Elements of Hypothesis Testing - 1

Prof. C. Emre Koksal

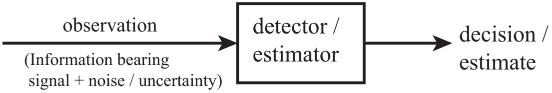
ECE 7001: Stochastic Processes, Detection, and Estimation

The Ohio State University

Outline

Introduction to Detection and Estimation, and Some Notations

Detection vs. Estimation



- detection: outcome is decision $\in$ discrete set
- estimator: outcome is estimate $\in$ continuous set

Detection Examples:

1. Digital Communication

$$\text{observe} \quad y(t) = \begin{cases} x(t) + N(t), & '0' \quad \text{is sent} \\ -x(t) + N(t), & '1' \quad \text{is sent} \end{cases}, \quad 0 \leq t \leq T$$

and decide whether "0" or "1" bit is transmitted.

2. Radar

$$\text{observe} \quad y(t) = \begin{cases} \nu \sin(\omega_c(t-\tau)+\phi) + N(t), & \text{signal is present} \\ N(t), & \text{otherwise} \end{cases}$$

$\{\nu, \tau, \phi\}$ are known, i.e. decide whether an aircraft is present or not.

Estimation Examples:

1. Pulse Amplitude Modulation:

$$\text{observe} \quad y(t) = Ab(t)\cos(\omega_C t) + N(t)$$

estimate $A \in \mathbb{R}$

2. Radar:

$$\text{observe} \quad y(t) = \nu \sin(\omega_c(t-\tau)+\phi) + N(t)$$

estimate $\{\nu, \tau, \phi\}$

Notation:

- observation: y or $\vec{y}$ or $y(t) \in \Gamma$
- outcome: $\theta \in \Lambda$: object of interest (decision or estimate),e.g., in binary hypothesis testing, $\Lambda = \{H_0, H_1\}$

Binary Hypothesis Testing

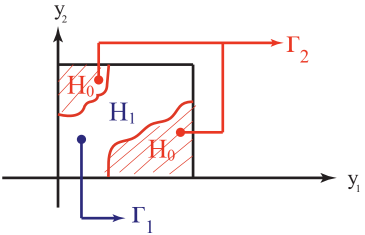
- Two choices: $|\Lambda| = 2 \Rightarrow H_0$ (Null hypothesis) or H_1 (alternate hypothesis)
- Observe: $\vec{Y} = \vec{y} \in \Gamma \subset \mathbb{R}^n$

1. Bayes: Apriori probabilities for H_0, H_1, costs for decision are known
2. Minimax: Unknown priors - Minimum cost for the worst case priors
3. Neyman-Pearson: Tradeoff between errors of different type
4. Composite: Unknown parameters in the system

Bayesian Hypothesis Testing

Components of the system:

- Apriori probabilities: $\Pi_0 = P(H = H_0), \quad \Pi_1 = P(H = H_1) = 1 - \Pi_0$
- Measurement Model: $p(\vec{y}|H_0), \quad p(\vec{y}|H_1)$
- Decision Rule: $\delta(\vec{y}) : \Gamma \to \Lambda = \{H_0, H_1\}$

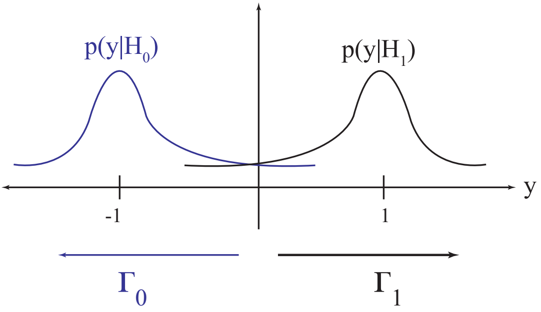


where $\Gamma_0 \cap \Gamma_1 = \emptyset$, $\Gamma_0 \cup \Gamma_1 = \Gamma$

Example:

Measurement model: $H_0 : Y \sim \mathcal{N}(-1, 1)$, $H_1 : Y \sim \mathcal{N}(1, 1)$

priors: $\Pi_0 = \Pi_1 = \frac{1}{2}$, $\Gamma = \mathbb{R}$, $\Lambda = \{H_0, H_1\}$



(may not be the best decision rule)

- Types of errors:

  $\delta(y) = 0$ given H_1 (misdetection)

  $\delta(y) = 1$ given H_0 (false alarm)

- Cost of decisions:

  C_{ij} is cost of choosing H_i given H_j. In general, $C_{ii} < C_{ij}$, $i \neq j$. In radar $C_{01} \gg C_{10}$.

