{if } \frac{a_1}{a_0} < \eta \le \frac{1-a_1}{1-a_0} \\[2mm] 0 & \text{if } \frac{1-a_1}{1-a_0} < \eta. \end{cases}
\tag{4.48}
$$

Similarly, using (2.32) we find

$$
P_M(d_\eta) \;=\; \begin{cases} 0 & \text{if } 0 < \eta \le \frac{a_1}{a_0} \\[2mm] a_1 & \text{if } \frac{a_1}{a_0} < \eta \le \frac{1-a_1}{1-a_0} \\[2mm] 1 & \text{if } \frac{1-a_1}{1-a_0} < \eta. \end{cases}
\tag{4.49}
$$

Thus, for each p in $[0, 1]$, we see from (4.7) 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 $(0, \frac{a_1}{a_0}]$,

$$
\begin{aligned}
J_p(d_\eta) &= pC(1,1) + (1-p)C(0,0) + \Gamma_0(1-p) \\
&= pC(1,1) + (1-p)C(0,1).
\end{aligned}
\tag{4.50}
$$

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

$$
\begin{aligned}
J_p(d_\eta) \\
&= pC(1,1) + (1-p)C(0,0) + \Gamma_0(1-p) \cdot (1-a_0) + \Gamma_1 p \cdot a_1 \\
&= pC(1,1) + (1-p)C(0,1) + \Gamma_1 p \cdot a_1 - \Gamma_0(1-p) \cdot a_0 \\
&= p\left(C(1,1) + \Gamma_1 a_1\right) + (1-p)\left(C(0,1) - \Gamma_0 a_0\right).
\end{aligned}
\tag{4.51}
$$

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

$$
\begin{aligned}
J_p(d_\eta) &= pC(1,1) + (1-p)C(0,0) + \Gamma_1 p \\
&= pC(1,0) + (1-p)C(0,0). \tag{4.52}
\end{aligned}
$$

Recall that

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

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

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

and

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

has a unique solution in $(0,1)$. These 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, and yield

$V(p)$

$$
= \begin{cases}
pC(1,0) + (1-p)C(0,0) & \text{if } p \in (0, p_-] \\[2ex]
p\left(C(1,1) + \Gamma_1 a_1\right) + (1-p)\left(C(0,1) - \Gamma_0 a_0\right) & \text{if } p \in (p_-, p_+] \\[2ex]
pC(1,1) + (1-p)C(0,1) & \text{if } p \in (p_+, 1).
\end{cases}
$$

It is plain that the function $V : [0,1] \to \mathbb{R}$ 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. The function $V : [0,1] \to \mathbb{R}$ is differentiable everywhere except at these kinks. However the maximum occurs at one of these points so that $p_\mathrm{m} \in \{p_-, p_+\}$.

Probability of error – In that case we find that

$$
V(p) \;=\; \begin{cases} p & \text{if } p \in (0, p_-] \\[2mm] pa_1 + (1-p)(1-a_0) & \text{if } p \in (p_-, p_+] \\[2mm] 1-p & \text{if } p \in (p_+, 1) \end{cases} \tag{4.53}
$$

with

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

and

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

It is a simple matter to check that $V(p_\pm -) = V(p_\pm +)$, establishing continuity at the kink points. As we compare $V(p_-)$ and $V(p_+)$, we readily conclude that $p_{\mathrm m} = p_-$ (resp. $p_{\mathrm m} = p_+$) iff $1 - p_+ < p_-$ (resp. $p_- < 1 - p_+$) iff $a_0 + a_1 < 1$ (resp. $1 < a_0 + a_1$). The minimax cost is then given by

$$
V_{\mathrm m} = \begin{cases} p_- = \frac{1-a_0}{(1-a_1)+(1-a_0)} & \text{if } a_0 + a_1 < 1 \\[2mm] 1 - p_+ = \frac{a_1}{a_1+a_0} & \text{if } 1 < a_0 + a_1 \end{cases}
$$

