

## Decision and Estimation in Data Processing

## Chapter II. Elements of Signal Detection Theory

## II.1 Introduction

# Introduction

- ▶ Signal detection = the problem of deciding which signal is present from 2 or more possibilities
  - ▶ one possibility may be that there is no signal
- ▶ Based on **noisy** observations
  - ▶ signals are affected by noise
  - ▶ noise is additive (added to the original signal)

# The model for signal detection

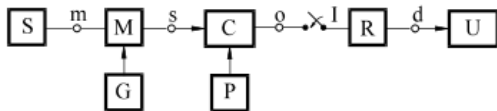


Figure 1: Signal detection model

## ► Contents:

- Information source: generates messages  $a_n$  with probabilities  $p(a_n)$
- Generator: generates different signals  $s_1(t), \dots, s_n(t)$
- Modulator: transmits a signal  $s_n(t)$  for message  $a_n$
- Channel: adds random noise
- Sampler: takes samples from the signal  $s_n(t)$
- Receiver: **decides** what message  $a_n$  has been transmitted
- User receives the recovered messages

# Practical scenarios

## ▶ Data transmission

- ▶ constant voltage levels (e.g.  $s_n(t) = \text{constant} = 0$  or  $5V$ )
- ▶ PSK modulation (Phase Shift Keying):  $s_n(t) = \text{cosine}$  with same frequency but various initial phases
- ▶ FSK modulation (Frequency Shift Keying):  $s_n(t) = \text{cosines}$  with different frequencies
- ▶ OFDM modulation (Orthogonal Frequency Division Multiplexing): particular case of FSK

## ▶ Radar

- ▶ a signal is emitted; if there is an obstacle, the signal gets reflected back
- ▶ the receiver waits for possible reflections of the signal and must decide
  - ▶ no reflection is present -> no object
  - ▶ reflected signal is present -> object detected

# Generalizations

- ▶ Decide between more than two signals
- ▶ Number of observations:
  - ▶ use only one sample
  - ▶ use multiple samples
  - ▶ observe the whole continuous signal for some time  $T$

## II.2 Detection of signals based on 1 sample



# Detection of a signal with 1 sample

- ▶ Simplest case: detection of a signal contaminated with noise using 1 sample
  - ▶ two messages  $a_0$  and  $a_1$
  - ▶ messages are encoded as signals  $s_0(t)$  and  $s_1(t)$ 
    - ▶ for  $a_0$ : send  $s(t) = s_0(t)$
    - ▶ for  $a_1$ : send  $s(t) = s_1(t)$
  - ▶ over the signals there is additive white noise  $n(t)$
  - ▶ receiver receives noisy signal  $r(t) = s(t) + n(t)$
  - ▶ receiver takes just 1 sample at time  $t_0$ ,  $r(t_0)$
  - ▶ decision: based on  $r(t_0)$ , which signal was it?

# Hypotheses and decisions

- ▶ There are **two hypotheses**:
  - ▶  $H_0$ : true signal is  $s(t) = s_0(t)$  ( $a_0$  has been transmitted)
  - ▶  $H_1$ : true signal is  $s(t) = s_1(t)$  ( $a_1$  has been transmitted)
- ▶ Receiver can take **two decisions**:
  - ▶  $D_0$ : receiver decides that signal was  $s(t) = s_0(t)$
  - ▶  $D_1$ : receiver decides that signal was  $s(t) = s_1(t)$

# Possible outcomes

- ▶ There are 4 possible situations:
  1. **Correct rejection**: true hypothesis is  $H_0$ , decision is  $D_0$ 
    - ▶ Probability is  $P_r = P(D_0 \cap H_0)$
  2. **False alarm** (false detection): true hypothesis is  $H_0$ , decision is  $D_1$ 
    - ▶ Probability is  $P_{fa}P(D_1 \cap H_0)$
  3. **Miss** (false rejection): true hypothesis is  $H_1$ , decision is  $D_0$ 
    - ▶ Probability is  $P_m = P(D_0 \cap H_1)$
  4. **Correct detection** (*hit*): true hypothesis is  $H_1$ , decision  $D_1$ 
    - ▶ Probability is  $P_d = P(D_1 \cap H_1)$

# Origin of terms

- ▶ Terms originate from radar application (first application of detection theory)
  - ▶ signal is emitted from source
  - ▶ received signal = possible reflection from a target, with lots of noise
  - ▶  $H_0$  = no target is present, no reflected signal (only noise)
  - ▶  $H_1$  = target is present, there is a reflected signal
  - ▶ hence the 4 scenarios refer to “has the target been detected”

# The noise

- ▶ In general we consider **additive, white, stationary** noise
  - ▶ additive = the noise is added to the signal
  - ▶ white = two samples from the noise are uncorrelated
  - ▶ stationary = has same statistical properties at all times
- ▶ The noise signal  $n(t)$  is unknown
  - ▶ it's random
  - ▶ we just know it's distribution, but not the actual values

# The sample

- ▶ The receiver receives  $r(t) = s(t) + n(t)$ 
  - ▶  $s(t)$  = original signal, either  $s_0(t)$  or  $s_1(t)$
  - ▶  $n(t)$  = unknown noise
- ▶ The value of the sample taken at  $t_0$  is  $r(t_0) = s(t_0) + n(t_0)$ 
  - ▶  $s(t_0)$  = either  $s_0(t_0)$  or  $s_1(t_0)$
  - ▶  $n(t_0)$  is a sample of the noise

# The sample

- ▶ The sample  $n(t_0)$  is a **random variable**
  - ▶ since it is a sample of noise (a sample from a random process)
  - ▶ assume is a continuous r.v., i.e. range of possible values is continuous
- ▶  $r(t_0) = s(t_0) + n(t_0)$  = a constant + a random variable
  - ▶ it is also a random variable
  - ▶  $s(t_0)$  is a constant, either  $s_0(t_0)$  or  $s_1(t_0)$
- ▶ What distribution does  $r(t_0)$  have?
  - ▶ a constant + a r.v. = has same distribution as r.v., but shifted with the constant

# The conditional distributions

- ▶ Assume the noise has known distribution  $w(x)$ 
  - ▶ this is the distribution of the r.v.  $n(t_0)$
- ▶ The distribution of  $r(t_0) = s(t_0) + n(t_0) = w(x)$  shifted by  $s(t_0)$
- ▶ In hypothesis  $H_0$ , the distribution is  $w(r|H_0) = w(x)$  shifted by  $s_0(t_0)$
- ▶ In hypothesis  $H_1$ , the distribution is  $w(r|H_1) = w(x)$  shifted by  $s_1(t_0)$
- ▶  $w(r|H_0)$  and  $w(r|H_1)$  are known as **conditional distributions** or **conditional likelihood functions**
  - ▶ “|” means “conditioned by”, “given that”
  - ▶ i.e. considering one hypothesis or the other one
  - ▶  $r$  is the unknown of the function



# Maximum Likelihood decision criterion

- ▶ How to decide what hypothesis is true based on the observed sample  $r = r(t_0)$ ?
- ▶ **Maximum Likelihood (ML) criterion**: choose the hypothesis that is **most likely** to have generated the observed sample value  $r = r(t_0)$ 
  - ▶ choose the higher value between  $w(r(t_0)|H_0)$  and  $w(r(t_0)|H_1)$
- ▶ ML expressed as a **Likelihood ratio** test:

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

- ▶ criterion is evaluated for our observed value  $r = r(t_0)$

## Example: gaussian noise

- ▶ Consider noise having a normal distribution
- ▶ At blackboard:
  - ▶ plot the two conditional distributions for  $w(r|H_0)$ ,  $w(r|H_1)$
  - ▶ discuss the decision taken for different values of  $r$
  - ▶ discuss the threshold value  $T$  for taking decisions

# Gaussian noise (AWGN)

- ▶ Particular case: the noise has normal distribution  $\mathcal{N}(0, \sigma^2)$ 
  - ▶ i.e. it is AWGN
- ▶ Likelihood ratio is  $\frac{w(r|H_1)}{w(r|H_0)} = \frac{e^{-\frac{(r-s_1(t_0))^2}{2\sigma^2}}}{e^{-\frac{(r-s_0(t_0))^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$
- ▶ For normal distribution, it is easier to apply **natural logarithm** to the terms
  - ▶ logarithm is a monotonic increasing function, so it won't change the comparison
  - ▶ if  $A < B$ , then  $\log(A) < \log(B)$
- ▶ The **log-likelihood** of an observation = the logarithm of the likelihood value
  - ▶ usually the natural logarithm, but any one can be used

# Log-likelihood test for ML

- ▶ Applying natural logarithm to both sides leads to:

$$-(r - s_1(t_0))^2 + (r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\geq}} 0$$

- ▶ Which means

$$|r - s_0(t_0)| \underset{H_0}{\overset{H_1}{\geq}} |r - s_1(t_0)|$$

- ▶ Note that  $|r - A|$  = distance from  $r$  to  $A$

- ▶  $|r|$  = distance from  $r$  to 0

- ▶ So we choose the smallest distance between  $r(t_0)$  and  $s_1(t_0)$  vs  $s_0(t_0)$

# Maximum Likelihood for gaussian noise

- ▶ ML criterion **for gaussian noise**: choose the hypothesis based on whichever of  $s_0(t_0)$  or  $s_1(t_0)$  is **nearest** to our observed sample  $r = r(t_0)$ 
  - ▶ also known as **nearest neighbor** principle / decision
  - ▶ very general principle, encountered in many other scenarios
  - ▶ because of this, a receiver using ML is also known as **minimum distance receiver**

## Steps for ML decision

1. Sketch the two conditional distributions  $w(r|H_0)$  and  $w(r|H_1)$
2. Find out which function is higher at the observed value  $r = r(t_0)$  given.

