

EE6617: 07 Statistical Decision Theory I

Major references:

- *Fundamentals of Statistical Processing, Volume II: Detection Theory*, 1/E, by Steven M. Kay, Prentice Hall, 1993
- *Statistical Signal Processing*, 1/E, by Louis L. Scharf, Prentice Hall, 1991

We now move to discuss about the basic detection theory. The method follows directly from the theory of hypothesis testing.

1 Detection Problem Setup

Consider a simple **example**. We observe realization of a random variable $x[0] \sim \mathcal{N}(\mu, \sigma^2)$.

- We don't know μ .
- We want to **determine** $\mu \in \{0, \mu_1\}$.

We have two hypothesis

$$\begin{aligned}\mathcal{H}_0 : \quad & \mu = 0 \text{ (null hypothesis)} \\ \mathcal{H}_1 : \quad & \mu = \mu_1 \text{ (alternative hypothesis)}\end{aligned}$$

There are two kinds of errors

Decide $\mathcal{H}_1$ when $\mathcal{H}_0$ is true (Type I error or **false alarm**)
Decide $\mathcal{H}_0$ when $\mathcal{H}_1$ is true (Type II error or **mis-detection**)

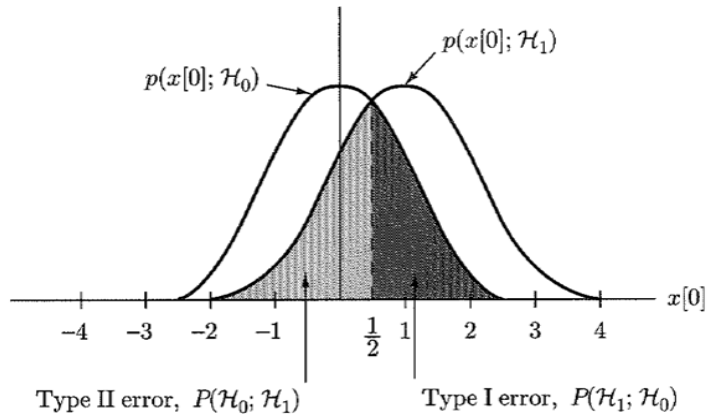
Outcome probability

$$p(\mathcal{H}_i; \mathcal{H}_j) : \text{probability that } \mathcal{H}_i \text{ decided when } \mathcal{H}_j \text{ is true}$$

Denote

$$\begin{aligned}p(\mathcal{H}_1; \mathcal{H}_0) : & \text{probability of false alarm } P_{FA} \\ p(\mathcal{H}_1; \mathcal{H}_1) : & \text{probability of detection } P_D \\ p(\mathcal{H}_0; \mathcal{H}_1) : & \text{probability of mis-detection } P_M\end{aligned}\tag{1}$$

Q: What is a reasonable way to decide which hypothesis is true?



From the figure we **may pick a threshold γ** such that

Decide $\mathcal{H}_0$ false if $x[0] > \gamma$
 $\mathcal{H}_0$ true if $x[0] < \gamma$.

Then,

$$\begin{aligned} P_{FA} &= p(\mathcal{H}_1; \mathcal{H}_0) = \int_{\gamma}^{\infty} p(t; \mathcal{H}_0) dt \\ P_D &= p(\mathcal{H}_1; \mathcal{H}_1) = \int_{\gamma}^{\infty} p(t; \mathcal{H}_1) dt \\ P_M &= 1 - P_D. \end{aligned}$$

Some questions arise

Q1: What is the **“optimal” γ** ?

Q2: Is **using a threshold the “best” way** to perform detection?

Ad hoc analysis doesn't always work.

Consider a general set-up. Given N observations $\mathbf{x} = [x[0], \dots, x[N-1]]^T$, decide whether $\mathcal{H}_0$ should be accepted or rejected. $\mathcal{H}_0$ & $\mathcal{H}_1$ are fixed but arbitrary.

What is a detector?

$$\text{Observation space } \mathbb{R}^N \xrightarrow{\text{detector}} \text{Hypothesis } \{\mathcal{H}_0, \mathcal{H}_1\} \quad (2)$$

Detector partitions $\mathbb{R}^N$ into regions

$$\begin{aligned} \mathcal{R}_0 &= \{\mathbf{x} : \text{accept } \mathcal{H}_0\} \\ \mathcal{R}_1 &= \{\mathbf{x} : \text{accept } \mathcal{H}_1\} \end{aligned}$$

Note $\mathcal{R}_0 \cup \mathcal{R}_1 = \mathbb{R}^N$. Then,

$$\begin{aligned} P_{FA} &= \int_{\mathcal{R}_1} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} \\ P_D &= \int_{\mathcal{R}_1} p(\mathbf{x}; \mathcal{H}_1) d\mathbf{x}. \end{aligned}$$

Designing the detector is equivalent to come up with $\mathcal{R}_0$ and $\mathcal{R}_1$.