Minimax strategy is necessarily randomized and is given by

$$
\delta_a = a d_+ + (1-a) d_-
$$

with the pure tests $d_-, d_+ : \mathbb{R} \to \{0,1\}$ given by

4.10 A proof of Lemma 4.2.1

The proof proceeds in several stages. We start with the fact that

$$
V(p) = \inf_{d \in \mathcal{D}} J_p(d), \quad p \in [0,1].
$$

Values at the boundary points – Consider a test d in $\mathcal{D}$. With $p = 0$ and $p = 1$ in (4.7) we get

$$
J_0(d) = C(0,0) + \Gamma_0 P_F(d)
$$

and

$$J_1(d) = C(1,1) + \Gamma_1 P_M(d).$$

Using the conditions $\Gamma_0 > 0$ and $\Gamma_1 > 0$, we conclude that

$$V(0) = \inf_{d \in \mathcal{D}} J_0(d) = C(0,0) + \Gamma_0 \cdot \inf_{d \in \mathcal{D}} P_F(d)$$

and

$$V(1) = \inf_{d \in \mathcal{D}} J_0(d) = C(1,1) + \Gamma_1 \cdot \inf_{d \in \mathcal{D}} P_M(d).$$

However, $P_F(d_\infty) = 0$ for the test $d_\infty : \mathbb{R}^k \to \{0,1\}$ which always selects the null hypothesis ($H = 0$) while $P_M(d_0) = 0$ for the test $d_0 : \mathbb{R}^k \to \{0,1\}$ which always selects the alternative ($H = 1$); see Section 1.6. It follows that $\inf_{d \in \mathcal{D}} P_F(d) = 0$ and $\inf_{d \in \mathcal{D}} P_M(d) = 0$, whence $V(0) = C(0,0)$ and $V(1) = C(1,1)$. ∎

Concavity on $[0,1]$ **and continuity on** $(0,1)$ **–**   Once the test d is selected, the probabilities $P_F(d)$ and $P_M(d)$ appearing in (4.7) do *not* depend on p, and are determined only through F_0 and F_1. Thus, the mapping $p \to J_p(d)$ is affine, hence concave in p. As a result, the mapping $V : [0,1] \to \mathbb{R}$ 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 $V : [0,1] \to \mathbb{R}$ is continuous on $(0,1)$ by virtue of Fact 9.4.2. ∎

Continuity at the boundary points –   We now turn to showing that the mapping $V : [0,1] \to \mathbb{R}$ 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.

For notational convenience here and below we write

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

Recall that $V(0) = C(0,0)$ by the first part of the proof. Thus, for each p in $(0,1]$ we get from the definition of $V(p)$ that

$$V(p) - V(0) \;\; = \;\; p\left(C(1,1) - C(0,0)\right) + \Delta(p) \tag{4.54}$$

by virtue of (4.7). The continuity of the mapping $V : [0, 1] \to \mathbb{R}$ at $p = 0$ is therefore equivalent to

$$\lim_{p \to 0} \Delta(p) = 0. \tag{4.55}$$

For any fixed p in $(0, 1]$, the conditions $\Gamma_0 > 0$ and $\Gamma_1 > 0$ yield the inequalities

$$0 \leq \Delta(p) \leq \Gamma_1 p \tag{4.56}$$

since under the test d_∞ (introduced earlier in the proof) we have $P_F(d_\infty) = 0$ and $P_M(d_\infty = 1$. The conclusion (4.55) is now immediate. ∎

Differentiability – The existence *and* finiteness of the right-derivative and left-derivative on the open interval $(0, 1)$ are simple consequences of Fact 9.4.4. The same argument also shows that the right-derivative (resp. left-derivative) does exist at $p = 0$ (resp. $p = 1$); however it may not necessarily be finite.

Instead, we provide a direct argument to show the existence and finiteness of the right-derivative (resp. left-derivative) at $p = 0$ (resp. $p = 1$). We carry out the discussion only for $p = 0$ as the case $p = 1$ is similar: For each p in $(0, 1]$, we note that

$$\frac{V(p) - V(0)}{p} = C(1, 1) - C(0, 0) + \frac{\Delta(p)}{p} \tag{4.57}$$

with

$$\frac{\Delta(p)}{p} = \inf_{d \in \mathcal{D}} \left(\Gamma_0 \left(\frac{1}{p} - 1 \right) \cdot P_F(d) + \Gamma_1 \cdot P_M(d) \right).$$