## Steps for ML decision in case of gaussian noise

- ▶ Only if the noise is Gaussian, identical for all hypotheses:
  1. Find  $s_0(t_0)$  = the value of the original signal, in absence of noise, in case of hypothesis  $H_0$
  2. Find  $s_1(t_0)$  = the value of the original signal, in absence of noise, in case of hypothesis  $H_1$
  3. Compare with observed sample  $r(t_0)$  and choose the nearest

# Thresholding based decision

- ▶ Choosing the nearest value = same thing as comparing  $r$  with a threshold  $T = \frac{s_0(t_0) + s_1(t_0)}{2}$ 
  - ▶ i.e. if the two values are 0 and 5, decide by comparing with 2.5 (like in laboratory)
- ▶ In general, the threshold = the cross-over point between the conditioned distributions



## Exercise

- ▶ A signal can have two possible values, 0 or 5. The signal is affected by white gaussian noise  $\mathcal{N}(\mu = 0, \sigma^2 = 2)$ . The receiver takes one sample with value  $r = 2.25$ 
  1. Write the expressions of the conditional probabilities and sketch them
  2. What is the decision based on the Maximum Likelihood criterion?
  3. What if the signal 0 is affected by gaussian noise  $\mathcal{N}(0, 0.5)$ , while the signal 5 is affected by uniform noise  $\mathcal{U}[-4, 4]$ ?
  4. Repeat b. and c. assuming the value 0 is replaced by  $-1$

# Decision regions

- ▶ The **decision regions** = the range of values of  $r$  for which a certain decision is taken
- ▶ Decision regions  $R_0$  = all the values of  $r$  which lead to decision  $D_0$
- ▶ Decision regions  $R_1$  = all the values of  $r$  which lead to decision  $D_1$
- ▶ The decision regions cover the whole  $\mathbb{R}$  axis
- ▶ Example: indicate the decision regions for the previous exercise:
  - ▶  $R_0 = [-\infty, 2.5]$
  - ▶  $R_1 = [2.5, \infty]$

# The likelihood function

- ▶ Call the hypotheses, generically,  $H_i$ , and the signals  $s_i(t)$ , where  $i$  is either 0 or 1
- ▶ Consider the conditional distribution  $w(r|H_i)$ 
  - ▶ think of the function in the previous example, e.g.:

$$w(r|H_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-s_i(t_0))^2}{2\sigma^2}}$$

- ▶ Which is the unknown in this function?
  - ▶ not  $r$ , since it is actually given in the exercise
  - ▶  $i$  is the unknown variable

# Terminology: probability vs likelihood

- ▶ In the same mathematical expression of a distribution function:
  - ▶ if we know the parameters (e.g.  $\mu, \sigma, H_i$ ), and the unknown is the value (e.g.  $r, x$ ) we call it **probability density function** (distribution)
  - ▶ if we know value (e.g.  $r, x$ ), and the unknown is some statistical parameter (e.g.  $\mu, \sigma, i$ ), we call it a **likelihood function**
- ▶ Hence the subtle distinction in terms: “probability” vs “likelihood”

# The likelihood function

- ▶ The function  $w(r|H_i) = f(i)$  is a likelihood function
  - ▶ the unknown is  $i$
- ▶ The function exists only in 2 points, for  $i = 0$  and  $i = 1$ 
  - ▶ or, in general, for  $i =$  how many hypotheses exist in the problem
- ▶ ML criterion = choose the  $i$  for which this function is maximum

$$\text{Decision } D_i = \arg \max_i w(r|H_i)$$

- ▶ Notation:
  - ▶  $\arg \max f(x)$  = the  $x$  for which the function  $f(x)$  is maximum
  - ▶  $\max f(x)$  = the maximum value of the function  $f(x)$
  - ▶ see graphical explanation at blackboard
- ▶ Maximum Likelihood criterion means “choose the  $i$  which maximizes the likelihood function  $f(i) = w(r|H_i)$ ”

# Generalizations

- ▶ What if the noise has another distribution?
  - ▶ Sketch the conditional distributions
  - ▶ Locate the given  $r = r(t_0)$
  - ▶ ML criterion = choose the highest function  $w(r|H_i)$  in that point
- ▶ The decision regions are defined by the cross-over points
  - ▶ There can be more cross-overs, so multiple thresholds

# Generalizations

- ▶ What if the noise has a different distribution in hypothesis  $H_0$  than in hypothesis  $H_1$ ?
- ▶ Same thing:
  - ▶ Sketch the conditional distributions
  - ▶ Locate the given  $r = r(t_0)$
  - ▶ ML decision = choose the highest function  $w(r|H_i)$  in that point

# Generalizations

- ▶ What if the two signals  $s_0(t)$  and  $s_1(t)$  are constant / not constant?
- ▶ We don't care about the shape of the signals
  - ▶ All we care about are the two values at the sample time  $t_0$ :
    - ▶  $s_0(t_0)$
    - ▶  $s_1(t_0)$



# Generalizations

- ▶ What if we have more than two hypotheses?
- ▶ Extend to  $n$  hypotheses
  - ▶ We have  $n$  possible signals  $s_0(t), \dots, s_{n-1}(t)$
  - ▶ We have  $n$  different values  $s_0(t_0), \dots, s_{n-1}(t_0)$
  - ▶ We have  $n$  conditional distributions  $w(r|H_i)$
  - ▶ For the given  $r = r(t_0)$ , choose the maximum value out of the  $n$  values  $w(r|H_i)$

# Generalizations

- ▶ What if we take more than 1 sample?
- ▶ Patience, we'll treat this later as a separate sub-chapter

## Exercise

- ▶ A signal can have four possible values: -6, -2, 2, 6. Each value lasts for 1 second. The signal is affected by white noise with normal distribution. The receiver takes 1 sample per second. Using ML criterion, decide what signal has been transmitted, if the received samples are:

4, 6.6, -5.2, 1.1, 0.3, -1.5, 7, -7, 4.4

# Conditional probabilities

- ▶ We compute the **conditional probabilities** of the 4 possible outcomes
- ▶ Consider the decision regions:
  - ▶  $R_0$ : when  $r \in R_0$ , decision is  $D_0$
  - ▶  $R_1$ : when  $r \in R_1$ , decision is  $D_1$
- ▶ Conditional probability of correct rejection
  - ▶ = probability to take decision  $D_0$  in the case that hypothesis is  $H_0$
  - ▶ = probability that  $r$  is in  $R_0$  computed from the distribution  $w(r|H_0)$

$$P(D_0|H_0) = \int_{R_0} w(r|H_0)dx$$

- ▶ Conditional probability of false alarm
  - ▶ = probability to take decision  $D_1$  in the case that hypothesis is  $H_0$
  - ▶ = probability that  $r$  is in  $R_1$  computed from the distribution  $w(r|H_0)$

$$P(D_1|H_0) = \int_{R_1} w(r|H_0)dx$$

# Conditional probabilities

- ▶ Conditional probability of miss

- ▶ = probability to take decision  $D_0$  in the case that hypothesis is  $H_1$
- ▶ = probability that  $r$  is in  $R_0$  computed from the distribution  $w(r|H_1)$

$$P(D_0|H_1) = \int_{R_0} w(r|H_1) dx$$

- ▶ Conditional probability of correct rejection

- ▶ = probability to take decision  $D_1$  in the case that hypothesis is  $H_1$
- ▶ = probability that  $r$  is in  $R_1$  computed from the distribution  $w(r|H_1)$

$$P(D_1|H_1) = \int_{R_1} w(r|H_1) dx$$

# Conditional probabilities

- ▶ Relation between them:
  - ▶ sum of correct rejection + false alarm = 1
  - ▶ sum of miss + correct detection = 1
  - ▶ Why? Prove this.

