ELEG 5633 Detection and Estimation
Detection Theory I

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

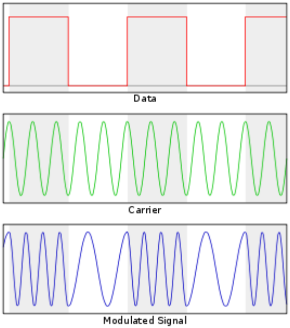
February 9, 2017

Detection Theory

- Binary Bayesian hypothesis test
  - Likelihood Ratio Test
  - Minimum probability of error
  - MAP: maximum a posteriori
  - ML: maximum likelihood
- Neyman-Pearson test
  - Receiving Operating Characteristics (ROC) curves
- M-ary hypothesis testing
- Composite hypothesis testing

A Binary Detection Example

010001000100



$$0 : s_0(t) = \sin(\omega_0 t)$$

$$1 : s_1(t) = \sin(\omega_1 t)$$

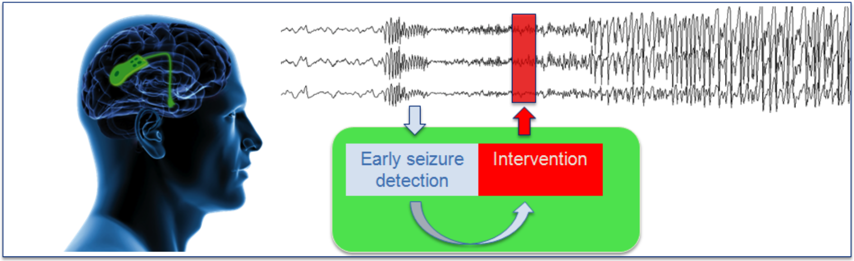
$$x(t) = \begin{cases} s_0(t) + n(t) & \text{if '0' sent} \\ s_1(t) + n(t) & \text{if '1' sent} \end{cases}$$

$$H_0 : \text{'0' is sent}$$

$$H_1 : \text{'1' is sent}$$

$$\hat{H}(x) = H_0? \quad H_1?$$

Another Binary Detection Example



Early seizure detection triggers an intervention.
k different channels: $x_i[n], i = 1, 2, \ldots, k$.

$$H_0 : \text{'no seisure'}$$
$$H_1 : \text{'seisure'}$$
$$\hat{H}(\cdot) : \quad \mathbb{R}^k \rightarrow \{H_0, H_1\}$$

Hypothesis Test

Each possible scenario is a hypothesis.

- M hypothesis, $M \geq 2$: $\{H_0, H_1, \ldots, H_{M-1}\}$

Observed data: $x = [x_1, x_2, \ldots, x_k]^T$

- scalar case: $k = 1$

Hypothesis Test Procedure

- Based on observation x, get an estimat of $\hat{H} \in \{H_0, H_1, \ldots, H_{M-1}\}$

$$\hat{H}(\cdot): \quad \mathbb{R}^k \to \{H_0, H_1, \ldots, H_{M-1}\} \tag{1}$$

Hypothesis Test: Classifications

Random hypothesis test

- H is a random variable, taking one of $\{H_0, H_1, \ldots, H_{M-1}\}$
- A priori probabilities: $\pi_m := \mathbb{P}[H = H_m]$
- Likelihood function: $X \sim p_{X|H}(x|H_i)$

Nonrandom Hypothesis test

- No proper priors about H
- Given $H = H_m$, X is still random: $X \sim p_X(x; H_m)$

Performance Criterion

- Generalized Bayesian risk
- Neyman-Pearson (NP)

Binary Random Hypothesis Testing: H_0 vs H_1

Given:

- Hypothesis

$$\begin{aligned} \text{null hypothesis:} \qquad & H_0 \\ \text{alternative hypothesis:} \qquad & H_1 \end{aligned}$$