This last expression shows that $p \to \frac{\Delta(p)}{p}$ is decreasing on $(0, 1]$, whence the limit $\lim_{p \downarrow 0} \frac{\Delta(p)}{p}$ always exists. This limit is finite by virtue of the bounds

$$0 \leq \frac{\Delta(p)}{p} \leq \Gamma_1, \quad p \in (0, 1]$$

which are inherited from the earlier bounds (4.56). This shows the existence of a finite right-derivative at $p = 0$. ∎

4.11   Exercises

4.1.
Let I denote an interval of $\mathbb{R}$, not necessarily finite, closed or open, and let A be an arbitrary index set. For each α in A, let $f_\alpha : I \to \mathbb{R}$ be a concave function. With the function $g : I \to \mathbb{R}$ defined by

$$g(x) = \inf\left(f_\alpha(x) : \ \alpha \in A\right), \quad x \in I$$

show that the mapping $g : I \to \mathbb{R}$ is concave.

4.2.
With $h > 0$ show that the equation

$$\Phi(x - h) + \Phi(x + h) = 1, \quad x \in \mathbb{R}$$

has a unique solution $x = 0$.

4.3.

4.12   References

Chapter 5

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. The Neyman-Pearson formulation of the binary hypothesis problem given next constitutes an approach to handle such situations.

5.1   A constrained optimization problem

Fix α in $(0, 1)$ (the limiting case $\alpha = 0$ being of little practical interest). Let $\mathcal{D}_\alpha$ denote the collection of admissible tests in $\mathcal{D}$ of *size* at most α, namely

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

The Neyman-Pearson formulation is based on solving the constrained optimization problem NP_α where

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

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

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

Such a test $d_\mathrm{NP}(\alpha)$, when it exists, is called a *Neyman–Pearson* test of size α, or

alternatively, an α-level Neyman–Pearson decision rule. Such decision rules may not be unique. Following the accepted terminology, its *power* $\beta(\alpha)$ is given by

$$\beta(\alpha) \equiv P_{\mathrm{D}}(d_{\mathrm{NP}}(\alpha)) = \sup_{d \in \mathcal{D}_\alpha} P_{\mathrm{D}}(d).$$

When reformulated as

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

the constrained optimization problem NP_α can be solved by standard Lagrangian arguments which are outlined in the next section. Throughout we assume that Condition (**A.1**) holds.

5.2   The Lagrangian arguments

Fix α in $(0,1)$. For each $\lambda \geq 0$ consider the Lagrangian functional $J_\lambda : \mathcal{D} \to \mathbb{R}$ given by

$$J_\lambda(d) = P_{\mathrm{M}}(d) + \lambda \left(P_{\mathrm{F}}(d) - \alpha \right), \quad d \in \mathcal{D}.$$

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

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

Solving LP_λ amounts to finding a test $d_\lambda^\star$ in $\mathcal{D}$ such that

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

Solving the Lagrangian problem LP_λ Fix $\lambda > 0$. For any test d in $\mathcal{D}$, we note that

$$
\begin{aligned}
J_\lambda(d) &= \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right] + \lambda \left(\mathbb{P}\left[d(\boldsymbol{Y}) = 1 | H = 0\right] - \alpha \right) \\
&= \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right] + \lambda \left(1 - \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 0\right] - \alpha \right) \\
&= \lambda(1 - \alpha) + \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 1\right] - \lambda \mathbb{P}\left[d(\boldsymbol{Y}) = 0 | H = 0\right] \\
&= \lambda(1 - \alpha) + \int_{C(d)} h_\lambda(\boldsymbol{y}) dF(\boldsymbol{y}) \tag{5.1}
\end{aligned}
$$

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

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

By the comparison arguments used in the proof of Theorem 2.2.1, the Lagrangian problem LP_λ is easily seen to be solved by the test $d_\lambda^\star : \mathbb{R}^k \to \{0, 1\}$ given by

$$d_\lambda^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad h_\lambda(\boldsymbol{y}) < 0, \tag{5.2}$$

or equivalently,

$$d_\lambda^\star(\boldsymbol{y}) = 0 \quad \text{iff} \quad f_1(\boldsymbol{y}) < \lambda f_0(\boldsymbol{y}). \tag{5.3}$$