# Computing conditional probabilities

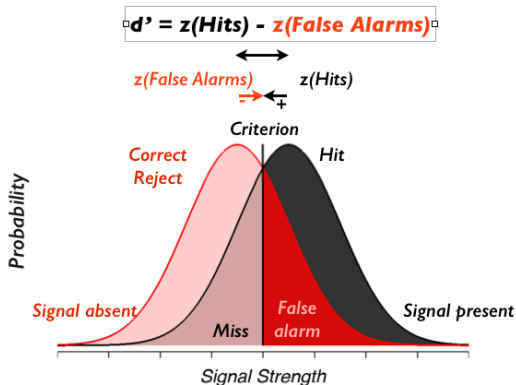


Figure 2: Conditional probabilities

- ▶ Ignore the text, just look at the nice colors
- ▶ [image from <http://gru.stanford.edu/doku.php/tutorials/sdt>]\*

## Probabilities of the 4 outcomes

- ▶ Conditional probabilities are computed **given that** one or the other hypothesis is true
- ▶ They do not account for the probabilities *of the hypotheses themselves*
  - ▶ i.e.  $P(H_0)$  = how many times does  $H_0$  happen?
  - ▶  $P(H_1)$  = how many times does  $H_1$  happen?
- ▶ To account for these, multiply with  $P(H_0)$  or  $P(H_1)$ 
  - ▶  $P(H_0)$  and  $P(H_1)$  are known as the **prior** (or **a priori**) probabilities of the hypotheses



# Reminder: the Bayes rule

- ▶ Reminder: the Bayes rule

$$P(A \cap B) = P(B|A) \cdot P(A)$$

- ▶ Interpretation

- ▶ The probability  $P(A)$  is taken out from  $P(B|A)$
- ▶  $P(B|A)$  gives no information on  $P(A)$ , the chances of  $A$  actually happening
- ▶ Example:  $P(\text{score} \mid \text{shoot}) = \frac{1}{2}$ . How many goals are scored?

- ▶ In our case:  $P(D_i \cap H_j) = P(D_i|H_j) \cdot P(H_j)$

- ▶ for all  $i$  and  $j$ , i.e. all 4 cases

## Exercise

- ▶ A constant signal can have two possible values,  $-2$  or  $5$ . The signal is affected by gaussian noise  $\mathcal{N}(\mu = 0, \sigma^2 = 2)$ . The receiver performs ML decision based on a single sample.
  1. Compute the conditional probability of a false alarm
  2. Compute the conditional probability of a miss
  3. If  $P(H_0) = \frac{1}{3}$  and  $P(H_1) = \frac{2}{3}$ , compute the actual probabilities of correct rejection and correct detection (not conditional)

# Pitfalls of ML decision criterion

- ▶ The ML criterion is based on comparing **conditional** distributions
  - ▶ conditioned by  $H_0$  or by  $H_1$
- ▶ Conditioning by  $H_0$  and  $H_1$  ignores the prior probabilities of  $H_0$  or  $H_1$ 
  - ▶ Our decision doesn't change if we know that  $P(H_0) = 99.99\%$  and  $P(H_1) = 0.01\%$ , or vice-versa
- ▶ But if  $P(H_0) > P(H_1)$ , we may want to move the threshold towards  $H_1$ , and vice-versa
  - ▶ because it is more likely that the true signal is  $s_0(t)$
  - ▶ and thus we want to “encourage” decision  $D_0$
- ▶ Looks like we want a more general criterion ...

# The minimum error probability criterion

- ▶ Takes into account the probabilities  $P(H_0)$  and  $P(H_1)$
- ▶ Goal is to **minimize the total probability of error**  $P_e = P_{fa} + P_m$ 
  - ▶ errors = false alarms and misses
- ▶ We need to find a new criterion (new decision regions  $R_0$  and  $R_1$ )

# Deducing the new criterion

- ▶ The probability of false alarm is:

$$\begin{aligned}P(D_1 \cap H_0) &= P(D_1|H_0) \cdot P(H_0) \\&= \int_{R_1} w(r|H_0) dx \cdot P(H_0) \\&= (1 - \int_{R_0} w(r|H_0) dx) \cdot P(H_0)\end{aligned}$$

- ▶ The probability of miss is:

$$\begin{aligned}P(D_0 \cap H_1) &= P(D_0|H_1) \cdot P(H_1) \\&= \int_{R_0} w(r|H_1) dx \cdot P(H_1)\end{aligned}$$

- ▶ The total error probability (their sum) is:

$$P_e = P(H_0) + \int_{R_0} [w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0)] dx$$

# Minimum probability of error

- ▶ We want to minimize  $P_e$ , i.e. to minimize the integral
- ▶ We can choose  $R_0$  as we want for this purpose
- ▶ We choose  $R_0$  such that for all  $r \in R_0$ , the term inside the integral is **negative**
  - ▶ because integrating over all the interval where the function is negative ensures minimum value of integral
- ▶ So, when  $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) < 0$  we have  $r \in R_0$ , i.e. decision  $D_0$
- ▶ Conversely, when  $w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) > 0$  we have  $r \in R_1$ , i.e. decision  $D_1$
- ▶ Therefore

$$w(r|H_1) \cdot P(H_1) - w(r|H_0) \cdot P(H_0) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

- ▶ **The minimum probability of error** criterion (MPE):

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

# Interpretation

- ▶ MPE criterion is more general than ML, depends on probabilities of the two hypotheses
  - ▶ Also expressed as a likelihood ratio test
- ▶ When one hypothesis has higher probability than the other, the threshold is **pushed in its favor**, towards the other one
- ▶ The ML criterion is a particular case of the MPE criterion, for  $P(H_0) = P(H_1) = \frac{1}{2}$



# Minimum probability of error - Gaussian noise

- Assuming the noise has normal distribution  $\mathcal{N}(0, \sigma^2)$

$$w(r|H_1) = e^{-\frac{(r-s_1(t_0))^2}{2\sigma^2}}$$

$$w(r|H_0) = e^{-\frac{(r-s_0(t_0))^2}{2\sigma^2}}$$

- Apply natural logarithm

$$-\frac{(r-s_1(t_0))^2}{2\sigma^2} + \frac{(r-s_0(t_0))^2}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- Equivalently

$$(r-s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r-s_1(t_0))^2 + 2\sigma^2 \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- or, after further processing:

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

## Interpretation 1: Comparing distance

- For ML criterion, we compare the (squared) distances:

$$|r - s_0(t_0)| \underset{H_0}{\overset{H_1}{\geq}} |r - s_1(t_0)|$$

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\geq}} (r - s_1(t_0))^2$$

- For MPE criterion, we compare the squared distances, but a supplementary term appears in favour of the most probable hypothesis:

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\geq}} (r - s_1(t_0))^2 + 2\sigma^2 \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- term depends on the ratio  $\frac{P(H_0)}{P(H_1)}$

## Interpretation 2: The threshold value

- ▶ For ML criterion, we compare  $r$  with a threshold  $T$

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2}$$

- ▶ For MPE criterion, the threshold is moved towards the less probable hypothesis:

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- ▶ depending on the ratio  $\frac{P(H_0)}{P(H_1)}$

- ▶ Consider the decision between two constant signals:  $s_0(t) = -5$  and  $s_1(t) = 5$ . The signals are affected by gaussian noise  $\mathcal{N}(0, \sigma^2 = 3)$ . The receiver takes one sample  $r$ .
  1. Find the decision regions  $R_0$  and  $R_1$  according to the MPE criterion
  2. What are the probabilities of false alarm and of miss?
  3. Repeat a) and b) considering that  $s_1(t)$  is affected by uniform noise  $\mathcal{U}[-4, 4]$

# Minimum risk criterion

- ▶ What if we care more about one type of errors (e.g. false alarms) than other kind (e.g. miss)?
  - ▶ MPE criterion treats all errors the same
  - ▶ Need a more general criterion
- ▶ Idea: assign a **cost** to each scenario, minimize average cost
- ▶  $C_{ij}$  = cost of decision  $D_i$  when true hypothesis was  $H_j$ 
  - ▶  $C_{00}$  = cost for good detection  $D_0$  in case of  $H_0$
  - ▶  $C_{10}$  = cost for false alarm (detection  $D_1$  in case of  $H_0$ )
  - ▶  $C_{01}$  = cost for miss (detection  $D_0$  in case of  $H_1$ )
  - ▶  $C_{11}$  = cost for good detection  $D_1$  in case of  $H_1$
- ▶ The idea of assigning “costs” and minimizing average cost is very general
  - ▶ e.g. IT: Shannon coding: “cost” of each message is the length of its codeword, we want to minimize average cost, i.e. minimize average length

# Minimum risk criterion

- ▶ Define the **risk** = **the average cost** value

$$R = C_{00}P(D_0 \cap H_0) + C_{10}P(D_1 \cap H_0) + C_{01}P(D_0 \cap H_1) + C_{11}P(D_1 \cap H_1)$$

- ▶ Minimum risk criterion: **minimize the risk R**
  - ▶ i.e. minimize the average cost
  - ▶ also known as “minimum cost criterion”

# Computations

- ▶ Proof on blackboard: (sorry, no time to put in on slides)
  - ▶ Use Bayes rule
  - ▶ Notations:  $w(r|H_j)$  (*likelihood*)
  - ▶ Probabilities:  $\int_{R_i} w(r|H_j) dV$
- ▶ Conclusion, **decision rule is**