- Priors

$$\pi_0 := \mathbb{P}[H = H_0]$$
$$\pi_1 := \mathbb{P}[H = H_1]$$

- Measurement model/ likelihood function

$$H_0 : X \sim p_0$$
$$H_1 : X \sim p_1$$

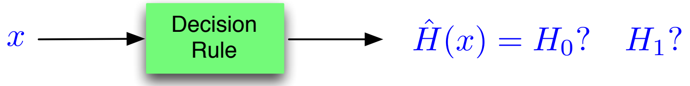
Example

Additive Gaussian White Noise (AWGN) communication channel. A bit, 0 or 1, is sent through a noisy channel. The receiver gets the bit plus noise, and the noise is modeled as a realization of a $\mathcal{N}(0, 1)$ random variable. Assume that both 0 and 1 are equally likely, i.e., $P_0 = P_1 = 1/2$. Once the receiver receives a signal x, it must decide between two hypotheses,

$$\begin{aligned} H_0 : &\quad X \sim \mathcal{N}(0, 1) \\ H_1 : &\quad X \sim \mathcal{N}(1, 1) \end{aligned}$$

Question: How to design the decision rule $\hat{H}(x)$?

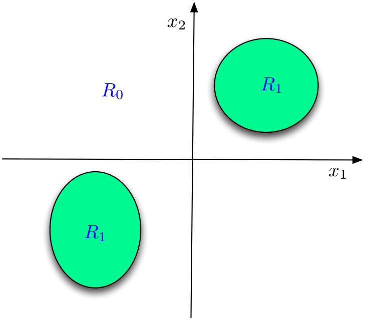
Decision Regions



A decision is made by partitioning the range of X into two disjoint regions.

$$R_0 = \{x : \hat{H}(x) = H_0\}$$

$$R_1 = \{x : \hat{H}(x) = H_1\}$$



Example

$x = [x_1, x_2]^T \in \mathbb{R}^2$

Question: How to design the decision regions R_0 and R_1?

Expected Cost Function Metric

- To optimize the choice of decision regions, we can specify a cost for decisions.
- Four possible outcomes in a test:

	H_0	H_1
$\hat{H} = H_0$	correct	incorrect
$\hat{H} = H_1$	incorrect	correct

- Denoted as $(i, j), i, j \in \{0, 1\}$, i: decision H_i; j: true distribution H_j
- Assign a cost $c_{i,j}$ for each outcome $(i, j), i, j \in \{0, 1\}$

Example

Cost Assignment example:
1) Digital communications: $c_{0,0} = 0$, $c_{0,1} = 1$, $c_{1,0} = 1$, $c_{1,1} = 0$.
2) Seizure detection: H_0 : ‘no seizure’; H_1 : 'seisure'

$$c_{0,1} \gg c_{1,0}$$

Bayes Cost

- Bayes Cost: the expected cost associated with a test

$$
\begin{aligned}
C &= \sum_{i,j=0}^{1} c_{i,j} \cdot \mathbb{P}(\text{decide } H_i, H_j \text{ is true}) \\
&= \sum_{i,j=0}^{1} c_{i,j}\pi_j \cdot \mathbb{P}(\text{decide } H_i | H_j \text{ is true})
\end{aligned}
$$

Bayes Cost

- Express Bayes cost directly in terms of the decision regions

$$
\begin{aligned}
C &= \sum_{i,j=0}^{1} c_{i,j} \pi_j \mathbb{P}(\text{decide } H_i | H_j \text{ is true}) \\
&= \sum_{i,j=0}^{1} c_{i,j} \pi_j \mathbb{P}(X \in R_i | H_j \text{ is true}) \\
&= \sum_{i,j=0}^{1} c_{i,j} \pi_j \int_{R_i} p_j(x) dx \\
&= \int_{R_0} \left[c_{0,0} \pi_0 p_0(x) + c_{0,1} \pi_1 p_1(x) \right] dx + \int_{R_1} \left[c_{1,0} \pi_0 p_0(x) + c_{1,1} \pi_1 p_1(x) \right] dx
\end{aligned}
$$

Objective: design R_0 and R_1 such that the Bayes cost C is minimized.

Optimum Decision Rule: Likelihood Ratio Test

To minimize the Bayes cost, we should design the decision regions as

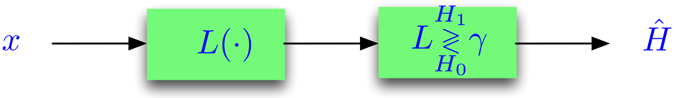
$$
\begin{aligned}
R_0 &:= \{x : c_{0,0}\pi_0 p_0(x) + c_{0,1}\pi_1 p_1(x) < c_{1,0}\pi_0 p_0(x) + c_{1,1}\pi_1 p_1(x)\} \\
R_1 &:= \{x : c_{0,0}\pi_0 p_0(x) + c_{0,1}\pi_1 p_1(x) > c_{1,0}\pi_0 p_0(x) + c_{1,1}\pi_1 p_1(x)\}
\end{aligned}
$$