Note that in the notation associated with the definition (**??**) we have $d_\lambda^\star = d_\lambda$. ∎

Meeting the constraint   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.,

$$P_{\mathrm{F}}(d_{\lambda(\alpha)}) = \alpha. \tag{5.4}$$

If such value $\lambda(\alpha)$ were to exist, then the optimality $d_{\lambda(\alpha)}$ implies

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

or equivalently,

$$P_{\mathrm{M}}(d_{\lambda(\alpha)}) \leq P_{\mathrm{M}}(d) + \lambda(\alpha)\left(P_{\mathrm{F}}(d) - \alpha\right), \quad d \in \mathcal{D}.$$

Consequently, for every test d *in* $\mathcal{D}_\alpha$ (and not merely in $\mathcal{D}$), it follows that

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

since then $P_{\mathrm{M}}(d) \leq \alpha$. The test $d_{\lambda(\alpha)}$ is a test in $\mathcal{D}_\alpha$ by virtue of (5.4), hence it solves NP_α – In other words, $d_{\mathrm{NP}}(\alpha)$ can be taken to be $d_{\lambda(\alpha)}$. ∎

A difficulty   The Lagrangian argument hinges upon the possibility of finding a value $\lambda(\alpha)$ of the Lagrange multiplier such that $P_{\mathrm{F}}(d_{\lambda(\alpha)}) = \alpha$. Unfortunately, this may not be always possible, unless additional assumptions are imposed. To see how this may indeed happen, note that

$$\begin{aligned} P_{\mathrm{F}}(d_\lambda) &= \mathbb{P}\left[d_\lambda(\boldsymbol{Y}) = 1 | H = 0\right] \\ &= \mathbb{P}\left[f_1(\boldsymbol{Y}) \geq \lambda f_0(\boldsymbol{Y}) | H = 0\right], \quad \lambda > 0. \end{aligned} \tag{5.5}$$

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

$$\lim_{\lambda\uparrow\lambda^\star} P_{\mathrm{F}}(d_\lambda) < \alpha < \lim_{\lambda\downarrow\lambda^\star} P_{\mathrm{F}}(d_\lambda).$$

In Section **??** we illustrate such situations on simple examples that involve discrete rvs. 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 (properly defined over randomized strategies) has a solution within the set of non-randomized policies $\mathcal{D}$.

5.3   The Neyman-Pearson Lemma

The discussion of Section 5.2 suggests the need to consider an extended version of the Neyman-Pearson formulation where randomized strategies are allowed.

Fix α in $(0, 1]$. Let $\mathcal{D}_\alpha^\star$ denote the collection of all randomized tests in $\mathcal{D}^\star$ of *size* at most α, namely

$$\mathcal{D}_\alpha^\star = \left\{ \delta \in \mathcal{D}^\star : \ P_{\mathrm{F}}(\delta) \leq \alpha \right\}.$$

The constrained optimization problem $\mathrm{NP}^\star$ is now replaced by the following constrained optimization problem $\mathrm{NP}_\alpha^\star$ where

$$\mathrm{NP}_\alpha^\star : \quad \text{Maximize } P_{\mathrm{D}}(\delta) \text{ over } \delta \text{ in } \mathcal{D}_\alpha^\star.$$

Solving $\mathrm{NP}_\alpha^\star$ amounts to finding a test $\delta_{\mathrm{NP}}(\alpha)$ in $\mathcal{D}_\alpha^\star$ with the property that

$$P_{\mathrm{D}}(\delta) \leq P_{\mathrm{D}}(\delta_{\mathrm{NP}}(\alpha)), \quad \delta \in \mathcal{D}_\alpha^\star.$$

Such a test $\delta_{\mathrm{NP}}(\alpha)$, when it exists, is also called a *Neyman–Pearson* test of size α, or alternatively, an α-level Neyman–Pearson decision rule. It may not be unique.

The existence of the Neyman–Pearson test $\delta_{\mathrm{NP}}(\alpha)$ of size α, its characterization and uniqueness are discussed below through three separate lemmas, known collectively as the Neyman-Pearson Lemma. Proofs are delayed until Section 5.4.