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)}$$

# Minimum risk criterion

**Minimum risk criterion (MR):**

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)}$$



- ▶ MR is a generalization of MPE criterion (which was itself a generalization of ML)
  - ▶ also expressed as a likelihood ratio test
- ▶ Both **probabilities** and the assigned **costs** can influence the decision towards one hypothesis or the other
- ▶ If  $C_{10} - C_{00} = C_{01} - C_{11}$ , MR reduces to MPE:
  - ▶ e.g. if  $C_{00} = C_{11} = 0$ , and  $C_{10} = C_{01}$

## Minimum Risk - gaussian noise

- ▶ If the noise is gaussian (normal), do like for the other criteria, apply logarithm
- ▶ Obtain:

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r - s_1(t_0))^2 + 2\sigma^2 \cdot \ln \left( \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)} \right)$$

▶ or

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln \left( \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)} \right)$$

## Interpretation 1: Comparing distance

- ▶ For MPE criterion, we compare the squared distances, but a supplementary term appears in favour of the most probable hypothesis:

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r - s_1(t_0))^2 + 2\sigma^2 \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- ▶ term depends on the ratio  $\frac{P(H_0)}{P(H_1)}$
- ▶ For MR criterion, besides the probabilities we also are influenced by the costs

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\gtrless}} (r - s_1(t_0))^2 + 2\sigma^2 \cdot \ln \left( \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)} \right)$$

## Interpretation 2: The threshold value

- ▶ For MPE criterion, the threshold is moved towards the less probable hypothesis:

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln \left( \frac{P(H_0)}{P(H_1)} \right)$$

- ▶ depending on the ratio  $\frac{P(H_0)}{P(H_1)}$
- ▶ For MR criterion, besides the probabilities we also are influenced by the costs

$$r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln \left( \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)} \right)$$

# Influence of costs

- ▶ The MR criterion pushes the decision towards **minimizing the high-cost scenarios**
- ▶ Example: from the equations:
  - ▶ what happens if cost  $C_{01}$  increases, while the others are unchanged?
  - ▶ what happens if cost  $C_{10}$  increases, while the others are unchanged?
  - ▶ what happens if both costs  $C_{01}$  and  $C_{10}$  increase, while the others are unchanged?

# General form of ML, MPE and MR criteria

- ▶ ML, MPE and MR criteria all have the following form

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ for ML:  $K = 1$
- ▶ for MPE:  $K = \frac{P(H_0)}{P(H_1)}$
- ▶ for MR:  $K = \frac{(C_{10} - C_{00})p(H_0)}{(C_{01} - C_{11})p(H_1)}$

# General form of ML, MPE and MR criteria

In gaussian noise, all criteria reduce to:

- ▶ Comparing squared distances:

$$(r - s_0(t_0))^2 \underset{H_0}{\overset{H_1}{\geq}} (r - s_1(t_0))^2 + 2\sigma^2 \cdot \ln(K)$$

- ▶ Comparing the sample  $r$  with a threshold  $T$ :

$$r \underset{H_0}{\overset{H_1}{\geq}} \underbrace{\frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln(K)}_T$$

## Exercise

- ▶ A vehicle airbag system detects a crash by evaluating a sensor which provides two values:  $s_0(t) = 0$  (no crash) or  $s_1(t) = 5$  (crashing)
- ▶ The signal is affected by gaussian noise  $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ .
- ▶ The costs of the scenarios are:  $C_{00} = 0$ ,  $C_{01} = 100$ ,  $C_{10} = 10$ ,  $C_{11} = -100$ 
  1. Find the decision regions  $R_0$  and  $R_1$ .



# Neyman-Pearson criterion

- ▶ An even more general criteria than all the others until now
- ▶ **Neyman-Pearson criterion:** maximize probability of correct detection ( $P(D_1 \cap H_1)$ ) while keeping probability of false alarms smaller then a limit ( $P(D_1 \cap H_0) \leq \lambda$ )
  - ▶ Deduce the threshold  $T$  from the limit condition  $P(D_1 \cap H_0) = \lambda$
- ▶ ML, MPE and MR criteria are particular cases of Neyman-Pearson, for particular values of  $\lambda$

# Exercise

- ▶ An information source provides two messages with probabilities  $p(a_0) = \frac{2}{3}$  and  $p(a_1) = \frac{1}{3}$ .
- ▶ The messages are encoded as constant signals with values  $-5$  ( $a_0$ ) and  $5$  ( $a_1$ ).
- ▶ The signals are affected by noise with uniform distribution  $U[-5, 5]$ .
- ▶ The receiver takes one sample  $r$ .
  1. Find the decision regions according to the Neyman-Pearson criterion, considering  $P_{fa} \leq 10^{-2}$
  2. What is the probability of correct detection, in this case?

# Application: Differential vs single-ended signalling

- ▶ Application: binary transmission with constant signals (e.g. constant voltage levels)
- ▶ Two common possibilities:
  - ▶ Single-ended signalling: one signal is 0, other is non-zero
    - ▶  $s_0(t) = 0, s_1(t) = A$
  - ▶ Differential signalling: use two non-zero levels with different sign, same absolute value
    - ▶  $s_0(t) = -\frac{A}{2}, s_1(t) = \frac{A}{2}$
- ▶ Find out which is better?

# Differential vs single-ended signalling

- ▶ Since difference between levels is the same, decision performance is the same
- ▶ Average power of a signal = average squared value
- ▶ For differential signal:  $P = \left(\pm \frac{A}{2}\right)^2 = \frac{A^2}{4}$
- ▶ For signal ended signal:  $P = P(H_0) \cdot 0 + P(H_1) (A)^2 = \frac{A^2}{2}$ 
  - ▶ assuming equal probabilities  $P(H_0) = P(H_1) = \frac{1}{2}$
- ▶ Differential uses half the power of single-ended (i.e. better), for same decision performance

## Summary of criteria

- ▶ We have seen decision based on 1 sample  $r$ , between 2 signals (mostly)
- ▶ All decisions are based on a likelihood-ratio test

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ Different criteria differ in the chosen value of  $K$  (likelihood threshold)
- ▶ Depending on the noise distributions, the real axis is partitioned into regions
  - ▶ region  $R_0$ : if  $r$  is in here, decide  $D_0$
  - ▶ region  $R_1$ : if  $r$  is in here, decide  $D_1$
- ▶ For gaussian noise, the boundary of the regions (threshold) is

$$T = \frac{s_0(t_0) + s_1(t_0)}{2} + \frac{\sigma^2}{s_1(t_0) - s_0(t_0)} \cdot \ln(K)$$

# Receiver Operating Characteristic

- ▶ The receiver performance is usually represented with “**Receiver Operating Characteristic**” (ROC) graph
- ▶ It is a graph of hit probability  $P_d = P(D_1 \cap H_1)$  (correct detection) as a function of false alarm probability  $P_{fa} = P(D_1 \cap H_0)$

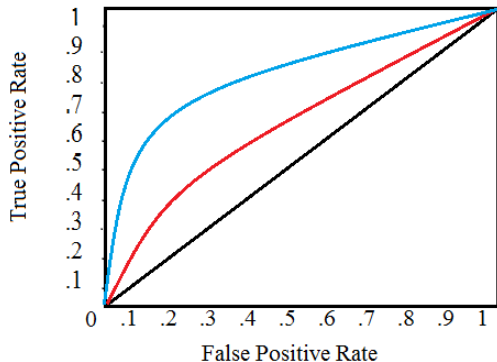


Figure 3: Sample ROC curves

# Receiver Operating Characteristic

- ▶ There is always a **tradeoff** between good  $P_d$  and bad  $P_{fa}$ 
  - ▶ to increase  $P_d$  one must also increase  $P_{fa}$
  - ▶ if we want to make sure we don't miss any real detections (increase  $P_d$ ), we pay by increasing the chances of false alarms
- ▶ Different criteria = different likelihood thresholds  $K$  = different points on the graph = different tradeoffs
  - ▶ but the tradeoff cannot be avoided
- ▶ How to improve the receiver?
  - ▶ i.e. increase  $P_D$  while keeping  $P_{fa}$  the same

# Performance of likelihood-ratio decoding in WGN

- ▶ WGN = “White Gaussian Noise”
- ▶ Assume equal probabilities  $P(H_0) = P(H_1) = \frac{1}{2}$
- ▶ All decisions are based on a likelihood-ratio test

$$\frac{w(r|H_1)}{w(r|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ Hit probability is

$$\begin{aligned} P_{hit} &= P(D_1|H_1)P(H_1) \\ &= P(H_1) \int_T^\infty w(r|H_1) \\ &= P(H_1)(F(\infty) - F(T)) \\ &= \frac{1}{4} \left( 1 - \operatorname{erf} \left( \frac{T - A}{\sqrt{2}\sigma} \right) \right) \\ &= Q \left( \frac{T - A}{\sqrt{2}\sigma} \right) \end{aligned}$$



# Performance of likelihood-ratio decoding in WGN

- False alarm probability is

$$\begin{aligned}P_{fa} &= P(D_1|H_0)P(H_0) \\&= P(H_0) \int_T^\infty w(r|H_0) \\&= P(H_0)(F(\infty) - F(T)) \\&= \frac{1}{4} \left( 1 - \operatorname{erf} \left( \frac{T - 0}{\sqrt{2}\sigma} \right) \right) \\&= Q \left( \frac{T}{\sqrt{2}\sigma} \right)\end{aligned}$$