Therefore, the optimal test takes the following form:

$$
\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0(c_{1,0} - c_{0,0})}{\pi_1(c_{0,1} - c_{1,1})}
$$

Likelihood ratio test (LRT)

$$L(x) := \frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \quad \gamma := \frac{\pi_0(c_{1,0} - c_{0,0})}{\pi_1(c_{0,1} - c_{1,1})} \text{ can be precomputed}$$



Special Cases: MAP Detector

- $c_{ij} = 1 - \delta[i-j]$, then Bayes cost = Probability of error, i.e., P_e
  The likelihood ratio test reduces to

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi_0}{\pi_1}$$

$$\pi_1 p_1(x) \underset{H_0}{\overset{H_1}{\gtrless}} \pi_0 p_0(x)$$

$$\mathbb{P}[H = H_1 | x] \underset{H_0}{\overset{H_1}{\gtrless}} \mathbb{P}[H = H_0 | x]$$

  This is called Max a Posteriori (MAP) Detector

Special Cases: ML Detector

- If $\pi_0 = \pi_1$, $c_{ij} = 1 - \delta[i-j]$, LRT reduces to

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$
$$p_1(x) \underset{H_0}{\overset{H_1}{\gtrless}} p_0(x)$$

This is called Maximum Likelihood (ML) Detector

Properties of LRT

- $L(x)$ is a sufficient statistic.
  It summarizes all relevant information in the observations.
- Any one-to-one transformation on L is a sufficient statistic.
  $g(\cdot)$: an invertible function

  $$l(x) := g(L(x))$$

  If $g(\cdot)$ is monotonically increasing

  $$l(x) \underset{H_0}{\overset{H_1}{\gtrless}} g(\gamma)$$

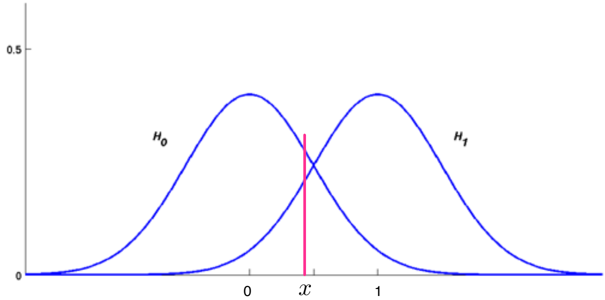
  Eg: $g(x) = \ln x$
- LRT indicates decision regions

  $$R_0 = \{x : \hat{H}(x) = H_0\} = \{x : L(x) < \gamma\}$$
  $$R_1 = \{x : \hat{H}(x) = H_1\} = \{x : L(x) > \gamma\}$$

Example

Additive Gaussian White Noise (AWGN) communication channel. A bit, 0 or 1, is sent through a noisy channel. The receiver gets the bit plus noise, and the noise is modeled as a realization of a $\mathcal{N}(0,1)$ random variable. Assume that both 0 and 1 are equally likely, i.e., $\pi_0 = \pi_1 = 1/2$. Let $c_{i,j} = 1$ if $i \neq j$; $c_{i,j} = 0$ if $i = j$. What is the optimal decision rule to minimize the expected cost?

$$H_0: \quad X \sim \mathcal{N}(0,1)$$
$$H_1: \quad X \sim \mathcal{N}(1,1)$$



Example

Additive Gaussian White Noise (AWGN) communication channel. Two messages $m = \{0, 1\}$ with probabilities $\pi_0 = \pi_1$. Given m, the received signal is a $N \times 1$ random vector

$$X = s_m + W$$

where $W \sim \mathcal{N}(0, \sigma^2 I_{N\times N})$. $\{s_0, s_1\}$ are known to the receiver.
Design an optimal detector w.r.t P_e criterion.

Neyman-Pearson Hypothesis Testing

- Two types of errors

	H_0	H_1
$\hat{H} = H_0$	correct	Miss Detection
$\hat{H} = H_1$	False Alarm	correct

$$\text{Type I: } P_{\mathsf{FA}} = \mathbb{P}[\hat{H} = H_1 | H_0] = \int_{R_1} p_0(x) dx$$

$$\text{Type II: } P_{\mathsf{MD}} = \mathbb{P}[\hat{H} = H_0 | H_1] = \int_{R_0} p_1(x) dx$$

Note: $P_{\mathsf{MD}} = 1 - P_{\mathsf{D}}$, where $P_{\mathsf{D}} = \int_{R_1} p_1(x) dx$.

