

Chapter 5

Signal detection and structure of optimal detectors

In previous chapters we discussed some basic optimality criteria and design methods for general hypothesis testing problems. In this chapter we apply them to derive optimal procedures for the detection of signals corrupted by some noise. We consider only the discrete case.

First, we summarize the Bayesian composite hypothesis testing and focus on the binary case. Then, we describe other related tests in this particular case. Finally, through some examples with different models for the signal and the noise, we derive the optimum detector structures.

At the end, we give some basic elements of robust, sequential and non parametric detection.

5.1 Bayesian composite hypothesis testing

Consider the following composite hypothesis testing:

$$\mathbf{X} \sim f_{\boldsymbol{\theta}}(\mathbf{x}) \quad (5.1)$$

and define the decision $\boldsymbol{\delta}(\mathbf{x})$, its associated cost function $c[\boldsymbol{\delta}(\mathbf{x}), \boldsymbol{\theta}]$. Then the conditional risk function is given by

$$\begin{aligned} R_{\boldsymbol{\theta}}(\boldsymbol{\delta}) &= \mathbb{E}_{\boldsymbol{\theta}} \{c[\boldsymbol{\delta}(\mathbf{X}), \boldsymbol{\theta}]\} = \mathbb{E} [c[\boldsymbol{\delta}(\mathbf{X}), \boldsymbol{\Theta}] | \boldsymbol{\Theta} = \boldsymbol{\theta}] \\ &= \int_{\Gamma} c[\boldsymbol{\delta}(\mathbf{x}), \boldsymbol{\theta}] f_{\boldsymbol{\theta}}(\mathbf{x}) \, d\mathbf{x} \end{aligned} \quad (5.2)$$

and the Bayes risk by

$$\begin{aligned} r(\boldsymbol{\delta}) &= \mathbb{E} [R_{\boldsymbol{\Theta}}(\boldsymbol{\delta}(\mathbf{X}))] = \mathbb{E} [\mathbb{E} [c[\boldsymbol{\delta}(\mathbf{x}), \boldsymbol{\Theta}] | \boldsymbol{\Theta} = \boldsymbol{\theta}]] \\ &= \int_{\mathcal{T}} \int_{\Gamma} c[\boldsymbol{\delta}(\mathbf{x}), \boldsymbol{\theta}] f_{\boldsymbol{\theta}}(\mathbf{x}) \pi(\boldsymbol{\theta}) \, d\mathbf{x} \, d\boldsymbol{\theta} \\ &= \int_{\Gamma} \int_{\mathcal{T}} c[\boldsymbol{\delta}(\mathbf{x}), \boldsymbol{\theta}] \pi(\boldsymbol{\theta} | \mathbf{x}) \, d\boldsymbol{\theta} \, d\mathbf{x} \\ &= \mathbb{E} [\mathbb{E} [c[\boldsymbol{\delta}(\mathbf{X}), \boldsymbol{\Theta}] | \mathbf{X} = \mathbf{x}]] \end{aligned} \quad (5.3)$$

From this relation, and the fact that in general the cost function is a positive function, we can deduce that minimizing $r(\delta)$ over δ is equivalent to minimize, for any $\mathbf{x} \in \Gamma$, the mean posterior cost

$$\bar{c}[\mathbf{x}|\boldsymbol{\theta}] = \mathbb{E}[c[\delta(\mathbf{X}), \boldsymbol{\Theta}]|\mathbf{X} = \mathbf{x}] = \int_{\mathcal{T}} c[\delta(\mathbf{x}), \boldsymbol{\theta}] \pi(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}. \quad (5.4)$$

5.1.1 Case of binary composite hypothesis testing

In this case, we have

$$\delta_B(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } \mathbb{E}[c[1, \boldsymbol{\theta}]|\mathbf{X} = \mathbf{x}] = \mathbb{E}[c[0, \boldsymbol{\theta}]|\mathbf{X} = \mathbf{x}] \\ 0 & < \end{cases} \quad (5.5)$$

If the two hypotheses correspond to two disjoint partitions of the parameter space $\tau = \{\mathcal{T}_0, \mathcal{T}_1\}$, we have

$$c[i, \boldsymbol{\theta}] = c_{ij} \quad \text{if } \boldsymbol{\theta} \in \mathcal{T}_j \quad (5.6)$$

and if we consider the uniform cost function, then we have

$$\delta_B(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } \frac{\Pr\{\boldsymbol{\theta} \in \mathcal{T}_1|\mathbf{X} = \mathbf{x}\}}{\Pr\{\boldsymbol{\theta} \in \mathcal{T}_0|\mathbf{X} = \mathbf{x}\}} = \frac{c_{10} - c_{00}}{c_{01} - c_{11}} \\ 0 & < \end{cases} \quad (5.7)$$

Now, using the Bayes' rule, we have

$$\Pr\{\boldsymbol{\theta} \in \mathcal{T}_i|\mathbf{X} = \mathbf{x}\} = \frac{\Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_i\} \Pr\{\boldsymbol{\theta} \in \mathcal{T}_i\}}{\sum_{i=0}^1 \Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_i\} \Pr\{\boldsymbol{\theta} \in \mathcal{T}_i\}}, \quad i = 0, 1. \quad (5.8)$$

The decision rule (5.7) becomes

$$\delta_B(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } L(\mathbf{x}) = \frac{\Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_1\}}{\Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_0\}} = \frac{\pi_0 c_{10} - c_{00}}{\pi_1 c_{01} - c_{11}} \\ 0 & < \end{cases} \quad (5.9)$$

where

$$\pi_i = \Pr\{\boldsymbol{\theta} \in \mathcal{T}_i\}, \quad i = 0, 1. \quad (5.10)$$

The conditional probability density functions of $\boldsymbol{\theta}$ are noted $r_i(\boldsymbol{\theta})$ and given by

$$r_i(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \notin \mathcal{T}_i \\ p(\boldsymbol{\theta})/\pi_i & \text{if } \boldsymbol{\theta} \in \mathcal{T}_i \end{cases} \quad (5.11)$$

The expression of $\Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_i\}$ is then given by

$$\begin{aligned} \Pr\{\mathbf{X} = \mathbf{x}|\boldsymbol{\theta} \in \mathcal{T}_i\} &= \int_{\mathcal{T}} f_{\boldsymbol{\theta}}(\mathbf{x}) r_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \frac{1}{\pi_i} \int_{\mathcal{T}_i} f_{\boldsymbol{\theta}}(\mathbf{x}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \frac{1}{\pi_i} \mathbb{E}_i\{f_{\boldsymbol{\Theta}}(\mathbf{x})\} \end{aligned} \quad (5.12)$$

where $E_i \{f_{\Theta}(\mathbf{x})\}$ stands for the expectation under the hypothesis H_i . We can then rewrite (5.9) as:

$$\delta_B(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } L(\mathbf{x}) = \frac{E_1 \{f_{\Theta}(\mathbf{x})\}}{E_0 \{f_{\Theta}(\mathbf{x})\}} = \frac{c_{10} - c_{00}}{c_{01} - c_{11}} \\ 0 & < \end{cases} \quad (5.13)$$

With such hypotheses, the false alarm and correct detection probabilities become respectively

$$P_F(\boldsymbol{\delta}, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \{\boldsymbol{\delta}(\mathbf{X})\} \quad \text{for } \boldsymbol{\theta} \in \mathcal{T}_0 \quad (5.14)$$

$$P_D(\boldsymbol{\delta}, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \{\boldsymbol{\delta}(\mathbf{X})\} \quad \text{for } \boldsymbol{\theta} \in \mathcal{T}_1 \quad (5.15)$$

5.2 Uniform most powerful (UMP) test

The α -level uniform most powerful (UMP) test is defined as:

$$\max P_D(\boldsymbol{\delta}, \boldsymbol{\theta}) \quad \text{s.t.} \quad P_F(\boldsymbol{\delta}, \boldsymbol{\theta}) \leq \alpha \quad (5.16)$$

Unfortunately, this optimization problem may not have a solution. In those cases, we can try to design a test by following the Neyman-Pearson scheme which is summarized below.