- Conditional Risk:

  The conditional expected cost of decision $\delta(\cdot)$ given H_j:

$$R_j(f) = \sum_i C_{ij} P_j(\Gamma_i)$$

where

$$P_j(\Gamma_i) = P(\vec{Y} \in \Gamma_i | H_j) = \int_{\Gamma_i} p(\vec{y}|H_j) d\vec{y} = \int_{\Gamma_i} \frac{P(H_j|\vec{Y} = \vec{y}) p(\vec{y})}{\Pi_j} d\vec{y}$$

- Bayes' risk:

$$r(\delta) = \sum_j \Pi_j R_j(\delta) = \sum_{i,j} C_{i,j} \int_{\Gamma_i} P(H_j|\vec{Y} = \vec{y})\, p(\vec{y})\, d\vec{y}$$

- Optimum decision rule: Minimize the Bayes' risk

$$\delta_{opt}(\cdot) = \arg\min_{\delta(\cdot)} \; r(\delta)$$

where

$$r(\delta) = \int_{\Gamma_0} \underbrace{\sum_j C_{0j} P(H_j|\vec{Y} = \vec{y})}_{*} \, p(\vec{y})\, d\vec{y} + \int_{\Gamma_1} \underbrace{\sum_j C_{1j} P(H_j|\vec{Y} = \vec{y})}_{**} \; p(\vec{y})\, d\vec{y}$$

objectives:

- For any given $\vec{y} \in \Gamma$, decide whether $\vec{y} \in \Gamma_0$ or $y \in \Gamma_1$ to minimize $r(\delta)$
- For any given $\vec{y}$, if $* < **$, choose $\vec{y} \in \Gamma_0$, if $** < *$, choose $\vec{y} \in \Gamma_1$
- $$C_{00}\ P(H = H_0|\vec{Y} = \vec{y}) + C_{01}\ P(H = H_1|\vec{Y} = \vec{y}) \underset{\delta_{opt}(\vec{y})=0}{\overset{\delta_{opt}(\vec{y})=1}{\gtrless}} C_{10}\ P(H = H_0|\vec{Y} = \vec{y}) + C_{11}\ P(H = H_1|\vec{Y} = \vec{y})$$
- $$\frac{P(H = H_1|\vec{Y} = \vec{y})}{P(H = H_0|\vec{Y} = \vec{y})} \underset{0}{\overset{1}{\gtrless}} \frac{C_{10} - C_{00}}{C_{01} - C_{11}}$$

- Since $P(H = H_i | \vec{Y} = \vec{y}) = \frac{p(\vec{y}|H_i)\Pi_i}{p(\vec{y})}$

  $\Rightarrow$ Likelihood Ratio Test (LRT)

$$\frac{p(\vec{y}|H_1)}{p(\vec{y}|H_0)} \underset{0}{\overset{1}{\gtrless}} \frac{\Pi_0(C_{10} - C_{00})}{\Pi_1(C_{01} - C_{11})}$$

  where $\frac{p(\vec{y}|H_1)}{p(\vec{y}|H_0)}$ is the likelihood ratio, $L(\vec{y})$

Notes:

- $L(\vec{y})$ is a non-negative random variable
- $L(\vec{y}) : \mathbb{R}^k \to \mathbb{R}$ (reduces the dimension of the problem)