Neyman-Pearson Hypothesis Testing

- Nyeman-Pearson Design Criterion:
  Minimizes one type of error subject to a constraint on the other type of error.

$$\begin{aligned} \text{minimize} \quad & P_{\mathsf{MD}} \\ \text{subject to} \quad & P_{\mathsf{FA}} \leq \alpha \end{aligned}$$

Such design does not require prior probabilities nor cost assignments.

Neyman-Pearson Theorem

Theorem

To maximize P_D with a given $P_{FA} \leq \alpha$, decide H_1 if

$$L(x) := \frac{p(x|H_1)}{p(x|H_0)} \geq \lambda$$

where λ is found from

$$P_{FA} = \int_{x:L(x)>\lambda} p(x|H_0)dx = \alpha$$

Example

Assume the scalar observation X is generated according to

$$\begin{aligned} H_0 :& X \sim \mathcal{N}(0, \sigma_0^2) \\ H_1 :& X \sim \mathcal{N}(0, \sigma_1^2), \quad \sigma_1^2 > \sigma_0^2 \end{aligned}$$

what is the NP test if we want to have $P_{\mathsf{FA}} \leq 10^{-3}$?

- Note: $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$
- in Matlab, $y = Q(x)$ is qfunc(x), and $x = Q^{-1}(y)$ is qfuncinv(y)

Example

Given N observations, $X_i, i = 1, 2, \ldots, N$, which are i.i.d and, depending on the hypothesis, generated according to

$$
\begin{aligned}
H_0 :& X_i \sim \mathcal{N}(0, \sigma_0^2) \\
H_1 :& X_i \sim \mathcal{N}(0, \sigma_1^2), \quad \sigma_1^2 > \sigma_0^2
\end{aligned}
$$

What is the NP test for $N = 20$ and $P_{\text{FA}} = 0.01$?

- Chi-squared distribution:
  - if $X_i \sim \mathcal{N}(0,1)$, then $Y := \sum_{i=1}^{k} X_i^2 \sim \chi^2(k)$
  - PDF: $p_Y(y;k) = \frac{1}{2\Gamma\left(\frac{k}{2}\right)} \left(\frac{y}{2}\right)^{\frac{k}{2}-1} e^{-\frac{y}{2}}, x > 0$
  - $\Pr(Y > \tau) = \Gamma\left(\frac{\tau}{2}, \frac{k}{2}\right)$, where $\Gamma(x,a) = \frac{1}{\Gamma(a)} \int_x^\infty t^{a-1} e^{-t} dt$ is the upper incomplete Gamma function.
  - in Matlab, $y = \Gamma(x,a)$ is gammainc(x, a, 'upper'). $x = \Gamma^{-1}(y,a)$ is gammaincinv(y, a, 'upper')
  - example: $\tau = 2\times$ gammaincinv(0.01, 20/2, 'upper') = 37.57

Example

Given N observations, $X_i, i = 1, 2, \ldots, N$, which are i.i.d and, depending on the hypothesis, generated according to

$$\begin{aligned} H_0 : &X_i = W_i \\ H_1 : &X_i = \mu + W_i \end{aligned}$$

where $W_i \sim \mathcal{N}(0, \sigma^2)$.

What is the NP test if we want to have $P_{\mathsf{FA}} \leq 10^{-3}$?

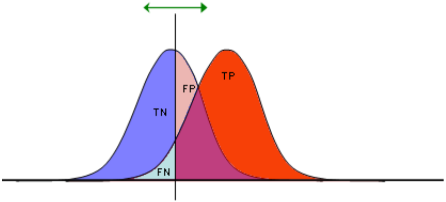
Receiver Operating Characteristics (ROC) of the LRT

ROC: plot P_{D} versus P_{FA}; summarizes the detection performance of NP-test.