Neyman-Pearson lemma:

Suppose the hypothesis H_0 is simple (\mathcal{T}_0 has only one component $\boldsymbol{\theta}_0$), the hypothesis H_1 is composite and we have a parametrically defined probability density function $p(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathcal{T}_1$. Then the most powerful α -level test for H_0 against H_1 has a unique critical region given by

$$\Gamma_{\boldsymbol{\theta}} = \left\{ \mathbf{x} \in \Gamma \mid f_{\boldsymbol{\theta}}(\mathbf{x}) > \tau f_{\boldsymbol{\theta}_0}(\mathbf{x}) \right\} \quad (5.17)$$

where τ depends on α . The corresponding test is given by

$$\delta(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } f_{\boldsymbol{\theta}}(\mathbf{x}) = \tau f_{\boldsymbol{\theta}_0}(\mathbf{x}) \\ 0 & < \end{cases} \quad (5.18)$$

5.3 Locally most powerful (LMP) test

The UMP test is too strong and the optimization problem may not have a unique solution in some more general cases. To illustrate this test, let consider the following case:

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta = \theta > \theta_0 \end{cases} \quad (5.19)$$

The α -level locally most powerful (LMP) test is based on the development of $P_D(\boldsymbol{\delta}, \theta)$ in Taylor series around the simple hypothesis parameter value θ_0

$$P_D(\delta, \theta) \simeq P_D(\delta, \theta_0) + (\theta - \theta_0) \left. \frac{\partial P_D(\delta, \theta)}{\partial \theta} \right|_{\theta=\theta_0} + \mathcal{O}(\theta - \theta_0)^2 \quad (5.20)$$

Noting that $P_F(\delta, \theta) = P_D(\delta, \theta_0)$, then the Neyman-Pearson test

$$\max P_D(\delta, \theta) \quad \text{s.t.} \quad P_F(\delta, \theta) \leq \alpha \quad (5.21)$$

becomes

$$\max \left. \frac{\partial P_D(\delta, \theta)}{\partial \theta} \right|_{\theta=\theta_0} \quad \text{s.t.} \quad P_F(\delta, \theta) \leq \alpha \quad (5.22)$$

Noting that

$$P_D(\delta, \theta) = \mathbb{E}_\theta \{ \delta(\mathbf{X}) \} = \int_{\Gamma} \delta(\mathbf{x}) f_\theta(\mathbf{x}) \, d\mathbf{x} \quad (5.23)$$

and assuming that $f_\theta(\mathbf{x})$ is sufficiently regular in the neighbourhood of θ_0 , we can calculate

$$P'_D(\delta, \theta_0) = \left. \frac{\partial P_D(\delta, \theta)}{\partial \theta} \right|_{\theta=\theta_0} = \int_{\Gamma} \delta(\mathbf{x}) \left. \frac{\partial f_\theta(\mathbf{x})}{\partial \theta} \right|_{\theta=\theta_0} \, d\mathbf{x} \quad (5.24)$$

In conclusion, the α -level locally most powerful (LMP) test is obtained in the same way that the α -level most powerful test by replacing $f_\theta(\mathbf{x})$ by $f'_{\theta_0}(\mathbf{x}) = \left. \frac{\partial f_\theta(\mathbf{x})}{\partial \theta} \right|_{\theta=\theta_0}$. The critical region of H_0 against H_1 is then given by

$$\Gamma_{\theta} = \left\{ \mathbf{x} \in \Gamma \mid f'_{\theta_0}(\mathbf{x}) > \tau f_{\theta_0}(\mathbf{x}) \right\} \quad (5.25)$$

and the test becomes

$$\delta(\mathbf{x}) = \begin{cases} 1 & > \\ 0/1 & \text{if } f'_{\theta_0}(\mathbf{x}) = \tau f_{\theta_0}(\mathbf{x}) \\ 0 & < \end{cases} \quad (5.26)$$

where τ and η depend on α .

5.4 Maximum likelihood test

In the absence of the aforementioned optimal tests, we can design a test just based on the likelihood ratios

$$\delta(\mathbf{x}) = \begin{cases} 1 & > \\ \eta & \text{if } \frac{\max_{\theta \in \mathcal{T}_1} f_\theta(\mathbf{x})}{\max_{\theta \in \mathcal{T}_0} f_\theta(\mathbf{x})} = \tau \\ 0 & < \end{cases} \quad (5.27)$$

5.5 Examples of signal detection schemes

To illustrate the common structure of these test designs, consider the following signal detection test:

$$\begin{cases} H_0 & : X_j = N_j, \\ H_1 & : X_j = S_j + N_j \end{cases}, \quad j = 1, \dots, n \quad (5.28)$$

Assume that the noise is i.i.d. with known probability density function $f(x)$

$$\begin{cases} H_0 & : X_j = N_j, & \longrightarrow f_0(\mathbf{x}) = \prod_j f(x_j) \\ H_1 & : X_j = S_j + N_j & \longrightarrow f_1(\mathbf{x}) = \prod_j f(x_j - s_j) \end{cases}, \quad j = 1, \dots, n \quad (5.29)$$

The likelihood ratio becomes

$$L(\mathbf{x}) = \prod_j L_j(x_j) \quad \text{with} \quad L_j(x_j) = \frac{f(x_j - s_j)}{f(x_j)} \quad (5.30)$$

and the test becomes

$$\delta(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_j \log L_j(x_j) > \log \tau \\ 0/1 & \text{if } \sum_j \log L_j(x_j) = \log \tau \\ 0 & \text{if } \sum_j \log L_j(x_j) < \log \tau \end{cases} \quad (5.31)$$

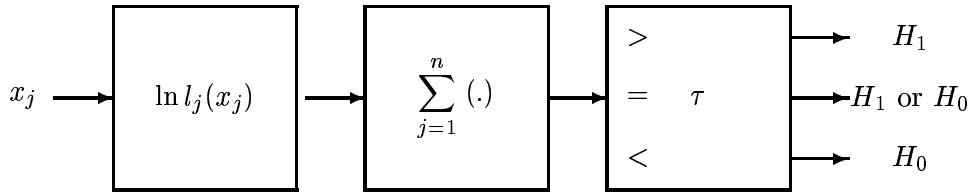


Figure 5.1: The structure of the optimal detector for an i.i.d. noise model.

