Chapter 4

Bayesian hypothesis testing

4.1 Introduction

Let start by reminding and precising the notations and definitions. First we consider a general case and we assume that there exists a parameter vector $\boldsymbol{\theta}$ of finite dimensionality m and M stochastic processes, such that for every fixed value of $\boldsymbol{\theta} \in \mathcal{T}$, the conditional distributions $\{F_{\boldsymbol{\theta},i}(\boldsymbol{x}) = F_i(\boldsymbol{x}|\boldsymbol{\theta}),\ i = 1, \cdots, M\}$ and their coresponding densities $\{f_{\boldsymbol{\theta},i}(\boldsymbol{x}) = f_i(\boldsymbol{x}|\boldsymbol{\theta}),\ i = 1, \cdots, M\}$ are well known, for all values $\boldsymbol{x} \in \Gamma$.

We also assume to know the conditional prior probability distributions

$$\{\pi_i(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|H_i), \quad i = 1, \cdots, M\},$$

their coresponding unconditional prior distributions

$$\{r_i(\boldsymbol{\theta}) = P(H = H_i)\, \pi_i(\boldsymbol{\theta}) = \pi_i\, \pi_i(\boldsymbol{\theta}), \quad i = 1, \cdots, M\}$$

and the penalty functions $\{c_{ki}(\boldsymbol{\theta})\}$.

For a given decision rule $\delta(\boldsymbol{x}) = \{\delta_j(\boldsymbol{x}), j = 1, \cdots, M\}$ we define the expected penalty

$$r(\delta) = \int_{\boldsymbol{\Gamma}} \mathrm{d}\boldsymbol{x} \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) \int_{\mathcal{T}} \sum_{i=1}^{M} c_{ki}(\boldsymbol{\theta}) f_{\boldsymbol{\theta},i}(\boldsymbol{x})\, r_i(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} \tag{4.1}$$

Note that, this general case reduces to the two following special cases:

- When the M stochastic processes coresponding to the M hypotheses are described through a single parametric stochastic process with M disjoint subdivisions $\{\mathcal{T}_1, \cdots, \mathcal{T}_M\}$ of the parameter space $\mathcal{T}$, then the quantities $\pi_i(\boldsymbol{\theta})$ and $r_i(\boldsymbol{\theta})$, both reduce to $\pi(\boldsymbol{\theta})$, and the quantities $\{f_{\boldsymbol{\theta},i}(\boldsymbol{x})\}$ and $\{c_{ki}(\boldsymbol{\theta})\}$ reduce to $\{f_{\boldsymbol{\theta}}(\boldsymbol{x})\}$ and $\{c_k(\boldsymbol{\theta})\}$. We then have

$$r(\delta) = \int_{\boldsymbol{\Gamma}} \mathrm{d}\boldsymbol{x} \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) \int_{\mathcal{T}} c_k(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{x}) \pi(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} \tag{4.2}$$

- When the M stochastic processes coresponding to the M hypotheses are all well known then we can eliminate $\boldsymbol{\theta}$ from the above equations and the quantities $f_{\boldsymbol{\theta},i}(\boldsymbol{x})$,

$r_i(\boldsymbol{\theta})$ and $\{c_{ki}(\boldsymbol{\theta})\}$ reduce respectively to $\{f_i(\boldsymbol{x})\}$, $\pi_i = \Pr\{H_i\}$ and $\{c_{ki}\}$. We then have

$$r(\delta) = \int_{\boldsymbol{\Gamma}} \mathrm{d}\boldsymbol{x} \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) \sum_{i=1}^{M} c_{ki}\, f_i(\boldsymbol{x})\, \pi_i \,\mathrm{d}\boldsymbol{x} \tag{4.3}$$