- Therefore  $\frac{T}{\sqrt{2}\sigma} = Q^{-1}(P_{fa})$
- Replacing in  $P_{hit}$  yields

$$P_{hit} = Q \left( \underbrace{Q^{-1}(P_{fa})}_{\text{constant}} - \frac{A}{\sqrt{2}\sigma} \right)$$

# Signal-to-noise ratio

- ▶ **Signal-to-noise ratio (SNR)** =  $\frac{\text{power of original signal}}{\text{power of noise}}$
- ▶ Average power of a signal = average squared value =  $\overline{X^2}$ 
  - ▶ Original signal power is  $\frac{A^2}{2}$
  - ▶ Noise power is  $\overline{X^2} = \sigma^2$  (when noise mean value  $\mu = 0$ )
- ▶ In our case,  $\text{SNR} = \frac{A^2}{2\sigma^2}$

$$P_{hit} = Q \left( \underbrace{Q^{-1}(P_{fa})}_{\text{constant}} - \sqrt{\text{SNR}} \right)$$

- ▶ For a fixed  $P_{fa}$ ,  $P_{hit}$  increases with SNR
  - ▶  $Q$  is a monotonic decreasing function

# Performance depends on SNR

- ▶ Receiver performance increases with SNR increase
  - ▶ high SNR: good performance
  - ▶ poor SNR: bad performance

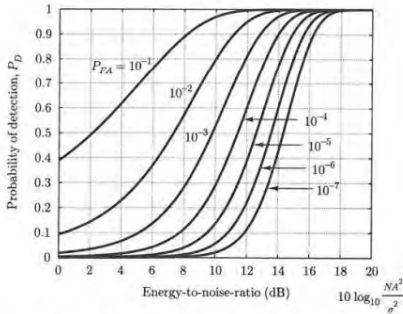


Figure 4: Detection performance depends on SNR

[source: *Fundamentals of Statistical Signal Processing*, Steven Kay]

# Applications of decision theory

- ▶ Can we apply these decision criteria in other engineering problems?
  - ▶ e.g. not for deciding between two signals, but for something else
- ▶ The core mathematical problem we solve is:
  - ▶ we have 2 (or more) possible distributions
  - ▶ we observe 1 value
  - ▶ we determine the most likely distribution, according to the value
- ▶ In our particular problem, we decide between two signals
- ▶ But this can be applied to many other statistical problems:
  - ▶ medicine: does this ECG signal look healthy or not?
  - ▶ business: will this client buy something or not?
  - ▶ Typically we use more than 1 value for these, but the mathematical principle is the same

# Applications of decision theory

Example (purely imaginary):

- ▶ A healthy person of weight =  $X$  kg has the concentration of thrombocytes per ml of blood distributed approximately as  $\mathcal{N}(\mu = 10 \cdot X, \sigma^2 = 20)$ .
- ▶ A person suffering from disease D has a much lower value of thrombocytes, distributed approximately as  $\mathcal{N}(100, \sigma^2 = 10)$ .
- ▶ The lab measures your blood and finds your value equal to  $r = 255$ . Your weight is 70 kg.
- ▶ Decide: are you most likely healthy, or ill?

## II.3 Detection of constant signals with multiple samples

# Multiple samples from a constant signal

- ▶ Suppose we have multiple samples, not just 1
- ▶ The samples are arranged in a **sample vector**

$$\mathbf{r} = [r_1, r_2, \dots, r_N]$$

- ▶ In each hypotheses, the signal is a **random process**
  - ▶  $H_0$ : random process with average value 0
  - ▶  $H_1$ : random process with average value A
- ▶ Assuming the noise is stationary and ergodic, the signal random process is stationary and ergodic (signal = constant + noise)
- ▶ The values of  $\mathbf{r}$  are described by the **distribution of order  $N$**  of the random processes,  $w_N(\mathbf{r}) = w_N(r_1, r_2, \dots, r_N)$
- ▶ Assuming the noise is white noise, the sample times don't matter

# Likelihood of vector samples

- ▶ We can apply **the same criteria** based on likelihood ratio as for 1 sample

$$\frac{w_N(\mathbf{r}|H_0)}{w_N(\mathbf{r}|H_1)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ Notes

- ▶  $\mathbf{r}$  is a vector; we consider the likelihood of all the samples
  - ▶ the hypotheses  $H_0$  and  $H_1$  are the same as for 1 sample
  - ▶  $w_N(\mathbf{r}|H_0)$  = likelihood of the whole vector  $\mathbf{r}$  being obtained in hypothesis  $H_0$
  - ▶  $w_N(\mathbf{r}|H_1)$  = likelihood of the whole vector  $\mathbf{r}$  being obtained in hypothesis  $H_1$
  - ▶ the value of  $K$  is given by the actual decision criterion used
- ▶ Interpretation: we choose the hypothesis that is most likely to have produced the observed data
  - ▶ the same, but now the data = multiple samples



- ▶ Assuming the noise is white noise, the samples  $r_i$  are **multiple independent realizations of the same distribution**
- ▶ In that case the joint distributions  $w_N(\mathbf{r}|H_j)$  can be decomposed as a product

$$w_N(\mathbf{r}|H_j) = w(r_1|H_j) \cdot w(r_2|H_j) \cdot \dots \cdot w(r_N|H_j)$$

- ▶ The  $w(r_i|H_j)$  are just the likelihoods of each individual sample
  - ▶ e.g. the likelihood of obtaining  $[5.1, 4.7, 4.9]$  = likelihood of obtaining 5.1  $\times$  likelihood of getting 4.7  $\times$  likelihood of getting 4.9

# Separation

- ▶ Then all likelihood ratio criteria can be written as

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{w(r_1|H_1)}{w(r_1|H_0)} \cdot \frac{w(r_2|H_1)}{w(r_2|H_0)} \cdots \frac{w(r_N|H_1)}{w(r_N|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ The likelihood ratio of a vector of samples = product of likelihood ratio for each sample

## Particular case: AWGN

- ▶ AWGN = “Additive White Gaussian Noise”
- ▶ In hypothesis  $H_1$ :  $w(r_i|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_i-A)^2}{2\sigma^2}}$
- ▶ In hypothesis  $H_0$ :  $w(r_i|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r_i^2}{2\sigma^2}}$
- ▶ Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i-A)^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}}$$

- ▶ We can interpret this likelihood ratio in three ways

# Interpretation 1: average value of samples

- Interpretation 1: average value of samples

$$\begin{aligned}\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} &= \frac{e^{-\frac{\sum (r_i - A)^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} \\ &= e^{-\frac{\sum (r_i - A)^2 - \sum (r_i)^2}{2\sigma^2}} \\ &= e^{-\frac{\sum (r_i^2 - 2r_i A + A^2) - \sum (r_i)^2}{2\sigma^2}} \\ &= e^{-\frac{\sum (-2r_i A + A^2)}{2\sigma^2}} \\ &= e^{-\frac{-2A \sum (r_i) + N A^2}{2\sigma^2}} \\ &= e^{-\frac{-2A \frac{\sum (r_i)}{N} + A^2}{2 \frac{\sigma^2}{N}}}\end{aligned}$$

# Average value of N gaussian random variables

- ▶ Let  $U_r$  = average value of the samples  $r_i$

$$U_r = \frac{1}{N} \sum r_i$$

- ▶ What distribution does it have?
- ▶ Consider the sum  $S_r = \sum r_i$  of the N samples  $r_i$ 
  - ▶ From chapter 1: the sum of normal r.v.  $\mathcal{N}(\mu, \sigma^2)$  has:
    - ▶ normal distribution  $\mathcal{N}(\mu_S, \sigma_S^2)$  with
    - ▶ average value:  $\mu_S = N \cdot \mu$
    - ▶ variance:  $\sigma_S^2 = N \cdot \sigma^2$
- ▶ Then  $U_r = \frac{1}{N} S_r$ , and from the properties of average values we have
  - ▶  $U_r$  has normal distribution with:
  - ▶ average value  $= \frac{1}{N} \mu_S = \frac{1}{N} N \mu = \mu$
  - ▶ variance  $= \left(\frac{1}{N}\right)^2 \sigma_S^2 = \left(\frac{1}{N}\right)^2 N \sigma^2 = \frac{1}{N} \sigma^2$

# Average value of $N$ gaussian random variables

- ▶ The mean value of  $N$  realizations of a normal distribution has a normal distribution with
  - ▶ same average value
  - ▶ variance  $N$  times smaller
- ▶ If  $N$  gets very large, the mean value is a very good **estimator** of the distribution's average value
  - ▶ its distribution gets very narrow around the average value