5.5.1 Case of Gaussian noise

In this case we have

$$f(x_j) \propto \exp \left[-\frac{1}{2\sigma^2} x_j^2 \right] \quad (5.32)$$

and the log likelihood $L_j(x_j)$ ratios become

$$\log L_j(x_j) = \log \frac{f(x_j - s_j)}{f(x_j)} = \frac{1}{\sigma^2} \left[s_j(x_j - \frac{s_j}{2}) \right] \quad (5.33)$$

Noting by $\tau_1 = \sigma^2 \log \tau$, we have the following structure for the optimal detector:

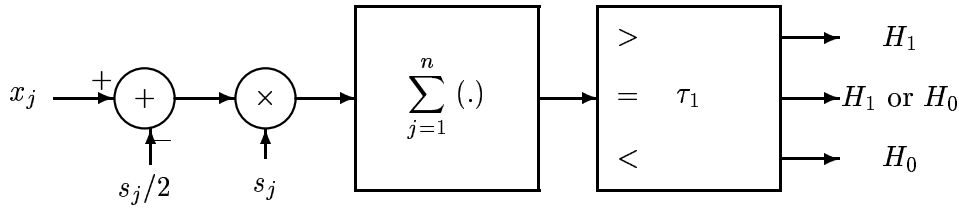


Figure 5.2: The structure of the optimal detector for an i.i.d. Gaussian noise model.

In reporting the constant value $\sum_j s_j^2$ in the threshold, we note $\tau_2 = \tau_1 + \frac{1}{2} \sum_{j=1}^n s_j^2$ and obtain the following scheme

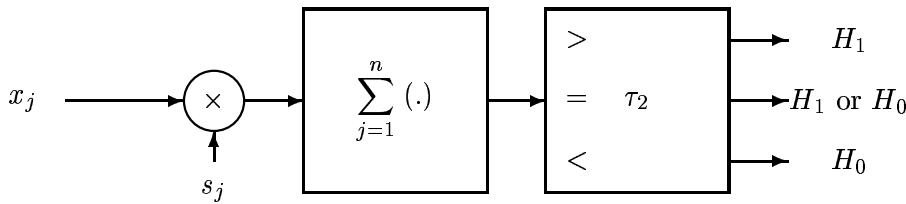


Figure 5.3: The simplified structure of the optimal detector for an i.i.d. Gaussian noise model.

5.5.2 Laplacian noise

In this case we have

$$f(x_j) \propto \exp[-\alpha|x_j|] \tag{5.34}$$

and the log likelihood $L_j(x_j)$ ratios become

$$\begin{aligned} \log L_j(x_j) &= \log \frac{f(x_j - s_j)}{f(x_j)} = -\alpha|x_j - s_j| + \alpha|x_j| \\ &= \begin{cases} -\alpha|s_j| & \text{if } \text{sgn}(s_j)x_j < 0 \\ \alpha \text{sgn}(x_j)|2x_j - s_j| & \text{if } 0 < \text{sgn}(s_j)x_j < |s_j| \\ \alpha|s_j| & \text{if } \text{sgn}(s_j)x_j > |s_j| \end{cases} \end{aligned} \tag{5.35}$$

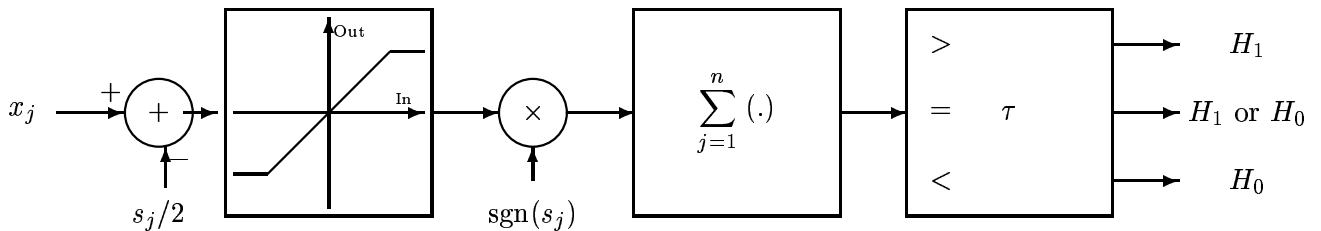


Figure 5.4: Bayesian detector in the case of i.i.d. Laplacian data.

5.5.3 Locally optimal detectors

Consider the following problem:

$$\begin{cases} H_0 & : X_j = N_j \\ H_1 & : X_j = N_j + \theta s_j, \quad \theta > 0 \end{cases} \quad (5.36)$$

We remember that the α -level uniformly optimal test for this problem is:

$$\delta(\mathbf{x}) = \begin{cases} 1 & > \\ \eta & \text{if } L_\theta(\mathbf{x}) = \tau \\ 0 & < \end{cases} \quad (5.37)$$

and the α -level locally optimal test for this problem is:

$$\delta(\mathbf{x}) = \begin{cases} 1 & > \\ \eta & \text{if } \left. \frac{\partial L_\theta(\mathbf{x})}{\partial \theta} \right|_{\theta=\theta_0} = \tau \\ 0 & < \end{cases} \quad (5.38)$$

where τ and η depend on α . For the case of i.i.d. noise model we have

$$L_\theta(\mathbf{x}) = \prod_{j=1}^n \frac{f(x_j - \theta s_j)}{f(x_j)} \quad (5.39)$$

Then, it is easy to show that

$$\left. \frac{\partial \log L_\theta(\mathbf{x})}{\partial \theta} \right|_{\theta=\theta_0} = \sum_{j=1}^n s_j g_{lo}(x_j) \quad (5.40)$$

where

$$g_{lo}(x) = -\frac{\frac{\partial f(x)}{\partial x}}{f(x)} = -\frac{f'(x)}{f(x)} \quad (5.41)$$

and the structure of a locally optimal detector is given in the following figure.

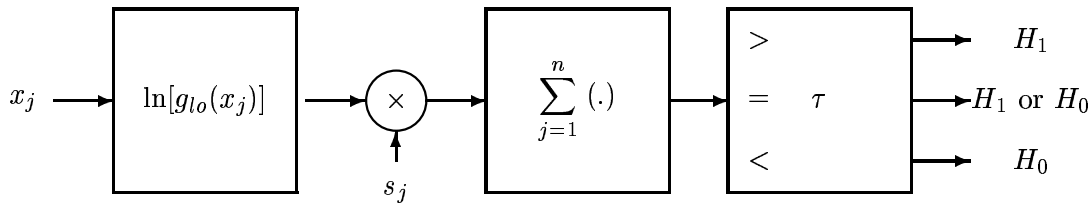


Figure 5.5: The structure of a locally optimal detector for an i.i.d. noise model.

