Detection Theory

Chapter 3. Statistical Decision Theory I.

Isael Diaz

Oct 26th 2010

Outline

- Neyman-Pearson Theorem
- Detector Performance
- Irrelevant Data
- Minimum Probability of Error
- Bayes Risk
- Multiple Hypothesis Testing

Chapter Goal

- State basic statistical groundwork for detectors design
- Address simple hypothesis , where PDFs are known (Chapter 6 will cover the case of unknown PDFs)
- Introduction to the classical approach based on Neyman-Pearson and Bayes Risk
  - Bayesian methods employ prior knowledge of the hypotheses
  - Neyman-Pearson has no prior knowledge

Bayes vs Newman-Pearson

- The particular method employed depends upon our willingness to incorporate prior knowledge about the probabilities of ocurrence of the various hypotheses

- Typical cases
  - Bayes : Communication systems, Pattern recognition
  - NP : Sonar, Radar

Nomenclature

$N(\mu, \sigma^2)$: Gaussian PDF with mean μ and variance σ^2

H_i : Hypothesis i

$p(x \mid H_i)$: Probability of observing x under assumprion H_i

$P(H_i; H_j)$: Probability of detecting H_i, when it is H_j

$P(H_i \mid H_j)$: Probability of detecting H_i, when H_j was observed

$L(x)$: Likelihood of x

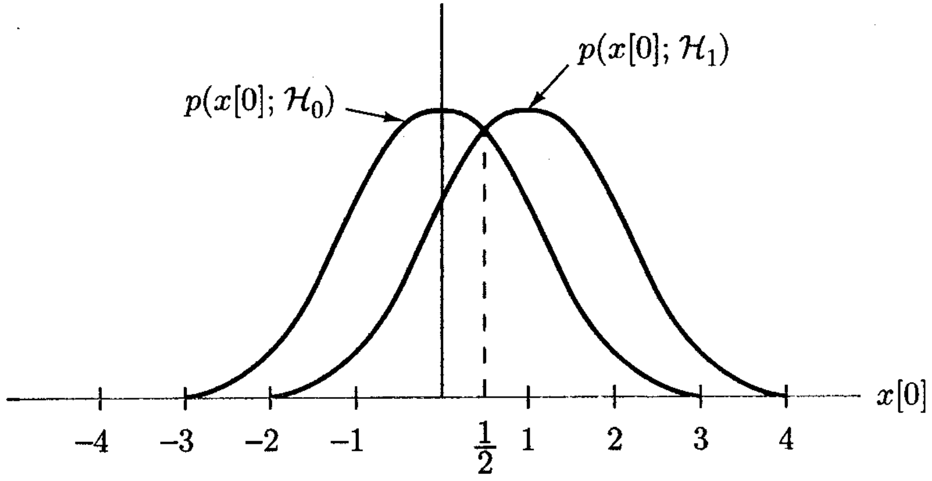
Nomenclature cont...

- Note that $P(H_i; H_j)$ is the probability of deciding $\mathcal{H}i$, if $\mathcal{H}j$ is true.

- While $P(H_i \mid H_j)$ assumes that the outcome of a probabilistic experiment is observed to be $\mathcal{H}j$ and that the probability of deciding $\mathcal{H}i$ is conditioned to that outcome.

Neyman-Pearson Theorem

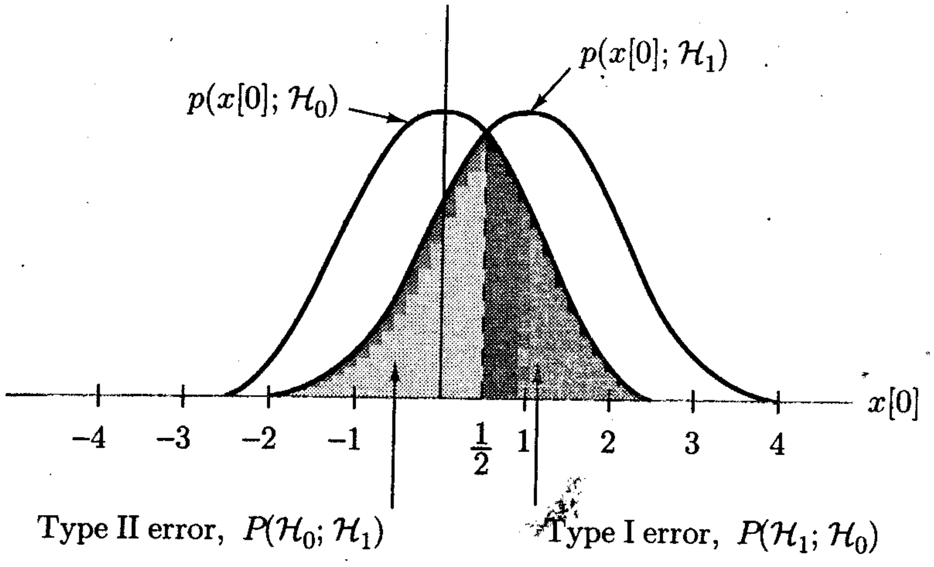
$$\mathcal{H}_0 : \mu = 0$$
$$\mathcal{H}_1 : \mu = 1$$