2 Neyman-Pearson Detector

To construct detector, set the **target false alarm rate** (or called significant level of the test)

$$P_{FA} = \alpha$$

and **pick $\mathcal{R}_1$ to maximize P_D such that $P_{FA} = \alpha$.**

Theorem 1 To **maximize P_D such that $P_{FA} = \alpha$, decide $\mathcal{H}_1$ if**

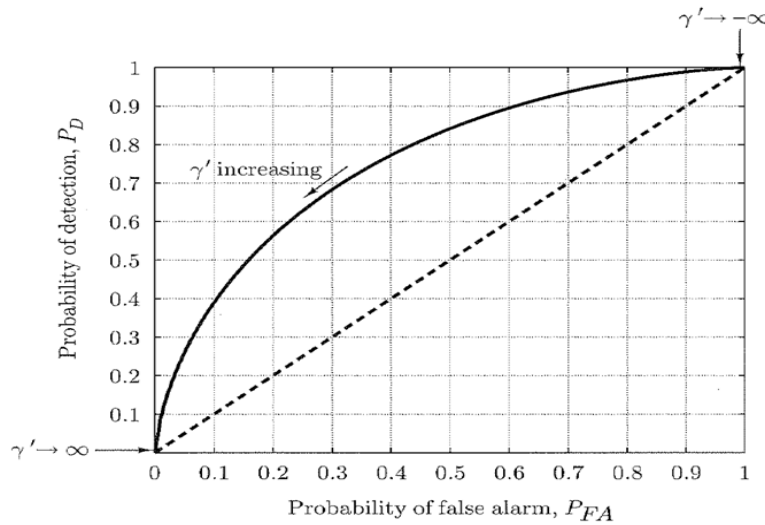
$$L(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} > \gamma$$

where γ **is found s.t.**

$$P_{FA} = \int_{\{\mathbf{x}: L(\mathbf{x}) > \gamma\}} p(\mathbf{x}; \mathcal{H}_0) d\mathbf{x} = \alpha \quad (3)$$

This detector is called **Neyman-Pearson (NP) detector**.

Remark 1 The P_D and P_{FA} monotonically increase as γ decreases, i.e., $P_D \nearrow$ **and** $P_{FA} \nearrow$ **as** $\gamma \searrow$. Thus, comparing different detectors, the NP detector is the one with the highest operating characteristic as shown below.



Example 1: Signal Detection in Noise

Want to decide if DC signal is present

$$\begin{aligned}\mathcal{H}_0 : & \quad x[n] = w[n] \\ \mathcal{H}_1 : & \quad x[n] = A + w[n]\end{aligned}$$

for $n = 0, \dots, N-1$ with $A > 0$, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, which is equivalent to

$$\begin{aligned}\mathcal{H}_0 : & \quad x[n] \sim \mathcal{N}(0, \sigma^2) \\ \mathcal{H}_1 : & \quad x[n] \sim \mathcal{N}(A, \sigma^2)\end{aligned}$$

The likelihoods are

$$\begin{aligned}p(\mathbf{x}; \mathcal{H}_0) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right) \\ p(\mathbf{x}; \mathcal{H}_1) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right).\end{aligned}$$

$$\begin{aligned}L(\mathbf{x}) &= \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\ &= \exp\left(\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - \sum_{n=0}^{N-1} (x[n] - A)^2\right)\right) \\ &= \exp\left(\frac{1}{2\sigma^2} \left(-NA^2 + 2A \sum_{n=0}^{N-1} x[n]\right)\right) > \gamma.\end{aligned}$$

Taking logs and moving terms

$$\begin{aligned}\frac{1}{2\sigma^2} \left(-NA^2 + 2A \sum_{n=0}^{N-1} x[n]\right) &> \ln(\gamma) \\ \sum_{n=0}^{N-1} x[n] &> \frac{\sigma^2}{A} \ln \gamma + \frac{NA}{2}\end{aligned}$$

or rather

$$\underbrace{T(\mathbf{x})}_{\text{Test statistics}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] > \underbrace{\frac{\sigma^2}{NA} \ln \gamma + \frac{A}{2}}_{\gamma'}$$