It is clear from (5.41) that, if we know the expression of the probability density function of the noise $f(x)$, we can easily determine the expression of $g_{lo}(x)$. For example:

- For a Gaussian noise model we have

$$f(x) \propto \exp \left[-\frac{1}{2\sigma^2} x^2 \right] \longrightarrow g_{lo}(x) = \frac{1}{\sigma^2} x \quad (5.42)$$

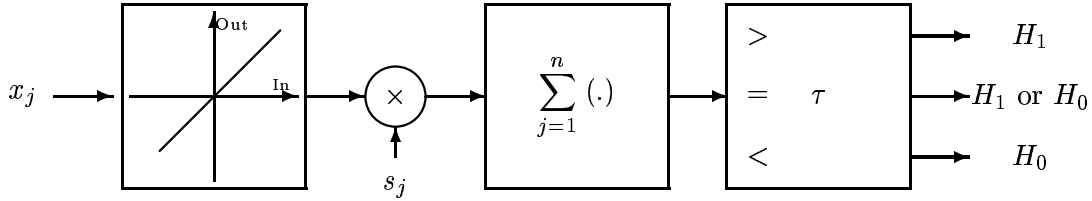


Figure 5.6: The structure of a locally optimal detector for an i.i.d. Gaussian noise model.

- For a Laplacian noise model we have

$$f(x) \propto \exp [-\alpha |x|] \longrightarrow g_{lo}(x) = \alpha \operatorname{sgn}(x) \quad (5.43)$$

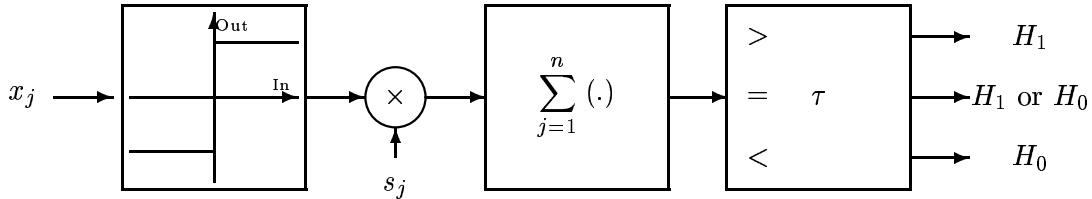


Figure 5.7: The structure of a locally optimal detector for an i.i.d. Laplacian noise model.

- For a Cauchy noise model we have

$$f(x) \propto \frac{1}{1+x^2} \longrightarrow g_{lo}(x) = \frac{2x}{1+x^2} \quad (5.44)$$

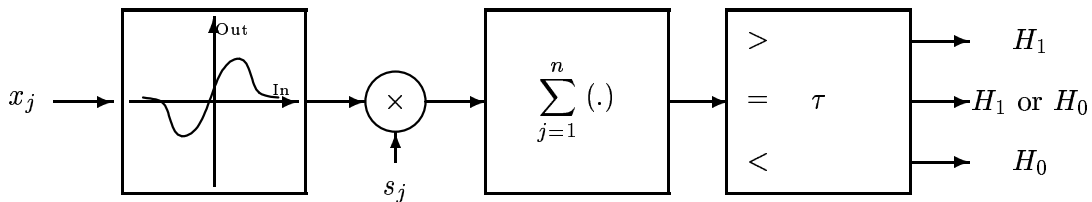


Figure 5.8: The structure of a locally optimal detector for an i.i.d. Cauchy noise model.

5.6 Detection of signals with unknown parameter

Here, we consider the case where the transmitted signals depend on some unknown parameter θ :

$$\begin{cases} H_0 & : X_j = N_j + S_{0j}(\theta) \\ H_1 & : X_j = N_j + S_{1j}(\theta) \end{cases}, \quad \mathbf{N} \sim f(\mathbf{n}) \quad (5.45)$$

The main quantity that we need to calculate is

$$L(\mathbf{x}) = \frac{\mathbb{E}_1 \{f(\mathbf{x} - \mathbf{s}_{1j}(\theta))\}}{\mathbb{E}_0 \{f(\mathbf{x} - \mathbf{s}_{0j}(\theta))\}} \quad (5.46)$$

In the following we consider the particular case of $\mathbf{s}_0 = \mathbf{0}$, $\mathbf{s}_1 = \mathbf{s}(\theta)$ and i.i.d. noise. Then we have

$$\begin{cases} H_0 & : X_j = N_j \\ H_1 & : X_j = N_j + S_j(\theta) \end{cases}, \quad N_j \sim f(n_j) \quad (5.47)$$

$$\begin{aligned} L(\mathbf{x}) &= \int_{\Gamma} \frac{f(\mathbf{x} - \mathbf{s}(\theta))}{f(\mathbf{x})} p(\theta) d\theta \\ &= \int_{\Gamma} \prod_{j=1}^n \frac{f(x_j - s_j(\theta))}{f(x_j)} p(\theta) d\theta \\ &= \int L_{\theta}(\mathbf{x}) p(\theta) d\theta \end{aligned} \quad (5.48)$$

Example: Non coherent detection:

Consider an amplitude modulated signal

$$s_j(\theta) = a_j \sin[(j-1)\omega_c T_s + \theta], \quad j = 1, \dots, n, \quad \omega_c T_s = m \frac{2\pi}{n} \quad (5.49)$$

where ω_c is the carrier frequency, T_s is the sampling rate, m is the number of periods in the observation time $[0, T = nT_s]$ and n/m is the number of samples per cycle. The carrier phase is unknown. With a uniform prior $p(\theta) = \frac{1}{2\pi}$ and the i.i.d. noise assumption we have

$$L(\mathbf{x}) = \frac{1}{2\pi} \int_0^{2\pi} \exp \left[\frac{1}{\sigma^2} \left[\sum_{j=1}^n x_j s_j(\theta) - \frac{1}{2} \sum_{j=1}^n s_j^2(\theta) \right] \right] d\theta \quad (5.50)$$

Using the trigonometric relations:

$$\sin(a+b) = \cos a \sin b + \sin a \cos b \quad (5.51)$$

$$\sin^2(a) = \frac{1}{2} - \frac{1}{2} \cos(2a) \quad (5.52)$$

we obtain:

$$\sum_{j=1}^n x_j s_j(\theta) = x_c \sin \theta + x_s \cos \theta \quad (5.53)$$

with

$$x_c \stackrel{\text{def}}{=} \sum_{j=1}^n a_j x_j \cos[(j-1)\omega_c T_s] \quad (5.54)$$

$$x_s \stackrel{\text{def}}{=} \sum_{j=1}^n a_j x_j \sin[(j-1)\omega_c T_s] \quad (5.55)$$

From (5.49) we have

$$\sum_{j=1}^n s_j^2(\theta) = \frac{1}{2} \sum_{j=1}^n a_j^2 + \frac{1}{2} \sum_{j=1}^n a_j^2 \cos[2(j-1)\omega_c T_s + 2\theta] \quad (5.56)$$

The second term is, in general either equal to zero or negligible with respect to the first term.