Neyman-Pearson Theorem cont...

$$H_0 : x[0] = w[0]$$

$$H_1 : x[0] = s[0] + w[0]$$



- It is no possible to reduce both error probabilities simultaneously

- In Neyman-Pearson the approach is to hold one error probability while minimizing the other.

Neyman-Pearson Theorem cont...

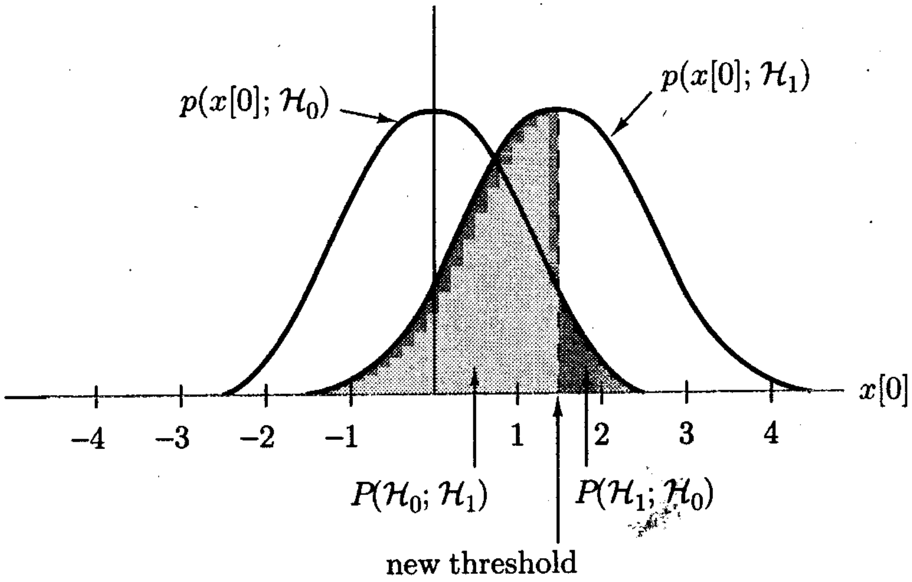
- We wish to maximize the probability of detection

$$P_D = P(H_1; H_1)$$

- And constraint the probability of false alarm

$$P_{FA} = P(H_1; H_0) = \alpha$$

- By choosing the threshold γ



Neyman-Pearson Theorem cont...

- To maximize the probability of detection for a given probability of false alarm decide for H1 if

$$L(x) = \frac{p(x; H_1)}{p(x; H_0)} > \gamma$$

Likelihood ratio test

- with

$$P_{FA} = \int_{\{x : L(x) > \gamma\}} p(x; H_0)\, dx = \alpha$$

Likelihood Ratio

- The LR test rejects the null hypothesis if the value of this statistic is too small.
- Intuition: large Pfa -> small gamma -> not much larger p(x,H1) than p(x;H0)

- Limiting case: Pfa->1
- Limiting case: Pfa->0

$$L(x) = \frac{p(x;H_1)}{p(x;H_0)} > \gamma$$

Neyman-Pearson Theorem cont...