Note that, in all the three above cases we can write

$$r(\delta) = \int_{\boldsymbol{\Gamma}} \mathrm{d}\boldsymbol{x} \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) g_k(\boldsymbol{x}) \tag{4.4}$$

where $g_k(\boldsymbol{x})$ is given by one of the following equations:

$$g_k(\boldsymbol{x}) = \int_{\mathcal{T}} \sum_{i=1}^{M} c_{ki}(\boldsymbol{\theta}) f_{\boldsymbol{\theta},i}(\boldsymbol{x})\, r_i(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \tag{4.5}$$

$$g_k(\boldsymbol{x}) = \int_{\mathcal{T}} c_k(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{x}) \pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \tag{4.6}$$

$$g_k(\boldsymbol{x}) = \sum_{i=1}^{M} c_{ki}\, f_i(\boldsymbol{x})\, \pi_i \,\mathrm{d}\boldsymbol{x} \tag{4.7}$$

4.2 Optimization problem

Now, we have all the ingredients to write down the optimization problem of the Bayesian hypothesis testing. Before starting, remember that for any decision rule $\delta = \{\delta_j(\boldsymbol{x}), j = 1, \cdots, M\}$, we have

$$\delta_k(\boldsymbol{x}) \geq 0,\ k = 1, \cdots, M \quad \text{and} \quad \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) = 1 \quad \forall \boldsymbol{x} \in \boldsymbol{\Gamma} \tag{4.8}$$

Now, consider the following optimization problem:

$$\text{Minimize} \qquad r(\delta) = \int_{\boldsymbol{\Gamma}} \mathrm{d}\boldsymbol{x} \sum_{k=1}^{M} \delta_k(\boldsymbol{x})\, g_k(\boldsymbol{x}) \tag{4.9}$$

$$\text{subject to} \quad \begin{cases} \delta_k(\boldsymbol{x}) \geq 0, & k = 1, \cdots, M, \\ \sum_{k=1}^{M} \delta_k(\boldsymbol{x}) = 1, \quad \forall \boldsymbol{x} \in \boldsymbol{\Gamma} \end{cases} \tag{4.10}$$

where $\{g_k(\boldsymbol{x}),\ k = 1, \cdots, M\}$ are non negative functions defined on $\boldsymbol{\Gamma}$. Their expression can be given by either (4.5), (4.6) or (4.7).

This optimization problem does not have, in general, a unique solution and there may exist a whole class $\mathcal{D}^*$ of equivalent decision rules. we remember that two decision rules δ_1^* and δ_2^* are equivalent if

$$r(\delta_1^*) = r(\delta_2^*) \ \leq\ r(\delta) \quad \forall \delta \in \mathcal{D} \tag{4.11}$$

The calss $\mathcal{D}^*$ includes both random and determinist decision rules. For the reason of simplicity, in general, one chooses the non random decision rules. Noting that δ_k are then either 0 or 1 depending on the conditions $\boldsymbol{x} \in \Gamma_i$ or $\boldsymbol{x} \notin \Gamma_i$. The expression (4.9) of $r(\delta)$ becomes

$$r(\delta) = \sum_k \int_{\Gamma_k} g_k(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{4.12}$$

Now, assuming that $g_k(\boldsymbol{x}) \geq 0\ \forall \boldsymbol{x} \in \Gamma_k$, the Bayesian hypothesis testing scheme consistes of the following steps:

- Given the observations $\boldsymbol{x}$, compute

$$t(\boldsymbol{x}) = \min_k g_k(\boldsymbol{x}); \tag{4.13}$$

- Select a single $k^* = k(\boldsymbol{x})$ such that $g_{k^*(\boldsymbol{x})}(\boldsymbol{x}) = t(\boldsymbol{x})$;
- Define

$$\delta_j^*(\boldsymbol{x}) = \begin{cases} 1 & j = k^*(\boldsymbol{x}) \\ 0 & j \neq k^*(\boldsymbol{x}) \end{cases} \tag{4.14}$$

The function $t(\boldsymbol{x})$ together with the index $k(\boldsymbol{x})$ are called the *test* and the statistic behavior of the pair $[t(\boldsymbol{X}), k(\boldsymbol{X})]$ is called the *test statistics*.

To go more in details we consider some examples from general cases to more specific ones.