To get the distribution of $T(\mathbf{x})$ for each hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, we write

$$\mathcal{H}_0 : T(\mathbf{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \frac{1}{N} \sum_{n=0}^{N-1} w[n], \quad w[n] \sim \mathcal{N}(0, \sigma^2) \quad (4)$$

$$\begin{aligned}
E[T(\mathbf{x})] &= \frac{1}{N} \sum_{n=0}^{N-1} E[w[n]] = 0 \\
E[T^2(\mathbf{x})] &= \frac{1}{N^2} \sum_{n=0}^{N-1} E[w^2[n]] = \frac{1}{N^2} \sum_{n=0}^{N-1} \sigma^2 = \frac{\sigma^2}{N} \\
\mathcal{H}_0 : \quad T(\mathbf{x}) &\sim \mathcal{N}(0, \frac{\sigma^2}{N})
\end{aligned}$$

and

$$\mathcal{H}_1 : T(\mathbf{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \frac{1}{N} \sum_{n=0}^{N-1} (A + w[n]), \quad x[n] \sim \mathcal{N}(A, \sigma^2) \quad (5)$$

$$\begin{aligned}
E[T(\mathbf{x})] &= \frac{1}{N} \sum_{n=0}^{N-1} E[x[n]] = A \\
E[(T(\mathbf{x}) - A)^2] &= \frac{1}{N^2} \sum_{n=0}^{N-1} E[(A + w[n] - A)^2] = \frac{\sigma^2}{N} \\
\mathcal{H}_1 : \quad T(\mathbf{x}) &\sim \mathcal{N}(A, \frac{\sigma^2}{N}).
\end{aligned}$$

We now **determine γ' to give $P_{FA} = \alpha$** .

$$\begin{aligned}
P_{FA} &= \Pr(T(\mathbf{x}) \geq \gamma'; \mathcal{H}_0) \\
&= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left(-\frac{1}{2\sigma^2/N} t^2\right) dt \\
&\quad \text{change variable } u = \frac{t}{\sqrt{\sigma^2/N}}, \quad du = \frac{dt}{\sqrt{\sigma^2/N}} \\
&= \int_{\frac{\gamma'}{\sqrt{\sigma^2/N}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\
&= Q\left(\frac{\gamma'}{\sqrt{\sigma^2/N}}\right) = \alpha
\end{aligned}$$

where $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du$. Then, we decide γ' such that $\frac{\gamma'}{\sqrt{\sigma^2/N}} = Q^{-1}(\alpha)$ and $\gamma' = \sqrt{\sigma^2/N} Q^{-1}(\alpha)$.

We can characterize the **detection performance**

$$\begin{aligned}
P_D &= \Pr(T(\mathbf{x}) \geq \gamma'; \mathcal{H}_1) \\
&= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left(-\frac{1}{2\sigma^2/N}(t-A)^2\right) dt \\
&\quad \text{change variable } u = \frac{t-A}{\sqrt{\sigma^2/N}}, du = \frac{dt}{\sqrt{\sigma^2/N}} \\
&= Q\left(\frac{\gamma' - A}{\sqrt{\sigma^2/N}}\right) = Q\left(Q^{-1}(\alpha) - \frac{A}{\sqrt{\sigma^2/N}}\right).
\end{aligned}$$

Q: Is P_D the maximum detection probability given $P_{FA} = \alpha$? ■

3 Minimum Error Probability Detector

This is different approach from NP. Now, we assume **the hypotheses are random variables**. We allocate probabilities to $\mathcal{H}_0$ and $\mathcal{H}_1$, i.e., priors $p(\mathcal{H}_0)$ and $p(\mathcal{H}_1)$. This comes from the **Bayesian approach to detection**. Notice that the **NP detector had no notion of $p(\mathcal{H}_0)$ and $p(\mathcal{H}_1)$** .

In some systems, the probabilities are known or can be well approximated.