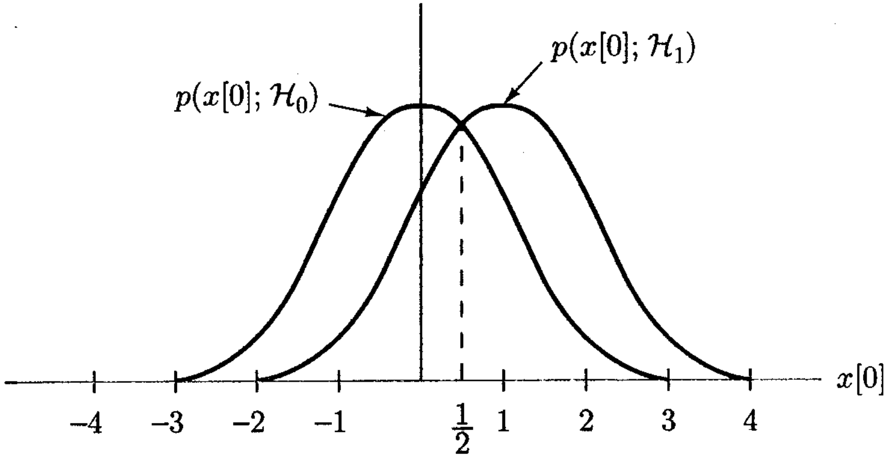
- Example 1
- Maximize

$$P_D = P(H_1; H_1)$$

- Calculate the threshold so that

$$P_{FA} = 10^{-3}$$

$$\mathcal{H}_0 : \mu = 0$$
$$\mathcal{H}_1 : \mu = 1$$



Neyman-Pearson Theorem Example 1

$$\frac{p(\mathbf{x};\mathcal{H}_1)}{p(\mathbf{x};\mathcal{H}_0)} = \frac{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x[0]-1)^2\right]}{\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^2[0]\right]} > \gamma \quad (1)$$

$$\exp\left[-\frac{1}{2}(x^2[0] - 2x[0] + 1 - x^2[0])\right] > \gamma \quad (2)$$

$$\exp\left(x[0] - \frac{1}{2}\right) > \gamma. \quad (3)$$

$$\gamma = \exp(\beta) \quad (4)$$

$$\exp\left(x[0] - \frac{1}{2}\right) > \exp(\beta) \quad (5)$$

$$x[0] > \beta + \frac{1}{2} = \underbrace{\boxed{\ln\gamma + \frac{1}{2}}}_{\gamma'}. \quad (6)$$

$$P_{FA} = \Pr\{x[0] > \gamma'; \mathcal{H}_0\} = 10^{-3} \quad (7)$$

$$\int_{\gamma'}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt = 10^{-3} \quad (8)$$

$$\gamma' = 3$$

$$\begin{aligned} P_D &= \Pr\{x[0] > 3; \mathcal{H}_1\} \\ &= \int_3^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-1)^2\right] dt = 0.023. \end{aligned}$$

Question

How can Pd be improved?

Answer is: Only if we allow for a larger Pfa.

NP gives the optimal detector!

By allowing a Pfa of 0.5, the Pd is increased to 0.85

System Model

- In the General case

$$\begin{array}{ll} \mathcal{H}_0 : x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 : x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{array}$$

- Where the signal s[n]=A for A>0 and w[n] is WGN with variance σ^2.
- It can be shown that the sum of x is a ***sufficient statistic***
- The summation of x[n] is sufficient.