Maximum A Posteriori and Maximum Likelihood Detection

Objective: Minimize the decision error

- Consider $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1 \rightarrow$ uniform costs
- $r(\delta) = P(\text{decide} H_1 | H = H_0)\, \Pi_0 + P(\text{decide} H_0 | H = H_1)\, \Pi_1 =$ probability of decision error $= P(E)$

Maximum A Posteriori (MAP) decision rule: (minimizes the prob. of error)

$$P(H = H_1 | \vec{Y} = \vec{y}) \underset{0}{\overset{1}{\gtrless}} P(H = H_0 | \vec{Y} = \vec{y})$$

where $P(H = H_1 | \vec{Y} = \vec{y})$ and $P(H = H_0 | \vec{Y} = \vec{y}) \rightarrow$ a posteriori probabilities

Maximum Likelihood (ML) decision rule:

- If $\Pi_0 = \Pi_1 = \frac{1}{2}$,

$$p(\vec{y}|H_1) \underset{0}{\overset{1}{\gtrless}} p(\vec{y}|H_0)$$

Note: If Π_0 and Π_1 are unknown, it may be reasonable to use $\Pi_0 = \Pi_1 = 1/2$ or other ideas (minimize decision rule).

Example: Digital communication system. Decide '0' or '1' based on Y. Under H_0, $Y \sim \mathbf{N}(-1, 1)$ and H_1, $Y \sim \mathbf{N}(1, 1)$.

Let $\Pi_0 = \Pi_1 = 1/2$. Find the detector, which minimizes the probability of error and find the associated $P(E)$.

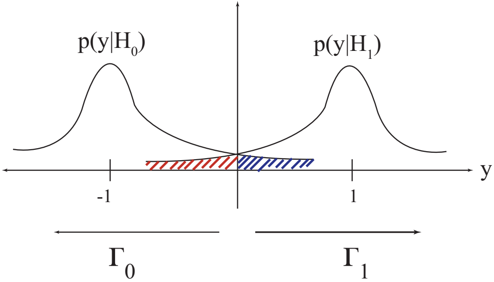
$$L(y) = \frac{p(y|H_1)}{p(y|H_0)} = \frac{\frac{1}{\sqrt{2\pi}}\exp(-\frac{(y-1)^2}{2})}{\frac{1}{\sqrt{2\pi}}\exp(-\frac{(y+1)^2}{2})} = \exp(2y)$$

ML decision:

$$L(y) \underset{0}{\overset{1}{\gtrless}} 1 \quad \Rightarrow \quad \underbrace{\log L(y)}_{\text{log likelihood ratio (LLR)}} \underset{0}{\overset{1}{\gtrless}} 0$$

$\Rightarrow y \underset{0}{\overset{1}{\gtrless}} 0$ where y: sufficient statistics

$$P(E) = r(\delta) = P(\delta(y) = H_1|H_0)\,\Pi_0 + P(\delta(y) = H_0|H_1)\,\Pi_1$$



$$\begin{aligned} P(E) &= P(Y > 0|H_0)\,\Pi_0 + P(Y < 0|H_1)\,\Pi_1 \\ &= \frac{1}{2}\left(1 - \Phi\left(\frac{0-(-1)}{\sqrt{1}}\right)\right) + \frac{1}{2}\Phi\left(\frac{0-1}{\sqrt{1}}\right) \\ &\stackrel{a}{=} \Phi(-1) \end{aligned}$$

where (a) follows from $\Phi(-x) = 1 - \Phi(x)$

Minimax Decision Rule

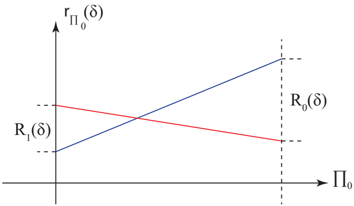
Unknown Π_0, Π_1, known C_{ij}.