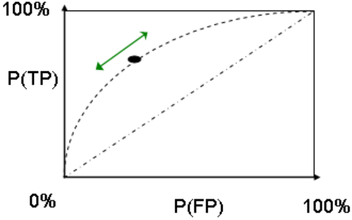
- LRT: $L(x) := \frac{p(x|H_1)}{p(x|H_0)} > \gamma$

$$P_{\mathsf{D}}(\gamma) = \int_{x:L(x)>\gamma} p(x|H_1)dx$$
$$P_{\mathsf{FA}}(\gamma) = \int_{x:L(x)>\gamma} p(x|H_0)dx$$

- $\gamma = 0$: $P_{\mathsf{D}} = P_{\mathsf{FA}} = 1$
- $\gamma = +\infty$: $P_{\mathsf{D}} = P_{\mathsf{FA}} = 0$
- As γ increases, both $P_{\mathsf{D}}, P_{\mathsf{FA}}$ decrease



TP	FP
FN	TN
1	1



TP: true positive; FP: false positive; TN: true negative; FN: false negative

Properties of ROC

- The ROC is a nondecreasing function.

- The ROC is above the line $P_D = P_{FA}$.

- The ROC is concave.

M-ary Hypothesis Testing

- Now, we wish to decide among M hypotheses $\{H_0, H_1 \ldots, H_{M-1}\}$.
- Focus on the Bayesian test.
  Priors: π_0, π_1, . . . , π_{M-1}.
  Likelihood: $p_{X|H_i}(x|H_i)$, denoted as $p_i(x)$.
- The Bayes cost

$$
\begin{aligned}
C &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} c_{ij} \pi_j \cdot \mathbb{P}(\text{decide } H_i | H_j \text{ is true}) \\
&= \sum_{i=0}^{M-1} \int_{R_i} \sum_{j=0}^{M-1} c_{ij} \pi_j p_j(x) dx \\
&= \sum_{i=0}^{M-1} \int_{R_i} h_i(x) dx
\end{aligned}
$$

- Define

$$
h_i(x) := \sum_{j=0}^{M-1} c_{ij} \pi_j p_j(x)
$$

- To minimize the Bayes cost, we let

$$R_i = \{x : h_i(x) < h_j(x), \forall j \neq i\}$$

- Equivalently, for a given x, we pick the hypothesis corresponding to the smallest $C_i(x)$

$$\hat{H}(x) = \underset{i \in \{0,1,\ldots,M-1\}}{\text{argmin}} \; h_i(x)$$

Special Cases: MAP detector

- $c_{ij} = 1 - \delta[i-j]$, then Bayes cost = Probability of error, i.e., P_e

$$h_i(x) = \sum_{j=0}^{M-1} c_{ij}\pi_j p_j(x) = \sum_{j\neq i} \pi_j p_j(x)$$
$$= \sum_{j=0}^{M-1} \pi_j p_j(x) - \pi_i p_i(x)$$

$$\hat{H}(x) = \underset{H_i\in\{H_0,H_1,\ldots,H_{M-1}\}}{\text{argmin}} h_i(x)$$
$$= \underset{H_i\in\{H_0,H_1,\ldots,H_{M-1}\}}{\text{argmax}} \pi_i p_i(x)$$

- Max A Posteriori (MAP) Detector for M-ary case $P(H_i|x) = \frac{\pi_i p_i(x)}{p(x)}$

Special Cases: ML detector

- If $\pi_0 = \ldots = \pi_{M-1} = \frac{1}{M}$, $c_{ij} = 1 - \delta[i-j]$, MAP reduces to

$$\begin{aligned}\hat{H}(x) &= \underset{H_i \in \{H_0, H_1, \ldots, H_{M-1}\}}{\text{argmax}} \pi_i p_i(x) \\ &= \underset{H_i \in \{H_0, H_1, \ldots, H_{M-1}\}}{\text{argmax}} p_i(x)\end{aligned}$$

- Maximum Likelihood (ML) Detector for M-ary case.

Example

Assume that we have three hypotheses

$$\begin{aligned} H_0 :& X_i = -s + W_i, \quad i = 1, \ldots, N \\ H_1 :& X_i = W_i, \qquad\quad i = 1, \ldots, N \\ H_2 :& X_i = s + W_i, \quad i = 1, \ldots, N \end{aligned}$$

where W_i are i.i.d $\mathcal{N}(0, 1)$, s is a positive constant.
What is the optimal decision rule to minimize P_e if $\pi_0 = \pi_1 = \pi_2 = 1/3$?

What if $N = 1$, i.e., a scalar observations?

What if $N > 1$, i.e., multiple samples?

What is the minimum P_e?