Detector Performance

- In a more general definition

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right).$$

Energy-to-noise-ratio

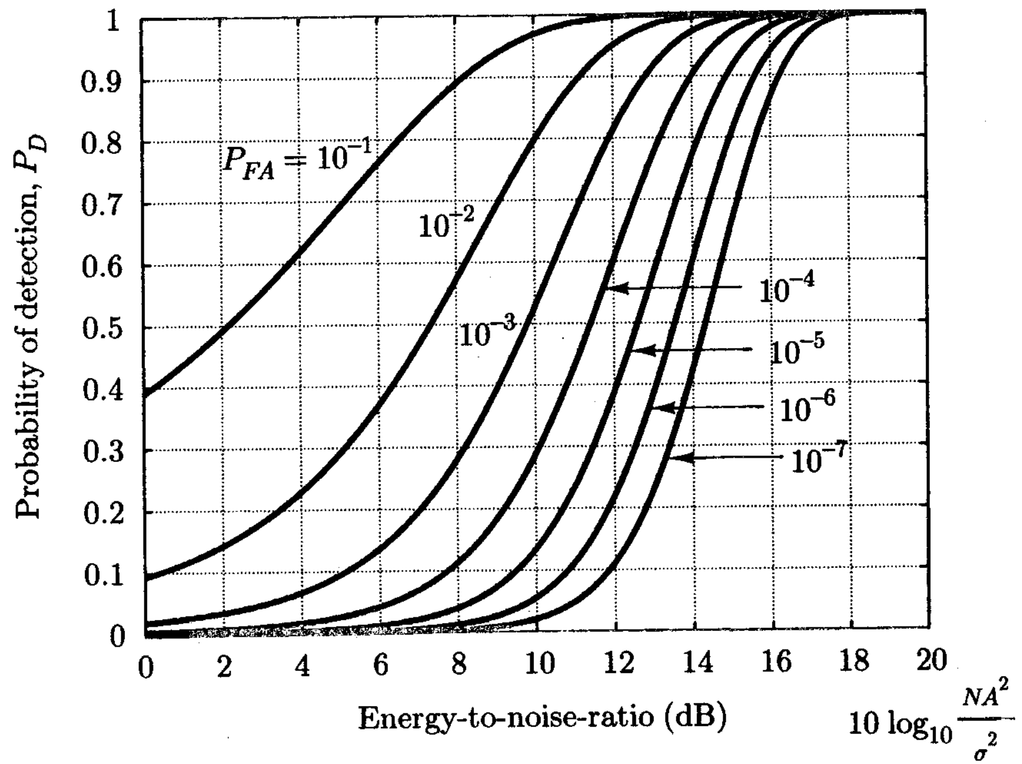
Or

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$

where

$$d^2 = \frac{(\mu_1 - \mu_0)^2}{\sigma^2}.$$

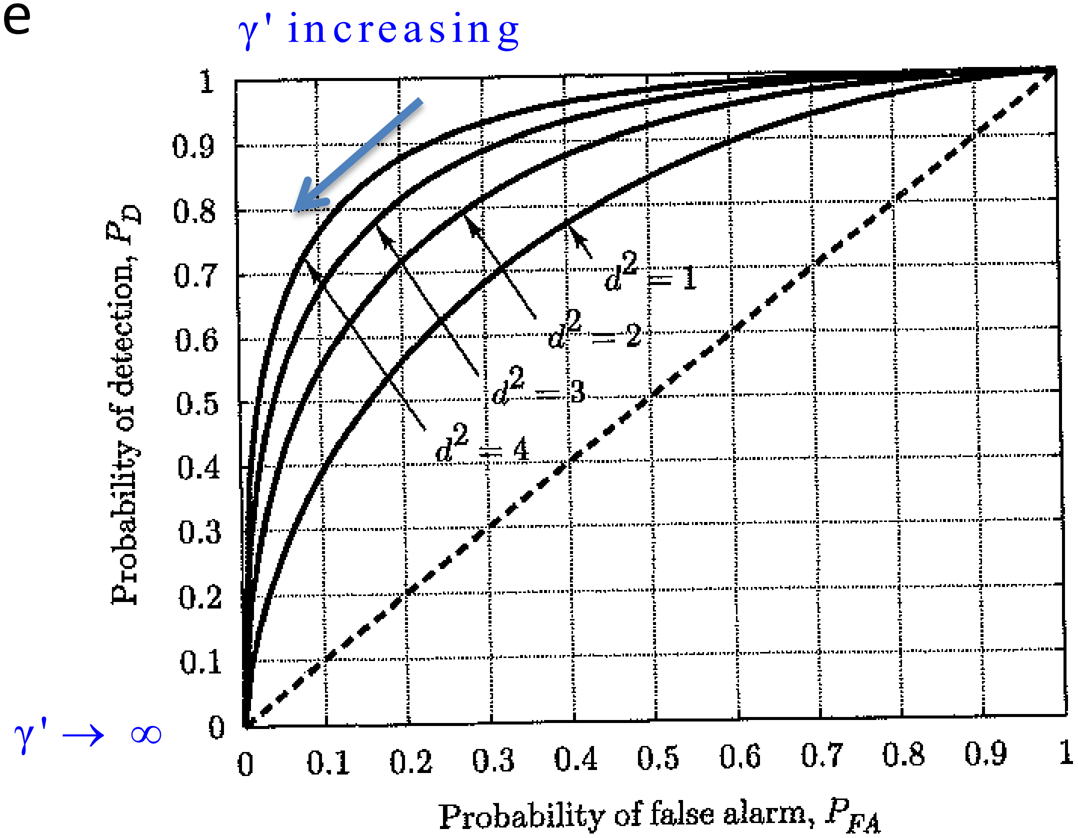
Deflection Coefficient



Detector Performance cont...