## Interpretation 1: average value of samples

$$\begin{aligned}\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} &= e^{\frac{-2AU_r + A^2}{2\frac{\sigma^2}{N}}} \\ &= \frac{e^{\frac{U_r^2 - 2AU_r + A^2}{2\frac{\sigma^2}{N}}}}{e^{\frac{U_r^2}{2\frac{\sigma^2}{N}}}} \\ &= \frac{e^{\frac{(U_r - A)^2}{2\frac{\sigma^2}{N}}}}{e^{\frac{U_r^2}{2\frac{\sigma^2}{N}}}} \\ &= \frac{w(U_r|H_1)}{w(U_r|H_0)}\end{aligned}$$

- The likelihood ratio of  $N$  gaussian samples = the likelihood ratio of **the mean of the samples**

## Interpretation 1: average value of samples

- ▶ The likelihood ratio of  $N$  gaussian samples = the likelihood ratio of **the mean of the samples**
  - ▶ the mean has smaller variance  $\frac{1}{N}\sigma^2$ , so is more accurate
  - ▶ it is like the noise distribution gets  $N$  times narrower (due to averaging)
- ▶ Detection of constant signals with  $N$  samples is the same as detection with 1 sample, but:
  - ▶ use the average value of the samples  $r_i$
  - ▶ its distributions are  $N$  times narrower (variance is  $N$  times smaller)
- ▶ As  $N$  increases, the probability of errors decrease  $\Rightarrow$  better performance



# Exercise

Exercise:

- ▶ A signal can have two values, 0 (hypothesis  $H_0$ ) or 6 (hypothesis  $H_1$ ). The signal is affected by AWGN  $\mathcal{N}(0, \sigma^2 = 1)$ . The receiver takes 5 samples with values  $\{1.1, 4.4, 3.7, 4.1, 3.8\}$ .
  1. What is decision according to Maximum Likelihood criterion?
  2. What is decision according to minimum probability of error criterion, assuming  $P(H_0) = 2/3$  and  $P(H_1) = 1/3$ ?

## Interpretation 2: geometrical

- ▶ Useful mainly for Maximum Likelihood criterion
- ▶ Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i - A)^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\geq}} K$$

- ▶ For Maximum Likelihood we compare to 1

$$\begin{aligned} \frac{e^{-\frac{\sum (r_i - A)^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} &\underset{H_0}{\overset{H_1}{\geq}} 1 \\ e^{-\frac{\sum (r_i - A)^2}{2\sigma^2} + \frac{\sum (r_i)^2}{2\sigma^2}} &\underset{H_0}{\geq} 1 \\ -\sum (r_i - A)^2 + \sum (r_i)^2 &\underset{H_0}{\geq} 0 \\ \sum (r_i)^2 &\underset{H_0}{\overset{H_1}{\geq}} \sum (r_i - A)^2 \end{aligned}$$

## Interpretation 2: geometrical

- ▶  $\sqrt{\sum (r_i)^2}$  is the geometrical (Euclidian) distance between point  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  and point  $\mathbf{0} = [0, 0, \dots, 0]$
- ▶  $\sqrt{\sum (r_i - A)^2}$  is the geometrical (Euclidian) distance between point  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  and point  $\mathbf{A} = [A, A, \dots, A]$
- ▶ ML decision chooses **the closest signal vector (point)** to the received vector (point), in a N-dimensional space
  - ▶ it is known as “minimum distance receiver”
  - ▶ same interpretation as in the 1-D case
- ▶ Question: what is the geometrical interpretation for the other criteria?

# Exercise

Exercise:

- ▶ A signal can have two values, 0 (hypothesis  $H_0$ ) or 6 (hypothesis  $H_1$ ). The signal is affected by AWGN  $\mathcal{N}(0, \sigma^2 = 1)$ . The receiver takes 2 samples with values  $\{1.1, 4.4\}$ .
  1. What is decision according to Maximum Likelihood criterion? Use the geometrical interpretation.

## Interpretation 3: cross-correlation

- Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i - A)^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$e^{-\frac{\sum (r_i - A)^2}{2\sigma^2} + \frac{\sum (r_i)^2}{2\sigma^2}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$-\sum (r_i - A)^2 + \sum (r_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$2\sum r_i A - NA^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$\frac{1}{N}\sum r_i A \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\frac{A^2}{2} + \frac{1}{N}\sigma^2 \ln K}_{L=\text{const}}$$

## Interpretation 3: cross-correlation

- ▶ The **cross-correlation** (sometimes just “the correlation”) of two signals  $x$  and  $y$  is

$$\langle x, y \rangle = \frac{1}{N} \sum x[n]y[n]$$

- ▶ It is the value of the correlation function in 0

$$\langle x, y \rangle = R_{xy}[0] = \overline{x[n]y[n+0]}$$

- ▶ For continuous signals

$$\langle x, y \rangle = \frac{1}{T} \int_{T/2}^{T/2} x(t)y(t)dt$$

- ▶  $\frac{1}{N} \sum r_i A = \langle \mathbf{r}, \mathbf{A} \rangle$  is the cross-correlation of the received samples  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  with the **target** samples  $\mathbf{A} = [A, A, \dots, A]$

## Interpretation 3: cross-correlation

- ▶ If the cross-correlation of the received samples with the target samples  $\mathbf{A} = [A, A, \dots A]$  is greater than a certain threshold  $L$ , we decide that signal is detected.
  - ▶ otherwise, the signal is rejected
- ▶ This is **similar to signal detection based on 1 sample**, with the sample value being  $\langle \mathbf{r}, \mathbf{A} \rangle$

# Cross-correlation as a measure of similarity

- ▶ Cross-correlation in signal processing measures **similarity** of two signals
- ▶ Interpretation: we check if the received samples look similar enough to the constant signal  $A$ 
  - ▶ If yes (high cross-correlation)  $\Rightarrow$  signal detected
  - ▶ If no (low cross-correlation)  $\Rightarrow$  no detection



## Generalization: two non-zero values

- ▶ Generalization: two non-zero signal values,  $B$  and  $A$ 
  - ▶ still with Gaussian noise
- ▶ Interpretation 1: average value of samples
  - ▶ use mean of samples, the two distributions are centered on  $B$  and  $A$
- ▶ Interpretation 2: geometric (Maximum Likelihood)
  - ▶ choose minimum Euclidean distance from  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  to points  $\mathbf{B} = [B, B, \dots]$  and  $\mathbf{A} = [A, A, \dots]$
- ▶ Interpretation 3: cross-correlation
  - ▶ compute  $\langle \mathbf{r}, \mathbf{B} \rangle$  and  $\langle \mathbf{r}, \mathbf{A} \rangle$ , cross-correlation of  $\mathbf{r}$  with  $\mathbf{B} = [B, B, \dots]$  and with  $\mathbf{A} = [A, A, \dots]$ .
  - ▶ see next slide

## Detection between two non-zero values with cross-correlation

$$e^{-\frac{\sum (r_i - A)^2}{2\sigma^2} + \frac{\sum (r_i - B)^2}{2\sigma^2}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$- \sum (r_i - A)^2 + \sum (r_i - B)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$2 \sum r_i A - NA^2 - 2 \sum r_i B + NB^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$\frac{1}{N} \sum r_i A - \frac{A^2}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{N} \sum r_i B - \frac{B^2}{2} + \frac{1}{N} \sigma^2 \ln K$$

# Detection between two non-zero values with cross-correlation

- ▶ For Maximum Likelihood ( $K = 1$ ):

$$\langle \mathbf{r}, \mathbf{A} \rangle - \frac{\langle \mathbf{A}, \mathbf{A} \rangle}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, \mathbf{B} \rangle - \frac{\langle \mathbf{B}, \mathbf{B} \rangle}{2}$$

- ▶ If the two values are opposite,  $B = -A$ , choose the most similar to  $\mathbf{r}$ :
  - ▶ cross-correlation measures similarity

$$\langle \mathbf{r}, \mathbf{A} \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, -\mathbf{A} \rangle$$

- ▶ For other criteria: with an extra offset factor  $\frac{1}{N} \sigma^2 \ln K$

# Exercise

Exercise:

- ▶ A signal can have two values,  $-4$  (hypothesis  $H_0$ ) or  $5$  (hypothesis  $H_1$ ). The signal is affected by AWGN  $\mathcal{N}(0, \sigma^2 = 1)$ . The receiver takes 3 samples with values  $\{1.1, 4.4, 2.2\}$ .
  1. What is decision according to Maximum Likelihood criterion? Use all three interpretations.

## II.4 Detection of general signals with multiple samples

## Multiple samples from a general (non-constant) signal

- ▶ We want to detect a **general (non-constant)** signal  $s(t)$
- ▶ The  $N$  samples are taken at times  $\mathbf{t} = [t_1, t_2, \dots, t_N]$  and are arranged in a **sample vector**

$$\mathbf{r} = [r_1, r_2, \dots, r_N]$$

- ▶ What changes compared to constant signals?