- Let $r_{\Pi_0}(\delta)$ be the expected risk for some (possibly suboptimal) $\delta(\cdot)$ and given the time prior Π_0;

  Let $\delta_{\Pi_0}(\cdot) = \arg\min_{\delta(\cdot)} r_{\Pi_0}(\delta)$ (optimal decision rule for prior Π_0)

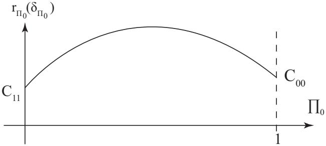
Objective: $\delta_M = \arg\min_{\delta(\cdot)} \left[\max_{\Pi_0} r_{\Pi_0}(\delta)\right]$

- $$\begin{aligned} r_{\Pi_0}(\delta) &= \Pi_0 \ R_0(\delta) + \Pi_1 \ R_1(\delta) \\ &= (R_0(\delta) - R_1(\delta))\, \Pi_0 + R_1(\delta) \geq r_{\Pi_0}(\delta_{\Pi_0}) \end{aligned}$$

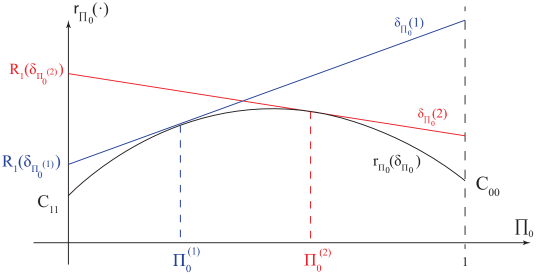


- $$\Pi_0 = 0 \Rightarrow \Gamma_0 = \emptyset \Rightarrow \delta_{\Pi_0}(\vec{y}) = 1$$
$$r_{\Pi_0}(\delta_{\Pi_0}) = C_{11}$$

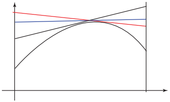
Similarly, if $\Pi_1 = 0 \Rightarrow r_{\Pi_0}(\delta_{\Pi_0}) = C_{00}$



- Combining, $r_{\Pi_0}(\delta_{\Pi_0})$ is concave, because
  1. $r_{\Pi_0}(\delta) \geq r_{\Pi_0}(\delta_{\Pi_0})$ with equality if and only if δ is the decision rule optimized for Π_0. In that case, line $r_{\Pi_0}(\delta)$ is tangent to $r_{\Pi_0}(\delta_{\Pi_0})$.
  2. Item (1) is true for all Π_0 $\Rightarrow$ Every tangent line is always above $r_{\Pi_0}(\delta_{\Pi_0})$ is concave.

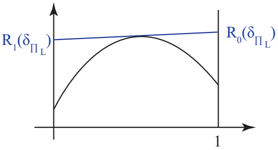


- $\delta_M = \arg\min_{\delta(\cdot)} \max_{\Pi_0} [r_{\Pi_0}(\delta)]$



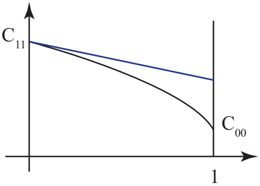
Three cases:

1. .



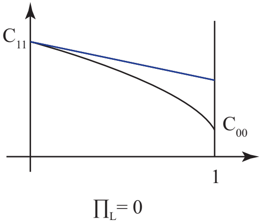
$\prod_L$: minimax prior

2. .



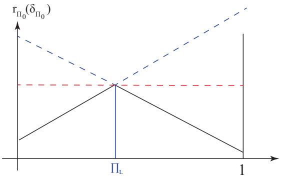
$\prod_L = 0$

3. .



- δ_M minimizes the "sensitivity" of $r_{\Pi_0}(\delta_M)$ with respect to variations in Π_0.

Note: Non-strictly concave $r_{\Pi_0}(\delta_{\Pi_0})$:



- $L(\vec{y})$ is a discrete r.v.
- One rule for $\Pi_0 < \Pi_L$ and another for $\Pi_L < \Pi_0$.
- Minimax: time share between two rules for the equalizer value.

Example: (In Poor referred to as Location Testing)

$H_0 : Y \sim \mathcal{N}(\mu_0, \sigma^2)$, $H_1 : Y \sim \mathcal{N}(\mu_1, \sigma^2)$, $\mu_1 > \mu_0$.

Find δ_M for unit costs ($C_{00} = C_{11} = 0$, $C_{10} = C_{01} = 1$).