- An alternative way to present the detection performance

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right)$$



Receiver Operating Characteristic (ROC)

Irrelevant data

- Some data not belonging to the desired signal might actually be useful.
- The observed data-set is $\{x[0], x[1], \ldots, x[N-1], w_R[0], w_R[1], \ldots, w_R[N-1]$
- If

$$H_0 : \begin{array}{l} x[n] = w[n] \\ w_R = w[n] \end{array} \qquad H_1 : \begin{array}{l} x[n] = A + w[n] \\ w_R = w[n] \end{array}$$

- Then the reference noise can be used to improve the detection. The perfect detector would become

$$T = x[0] - w_R[0] > \frac{A}{2}$$

Irrelevant data cont...

- Generalizing, using the NP theorem

$$L(\mathbf{x}_1, \mathbf{x}_2) = \frac{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_1)}{p(\mathbf{x}_1, \mathbf{x}_2; \mathcal{H}_0)} = \frac{p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_1)p(\mathbf{x}_1; \mathcal{H}_1)}{p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_0)p(\mathbf{x}_1; \mathcal{H}_0)}.$$

- If they are un-correlated

$$p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_1) = p(\mathbf{x}_2|\mathbf{x}_1; \mathcal{H}_0)$$

- Then, x2 is irrelevant and

$$L(x_1, x_2) = L(x_1)$$

Minimum Probability Error

- The probability of error is defined with Bayesian paradigm as

$$\begin{aligned} P_e &= \Pr\{\text{decide } \mathcal{H}_0, \mathcal{H}_1 \text{ true}\} + \Pr\{\text{decide } \mathcal{H}_1, \mathcal{H}_0 \text{ true}\} \\ &= P(\mathcal{H}_0|\mathcal{H}_1)P(\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)P(\mathcal{H}_0) \end{aligned}$$

- Minimizing the probability of error, we decide H1 if

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{P(\mathcal{H}_0)}{P(\mathcal{H}_1)} = \gamma.$$

- If both prior probabilities are equal we have

$$p(\mathbf{x}|\mathcal{H}_1) > p(\mathbf{x}|\mathcal{H}_0).$$

Minimum Probability Error example

- An On-off key communication system , where we transmit either

$S_0[n] = 0$ or $S_1[n] = A$

$$\begin{aligned} \mathcal{H}_0 &: x[n] = w[n] & n = 0, 1, \ldots, N-1 \\ \mathcal{H}_1 &: x[n] = A + w[n] & n = 0, 1, \ldots, N-1 \end{aligned}$$

- The probability of transmitting 0 or A is the same ½. Then, in order to minimize the probability of error $\gamma=1$

Minimum Probability Error example ...

$$\frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right]} > 1.$$

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

Analogous to NP, we decide H_1 if $\bar{x} > A/2$

Then we calculate the Pe

$$\bar{x} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_0 \\ \mathcal{N}(A, \frac{\sigma^2}{N}) & \text{conditioned on } \mathcal{H}_1. \end{cases}$$

$$\begin{aligned} P_e &= \frac{1}{2}\left[P(\mathcal{H}_0|\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_0)\right] \\ &= \frac{1}{2}\left[\Pr\{\bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_0\}\right] \\ &= \frac{1}{2}\left[\left(1 - Q\left(\frac{A/2 - A}{\sqrt{\sigma^2/N}}\right)\right) + Q\left(\frac{A/2}{\sqrt{\sigma^2/N}}\right)\right] \end{aligned}$$

$$P_e = Q\left(\sqrt{\boxed{\frac{NA^2}{4\sigma^2}}}\right).$$

Deflection Coefficient

The probability error decreases monotonically with respect of the deflection coefficient

Bayes Risk

- Bayes Risk assigns costs to the type of errors, C_{ij} denotes the cost if we decide H_i but H_j is true. If no error is made, no cost is assigned.