Noting $\overline{a^2} = \frac{1}{n} \sum_{j=1}^n a_j^2$, we obtain

$$L(\mathbf{x}) = \exp\left[-\frac{n\overline{a^2}}{4\sigma^2}\right] \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{1}{\sigma^2} [x_c \sin \theta + x_s \cos \theta]\right] d\theta \quad (5.57)$$

$$= \exp\left[-\frac{n\overline{a^2}}{4\sigma^2}\right] \frac{1}{2\pi} I_0\left(\frac{r}{\sigma^2}\right) \quad (5.58)$$

where I_0 is the zeroth-order Bessel function, which is monotone. Then the detection rule becomes

$$\begin{cases} 1 & > \\ \gamma & \text{if } L(\mathbf{x}) = 1 \\ 0 & < \end{cases} \longrightarrow \begin{cases} 1 & > \\ \gamma & \text{if } r = \tau' = \sigma^2 I_0^{-1}\left(\tau \exp\left[\frac{n\overline{a^2}}{4\sigma^2}\right]\right) \\ 0 & < \end{cases} \quad (5.59)$$

The structure of this detector is given in the following figure.

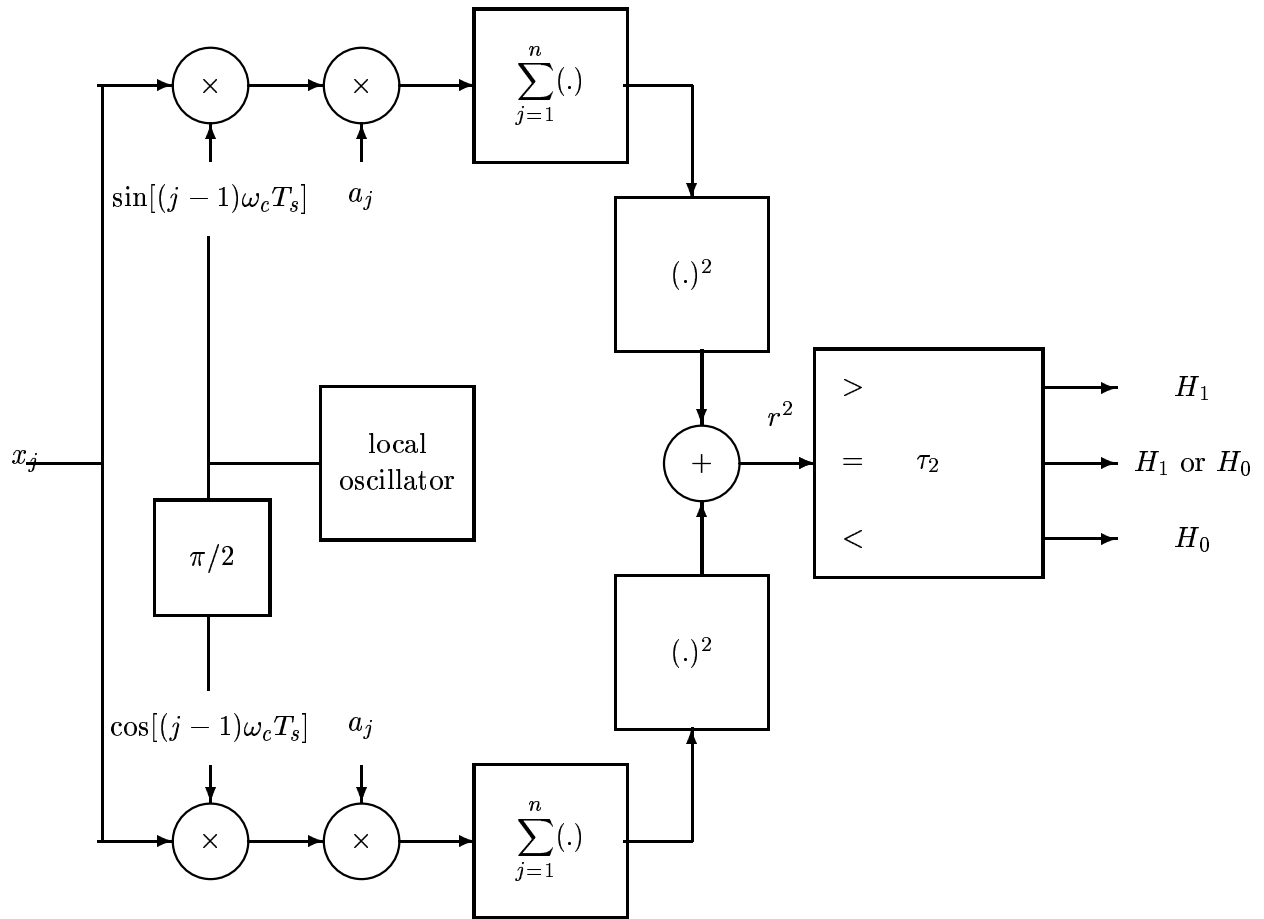


Figure 5.9: Coherent detector.

Performance analysis:

We need to calculate the probabilities such as

$$P_j(R > \tau') = P_j(R^2 > \tau'^2), \quad j = 0, 1 \quad (5.60)$$

with $R^2 = X_c^2 + X_s^2$.

Note that X_c and X_s are linear combinations of X_j . So, if X_j are Gaussian, X_c and X_s are Gaussian too.

Under the hypothesis H_0 , we have

$$\mathbb{E}[X_c|H_0] = \mathbb{E}[X_s|H_0] = 0 \quad (5.61)$$

$$\text{Var}[X_s|H_0] = \text{Var}[X_c|H_0] = \frac{n\sigma^2\bar{a}^2}{2} \quad (5.62)$$

$$\text{Cov}\{X_s, X_c|H_0\} = 0 \quad (5.63)$$

and

$$P_0(\Gamma_1) = \iint_{r^2=x_c^2+x_s^2 \geq \tau'^2} \frac{1}{n\pi\sigma^2\bar{a}^2} \exp\left[-\frac{1}{n\pi\sigma^2\bar{a}^2}(x_c^2 + x_s^2)\right] dx_c dx_s \quad (5.64)$$

With the cartesian to polar coordinate change $(x_c, x_s) \rightarrow (r, \theta)$ we obtain

$$\begin{aligned} P_0(\Gamma_1) &= \frac{1}{n\pi\sigma^2\bar{a}^2} \int_0^{2\pi} \int_{\tau'}^{\infty} r \exp\left[-\frac{r^2}{n\sigma^2\bar{a}^2}\right] dr d\theta \\ &= \frac{1}{n\pi\sigma^2\bar{a}^2} \exp\left[-\frac{\tau'^2}{n\sigma^2\bar{a}^2}\right] \end{aligned} \quad (5.65)$$

Under the hypothesis H_1 , noting that for a given value of θ $\mathbf{x}|\theta \sim \mathcal{N}(\mathbf{s}(\theta), \sigma^2\mathbf{I})$ we have

$$\mathbb{E}[X_c|H_1, \Theta = \theta] = \frac{n\bar{a}^2}{2} \sin \theta \quad (5.66)$$

$$\mathbb{E}[X_s|H_1, \Theta = \theta] = \frac{n\bar{a}^2}{2} \cos \theta \quad (5.67)$$

$$\text{Var}[X_s|H_1, \Theta = \theta] = \text{Var}[X_c|H_1, \Theta = \theta] = \frac{n\sigma^2\bar{a}^2}{2} \quad (5.68)$$

$$\text{Cov}\{X_s, X_c|H_1, \Theta = \theta\} = 0 \quad (5.69)$$

and

$$p(x_c, x_s|H_1) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{n\pi\sigma^2\bar{a}^2} \exp\left[-\frac{1}{n\sigma^2\bar{a}^2} q(x_c, x_s; n\bar{a}^2/2, \theta)\right] d\theta \quad (5.70)$$

$$= p(x_c, x_s|H_0) \exp\left[-\frac{n\bar{a}^2}{4\sigma^2}\right] I_0\left(\frac{r}{\sigma^2}\right) \quad (5.71)$$