Let start with a general case. We here we assume that during any n observations the transmitted signal is exactly one of the M possibles $\{\boldsymbol{s}_i,\ i = 0, \cdots, M-1\}$. Then we have M hypotheses:

$$H_i: \quad \boldsymbol{x} \sim f_i(\boldsymbol{x}), \quad i = 0, \cdots, M-1 \tag{4.15}$$

Then, we have

$$g_k(\boldsymbol{x}) = \sum_{i=1}^{M} c_{ki}\ f_i(\boldsymbol{x})\ \pi_i \tag{4.16}$$

and

$$t(\boldsymbol{x}) = \min_k g_k(\boldsymbol{x}) = \min_k \sum_{i=1}^{M} c_{ki}\ f_i(\boldsymbol{x})\ \pi_i \tag{4.17}$$

Given the observation $\boldsymbol{x}$, the search for some index $k(\boldsymbol{x})$ that satisfies (4.17) can be realized via the differences

$$\sum_{i=1}^{M} (c_{ki} - c_{li})\ f_i(\boldsymbol{x})\ \pi_i \tag{4.18}$$

The optimal index $k^*(\boldsymbol{x})$ is such that

$$k^*(\boldsymbol{x}): \quad \sum_{i=1}^{M} (c_{k^*i} - c_{li})\ f_i(\boldsymbol{x})\ \pi_i \leq 0 \quad \forall l \tag{4.19}$$

If $\boldsymbol{x}$ is such that $f_0(\boldsymbol{x}) > 0$ and if $\pi_0 > 0$, then we have

$$k^*(\boldsymbol{x}): \quad \sum_{i=1}^{M} (c_{k^*i} - c_{li})\ \frac{\pi_i}{\pi_0}\ \frac{f_i(\boldsymbol{x})}{f_0(\boldsymbol{x})} \leq 0 \quad \forall l \tag{4.20}$$

The ratio $\left\{\frac{f_i(\boldsymbol{x})}{f_0(\boldsymbol{x})}\right\}$ are called the *likelihood ratios*, so that, the procedure to obtain the decisions now consists in comparing the likelihood ratios against some thresholds. The test induced by (4.20) thus consists of a weighted sum of the likelihood ratios. This weigthed sum is called the *test function.* So, in general, the test function is compared with the threshold zero which is independent of the observation.

Note also that we can rewrite (4.20) in the two following other forms:

$$k^*(\boldsymbol{x}): \quad \sum_{i=1}^{M}(c_{k^*i} - c_{li}) \, \frac{\pi_i(\boldsymbol{x})}{\pi_0(\boldsymbol{x})} \leq 0 \quad \forall l \tag{4.21}$$

or still

$$k^*(\boldsymbol{x}): \quad \sum_{i=1}^{M} c_{k^*i} \, \pi_i(\boldsymbol{x}) \leq \sum_{i=1}^{M} c_{li} \; \pi_i(\boldsymbol{x}) \quad \forall l \tag{4.22}$$

The fractions $\dfrac{\pi_i(\boldsymbol{x})}{\pi_0(\boldsymbol{x})}$ are the *posterior likelihood ratios* and $\bar{c}_l(\boldsymbol{x}) = \sum_{i=1}^{M} c_{li} \, \pi_i(\boldsymbol{x})$ are the expected posterior penalties.

Further simplifications can be acheived with uniform cost functions

$$c_{ki} = \begin{cases} 0 & \text{if } k = i \\ 1 & \text{if } k \neq i \end{cases} \tag{4.23}$$

and with uniform priors

$$\pi_1 = \pi_2 = \cdots = \pi_M = \frac{1}{M}. \tag{4.24}$$

With these assumptions, the Bayesian decision rule becomes

$$k^*(\boldsymbol{x}): \quad L_{k^*}(\boldsymbol{x}) \geq L_l(\boldsymbol{x}) \quad \forall l \tag{4.25}$$

where

$$L_i(\boldsymbol{x}) = \frac{f_i(\boldsymbol{x})}{f_0(\boldsymbol{x})} \tag{4.26}$$

Now, let consider some special cases.