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

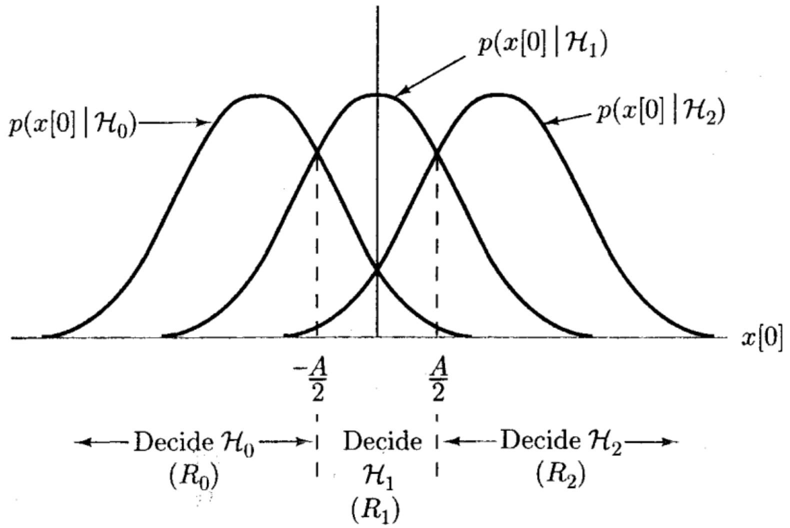
- The detector that minimizes Bayes risk is the following

$$\frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} > \frac{(C_{10} - C_{00}) P(\mathcal{H}_0)}{(C_{01} - C_{11}) P(\mathcal{H}_1)} = \gamma.$$

Multiple Hyphoteses testing

- If we have a communication system, M-ary PAM (pulse amplitud modulation) with M=3.

$$\begin{aligned}
\mathcal{H}_0 &: x[n] = -A + w[n] \quad && n = 0, 1, \ldots, N-1 \\
\mathcal{H}_1 &: x[n] = w[n] && n = 0, 1, \ldots, N-1 \\
\mathcal{H}_2 &: x[n] = A + w[n] && n = 0, 1, \ldots, N-1
\end{aligned}$$



Multiple Hyphoteses testing cont...

- The conditional PDF describing the communication system is given by

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A_i)^2\right]$$

- In order to maximize *p(x|Hi)*, we can minimize

$$D_i^2 = \sum_{n=0}^{N-1}(x[n]-A_i)^2.$$

Multiple Hyphoteses testing cont...

- Then

$$
\begin{aligned}
D_i^2 &= \sum_{n=0}^{N-1} (x[n] - \bar{x} + \bar{x} - A_i)^2 \\
&= \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 + 2(\bar{x} - A_i) \sum_{n=0}^{N-1} (x[n] - \bar{x}) + N(\bar{x} - A_i)^2 \\
&= \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 + N(\bar{x} - A_i)^2.
\end{aligned}
$$

- Therefore

$$
\begin{array}{ll}
\mathcal{H}_0 & \text{if } \bar{x} < -A/2 \\
\mathcal{H}_1 & \text{if } -A/2 < \bar{x} < A/2 \\
\mathcal{H}_2 & \text{if } \bar{x} > A/2.
\end{array}
$$

Multiple Hyphoteses testing cont...

- There are six types of $\mathcal{P}e$, so Instead of calculating $\mathcal{P}e$, we calculate $\mathcal{P}c=1-\mathcal{P}e$, the probability of correct desicion

$$\begin{aligned} P_c &= \sum_{i=0}^{2} P(\mathcal{H}_i|\mathcal{H}_i)P(\mathcal{H}_i) \\ &= \frac{1}{3}\sum_{i=0}^{2} P(\mathcal{H}_i|\mathcal{H}_i) \\ &= \frac{1}{3}\left[\Pr\{\bar{x} < -A/2|\mathcal{H}_0\} + \Pr\{-A/2 < \bar{x} < A/2|\mathcal{H}_1\} + \Pr\{\bar{x} > A/2|\mathcal{H}_2\}\right]. \end{aligned}$$

- And finally

$$\begin{aligned} P_c &= \frac{1}{3}\left[1 - Q\left(\frac{-\frac{A}{2}+A}{\sqrt{\sigma^2/N}}\right) + Q\left(\frac{-\frac{A}{2}}{\sqrt{\sigma^2/N}}\right) - Q\left(\frac{\frac{A}{2}}{\sqrt{\sigma^2/N}}\right) + Q\left(\frac{\frac{A}{2}-A}{\sqrt{\sigma^2/N}}\right)\right] \\ &= 1 - \frac{4}{3}Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right) \end{aligned}$$

Multiple Hyphoteses testing cont...

Binary Case

$$P_e = Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right)$$

PAM Case

$$P_e = \frac{4}{3}Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right)$$

In the digital communications world, there is a difference in the energy per symbol, which is not considered in this example.

Chapter Problems

- Warm up problems: 3.1, 3.2
- Problems to be solved in class: 3.4, 3.8, 3.15, 3.18, 3.21