By Bayesian framework, we can define a **probability of error P_e** as

$$\begin{aligned}
P_e &= p(\mathcal{H}_1|\mathcal{H}_0)p(\mathcal{H}_0) + p(\mathcal{H}_0|\mathcal{H}_1)p(\mathcal{H}_1) \\
p(\mathcal{H}_0|\mathcal{H}_1) &= \int_{\mathcal{R}_0} p(\mathbf{x}|\mathcal{H}_1)d\mathbf{x} = 1 - \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_1)d\mathbf{x} \\
p(\mathcal{H}_1|\mathcal{H}_0) &= \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_0)d\mathbf{x}.
\end{aligned}$$

The P_e can be rewritten **in terms of $\mathcal{R}_1$** by

$$\begin{aligned}
P_e &= p(\mathcal{H}_0) \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_0)d\mathbf{x} + p(\mathcal{H}_1) \left(1 - \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_1)d\mathbf{x}\right) \\
&= p(\mathcal{H}_1) + \int_{\mathcal{R}_1} (p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) - p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1)) d\mathbf{x}
\end{aligned} \tag{6}$$

To minimize P_e , we want the second term of (6) to be as negative as possible. This occurs when the integrand is negative for all $\mathbf{x} \in \mathcal{R}_1$, i.e.,

$$p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) - p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1) < 0 \Rightarrow p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) < p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1), \quad \forall \mathbf{x} \in \mathcal{R}_1$$

Thus, the **minimum error probability detection criterion** is

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \underset{0}{\geqslant}_1 \frac{p(\mathcal{H}_0)}{p(\mathcal{H}_1)}.$$

4 Bayes Risk Detection

More general cost function than probability of error is **Bayes Risk** which is defined as

$$R = \underbrace{E[c]}_{\text{average cost}} = \sum_{i=0}^1 \sum_{j=0}^1 C_{ij} p(\mathcal{H}_i | \mathcal{H}_j) p(\mathcal{H}_j)$$

Notice that $R = P_e$ if $C_{01} = C_{10} = 1$ and $C_{11} = C_{00} = 0$ for $\{\mathcal{H}_0, \mathcal{H}_1\}$.

Consider

$$\begin{aligned} R &= C_{00}p(\mathcal{H}_0)p(\mathcal{H}_0|\mathcal{H}_0) + C_{01}p(\mathcal{H}_1)p(\mathcal{H}_0|\mathcal{H}_1) + C_{10}p(\mathcal{H}_0)p(\mathcal{H}_1|\mathcal{H}_0) + C_{11}p(\mathcal{H}_1)p(\mathcal{H}_1|\mathcal{H}_1) \quad (7) \\ &= C_{00}p(\mathcal{H}_0) \int_{\mathcal{R}_0} p(\mathbf{x}|\mathcal{H}_0) d\mathbf{x} + C_{01}p(\mathcal{H}_1) \int_{\mathcal{R}_0} p(\mathbf{x}|\mathcal{H}_1) d\mathbf{x} \\ &\quad + C_{10}p(\mathcal{H}_0) \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_0) d\mathbf{x} + C_{11}p(\mathcal{H}_1) \int_{\mathcal{R}_1} p(\mathbf{x}|\mathcal{H}_1) d\mathbf{x} \end{aligned}$$

The R can be rewritten in terms of $\mathcal{R}_1$ by

$$R = C_{00}p(\mathcal{H}_0) + C_{01}p(\mathcal{H}_1) + \int_{\mathcal{R}_1} ((C_{10} - C_{00})p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) + (C_{11} - C_{01})p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1)) d\mathbf{x}$$

Make the second term as negative as possible. This occurs when the integrand is negative for all $\mathbf{x} \in \mathcal{R}_1$.

$$\begin{aligned} (C_{10} - C_{00})p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) + (C_{11} - C_{01})p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1) &< 0 \\ \Rightarrow (C_{10} - C_{00})p(\mathcal{H}_0)p(\mathbf{x}|\mathcal{H}_0) &< (C_{01} - C_{11})p(\mathcal{H}_1)p(\mathbf{x}|\mathcal{H}_1) \end{aligned}$$

If $C_{01} > C_{11}$ and $C_{10} > C_{00}$,

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \gtrless_0^1 \frac{(C_{10} - C_{00})p(\mathcal{H}_0)}{(C_{01} - C_{11})p(\mathcal{H}_1)}.$$

This is the minimum P_e detector when $\frac{C_{10}-C_{00}}{C_{01}-C_{11}} = 1$.