4.3 Examples

4.3.1 Radar applications

One of the oldest area where the detection-estimation theory has been used is the radar applications. The main problem is to detect the presence of M knwon signals transmitted through the atmosphere. The transmission chanel is the atmosphere and it is assumed to be statistically well knwon. A simple model for the received signal is then

$$X(t) = S(t) + N(t) \tag{4.27}$$

where $N(t)$ is the additive noise due to the chanel. In the discrete case, where we assume to observe n samples of the received signal in the time period $[0, T]$, this model becomes

$$X_j = S_j + N_j, \quad j = 1, \cdots, n \tag{4.28}$$

or still

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n}. \tag{4.29}$$

Consider now the case where one of the signals $\boldsymbol{s}_i$ is null. Then the M hypotheses become:

$$\begin{cases} H_0 & : \quad \boldsymbol{x} = \boldsymbol{n} \\ H_i & : \quad \boldsymbol{x} = \boldsymbol{s}_i + \boldsymbol{n}, \quad i = 1, \ldots, M-1 \end{cases} \tag{4.30}$$

Assume that we know also the conditional probability density functions $f_0(\boldsymbol{x})$ and $f_i(\boldsymbol{x})$ of the received signal under the hypotheses H_0 and H_i. The likelihood ratios $L_i(\boldsymbol{x})$ then become

$$L_i(\boldsymbol{x}) = \frac{f_i(\boldsymbol{x})}{f_0(\boldsymbol{x})} = \frac{\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{s}_i)^t(\boldsymbol{x} - \boldsymbol{s}_i)\right]}{\exp\left[\frac{-1}{2\sigma^2}\boldsymbol{x}^t\boldsymbol{x}\right]} = \exp\left[\frac{-1}{2\sigma^2}(-2\boldsymbol{s}_i^t\boldsymbol{x} + \boldsymbol{s}_i^t\boldsymbol{s}_i)\right] \tag{4.31}$$

and $k^*(\boldsymbol{x})$ satisfies:

$$k^*(\boldsymbol{x}) : \quad \boldsymbol{s}_{k^*}^t\boldsymbol{x} + \frac{1}{2}\boldsymbol{s}_{k^*}^t\boldsymbol{s}_{k^*} > \boldsymbol{s}_l^t\boldsymbol{x} + \frac{1}{2}\boldsymbol{s}_l^t\boldsymbol{s}_l \quad \forall l \tag{4.32}$$

Figure 4.3.1 shows the structure of this optimal test.

Indeed, if we assume that all the signals have the same energies $|\boldsymbol{s}_i|^2 = \boldsymbol{s}_i^t\boldsymbol{s}_i$, then we have

$$k^*(\boldsymbol{x}) : \quad \boldsymbol{s}_{k^*}^t\boldsymbol{x} > \boldsymbol{s}_l^t\boldsymbol{x} \quad \forall l \tag{4.33}$$

Figure 4.3.1 shows the structure of this optimal test.

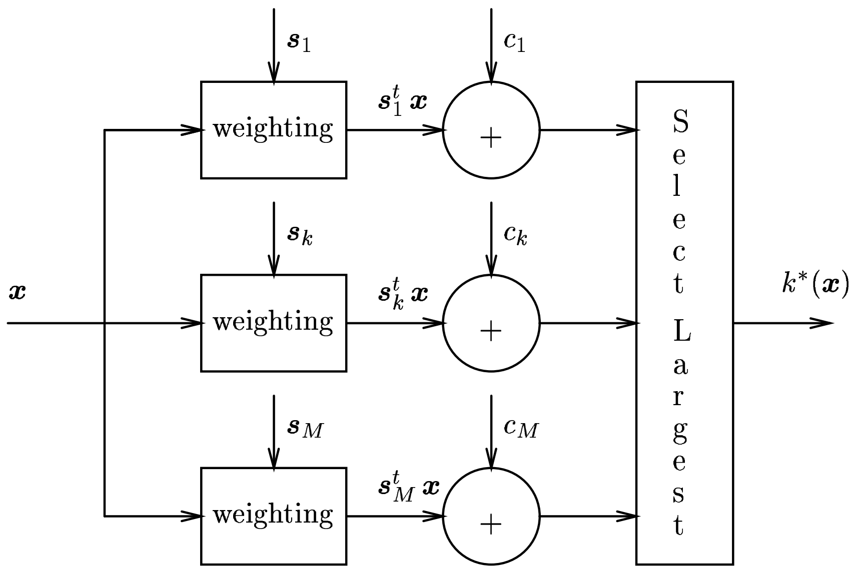

Figure 4.1: General structure of a Bayesian optimal detector.