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

$$
\delta^\star(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \eta f_0(\boldsymbol{y}) < f_1(\boldsymbol{y}) \\[2mm] \gamma(\boldsymbol{y}) & \text{if } f_1(\boldsymbol{y}) = \eta f_0(\boldsymbol{y}) \\[2mm] 0 & \text{if } f_1(\boldsymbol{y}) < \eta f_0(\boldsymbol{y}). \end{cases} \tag{5.6}
$$

The inequality discussed next lays the groundwork for identifying the Neyman–Pearson test $\delta_{\mathrm{NP}}(\alpha)$.

Lemma 5.3.1 *For any test $\delta : \mathbb{R}^k \to [0,1]$ in $\mathcal{D}^\star$, the inequality*

$$
P_{\mathrm{D}}(\delta^\star) - P_{\mathrm{D}}(\delta) \geq \eta \left(P_{\mathrm{F}}(\delta^\star) - P_{\mathrm{F}}(\delta) \right) \tag{5.7}
$$

holds where the randomized test $\delta^\star : \mathbb{R}^k \to [0,1]$ in $\mathcal{D}^\star$ is given by (5.7).

If we select $\eta \geq 0$ and $\gamma : \mathbb{R}^k \to [0,1]$ so that $\delta^\star$ satisfies the equality

$$
P_{\mathrm{F}}(\delta^\star) = \alpha, \tag{5.8}
$$

then the inequality (5.7) reads

$$
P_{\mathrm{D}}(\delta^\star) - P_{\mathrm{D}}(\delta) \geq \eta \left(\alpha - P_{\mathrm{F}}(\delta) \right), \quad \delta \in \mathcal{D}^\star. \tag{5.9}
$$

For any test $\delta : \mathbb{R}^k \to [0,1]$ in $\mathcal{D}_\alpha^\star$, we then conclude that

$$
P_{\mathrm{D}}(\delta^\star) - P_{\mathrm{D}}(\delta) \geq \eta \left(\alpha - P_{\mathrm{F}}(\delta) \right) \geq 0 \tag{5.10}
$$

since $P_{\mathrm{F}}(\delta) \leq \alpha$. In other words,

$$
P_{\mathrm{D}}(\delta) \leq P_{\mathrm{D}}(\delta^\star), \quad \delta \in \mathcal{D}_\alpha^\star
$$

and the test $\delta^\star$ solves the constrained problem $\mathrm{NP}_\alpha^\star$.

We now show that the parameter $\eta \geq 0$ and the Borel mapping $\gamma : \mathbb{R}^k \to [0,1]$ can indeed be selected so that a test $\delta^\star$ of the form (5.7) indeed satisfies (5.8).

Lemma 5.3.2 *For every α in $(0,1]$ it is always possible to select $\eta \geq 0$ and a Borel mapping $\gamma : \mathbb{R}^k \to [0,1]$ in (5.6) so that (5.8) holds.*

Finally uniqueness is shown to hold in the following sense.

Lemma 5.3.3 *For every α in $(0,1]$, if $\delta_{\mathrm{NP}}(\alpha)$ is a Neyman–Pearson test (possibly in $\mathcal{D}^\star$) of size α, then it necessarily holds that*

$$
\mathbb{P}\left[\delta_{\mathrm{NP}}(\alpha)(\boldsymbol{Y}) = \delta^\star(\boldsymbol{Y}) | H = h \right] = 1, \quad h = 0, 1 \tag{5.11}
$$

where the test $\delta^\star$ is given by (5.6) with $\eta \geq 0$ and Borel mapping $\gamma : \mathbb{R}^k \to [0,1]$ selected so that (5.8) holds.

5.4   Proofs

Throughout the discussion α is given in $(0, 1]$ and held fixed.