# Hypotheses

- ▶ In each hypothesis, the signal is a **random process**
  - ▶  $H_0$ : random process with average value 0
  - ▶  $H_1$ : random process with average value  $s(t)$
- ▶ The sample  $r_i$ , at time  $t_i$ , is:
  - ▶  $0 + \text{noise}$ , in hypothesis  $H_0$
  - ▶  $s(t_i) + \text{noise}$ , in hypothesis  $H_1$
- ▶ The whole sample vector  $\mathbf{r}$  is
  - ▶  $0 + \text{noise}$ , in hypothesis  $H_0$
  - ▶  $s(t) + \text{noise}$ , in hypothesis  $H_1$ , for  $t$  being all the sample times  $t_i$
- ▶ The distribution of the whole vector  $\mathbf{r}$  is described by a function  $w_N(\mathbf{r})$

# Likelihood of vector samples

- ▶ We can apply **the same criteria** based on likelihood ratio as for constant signals:

$$\frac{w_N(\mathbf{r}|H_0)}{w_N(\mathbf{r}|H_1)} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

- ▶ The difference is that the “true” underlying signals are now
  - ▶  $[0, 0, \dots, 0]$  in hypothesis  $H_0$
  - ▶  $[s(t_1), s(t_2), \dots, s(t_N)]$  in hypothesis  $H_1$



# Separation

- ▶ The joint distribution  $w_N(\mathbf{r}|H_j)$  can be decomposed as a product

$$w_N(\mathbf{r}|H_j) = w(r_1|H_j) \cdot w(r_2|H_j) \cdot \dots \cdot w(r_N|H_j)$$

- ▶ Then all likelihood ratio criteria can be written as

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{w(r_1|H_1)}{w(r_1|H_0)} \cdot \frac{w(r_2|H_1)}{w(r_2|H_0)} \dots \frac{w(r_N|H_1)}{w(r_N|H_0)} \underset{H_0}{\overset{H_1}{\geq}} K$$

- ▶ The likelihood ratio of a sample  $r_i$  is computed considering the two possible values of the underlying signal, 0 and  $s(t_i)$ 
  - ▶ for constant signals, the two values were 0 and  $A$  all the time
  - ▶ now they are 0 and  $s(t_i)$ , depending on the sample times  $t_i$
  - ▶ the sample times  $t_i$  should be chosen such as to maximize the performance of detection

## Particular case: AWGN

- ▶ AWGN = “Additive White Gaussian Noise”
- ▶ In hypothesis  $H_1$ :  $w(r_i|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r_i-s(t_i))^2}{2\sigma^2}}$
- ▶ In hypothesis  $H_0$ :  $w(r_i|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{r_i^2}{2\sigma^2}}$
- ▶ Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i-s(t_i))^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}}$$

- ▶ We can interpret this likelihood ratio in two ways

## Interpretation 1: average value of samples

- ▶ Interpretation 1: average value of samples
- ▶ Cannot be used anymore, since the values  $s(t_i)$  are not the same

## Interpretation 2: geometrical

- ▶ Useful mainly for Maximum Likelihood criterion
- ▶ Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i - s(t_i))^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\geq}} K$$

- ▶ For Maximum Likelihood we compare to 1

$$\begin{aligned} \frac{e^{-\frac{\sum (r_i - s(t_i))^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} &\underset{H_0}{\overset{H_1}{\geq}} 1 \\ e^{-\frac{\sum (r_i - s(t_i))^2}{2\sigma^2} + \frac{\sum (r_i)^2}{2\sigma^2}} &\underset{H_0}{\overset{H_1}{\geq}} 1 \\ -\sum (r_i - s(t_i))^2 + \sum (r_i)^2 &\underset{H_0}{\overset{H_1}{\geq}} 0 \\ \sum (r_i)^2 &\underset{H_0}{\overset{H_1}{\geq}} \sum (r_i - s(t_i))^2 \end{aligned}$$

## Interpretation 2: geometrical

- ▶  $\sqrt{\sum (r_i)^2}$  is the geometrical (Euclidian) distance between point  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  and point  $\mathbf{0} = [0, 0, \dots, 0]$
- ▶  $\sqrt{\sum (r_i - s(t_i))^2}$  is the geometrical (Euclidian) distance between point  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  and point  $\mathbf{s}(\mathbf{t}) = [s(t_1), s(t_2), \dots, s(t_N)]$
- ▶ ML decision chooses **the closest signal vector (point)** to the received vector (point), in a N-dimensional space
  - ▶ it is known as “minimum distance receiver”
  - ▶ same interpretation as in the 1-D case
- ▶ Question: what is the geometrical interpretation for the other criteria?

# Exercise

## Exercise:

- ▶ Consider detecting a signal  $s(t) = 3 \sin(2\pi f_1 t)$  that can be present (hypothesis  $H_1$ ) or not (hypothesis  $H_0$ ). The signal is affected by AWGN  $\mathcal{N}(0, \sigma^2 = 1)$ . The receiver takes 2 samples.
  1. What are the best sample times  $t_1$  and  $t_2$  to maximize detection performance?
  2. The receiver takes 2 samples with values  $\{1.1, 4.4\}$ , at sample times  $t_1 = \frac{0.125}{f_1}$  and  $t_2 = \frac{0.625}{f_1}$ . What is decision according to Maximum Likelihood criterion? Use the geometrical interpretation.
  3. What if the receiver takes an extra third sample at time  $t_3 = \frac{0.5}{f_1}$ . Will the detection be improved?

## Interpretation 3: cross-correlation

- Likelihood ratio for vector  $\mathbf{r}$

$$\frac{w_N(\mathbf{r}|H_1)}{w_N(\mathbf{r}|H_0)} = \frac{e^{-\frac{\sum (r_i - s(t_i))^2}{2\sigma^2}}}{e^{-\frac{\sum (r_i)^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$e^{-\frac{\sum (r_i - s(t_i))^2}{2\sigma^2} + \frac{\sum (r_i)^2}{2\sigma^2}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$-\sum (r_i - s(t_i))^2 + \sum (r_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$2\sum r_i s(t_i) - \sum s(t_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$\frac{1}{N} \sum r_i s(t_i) \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\frac{1}{2} \frac{\sum s(t_i)^2}{N} + \frac{1}{N} \sigma^2 \ln K}_{L = \text{const}}$$

## Interpretation 3: cross-correlation

- ▶  $\frac{1}{N} \sum r_i s(t_i)$  is the cross-correlation of the received samples  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  with the **target** samples  $\mathbf{s}(\mathbf{t}_i) = [s(t_1), s(t_2), \dots, s(t_N)]$
- ▶ If the cross-correlation of the received samples with the target samples  $\mathbf{s}(\mathbf{t}_i)$  is greater than a certain threshold  $L$ , we decide that signal is detected.
  - ▶ otherwise, the signal is rejected
  - ▶ cross-correlation is a measure of similarity



## Generalization: two non-zero signals

- ▶ Generalization: decide between **two different signals**  $s_0(t)$  and  $s_1(t)$ 
  - ▶ still with Gaussian noise
- ▶ Interpretation 2: geometric
  - ▶ choose minimum Euclidean distance from  $\mathbf{r} = [r_1, r_2, \dots, r_N]$  to points  $\mathbf{s}_0(\mathbf{t}) = [s_0(t_1), s_0(t_2), \dots]$  and  $\mathbf{s}_1(\mathbf{t}) = [s_1(t_1), s_1(t_2), \dots]$
- ▶ Interpretation 3: cross-correlation
  - ▶ compute cross-correlation of  $\mathbf{r}$  with  $\mathbf{s}_0(\mathbf{t}) = [s_0(t_1), s_0(t_2), \dots]$  and with  $\mathbf{s}_1(\mathbf{t}) = [s_1(t_1), s_1(t_2), \dots]$ ,  $\langle \mathbf{r}, \mathbf{s}_0 \rangle$  and  $\langle \mathbf{r}, \mathbf{s}_1 \rangle$ .
  - ▶ see next slide

## Detection between two non-zero signals with cross-correlation

$$e^{-\frac{\sum (r_i - s_1(t_i))^2}{2\sigma^2} + \frac{\sum (r_i - s_0(t_i))^2}{2\sigma^2}} \underset{H_0}{\overset{H_1}{\gtrless}} K$$

$$- \sum (r_i - s_1(t_i))^2 + \sum (r_i - s_0(t_i))^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$2 \sum r_i s_1(t_i) - \sum s_1(t_i)^2 - 2 \sum r_i s_0(t_i) + \sum s_0(t_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} 2\sigma^2 \ln K$$

$$\frac{1}{N} \sum r_i s_1(t_i) - \sum s_1(t_i)^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{N} \sum r_i s_0(t_i) - \sum s_0(t_i)^2 + \frac{1}{N} \sigma^2 \ln K$$

# Detection between two non-zero signals with cross-correlation

- ▶ For Maximum Likelihood ( $K = 1$ ):

$$\langle \mathbf{r}, \mathbf{s}_1 \rangle - \frac{\langle \mathbf{s}_1, \mathbf{s}_1 \rangle}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, \mathbf{s}_0 \rangle - \frac{\langle \mathbf{s}_0, \mathbf{s}_0 \rangle}{2}$$