The detection probability becomes then

$$P_D(\delta) = P_1(\Gamma_1) = \iint_{r^2=x_c^2+x_s^2 \geq \tau'^2} p(x_c, x_s|H_1) dx_c dx_s \quad (5.72)$$

$$\frac{\exp\left[-\frac{n\bar{a}^2}{4\sigma^2}\right]}{n\pi\sigma^2\bar{a}^2} \int_0^{2\pi} \int_{\tau'}^{\infty} r \exp\left[-\frac{r^2}{n\sigma^2\bar{a}^2}\right] I_0\left(\frac{r}{\sigma^2}\right) dr d\theta \quad (5.73)$$

Noting $b^2 = \frac{\overline{na^2}}{2\sigma^2}$ and $\tau_0 = \frac{\tau}{b\sigma^2}$ and changing the variable $x = \frac{r}{b\sigma^2}$ we obtain

$$P_D(\delta) = P_1(\Gamma_1) = \int_{r_0}^{\infty} x \exp\left[-\frac{1}{2}(x^2 + b^2)\right] I_0(bx) dx \stackrel{\text{def}}{=} Q(b, \tau_0) \quad (5.74)$$

$Q(b, \tau_0)$ is called Marcum's Q-function.

Note also that $P_f(\delta) = Q(0, \tau_0)$. So for a α -level Neyman-Pearson detection test we have $\tau' = \left[n\sigma^2\overline{a^2} \log(1/\alpha)\right]^{\frac{1}{2}}$ and the probability of detection is given by

$$P_D(\delta) = Q\left[b, 2[\log(1/\alpha)]^{\frac{1}{2}}\right] \quad (5.75)$$

Note also that

$$\text{E}\left[\frac{1}{n} \sum_{j=1}^n s_j^2(\theta)\right] = \overline{a^2}/2 \quad (5.76)$$

so, $b^2 = \frac{\overline{na^2}}{2\sigma^2}$ is a measure of S/N ratio.

5.7 sequential detection

$$\Gamma = \{X_j, j = 1, 2, \dots\} \quad (5.77)$$

$$\begin{cases} H_0 & : X_j \sim P_0 \\ H_1 & : X_j \sim P_1 \end{cases} \quad (5.78)$$

A sequential decision rule is a pair of sequences (Δ, δ) where :

- $\Delta = \{\Delta_j, j = 0, 1, 2, \dots\}$ is the sequence of stopping rules ;
- $\delta = \{\delta_j, j = 0, 1, 2, \dots\}$ is the sequence of decision rules ;
- $\Delta_j(x_1, \dots, x_j)$ is a function from \mathbf{R}^j to $\{0, 1\}$;
- $\delta_j(x_1, \dots, x_j)$ is a decision rule on $(\mathbf{R}^j, \mathcal{B}^j)$;
- If $\Delta_n(x_1, \dots, x_n) = 0$ we take another sample ;
- If $\Delta_n(x_1, \dots, x_n) = 1$ we stop sampling and make a decision.
- $N = \min\{n | \Delta_n(x_1, \dots, x_n) = 1\}$ is the stopping time ;
- $\Delta_N(x_1, \dots, x_N)$ is the terminal decision rule.
- (Δ_0, δ_0) correspond to the situation where we have not yet observed any data. $\Delta_0 = 0$ means take at least one sample before making a decision. $\Delta_0 = 1$ means make a decision without taking any sample.

Note that N is a random variable depending on the data sequence. The terminal decision rule $\delta_N(x_1, \dots, x_N)$ tells us which decision to make when we stop sampling.

The fixed-sample-size N decision rule can be defined as the following sequential detection rule:

$$\Delta_j(x_1, \dots, x_j) = \begin{cases} 0 & \text{if } j \neq N \\ 1 & \text{if } j = N \end{cases} \quad (5.79)$$

$$\delta_j(x_1, \dots, x_j) = \begin{cases} \delta(x_1, \dots, x_n) & \text{if } j = N \\ \text{arbitrary} & \text{if } j \neq N \end{cases} \quad (5.80)$$

In the following we consider only the binary hypothesis testing and we analyse the Bayesian approach with the prior distribution $\{\pi_0 = 1 - \pi_1, \pi_1\}$ and the uniform cost function. We assume that we can have an infinite number of i.i.d. observations at our disposal. However, we should assign a cost $c > 0$ to each sample, so that the cost of taking n samples is nc .

With these assumptions, the conditional risks for a given sequential decision rule are:

$$R_0(\Delta, \delta) = E_0 \{\delta(x_1, \dots, x_n)\} + E_0 \{cN\} \quad (5.81)$$

$$R_1(\Delta, \delta) = 1 - E_1 \{\delta(x_1, \dots, x_n)\} + E_1 \{cN\} \quad (5.82)$$

where the subscripts denote the hypotheses under which the expectation is computed and N is the stopping time. The Bayes risk is thus given by

$$r(\Delta, \delta) = (1 - \pi_1)R_0(\Delta, \delta) + \pi_1R_1(\Delta, \delta) \quad (5.83)$$

and the sequential Bayesian rule is the one which minimizes $r(\Delta, \delta)$.

To analyse the structure of this optimal rule we define

$$V^*(\pi_1) \stackrel{\text{def}}{=} \min_{\substack{\Delta, \delta \\ \Delta_0 = 0}} r(\Delta, \delta), \quad 0 \leq \pi_1 \leq 1. \quad (5.84)$$

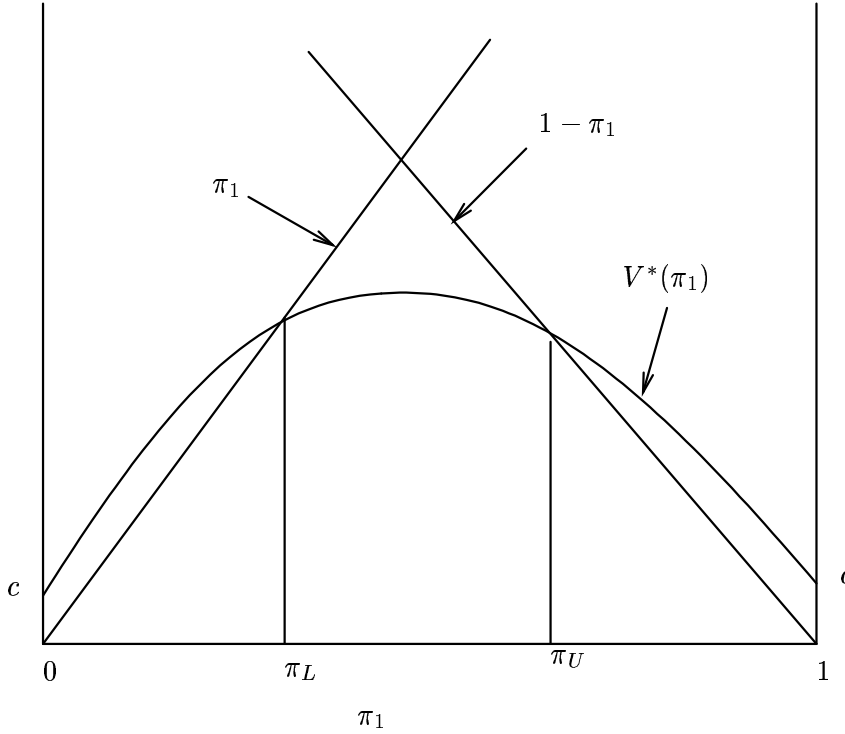


Figure 5.10: Sequential detection.