A proof of Lemma 5.3.1   Let $\delta : \mathbb{R}^k \rightarrow [0, 1]$ be an arbitrary test in $\mathcal{D}^\star$. As discussed in Section 3.3 recall that

$$P_\mathrm{F}(\delta) = \int_{\mathbb{R}^k} \delta(\boldsymbol{y}) f_0(\boldsymbol{y}) dF(\boldsymbol{y}) \quad \text{and} \quad P_\mathrm{F}(\delta^\star) = \int_{\mathbb{R}^k} \delta^\star(\boldsymbol{y}) f_0(\boldsymbol{y}) dF(\boldsymbol{y}),$$

while

$$P_\mathrm{D}(\delta) = \int_{\mathbb{R}^k} \delta(\boldsymbol{y}) f_1(\boldsymbol{y}) dF(\boldsymbol{y}) \quad \text{and} \quad P_\mathrm{D}(\delta^\star) = \int_{\mathbb{R}^k} \delta^\star(\boldsymbol{y}) f_1(\boldsymbol{y}) dF(\boldsymbol{y}).$$

It follows that

$$
\begin{aligned}
& P_\mathrm{D}(\delta^\star) - P_\mathrm{D}(\delta) - \eta \left(P_\mathrm{F}(\delta^\star) - P_\mathrm{F}(\delta) \right) \\
&= \int_{\mathbb{R}^k} \left(\delta^\star(\boldsymbol{y}) - \delta(\boldsymbol{y}) \right) f_1(\boldsymbol{y}) dF(\boldsymbol{y}) - \eta \int_{\mathbb{R}^k} \left(\delta^\star(\boldsymbol{y}) - \delta(\boldsymbol{y}) \right) f_0(\boldsymbol{y}) dF(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^k} \left(\delta^\star(\boldsymbol{y}) - \delta(\boldsymbol{y}) \right) \left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y}) \right) dF(\boldsymbol{y}) \\
&= \int_{\mathbb{R}^k} P_\eta(\boldsymbol{y}) dF(\boldsymbol{y}) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (5.12)
\end{aligned}
$$

where we have set

$$P_\eta(\boldsymbol{y}) \equiv \left(\delta^\star(\boldsymbol{y}) - \delta(\boldsymbol{y}) \right) \left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y}) \right), \quad\quad \boldsymbol{y} \in \mathbb{R}^k.$$

Direct inspection shows that we always have

$$P_\eta(\boldsymbol{y}) \geq 0, \quad \boldsymbol{y} \in \mathbb{R}^k. \quad\quad\quad\quad (5.13)$$

Obviously, we have $P_\eta(\boldsymbol{y}) = 0$ when $f_1(\boldsymbol{y}) = \eta f_0(\boldsymbol{y})$. When $\eta f_0(\boldsymbol{y}) < f_1(\boldsymbol{y})$, then

$$P_\eta(\boldsymbol{y}) = \left(1 - \delta(\boldsymbol{y}) \right) \left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y}) \right) \geq 0,$$

while when $f_1(\boldsymbol{y}) < \eta f_0(\boldsymbol{y})$, then

$$P_\eta(\boldsymbol{y}) = -\delta(\boldsymbol{y}) \left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y}) \right) \geq 0.$$

It is now plain from (5.12) and (5.13) that

$$P_\mathrm{D}(\delta^\star) - P_\mathrm{D}(\delta) - \eta \left(P_\mathrm{F}(\delta^\star) - P_\mathrm{F}(\delta) \right) \geq 0 \quad\quad\quad\quad (5.14)$$

and the inequality (5.7) follows.   ∎

A proof of Lemma 5.3.2 Using the definition (5.6) of the randomized test $\delta^\star$, we note that

$$
\begin{aligned}
&P_{\mathrm{F}}(\delta^\star) \\
={}& \int_{\mathbb{R}^k} \delta^\star(\boldsymbol{y}) f_0(\boldsymbol{y}) dF(\boldsymbol{y}) \\
={}& \int_{\{\boldsymbol{y}\in\mathbb{R}^k:f_1(\boldsymbol{y})=\eta f_0(\boldsymbol{y})\}} \gamma(\boldsymbol{y}) f_0(\boldsymbol{y}) dF(\boldsymbol{y}) + \int_{\{\boldsymbol{y}\in\mathbb{R}^k:f_1(\boldsymbol{y})>\eta f_0(\boldsymbol{y})\}} f_0(\boldsymbol{y}) dF(\boldsymbol{y}) \\
={}& \int_{\{\boldsymbol{y}\in\mathbb{R}^k:f_1(\boldsymbol{y})=\eta f_0(\boldsymbol{y})\}} \gamma(\boldsymbol{y}) f_0(\boldsymbol{y}) dF(\boldsymbol{y}) + \mathbb{P}\left[f_1(\boldsymbol{Y}) > \eta f_0(\boldsymbol{Y})|H=0\right].
\end{aligned}
$$