Example 2: BPSK Detection in Noise

Consider a Bayes Risk detector with $C_{10} = 0.7, C_{00} = 0.3, C_{01} = 0.8, C_{11} = 0.2$. We want to detect if a noisy DC signal is $+A$ or $-A$ using N samples.

$$\begin{aligned} \mathcal{H}_0 : x[n] &= -A + w[n] \\ \mathcal{H}_1 : x[n] &= A + w[n] \end{aligned}$$

for $n = 0, \dots, N-1$ and $w[n] \sim \mathcal{N}(0, \sigma^2)$. Assume $p(+A) = p(\mathcal{H}_1) = 0.4$ and $p(-A) = p(\mathcal{H}_0) = 0.6$.

We have

$$\begin{aligned}
p(\mathbf{x}|\mathcal{H}_0) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] + A)^2\right) \\
p(\mathbf{x}|\mathcal{H}_1) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right). \\
\\
L(\mathbf{x}) &= \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \\
&= \exp\left(\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} (x[n] + A)^2 - \sum_{n=0}^{N-1} (x[n] - A)^2\right)\right) \\
&= \exp\left(\frac{2A}{\sigma^2} \sum_{n=0}^{N-1} x[n]\right) > \gamma.
\end{aligned}$$

where

$$\begin{aligned}
\gamma &= \frac{(C_{10} - C_{00})p(\mathcal{H}_0)}{(C_{01} - C_{11})p(\mathcal{H}_1)} \\
&= \frac{(0.7 - 0.3)0.6}{(0.8 - 0.2)0.4} = 1
\end{aligned}$$

Taking logs and moving terms

$$\begin{aligned}
\frac{2A}{\sigma^2} \sum_{n=0}^{N-1} x[n] &\geq_0^1 0 \\
T(\mathbf{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] &\geq_0^1 0
\end{aligned}$$

and $T(\mathbf{x}) \sim \mathcal{N}(-A, \frac{\sigma^2}{N})$ under $\mathcal{H}_0$ and $T(\mathbf{x}) \sim \mathcal{N}(A, \frac{\sigma^2}{N})$ under $\mathcal{H}_1$.

We compute the conditional error probabilities

$$\begin{aligned}
p(\mathcal{H}_1|\mathcal{H}_0) &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left(-\frac{1}{2\sigma^2/N}(t+A)^2\right) dt \\
&\quad \text{change variable } u = \frac{t+A}{\sqrt{\sigma^2/N}}, \quad du = \frac{dt}{\sqrt{\sigma^2/N}} \\
&= \int_{\frac{A}{\sqrt{\sigma^2/N}}}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\
&= Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right) \\
p(\mathcal{H}_0|\mathcal{H}_1) &= Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right)
\end{aligned}$$

From (7),

$$\begin{aligned}
R &= C_{01}p(\mathcal{H}_1)p(\mathcal{H}_0|\mathcal{H}_1) + C_{10}p(\mathcal{H}_0)p(\mathcal{H}_1|\mathcal{H}_0) + C_{00}p(\mathcal{H}_0)(1 - p(\mathcal{H}_1|\mathcal{H}_0)) + C_{11}p(\mathcal{H}_1)(1 - p(\mathcal{H}_0|\mathcal{H}_1)) \\
&= [0.7(0.6) + 0.8(0.4)]Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right) + [0.3(0.6) + 0.2(0.4)]\left(1 - Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right)\right) \\
&= 0.48Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right) + 0.26.
\end{aligned}$$

What is the average P_e ?

$$P_e = 1(0.6)Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right) + 1(0.4)Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right) = Q\left(\sqrt{\frac{NA^2}{\sigma^2}}\right).$$

5 Multiple Hypothesis Testing

So far we looked at binary hypothesis $\mathcal{H}_0$ & $\mathcal{H}_1$.

Many practical problem have $\{\mathcal{H}_0, \dots, \mathcal{H}_{M-1}\}$, $M > 2$ hypotheses (e.g., communication system M-ary constellation, character recognition).

For **Bayes risk**,

$$R = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij}p(\mathcal{H}_i|\mathcal{H}_j)p(\mathcal{H}_j)$$

Note that $R = P_e$ when $C_{ij} = 0$ if $i = j$ and $C_{ij} = 1$ if $i \neq j$.