- ▶ If the two signals have the same energy:  $\sum s_1(t_i)^2 = \sum s_0(t_i)^2$ , then  $\langle \mathbf{s}_1, \mathbf{s}_1 \rangle = \langle \mathbf{s}_0, \mathbf{s}_0 \rangle$ , so we choose **the signal most similar to  $\mathbf{r}$** :
  - ▶ cross-correlation measures similarity

$$\langle \mathbf{r}, \mathbf{s}_1 \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, \mathbf{s}_0 \rangle$$

- ▶ Examples:

- ▶ BPSK modulation:  $s_1 = A \cos(2\pi ft)$ ,  $s_0 = -A \cos(2\pi ft)$
- ▶ 4-PSK modulation:  $s_{n=0,1,2,3} = A \cos(2\pi ft + n\frac{\pi}{4})$

## Detection with correlator circuit

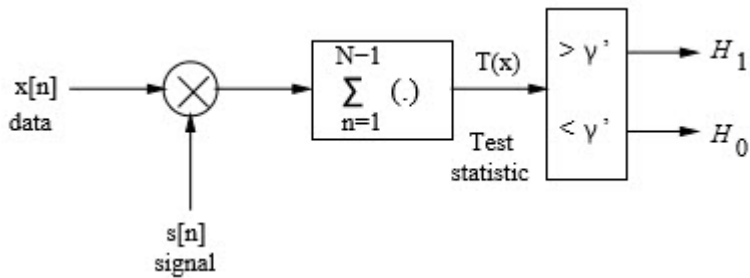


Figure 5: Signal detection using a correlator

[image from <http://nptel.ac.in/courses/117103018/43>]

# Detection of two signals

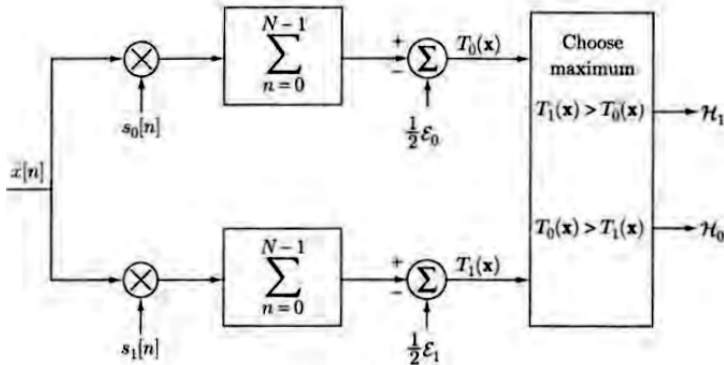


Figure 6: Decision between two signals

[source: *Fundamentals of Statistical Signal Processing*, Steven Kay]

# Matched filters

- ▶ How to compute the cross-correlation of two signals  $r[n]$  and  $s[n]$  of length  $N$ ?

$$\langle r, s \rangle = \frac{1}{N} \sum r_i s(t_i)$$

- ▶ Let  $h[n]$  be the signal  $s[n]$  **flipped** / **mirrored** (“oglindit”)
  - ▶ still starting from time 0 onwards, we want causality

$$h[n] = s[N - 1 - n]$$

- ▶ The convolution of  $r[n]$  with  $h[n]$  is

$$y[n] = \sum_k r[k] h[n - k] = \sum_k r[k] s[N - 1 - n + k]$$

- ▶ The convolution sampled at the end of the signal,  $y[N - 1]$  ( $n = N - 1$ ), is the cross-correlation
  - ▶ up to a scaling constant  $\frac{1}{N}$

$$y[N - 1] = \sum_k r[k] s[k]$$

# Matched filters

- ▶ To detect a signal  $s[n]$  we can use a **filter with impulse response = mirrored version of  $s[n]$** , and take the final sample of the output
  - ▶ it is identical to computing the cross-correlation

$$h[n] = s[N - 1 - n]$$

- ▶ **Matched filter** = a filter designed to have the impulse response the flipped version of a signal we search for
  - ▶ the filter is *matched* to the signal we want to detect
  - ▶ rom. “filtru adaptat”

# Matched filters

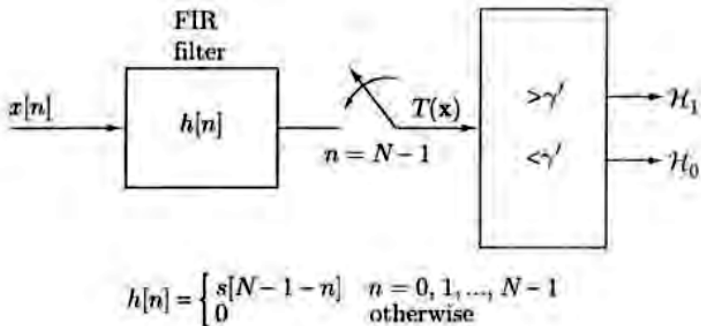


Figure 7: Signal detection with matched filter

[source: *Fundamentals of Statistical Signal Processing*, Steven Kay]



## II.5 Detection of general signals with continuous observations

# Continuous observation of a general signal

- ▶ Continuous observation = we don't take samples anymore, we use **all the continuous signal**
  - ▶ like taking  $N$  samples but with  $N \rightarrow \infty$
- ▶ Received signal is  $r(t)$
- ▶ Target signal is  $s(t)$
- ▶ Assume Gaussian noise only
- ▶ How to detect?

- ▶ Extend the previous case of  $N$  samples to the case a full continuous signal
- ▶ Interpretation 1: average value of samples
  - ▶ Cannot be used anymore, since the values  $s(t_i)$  are not the same

## Interpretation 2: geometrical

- ▶ Interpretation 2: geometrical
- ▶ Each signal  $r(t)$ ,  $s(t)$  or 0 is a data point in an infinite-dimensional Euclidean space
- ▶ Distance between two signals is

$$d(r, s) = \sqrt{\int (r(t) - s(t))^2 dt}$$

- ▶ Similar with the N dimensional case, but with integral instead of sum
- ▶ Maximum Likelihood criterion:
  - ▶ compute distance  $d(r, s)$  from  $r(t)$  to  $s(t)$
  - ▶ compute distance  $d(r, 0)$  from  $r(t)$  to 0
  - ▶ choose the minimum

## Interpretation 3: cross-correlation

- ▶ The cross correlation of a continuous signal  $r(t)$  with a target signal  $s(t)$  of length  $T$

$$\langle \mathbf{r}, \mathbf{s} \rangle = \frac{1}{T} \int_0^T r(t) \cdot s(t) dt$$

- ▶ If the cross-correlation of the received signal with the true signal  $\mathbf{s}(\mathbf{t}_i)$  is greater than a certain threshold  $L$ , we decide that signal is detected.
  - ▶ otherwise, the signal is rejected
  - ▶ cross-correlation is a measure of similarity

# Generalizations

- ▶ Detection **between two signals**  $s_0(t)$  and  $s_1(t)$ 
  - ▶ still with Gaussian noise
- ▶ Interpretation 2: geometric
  - ▶ choose minimum Euclidean distance from point  $\mathbf{r}(\mathbf{t})$  to points  $\mathbf{s}_0(\mathbf{t})$  and  $\mathbf{s}_1(\mathbf{t})$ 
    - ▶ using the specified distance formula
- ▶ Interpretation 3: cross-correlation
  - ▶ compute cross-correlation of  $\mathbf{r}(\mathbf{t})$  with  $\mathbf{s}_0(\mathbf{t})$  and with  $\mathbf{s}_1(\mathbf{t})$ .

# Detection between two non-zero signals with cross-correlation

- ▶ For Maximum Likelihood ( $K = 1$ ):

$$\langle \mathbf{r}, \mathbf{s}_1 \rangle - \frac{\langle \mathbf{s}_1, \mathbf{s}_1 \rangle}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, \mathbf{s}_0 \rangle - \frac{\langle \mathbf{s}_0, \mathbf{s}_0 \rangle}{2}$$

- ▶ If the two signals have the same energy:  $\int s_1(t)^2 dt = \int s_0(t)^2 dt$ , then  $\langle \mathbf{s}_1, \mathbf{s}_1 \rangle = \langle \mathbf{s}_0, \mathbf{s}_0 \rangle$ , so we choose **the signal most similar to  $\mathbf{r}$** :
  - ▶ cross-correlation measures similarity

$$\langle \mathbf{r}, \mathbf{s}_1 \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \langle \mathbf{r}, \mathbf{s}_0 \rangle$$

- ▶ Examples:

- ▶ BPSK modulation:  $s_1 = A \cos(2\pi ft)$ ,  $s_0 = -A \cos(2\pi ft)$
- ▶ 4-PSK modulation:  $s_{n=0,1,2,3} = A \cos(2\pi ft + n\frac{\pi}{4})$

# Matched filters

- ▶ Cross-correlation of signals can be computed with **matched filters**
- ▶ **Matched filter** = a filter designed to have the impulse response the flipped version of a signal we search for
  - ▶ the filter is *matched* to the signal we want to detect
  - ▶ rom. “filtru adaptat”
  - ▶ filter is continuous, continuous impulse response
- ▶ To detect a signal  $s(t)$  we use a matched filter and take the sample of the output at the final moment of the input signal
  - ▶ it is identical with computing cross-correlation