Since $\Delta_0 = 0$ means that the test does not stop with zero observation, $V^*(\pi_1)$ corresponds then to the minimum Bayes risk over all sequential tests that take at least one sample. $V^*(\pi_1)$ is in general concave and continuous and $V^*(0) = V^*(1) = c$. Now, let plot this function as well as these two specific sequential tests:

- Take no sample and decide H_0 , i.e., $\Delta_0 = 1, \delta_0 = 0$ and
- Take no sample and decide H_1 , i.e., $\Delta_0 = 1, \delta_0 = 1$.

Note that the Bayes risks for these tests are

$$r(\Delta, \delta)|_{\Delta_0=1, \delta_0=0} = 1 - \pi_1$$

$$r(\Delta, \delta)|_{\Delta_0=1, \delta_0=1} = \pi_1$$

These tests are the only two Bayesian tests that are not included in the minimization of (5.84). We note, respectively by π_U and π_L the abscisses of the intersections of the lines $r(\Delta, \delta)|_{\Delta_0=1, \delta_0=0}$ and $r(\Delta, \delta)|_{\Delta_0=1, \delta_0=1}$ with $V^*(\pi_1)$.

Now, by inspection of these plots, we see that the Bayes rule with a fixed given prior π_1 is:

- $(\Delta_0 = 1, \delta_0 = 0)$ if $\pi_1 \leq \pi_L$;
- $(\Delta_0 = 1, \delta_0 = 1)$ if $\pi_1 \geq \pi_U$;
- The decision rule which minimizes the Bayes risk among all the tests such that $(\Delta_0 = 0)$ corresponds to a point such that $\pi_L \leq \pi_1 \leq \pi_U$.

In the two first cases the test is stopped. In the third one, we know that the optimal test takes at least one more sample. After doing so, we are faced to a similar situation as before except that we now have more information due to the additional sample. In particular, the prior π_1 is replaced by $\pi_1(x_1) = \Pr\{H = H_1 | X_1 = x_1\}$ which is the posterior probability of H_1 given the observation $X_1 = x_1$. We can apply this method to any arbitrary number of samples. We then have the following rules:

- **Stopping rule:**

$$\Delta_n(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } \pi_L < \pi_1(x_1, \dots, x_n) < \pi_U \\ 1 & \text{otherwise.} \end{cases} \quad (5.85)$$

- **Terminal decision rule:**

$$\delta_n(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } \pi_1(x_1, \dots, x_n) \leq \pi_L \\ 1 & \text{if } \pi_1(x_1, \dots, x_n) \geq \pi_U. \end{cases} \quad (5.86)$$

It has been proved that under mild conditions the posterior probability $\pi_1(x_1, \dots, x_n)$ converges almost surely to 1 under H_1 and to 0 under H_0 . Thus the test terminates with probability one. The only knowledge of the probabilities π_L and π_U and an algorithm to compute $\pi_1(x_1, \dots, x_n)$ are sufficient to define this rule. The computation of $\pi_1(x_1, \dots, x_n)$ is quite easy, but unfortunately, it is very difficult to obtain exactly π_L and π_U .

Now consider the case where the two processes P_0 and P_1 have densities f_0 and f_1 . Then the Baye formula yields

$$\begin{aligned} \pi_1(x_1, \dots, x_n) &= \frac{\pi_1 \prod_{j=1}^n f_1(x_j)}{\pi_0 \prod_{j=1}^n f_0(x_j) + \pi_1 \prod_{j=1}^n f_1(x_j)} \\ &= \frac{\pi_1 \lambda_n(x_1, \dots, x_n)}{\pi_0 + \pi_1 \lambda_n(x_1, \dots, x_n)} \end{aligned} \quad (5.87)$$

where

$$\lambda_n(x_1, \dots, x_n) = \prod_{j=1}^n \frac{f_1(x_j)}{f_0(x_j)} \quad (5.88)$$

is the likelihood ratio based on n samples.

Noting that $\pi_1(x_1, \dots, x_n)$ is an increasing function of $\lambda_n(x_1, \dots, x_n)$ we can rewrite (5.85) and (5.86) as:

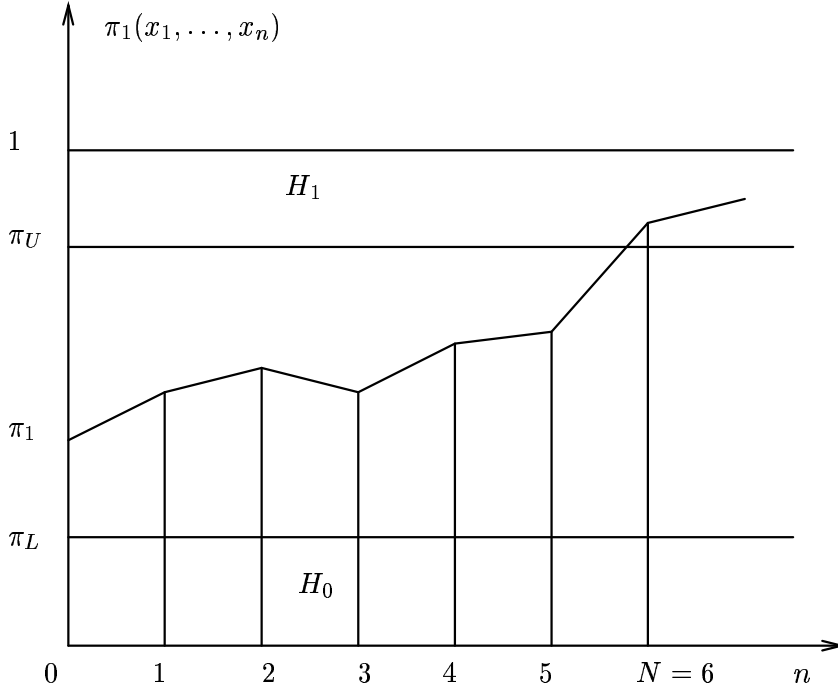


Figure 5.11: Stopping rule in sequential detection.