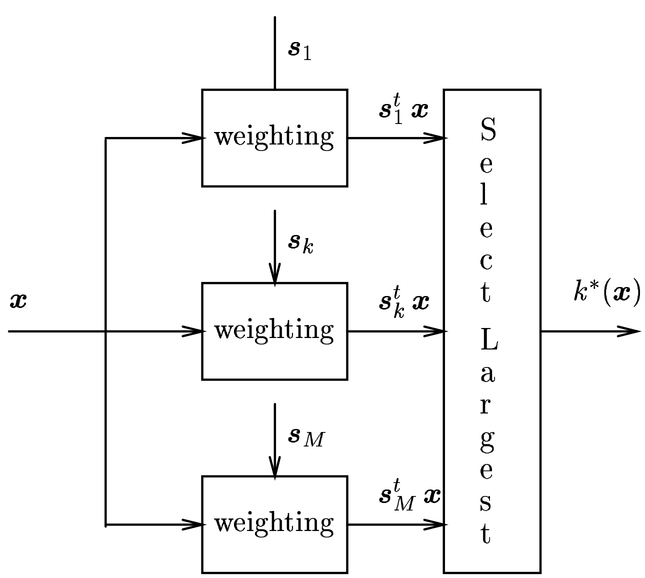

Figure 4.2: Simplified structure of a Bayesian optimal detector.

4.3.2 Detection of a known signal in an additive noise

Consider now the case where there is only one signal. So that, we have a binary detection problem:

$$\begin{cases} H_0 & : \quad \boldsymbol{x} = \boldsymbol{n} & \longrightarrow f_0(\boldsymbol{x}) = f(\boldsymbol{x}) \\ H_1 & : \quad \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} & \longrightarrow f_1(\boldsymbol{x}) = f(\boldsymbol{x} - \boldsymbol{s}) \end{cases} \tag{4.34}$$

Then, we have:

$$L(\boldsymbol{x}) = \frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} \tag{4.35}$$

General case:

The general optimal Bayesian detector structure becomes:

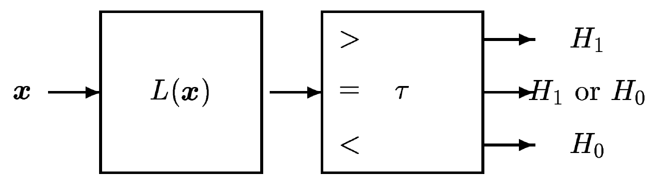

Figure 4.3: General Bayesian detector.

Case of white noise:

Now, if we assume that the noise is white, we have:

$$\begin{cases} H_0 & : \quad \boldsymbol{x} = \boldsymbol{n} & \longrightarrow f_0(\boldsymbol{x}) = \prod_j f(x_j) \\ H_1 & : \quad \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} & \longrightarrow f_1(\boldsymbol{x}) = \prod_j f(x_j - s_j) \end{cases} \tag{4.36}$$

and consequently

$$L(\boldsymbol{x}) = \prod_j L_j(x_j) = \prod_j \frac{f(x_j - s_j)}{f(x_j)} \tag{4.37}$$

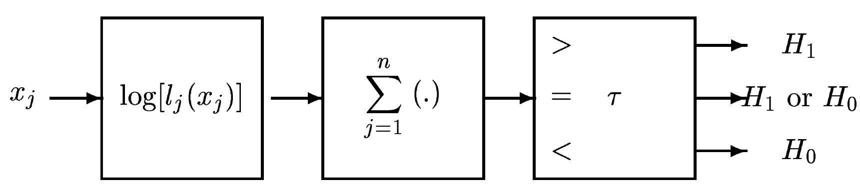

Figure 4.4: Bayesian detector in the case of i.i.d. data.