As we seek to satisfy (5.8), we need to select $\eta \geq 0$ and a Borel mapping $\gamma : \mathbb{R}^k \to [0,1]$ such that

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

This last relation suggests introducing the quantity $\eta(\alpha)$ defined by

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

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

Two cases are possible: If

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

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

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

In that case, the constant $\gamma(\alpha)$ satisfies

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

whence

$$
\gamma(\alpha) = \frac{\alpha - \mathbb{P}\left[f_1(\boldsymbol{Y}) > \eta(\alpha) f_0(\boldsymbol{Y})|H=0\right]}{\mathbb{P}\left[f_1(\boldsymbol{Y}) = \eta(\alpha) f_0(\boldsymbol{Y})|H=0\right]}. \tag{5.15}
$$

If

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

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

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

This can be achieved by taking the constant mapping given by

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

∎

A proof of Lemma 5.3.3   The test $\delta_{\mathrm{NP}}(\alpha)$ being a Neyman–Pearson test of size α, the equality

$$P_{\mathrm{D}}(\delta_{\mathrm{NP}}(\alpha)) = P_{\mathrm{D}}(\delta^\star)$$

must hold where the test $\delta^\star$ is given by (5.6) with $\eta > 0$ and Borel mapping $\gamma : \mathbb{R}^k \to [0, 1]$ selected so that (5.8) holds. This a consequence of the fact that both $\delta_{\mathrm{NP}}(\alpha)$ and $\delta^\star$ solve the problem $\mathrm{NP}^\star_\alpha$.

It then follows from (5.7) that

$$0 = P_{\mathrm{D}}(\delta^\star) - P_{\mathrm{D}}(\delta_{\mathrm{NP}}(\alpha)) \geq \eta\left(\alpha - P_{\mathrm{F}}(\delta_{\mathrm{NP}}(\alpha))\right) \geq 0 \qquad (5.16)$$

since $P_{\mathrm{F}}(\delta^\star) = \alpha$ under the choice of $\eta > 0$ and the Borel mapping $\gamma : \mathbb{R}^k \to [0, 1]$, whence $P_{\mathrm{F}}(\delta_{\mathrm{NP}}(\alpha)) = \alpha$.

In other words, $P_{\mathrm{D}}(\delta_{\mathrm{NP}}(\alpha)) = P_{\mathrm{D}}(\delta^\star)$ and $P_{\mathrm{F}}(\delta_{\mathrm{NP}}(\alpha)) = P_{\mathrm{F}}(\delta^\star)$. Using these facts in the expression (5.12) (with the strategy $\delta_{\mathrm{NP}}(\alpha)$) we find that

$$\begin{aligned} 0 &= P_{\mathrm{D}}(\delta^\star) - P_{\mathrm{D}}(\delta_{\mathrm{NP}}(\alpha)) - \eta\left(P_{\mathrm{F}}(\delta^\star) - P_{\mathrm{F}}(\delta_{\mathrm{NP}}(\alpha))\right) \\ &= \int_{\mathbb{R}^k} \left(\delta^\star(\boldsymbol{y}) - \delta_{\mathrm{NP}}(\alpha)(\boldsymbol{y})\right)\left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y})\right) dF(\boldsymbol{y}) \qquad (5.17) \end{aligned}$$

with

$$\left(\delta^\star(\boldsymbol{y}) - \delta_{\mathrm{NP}}(\alpha)(\boldsymbol{y})\right)\left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y})\right) \geq 0, \quad \boldsymbol{y} \in \mathbb{R}^k$$

by virtue of (5.13). It immediately follows that

$$\left(\delta^\star(\boldsymbol{y}) - \delta_{\mathrm{NP}}(\alpha)(\boldsymbol{y})\right)\left(f_1(\boldsymbol{y}) - \eta f_0(\boldsymbol{y})\right) = 0 \quad F - a.e. \qquad (5.18)$$

on $\mathbb{R}^k$. Therefore,

$$\delta_{\mathrm{NP}}(\alpha)(\boldsymbol{y}) = \delta^\star(\boldsymbol{y}) \quad F - a.e. \tag{5.19}$$

on $\{\boldsymbol{y} \in \mathbb{R}^k : f_1(\boldsymbol{y}) \neq \eta f_0(\boldsymbol{y})\}$. ∎