To get the detector

$$\begin{aligned}
R &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} p(\mathcal{H}_i | \mathcal{H}_j) p(\mathcal{H}_j) d\mathbf{x} \\
&= \sum_{i=0}^{M-1} \int_{\mathcal{R}_i} \sum_{j=0}^{M-1} C_{ij} p(\mathbf{x} | \mathcal{H}_j) p(\mathcal{H}_j) d\mathbf{x} \\
&= \sum_{i=0}^{M-1} \int_{\mathcal{R}_i} \underbrace{\sum_{j=0}^{M-1} C_{ij} p(\mathcal{H}_j | \mathbf{x})}_{=C_i(\mathbf{x})} p(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

The **risk contribution by assigning $\mathbf{x}$ to some hypothesis $\mathcal{H}_i$** is

$$\sum_{j=0}^{M-1} C_{ij} p(\mathcal{H}_j | \mathbf{x}) p(\mathbf{x}) d\mathbf{x} = C_i(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

Then the **optimal assignment that minimizes the Bayes risk for a given $\mathbf{x}$** is to assign $\mathbf{x}$ to $\mathcal{H}_i$ such that

$$i = \underset{0 \leq \ell \leq M-1}{\operatorname{argmin}} C_\ell(\mathbf{x})$$

5.1 Optimal Bayes Risk Detector

Given $\mathbf{x}$ declare $\mathcal{H}_i$ is true when

$$i = \underset{0 \leq \ell \leq M-1}{\operatorname{argmin}} \sum_{j=0}^{M-1} C_{\ell j} p(\mathcal{H}_j | \mathbf{x}).$$

5.2 Optimal M-ary P_e Detector: Maximum A Posteriori Probability (MAP) detector

For $R = P_e$, $C_{ij} = \begin{cases} 1 & i \neq j \\ 0 & i = j \end{cases}$

$$\begin{aligned}
i &= \underset{0 \leq \ell \leq M-1}{\operatorname{argmin}} \sum_{j \neq \ell}^{M-1} p(\mathcal{H}_j | \mathbf{x}) \\
&= \underset{0 \leq \ell \leq M-1}{\operatorname{argmin}} \left(\underbrace{\sum_{j=0}^{M-1} p(\mathcal{H}_j | \mathbf{x})}_{=1} - p(\mathcal{H}_\ell | \mathbf{x}) \right) \\
&= \underset{0 \leq \ell \leq M-1}{\operatorname{argmax}} p(\mathcal{H}_\ell | \mathbf{x})
\end{aligned}$$

Given $\mathbf{x}$ declare $\mathcal{H}_i$ is true when

$$i = \underset{0 \leq \ell \leq M-1}{\operatorname{argmax}} p(\mathcal{H}_\ell | \mathbf{x})$$

This is a **maximum a posteriori probability (MAP) detector**.

5.3 Optimal M-ary P_e Detector: Maximum Likelihood (ML) detector

In many problems, especially for **communication sources**, it is reasonable to assume

$$p(\mathcal{H}_0) = p(\mathcal{H}_1) = \cdots = \frac{1}{M}.$$

By **Bayes Rule**, we can write

$$p(\mathcal{H}_i | \mathbf{x}) = \frac{p(\mathbf{x} | \mathcal{H}_i) p(\mathcal{H}_i)}{p(\mathbf{x})} = p(\mathbf{x} | \mathcal{H}_i) \underbrace{\frac{1}{Mp(\mathbf{x})}}_{\text{constant}}.$$

Thus,

$$\begin{aligned} i &= \underset{0 \leq \ell \leq M-1}{\operatorname{argmax}} p(\mathcal{H}_\ell | \mathbf{x}) \\ &= \underset{0 \leq \ell \leq M-1}{\operatorname{argmax}} p(\mathbf{x} | \mathcal{H}_\ell). \end{aligned}$$

This is a **maximum likelihood (ML) detector**.

The MAP detector is equivalent to the ML detector if the priors are equally likely, i.e., $p(\mathcal{H}_0) = p(\mathcal{H}_1) = \cdots = \frac{1}{M}$.

6 Known Signals in Noise: Matched Filter

We are trying to detect if a predetermined signal $s[0], \dots, s[N-1]$ is present in a noisy received signal $x[0], \dots, x[N-1]$.

$$\begin{aligned} \mathcal{H}_0 : & x[n] = w[n] \\ \mathcal{H}_1 : & x[n] = s[n] + w[n] \end{aligned}$$

for $n = 0, \dots, N-1$ and $w[n] \sim \mathcal{N}(0, \sigma^2)$.

Q: When the detector knows $s[0], \dots, s[N-1]$, what is the optimal detector?

A: We already know the **Neyman-Pearson detector is optimal**.