- **Stopping rule:**

$$\Delta_n(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } \underline{\pi} < \lambda_n(x_1, \dots, x_n) < \bar{\pi} \\ 1 & \text{otherwise.} \end{cases} \quad (5.89)$$

- **Terminal decision rule:**

$$\delta_n(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } \lambda_n(x_1, \dots, x_n) \leq \underline{\pi} \\ 1 & \text{if } \lambda_n(x_1, \dots, x_n) \geq \bar{\pi}. \end{cases} \quad (5.90)$$

where

$$\underline{\pi} \stackrel{\text{def}}{=} \frac{\pi_0 \pi_L}{\pi_1 (1 - \pi_L)} \quad \text{and} \quad \bar{\pi} \stackrel{\text{def}}{=} \frac{\pi_0 \pi_U}{\pi_1 (1 - \pi_U)}. \quad (5.91)$$

In conclusion, the Bayesian sequential test takes samples until the likelihood ratio falls outside the interval $[\underline{\pi}, \bar{\pi}]$ and decides H_0 or H_1 if $\lambda_n(x_1, \dots, x_n)$ falls outside of this interval.

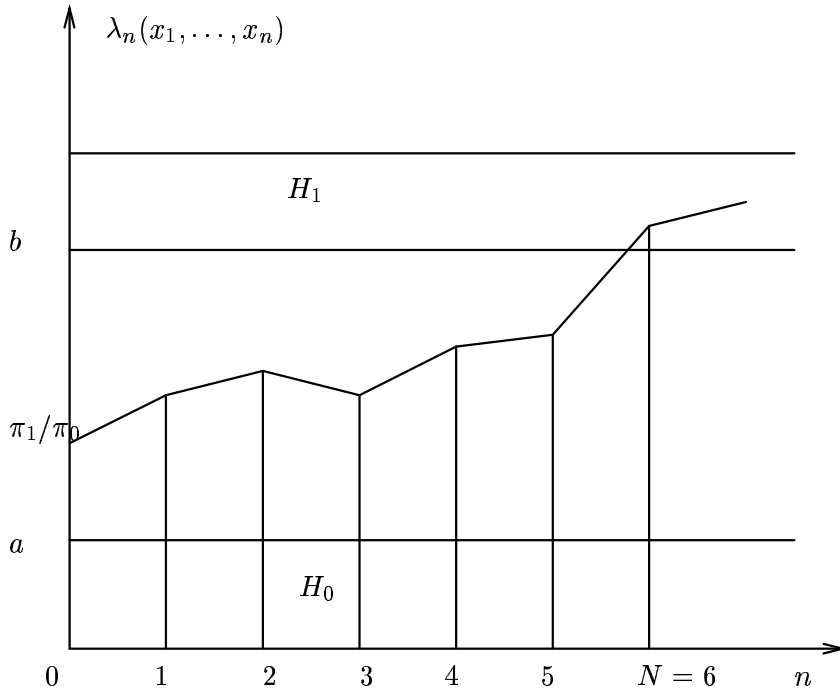
The main problem in practical situations is to fix the values of the boundaries $a = \underline{\pi}$ and $b = \bar{\pi}$. This test is called the *sequential probability ratio test* with the boundaries a and b and is noted $SPART(a, b)$.

The following theorem gives some of the optimality properties of $SPART(a, b)$.

Wald-Wolfowitz theorem :

Note by

$$N(\Delta) = \min\{n | \Delta_n(x_1, \dots, x_n) = 1\}$$

Figure 5.12: Stopping rule in $SPART(a, b)$.

$$P_F(\Delta, \delta) = \Pr\{\delta_N(x_1, \dots, x_N) = 1 | H = H_0\}$$

$$P_M(\Delta, \delta) = \Pr\{\delta_N(x_1, \dots, x_N) = 0 | H = H_1\}$$

and (Δ^*, δ^*) the $SPART(a, b)$. Then, for any sequential decision rule (Δ, δ) for which

$$P_F(\Delta, \delta) \leq P_F(\Delta^*, \delta^*)$$

$$P_M(\Delta, \delta) \leq P_M(\Delta^*, \delta^*)$$

we have

$$\mathbb{E}[N(\Delta) | H = H_j] \geq \mathbb{E}[N(\Delta^*) | H = H_j], \quad j = 0, 1$$

The validity of Wald-Wolfowitz theorem is a consequence of the Bayes optimality of $SPART(a, b)$. The results of this theorem and other related theorems are summarized in the following items:

- For a given performance, there is no other sequential decision rule with a smaller expected sample size than the $SPART(a, b)$ with the same performance.
- The average sample size of $SPART(a, b)$ is not greater than the sample size of a fixed-sample-size test with the same performance.
- For a given expected sample size, no sequential decision rule has smaller error probabilities than the $SPART(a, b)$.

Two main questions remains:

- How to choose a and b to yield a desired level of performance?
- How to evaluate the expected sample size of a sequential detector?

The following result gives an answer to the first one.

Let $(\Delta, \delta) = SPART(a, b)$ with $a < 1 < b$ and $\alpha = P_F(\Delta, \delta)$, $\gamma = 1 - \beta = P_M(\Delta, \delta)$ and $N = N(\Delta)$. Then the rejection region of (Δ, δ) is

$$\Gamma_1 = \left\{ \mathbf{x} \in \mathbb{R}^\infty \mid \lambda_N(x_1, \dots, x_N) \geq b \right\} = \cup_{n=1}^{\infty} Q_n \quad (5.92)$$

with

$$Q_n = \left\{ \mathbf{x} \in \mathbb{R}^\infty \mid N = n, \lambda_n(x_1, \dots, x_N) \geq b \right\} = \cup_{n=1}^{\infty} Q_n \quad (5.93)$$

Since Q_n and Q_m are mutually exclusive sets for $m \neq n$, we have

$$\alpha = \Pr \{ \lambda_n(x_1, \dots, x_N) \geq b \mid H = H_0 \} = \sum_{n=1}^{\infty} \int_{Q_n} \prod_{j=1}^n f_0(x_j) dx_j \quad (5.94)$$

On Q_n we have

$$\prod_{j=1}^n f_0(x_j) dx_j \leq \frac{1}{b} \prod_{j=1}^n f_1(x_j) dx_j \quad (5.95)$$

So, we have

$$\alpha \leq \frac{1}{b} \sum_{n=1}^{\infty} \int_{Q_n} \prod_{j=1}^n f_1(x_j) dx_j = \frac{1}{b} \Pr \{ \lambda_n(x_1, \dots, x_N) \geq b \mid H = H_0 \} = \frac{1}{b} (1 - \gamma) \quad (5.96)$$

and in the same manner we obtain

$$\gamma = \Pr \{ \lambda_n(x_1, \dots, x_N) \leq a \mid H = H_1 \} \leq a (1 - \alpha) \quad (5.97)$$

From these two relations we deduce

$$\begin{cases} b < \frac{1-\gamma}{\alpha} \\ a > \frac{\gamma}{1-\alpha} \end{cases} \quad (5.98)$$

The following choice is called the Wald's approximation:

$$\begin{cases} b \simeq \frac{1-\gamma}{\alpha} \\ a \simeq \frac{\gamma}{1-\alpha} \end{cases} \quad (5.99)$$

To be completed later

5.8 Robust detection

To be completed later