5.5 Examples

The Gaussian case Consider again the situation discussed in Section 2.6 where the observation rv $\boldsymbol{Y}$ is conditionally Gaussian given H, i.e.,

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

where $\boldsymbol{m}_1$ and $\boldsymbol{m}_0$ are distinct elements in $\mathbb{R}^k$, and the $k \times k$ *symmetric* matrix $\boldsymbol{R}$ is positive definite (thus invertible). From the discussion given in Section 2.6, it follows for each $\lambda > 0$ the test d_λ takes the form

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

with $\Delta \boldsymbol{m}$ and $\phi(\lambda)$ given by (2.23) and (2.24), respectively. We also have

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

where d^2 is given by (2.29) – It is plain that the function $\lambda \to P_{\mathrm{F}}(d_\lambda)$ is continuous on $\mathbb{R}_+$ with $\{P_{\mathrm{F}}(d_\lambda), \ \lambda > 0\} = (0, 1)$. 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 2.5 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. We have shown that

$$P_{\mathrm{F}}(d_\lambda) \;\; = \;\; \begin{cases} 1 & \text{if } 0 < \lambda \le \frac{a_1}{a_0} \\[2mm] 1 - a_0 & \text{if } \frac{a_1}{a_0} < \lambda \le \frac{1-a_1}{1-a_0} \\[2mm] 0 & \text{if } \frac{1-a_1}{1-a_0} < \lambda \end{cases} \tag{5.20}$$

as λ ranges over $(0, \infty)$.

Note that $\lambda \to P_{\mathrm{F}}(d_\lambda)$ is left-continuous but not continuous with

$$\{P_{\mathrm{F}}(d_\lambda), \quad \lambda > 0\} = \{0, 1 - a_0, 1\}.$$

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

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

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

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

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

$$\begin{aligned} d_\lambda(k) = 0 \quad &\text{iff} \quad \frac{(m_1)^k}{k!} e^{-m_1} < \lambda \frac{(m_0)^k}{k!} e^{-m_0} \\[2mm] &\text{iff} \quad \left(\frac{m_1}{m_0}\right)^k < \lambda e^{-(m_0 - m_1)} \end{aligned} \tag{5.21}$$

with $k = 0, 1, \ldots.$

If $m_0 < m_1$, then

$$\begin{aligned} d_\lambda(k) = 0 \quad &\text{iff} \quad \frac{(m_1)^k}{k!} e^{-m_1} < \lambda \frac{(m_0)^k}{k!} e^{-m_0} \\[2mm] &\text{iff} \quad \left(\frac{m_1}{m_0}\right)^k < \lambda e^{-(m_0 - m_1)} \\[2mm] &\text{iff} \quad k < \eta(\lambda) \end{aligned} \tag{5.22}$$

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

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

It follows that

$$\begin{aligned}
P_{\mathrm{F}}(d_\lambda) &= \mathbb{P}\left[d_\lambda(Y) = 1 | H = 0\right] \\
&= \mathbb{P}\left[Y \geq \eta(\lambda) | H = 0\right] \\
&= \sum_{k=0:\ \eta(\lambda) \leq k}^{\infty} \frac{(m_0)^k}{k!} e^{-m_0}.
\end{aligned} \qquad (5.23)$$

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

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

As a result, the mapping $\lambda \to P_{\mathrm{F}}(d_\lambda)$ is easily seen to be a *left-continuous piecewise constant* mapping with

$$P_{\mathrm{F}}(d_\lambda) = P_{\mathrm{F}}(d_{\lambda_n}), \qquad \begin{array}{l} \lambda_n < \lambda \leq \lambda_{n+1} \\ n = 0, 1, \ldots \end{array}$$

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

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

or equivalently,

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

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

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