Case of Gaussian noise:

In this case we have

$$\begin{cases} H_0 & : \quad \boldsymbol{x} = \boldsymbol{n} & \longrightarrow f_0(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma^2}\boldsymbol{x}^t\boldsymbol{x}\right] \\ H_1 & : \quad \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} & \longrightarrow f_1(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2\sigma^2}[\boldsymbol{x} - \boldsymbol{s}]^t[\boldsymbol{x} - \boldsymbol{s}]\right] \end{cases} \tag{4.38}$$

$$L(\boldsymbol{x}) = \frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} = \frac{\exp\left[\frac{-1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{s})^t(\boldsymbol{x}-\boldsymbol{s})\right]}{\exp\left[\frac{-1}{2\sigma^2}\boldsymbol{x}^t\boldsymbol{x}\right]} = \exp\left[\frac{-1}{2\sigma^2}(-2\boldsymbol{s}^t\boldsymbol{x}+\boldsymbol{s}^t\boldsymbol{s})\right] \tag{4.39}$$

We then have

$$\delta_B(\boldsymbol{x}) = \begin{cases} 1 & & > \\ 0/1 & \text{if } \boldsymbol{s}^t(\boldsymbol{x}-\boldsymbol{s}) & = \tau \\ 1 & & < \end{cases} \tag{4.40}$$

The detector has the following structure:

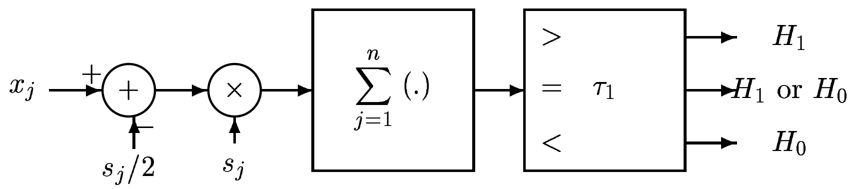

Figure 4.5: General Bayesian detector in the case of i.i.d. Gaussian data.

Note that we can rewrite (4.32) as:

$$\delta_B(\boldsymbol{x}) = \begin{cases} 1 & & > \\ 0/1 & \text{if } \boldsymbol{s}^t\boldsymbol{x} & = \tau' \\ 1 & & < \end{cases} \tag{4.41}$$

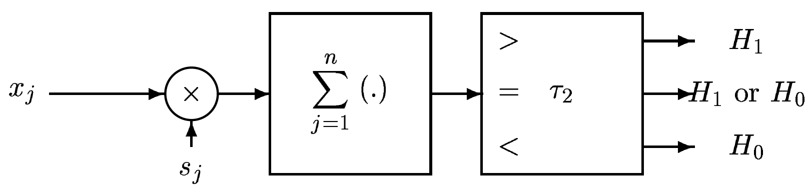

Figure 4.6: Simplified Bayesian detector in the case of i.i.d. Gaussian data.

Case of Laplacian noise:

$$H_0 : \boldsymbol{x} = \boldsymbol{n} \quad \longrightarrow \quad f_0(\boldsymbol{x}) = \prod_{j=1}^{n} \frac{\alpha}{2} \exp\left[-\alpha \sum_{j=1}^{n} |x_j|\right] \tag{4.42}$$

$$H_i : \boldsymbol{x} = \boldsymbol{s} + \boldsymbol{n} \quad \longrightarrow \quad f_1(\boldsymbol{x}) = \prod_{j=1}^{n} \frac{\alpha}{2} \exp\left[-\alpha \sum_{j=1}^{n} |x_j - s_j|\right] \tag{4.43}$$

$$L(\boldsymbol{x}) = \prod_{j=1}^{n} L_j(x_j) \tag{4.44}$$

with

$$
\begin{aligned}
L_j(x_j) &= \ = \frac{f_1(x_j)}{f_0(x_j)} = \exp\left[\alpha|x_j - s_j| + \alpha|x_j|\right] \\
&= \begin{cases} \exp\left[-\alpha|s_j|\right] & \text{if} \quad \operatorname{sgn}(s_j)x_j < 0 \\ exp f \alpha|2x_j - s_j| & \text{if} \quad 0 < \operatorname{sgn}(s_j)x_j < |s_j| \\ exp f -\alpha|s_j| & \text{if} \quad \operatorname{sgn}(s_j)x_j > |s_j| \end{cases} \qquad (4.45)
\end{aligned}
$$

where

$$
\operatorname{sgn}(x) = \begin{cases} +1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -1 & \text{if} \quad x < 0 \end{cases} \qquad (4.46)
$$