The LRT:

$$L(y) = \exp\left(\frac{\mu_1 - \mu_0}{\sigma^2}\left(y - \frac{\mu_0 + \mu_1}{2}\right)\right) \underset{0}{\overset{1}{\gtrless}} \tau_L = \frac{\Pi_L}{1 - \Pi_L}$$

The log of both sides:

$$y \underset{0}{\overset{1}{\gtrless}} \underbrace{\frac{\mu_0 + \mu_1}{2} + \frac{\sigma^2}{\mu_0 - \mu_1} \log \tau_L}_{y_{\tau_L}}$$

- To find the equalizer rule, solve the equation: $R_1(\delta_L) = R_0(\delta_L) (*)$

$$R_j(\delta_L) = \sum_i C_{ij} P_j(\Gamma_i^{(L)}) \Rightarrow R_1(\delta_L) = P_1(\Gamma_0^{(L)}) = P(Y < y_{\tau_L} | H_1)$$

$$= \Phi\left(\frac{y_{\tau_L} - \mu_1}{\sigma}\right)$$

Similarly,

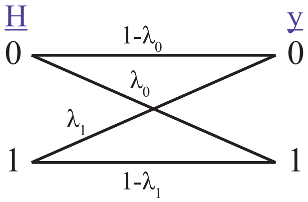
$$R_0(\delta_L) = P_0(\Gamma_1^{(L)}) = P(Y < y_{\tau_L} | H_0) = 1 - \Phi\left(\frac{y_{\tau_L} - \mu_0}{\sigma}\right) = \Phi\left(-\frac{y_{\tau_L} - \mu_0}{\sigma}\right)$$

- Solving $(*)$, we obtain

$$\frac{y_{\tau_L} - \mu_1}{\sigma} = \frac{y_{\tau_L} - \mu_0}{\sigma} \quad \text{(since } \Phi \text{ is one-to-one)}$$

$$\Rightarrow y_{\tau_L} = \frac{\mu_1 + \mu_0}{2} \Rightarrow \tau_L = 1 \Rightarrow \Pi_L = \frac{1}{2}$$

Example: Binary Channel: Simplified view of a communication channel



Assume $\Pi_0 = \Pi_1 = \frac{1}{2}$ and unit costs $\Rightarrow$ ML decision rule

$$L(y) = \frac{P(Y = y|H_1)}{P(Y = y|H_0} = \begin{cases} \frac{\lambda_1}{1-\lambda_0}, & y = 0 \\ \frac{1-\lambda_1}{\lambda_0}, & y = 1 \end{cases}$$

decision rule: $L(y) \underset{0}{\overset{1}{\gtrless}} 1$

1. $y = 0,\ \lambda_1 + \lambda_0 > 1 \Rightarrow \delta(y) = 1$
2. $y = 0,\ \lambda_1 + \lambda_0 < 1 \Rightarrow \delta(y) = 0$
3. $y = 1,\ \lambda_1 + \lambda_0 < 1 \Rightarrow \delta(y) = 1$
4. $y = 1,\ \lambda_1 + \lambda_0 > 1 \Rightarrow \delta(y) = 0$

$$\delta(y) = \begin{cases} y, & \lambda_0 + \lambda_1 \le 1 \\ 1 - y, & \lambda_0 + \lambda_1 > 1 \end{cases}$$

- What if $\lambda_0 + \lambda_1 = 1$

  1. $$\text{if} \quad \delta(y) = y \Rightarrow r(\delta) = \frac{1}{2} P_0(\Gamma_1) + \frac{1}{2} P_1(\Gamma_0)$$
     $$= \frac{1}{2} P(Y = 1 | H_0) + \frac{1}{2} P(Y = 0 | H_1)$$
     $$= \frac{1}{2} (\lambda_0 + \lambda_1) = \frac{1}{2}$$

ii

$$\text{if} \quad \delta(y) = 1 - y \Rightarrow r(\delta) = \frac{1}{2}P(Y = 0|H_0) + \frac{1}{2}P(Y = 1|H_1)$$
$$= \frac{1}{2}(1 - \lambda_0 + 1 - \lambda_1)$$
$$= 1 - \frac{1}{2}(\lambda_0 + \lambda_1) = \frac{1}{2}$$