6.1 Optimal Neyman-Pearson Detector

The Neyman-Pearson detector uses

$$\begin{aligned}
 L(\mathbf{x}) &= \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \\
 &= \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} (x[n] - s[n])^2 - \sum_{n=0}^{N-1} x^2[n] \right) \right) \\
 &= \exp \left(-\frac{1}{2\sigma^2} \left(-2 \sum_{n=0}^{N-1} x[n]s[n] + \sum_{n=0}^{N-1} s^2[n] \right) \right) \gtrless_0^1 \gamma.
 \end{aligned}$$

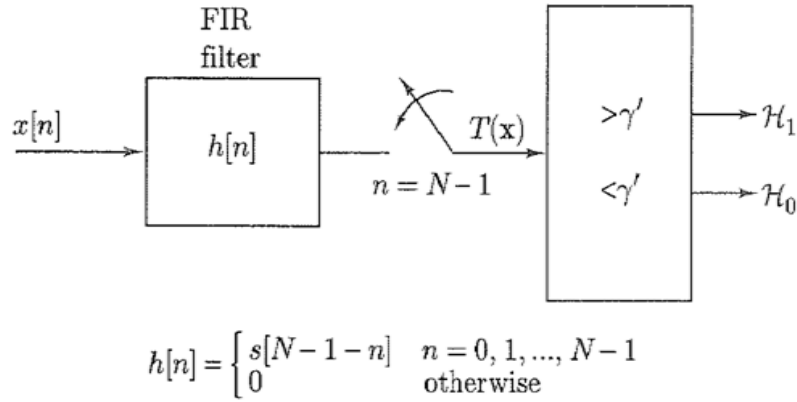
Using the log-likelihood

$$\begin{aligned}
 &\sum_{n=0}^{N-1} (2x[n]s[n] - s^2[n]) \gtrless_0^1 2\sigma^2 \ln(\gamma) \\
 \Rightarrow &\sum_{n=0}^{N-1} x[n]s[n] \gtrless_0^1 \underbrace{\sigma^2 \ln(\gamma) + \frac{1}{2} \sum_{n=0}^{N-1} s^2[n]}_{=\gamma'}
 \end{aligned}$$

The test statistic is

$$\begin{aligned}
 T(\mathbf{x}) &= \sum_{n=0}^{N-1} x[n]s[n] = \mathbf{x}^T \mathbf{s} \\
 &= \sum_{k=0}^{N-1} x[k]s[N-1-(N-1-k)] \\
 &= \sum_{k=0}^{N-1} x[k]h[N-1-k] \\
 &= \sum_{k=0}^{N-1} x[k]h[n-k] \Big|_{n=N-1} \\
 &= x[n] * h[n] \Big|_{n=N-1}
 \end{aligned}$$

where $h[n] = s[N-1-n]$. This is the well-known **matched filter**.



The **matched filter is the optimal Neyman-Pearson detector** to determine the presence of a known signal in a noisy observation (with i.i.d. $\mathcal{N}(0, \sigma^2)$ noise).

Q: Are there other detectors better than the matched filter?
We may try some other design criterion.

6.2 Maximum SNR Detector

One other way of determining detector is to use the **signal-to-noise ratio (SNR)**.

Q: What is the optimal $h[n]$, $n = 0, \dots, N-1$ that maximize SNR?

Try to maximize the SNR under $\mathcal{H}_1$.

$$SNR = \frac{\text{signal power under } \mathcal{H}_1}{\text{noise power under } \mathcal{H}_1}.$$

The signal power under $\mathcal{H}_1$ is

$$E^2 \left[\sum_{k=0}^{N-1} h[N-1-k]s[k] \right] = \left(\sum_{k=0}^{N-1} h[N-1-k]s[k] \right)^2$$

Let $y(N-1) = x[n] * h[n]|_{n=N-1}$. The noise power under $\mathcal{H}_1$ is

$$\begin{aligned}
\text{var}(y[N-1]) &= E[y^2(N-1)] - E[y(N-1)]^2 \\
&= \sum_{\ell=0}^{N-1} \sum_{k=0}^{N-1} h[N-1-\ell]h[N-1-k]E[x[\ell]x[k]] - \left(\sum_{k=0}^{N-1} h[N-1-k]s[k]\right)^2 \\
&= E\left[\left(\sum_{k=0}^{N-1} h[N-1-k]w[k]\right)^2\right] \\
&= \sum_{k=0}^{N-1} h^2[N-1-k]\sigma^2.
\end{aligned}$$

Thus, the SNR is

$$SNR(\mathbf{h}) = \frac{\left(\sum_{k=0}^{N-1} h[N-1-k]s[k]\right)^2}{\sum_{k=0}^{N-1} h^2[N-1-k]\sigma^2}.$$

The goal is to **design $\mathbf{h}$ to maximize $SNR(\mathbf{h})$** .