Considering then two cases of $s_j < 0$ and $s_j > 0$ we obtain

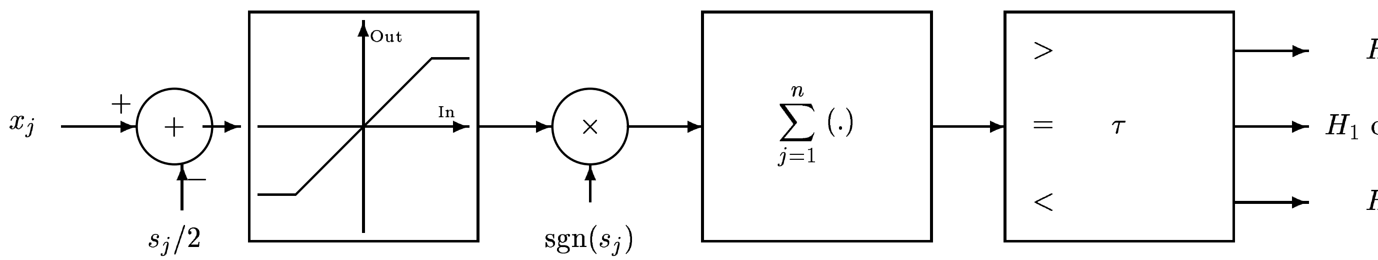

Figure 4.7: Bayesian detector in the case of i.i.d. Laplacian data.

4.4 Binary chanel transmission

In numeric signal transmission, in general, we have to transmit binary sequences. If we assume that the chanel transmits each bit separately in a memoryless fashion and that each bit s_j is transmitted correctly with probability q, then we can describe this chanel graphically as follows

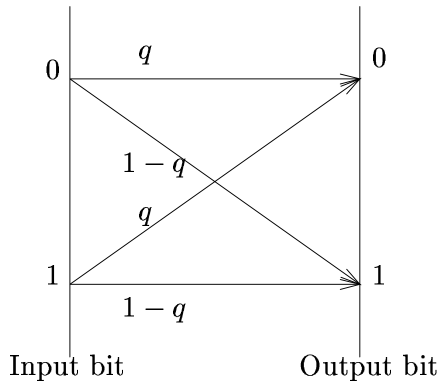

Figure 4.8: A binary chanel.

A useful binary, memoryless and symetric chanel should have a probability of correct transmission higher than the probability of incorrect transmission, $i.e. q > 0.5$ and $1-q < 0.5$.

With these assumptions on the chanel we can easily calculate the probability of observing $\boldsymbol{x}$ conditional to the transmitted sequence $\boldsymbol{s}$

$$\begin{aligned}\Pr\{\boldsymbol{x}|\boldsymbol{s}\} &= \prod_{j=1}^{n} \Pr\{x[j]|s[j]\} = \prod_{i=1}^{n} q^{1-(x[j]\oplus s[j])}(1-q)^{(x[j]\oplus s[j])} \\ &= q^n \left(\frac{1-q}{q}\right)^{\sum_{i=1}^{M}(x[j]\oplus s[j])} \end{aligned} \tag{4.47}$$

where $\oplus$ signifies binary sum.

Now, assume that, during each observation period, only one of the M well knwon binary sequences $\boldsymbol{s}_k$ (called codewords) are transmitted. Now, we have received the binary sequence $\boldsymbol{x}$ and we want to know which one of them has been transmitted.