Both decisions are equally good if $\lambda_0 + \lambda_1 = 1$

- Binary symmetric channel: $\lambda_0 = \lambda_1 = \lambda$

$$\Rightarrow \delta(y) = \begin{cases} y, & \lambda \leq \frac{1}{2} \\ 1 - y, & \lambda > \frac{1}{2} \end{cases}$$

Signal detection under additive noise

Measurement model: $H_i : \vec{s_i} + \vec{N} = \vec{Y}$

Q: How do we find $p(\vec{y}|H_i)$, given $p_{\vec{N}}(\vec{n})$?

$$p(\vec{y}|H_i) = P_{\vec{N}}(\vec{y} - \vec{s_i})$$

$$L(\vec{y}) = \frac{P_{\vec{N}}(\vec{y} - \vec{s_1})}{P_{\vec{N}}(\vec{y} - \vec{s_0})} = \frac{P_{N_1,N_2,\ldots,N_n}(y_1 - s_{11}, y_2 - s_{12}, \ldots, y_n - s_{1n})}{P_{N_1,N_2,\ldots,N_n}(y_1 - s_{01}, y_2 - s_{02}, \ldots, y_n - s_{0n})}$$

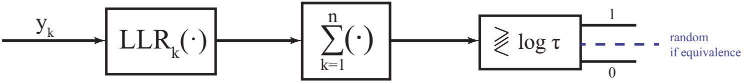
let $\vec{N}$ be iid:

$$L(\vec{y}) = \frac{\prod_{k=1}^n P_{N_k}(y_k - s_{1k})}{\prod_{k=1}^n P_{N_k}(y_k - s_{0k})} = \prod_{k=1}^n L_k(y_k)$$

$$LLR(\vec{y}) \triangleq \log L(\vec{y}) = \underbrace{\sum_{k=1}^n LLR_k(y_k) \underset{0}{\overset{1}{\gtrless}} \log \tau}_{LRT}$$

signal is known

$\Rightarrow$ Coherent detector



Example:

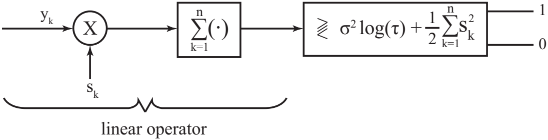
$H_1 : Y_k = s_k + N_k,\ H_0 : Y_k = N_k,\ N_k \sim \mathcal{N}(0, \sigma^2)$, iid

$$L_k(y_k) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_k - s_k)^2}{\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{y_k^2}{\sigma^2}\right)} = \exp\left(-\frac{1}{2\sigma^2}(s_k^2 - 2s_k y_k)\right)$$

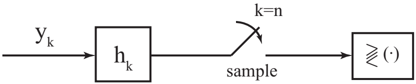
$$\sum_{k=1}^{n} LLR_k(y_k) = \frac{1}{\sigma^2}\sum_{k=1}^{n}\left(s_k y_k - \frac{1}{2}s_k^2\right) \underset{0}{\overset{1}{\gtrless}} \log \tau$$

$$\Rightarrow \quad \underbrace{\sum_{k=1}^{n} s_k y_k}_{\substack{\text{projection of } \vec{y} \\ \text{on to } \vec{s}}} \quad \underset{0}{\overset{1}{\gtrless}} \sigma^2 \log \tau + \frac{1}{2} \sum_{k=1}^{n} s_k^2$$

detector: linear correlator:



- The linear operator can be represented with a discrete-time filter, called the matched filter:



$$h_k = \begin{cases} s_{n-k}, & 0 \leq k \leq n-1 \\ 0, & \text{otherwise} \end{cases} \Rightarrow h_n \circledast y_n = \sum_{k=1}^{n} h_{n-k} y_k = \sum_{k=1}^{n} s_k y_k$$

<u>read:</u> detection in iid Laplacian noise in Poor.