Let $\mathbf{h} = [h[N-1], h[N-2], \dots, h[0]]^T$ and the signal power $\|\mathbf{s}\|^2 = \xi$, then

$$SNR(\mathbf{h}) = \frac{(\mathbf{h}^T \mathbf{s})^2}{\sigma^2 \|\mathbf{h}\|^2} = \left(\frac{\mathbf{h}^T}{\|\mathbf{h}\|} \mathbf{s}\right)^2 \frac{1}{\sigma^2} \leq \|\mathbf{s}\|^2 / \sigma^2.$$

$$\max_{\mathbf{h}} SNR(\mathbf{h}) \leq \frac{\xi}{\sigma^2}.$$

This is achieved if $\mathbf{h} = c\mathbf{s}$ and without loss of any generality we set $\mathbf{h} = [h[N-1], h[N-2], \dots, h[0]]^T = [s[0], s[1], \dots, s[N-1]]^T$,

$$h[n] = s[N-1-n] \Rightarrow \text{matched filter}.$$

The matched filter is also optimal in terms of maximizing received SNR over all FIR filters and the matched filter SNR is $SNR = \frac{\xi}{\sigma^2}$.

The NP detector and max SNR detector are only the same when noise is i.i.d. $\mathcal{N}(0, \sigma^2)$.

6.3 Detector Performance

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]s[n] = \mathbf{x}^T \mathbf{s} = \begin{cases} \mathbf{w}^T \mathbf{s} & \text{if } \mathcal{H}_0 \\ (\mathbf{s} + \mathbf{w})^T \mathbf{s} & \text{if } \mathcal{H}_1 \end{cases}.$$

Under $\mathcal{H}_0$

$$\begin{aligned} E[T(\mathbf{x})] &= 0 \\ \text{var}(T(\mathbf{x})) &= E[\mathbf{s}^T \mathbf{w} \mathbf{w}^T \mathbf{s}] = \sigma^2 \|\mathbf{s}\|^2 = \sigma^2 \xi \\ T(\mathbf{x}) &\sim \mathcal{N}(0, \sigma^2 \xi). \end{aligned}$$

Under $\mathcal{H}_1$,

$$\begin{aligned} E[T(\mathbf{x})] &= E[(\mathbf{s} + \mathbf{w})^T \mathbf{s}] = \xi \\ \text{var}(T(\mathbf{x})) &= E[T(\mathbf{x})^2] - E[T(\mathbf{x})]^2 = E[\mathbf{s}^T \mathbf{w} \mathbf{w}^T \mathbf{s}] = \sigma^2 \xi \\ T(\mathbf{x}) &\sim \mathcal{N}(\xi, \sigma^2 \xi). \end{aligned}$$

For the target false alarm rate $P_{FA} = \alpha$, the threshold γ' can be determined such that

$$\begin{aligned} P_{FA} &= p(T(\mathbf{x}) > \gamma'; \mathbf{H}_0) \\ &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi\xi\sigma^2}} \exp\left(-\frac{1}{2\xi\sigma^2}t^2\right) dt \\ &= \int_{\frac{\gamma'}{\sqrt{\xi\sigma^2}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \\ &= Q\left(\frac{\gamma'}{\sqrt{\xi\sigma^2}}\right) = \alpha \end{aligned}$$

and $\gamma' = \sqrt{\xi\sigma^2}Q^{-1}(\alpha)$.

We can characterize the detection performance

$$\begin{aligned} P_D &= \Pr(T(\mathbf{x}) \geq \gamma'; \mathcal{H}_1) \\ &= \int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi\xi\sigma^2}} \exp\left(-\frac{1}{2\xi\sigma^2}(t - \xi)^2\right) dt \\ &= Q\left(\frac{\gamma' - \xi}{\sqrt{\xi\sigma^2}}\right) = Q\left(Q^{-1}(\alpha) - \frac{\xi}{\sqrt{\xi\sigma^2}}\right). \end{aligned}$$

This P_D is the maximum detection probability given $P_{FA} = \alpha$ by the Neyman-Pearson theorem.

Q: What if we want to choose $\mathbf{s}$ to maximize the detector performance subject to a fixed $\|\mathbf{s}\|^2 = \xi$?

Answer is any vector on the N -dimensional sphere of radius $\sqrt{\xi}$ is optimal.