Indeed, if we assume $p_1 = p_2 = \cdots = p_M = 1/M$ and if we note by

$$H(\boldsymbol{s}_k, \boldsymbol{s}_l) = \frac{1}{n}\sum_{j=1}^{n} s_k[j] \oplus s_l[j] \tag{4.48}$$

the Haming distance between the two binary words $\boldsymbol{s}_k$ and $\boldsymbol{s}_l$, then, the likelihood ratios have the following form:

$$\begin{aligned}\frac{\Pr\{\boldsymbol{x}|\boldsymbol{s}_k\}}{\Pr\{\boldsymbol{x}|\boldsymbol{s}_l\}} &= \prod_{j=1}^{n} q^{1-(x[j]\oplus s_k[j])}(1-q)^{(x[j]\oplus s_k[j])} \\ &= q^n \left(\frac{1-q}{q}\right)^{\sum_{i=1}^{M}(x[j]\oplus s_k[j])} \end{aligned} \tag{4.49}$$

and the Bayesian optimal test becomes:

$$k^*(\boldsymbol{x}): \quad \left(\frac{1-q}{q}\right)^{\sum_{j=1}^{n}(x[j]\oplus s_{k*}[j])-\sum_{j=1}^{n}(x[j]\oplus s_l[j])} \geq 1 \quad \forall l = 1,\dots,M \tag{4.50}$$

Taking the logarithm of both parts we obtain the following condition on $k^*(\boldsymbol{x})$:

$$k^*(\boldsymbol{x}): \quad \sum_{j=1}^{n}(x[j]\oplus s_{k*}[j]) - \sum_{j=1}^{n}(x[j]\oplus s_l[j]) \;\log\left(\frac{1-q}{q}\right) \geq 0 \quad \forall l = 1,\dots,M \tag{4.51}$$

We can then discriminate two cases:

- Case 1: Let $q > .5$, which means that the transmission chanel has a higher probability of transmitting correctly than incorrectly. Then $\frac{1-q}{q} < 1$ and $k^*(\boldsymbol{x})$ satisfies

  $$k^*(\boldsymbol{x}): \quad \sum_{j=1}^{n}(x[j]\oplus s_{k*}[j]) \leq \sum_{j=1}^{n}(x[j]\oplus s_l[j]), \quad \forall l = 1,\dots,M \tag{4.52}$$

  or still

  $$k^*: \quad H(\boldsymbol{x}, \boldsymbol{s}_{k*}) \leq H(\boldsymbol{x}, \boldsymbol{s}_l), \quad \forall l = 1,\dots,M \tag{4.53}$$

  The test clearly decides in favor of the codeword $\boldsymbol{s}_{k*}$ whose Hamming distance from the observed sequence $\boldsymbol{x}$ is the minimum one. This is why this detector is called *the minimum distance decoding scheme.*

  Let now the M codewords be designed so that the Haming distance between any two such codewords equals $(2d+1)/n$, where d is a positive integer, i.e.

  $$H(\boldsymbol{s}_k, \boldsymbol{s}_l) = (2d+1)/n, \quad \forall k \neq l,\; k,l = 1,\dots,M \tag{4.54}$$

  and d such that

  $$\sum_{i=0}^{d}\binom{n}{i} \leq 2^n/M \tag{4.55}$$

  Then, via the minimum distance decoding scheme, if the distance between the received word $\boldsymbol{x}$ and the codeword $\boldsymbol{s}_k$ is at most d/n, then the codeword $\boldsymbol{s}_k$ is correctly detected and we have

  $$P_d(\boldsymbol{s}_k) \;\geq\; \sum_{i=0}^{d}\binom{n}{i} q^{n-i}(1-q)^i \tag{4.56}$$

  $$P_d = \sum_{k=1}^{M}(1/M)P_d(\boldsymbol{s}_k) \;\geq\; \sum_{i=0}^{d}\binom{n}{i} q^{n-i}(1-q)^i \tag{4.57}$$

  $$P_e = 1 - Pd \;\leq\; 1 - \sum_{i=0}^{d}\binom{n}{i} q^{n-i}(1-q)^i \tag{4.58}$$

  $$\tag{4.59}$$

- Case 2: In the case of $q < 0.5$, by the same analysis, the Bayesian detection scheme decides in favor of the codeword whose Hamming distance from the observed sequence is the maximum. This is not surprising because if $q < 0.5$, then with probability $1-q > 0.5$, more than half of the codeword bits are changed in the transmission.