6.434/16.391 Statistics for Engineers and Scientists

Lecture 15 11/14/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 15a 11/14/2012

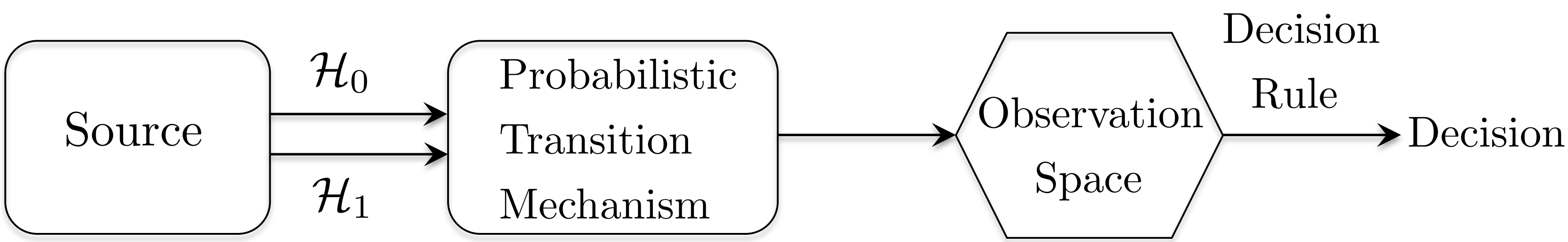
HYPOTHESIS TESTING

WHAT IS HYPOTHESIS TESTING

- Consider a source with two possible states often denoted as "hypotheses" $\mathcal{H}_0, \mathcal{H}_1$
- The source produces, through a probabilistic transition mechanism some quantities denoted as "observables" for instance arranged in a vector $\boldsymbol{r}$.
- Due to the probabilistic transition mechanism there is not a one-to-one relation between $\boldsymbol{r}$ and $\mathcal{H}_i$
- The problem of hypothesis testing: given an observed r, decide about the state of the source, among $\mathcal{H}_0, \mathcal{H}_1$, that most probably (in a way that will be discussed) produced $\boldsymbol{r}$.
- This is known as a **binary** hypothesis test. We will see later a generalization.

Simple Binary Hypothesis Test

- "Simple" means not composite. We will discuss this aspect later.



- Example 1: Simple binary test, where a binary (antipodal) sample is disturbed by Gaussian noise:

$$\begin{cases} \mathcal{H}_0 : & r & = & -1 + n \\ \mathcal{H}_1 : & r & = & +1 + n \end{cases}$$

- After observing a realization r, we want to make a decision for $\mathcal{H}_0$ or $\mathcal{H}_1$.

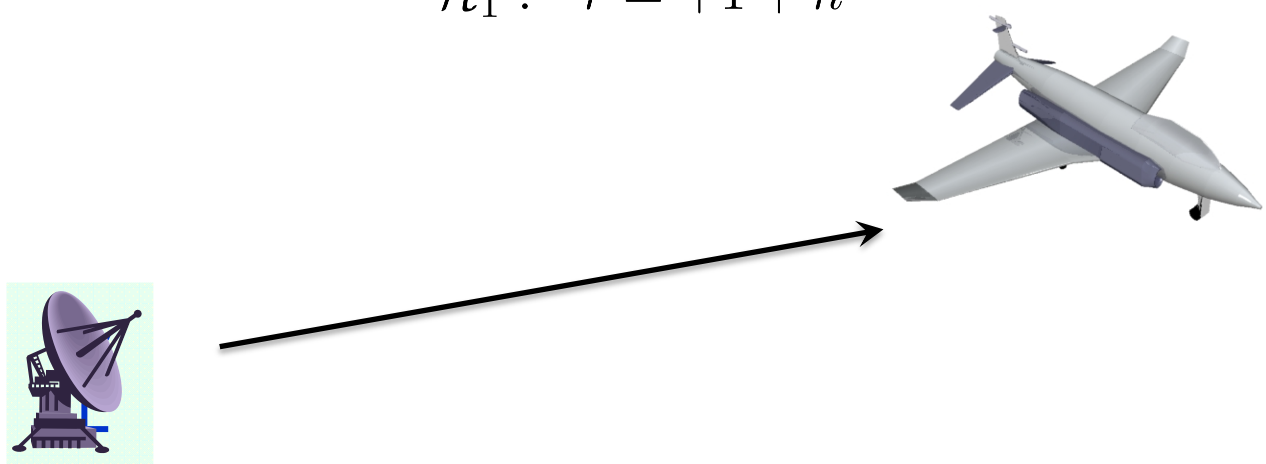
Simple Binary Hypothesis Test

- Example 2: We try to detect an aircraft using a radar based on the received signal. Under hypothesis $\mathcal{H}_0$, i.e., the aircraft does not exist, the received signal is

$$\mathcal{H}_0 : \quad r = n$$

where n is the noise. Under hypothesis $\mathcal{H}_1$, i.e., the aircraft exists, the received signal is

$$\mathcal{H}_1 : \quad r = +1 + n$$



Some Definitions

- Observation space $Z = \{\boldsymbol{r}\}$: ensemble of all observations

$$\boldsymbol{r} = [\, r_1, \, r_2, \, \cdots \, , \, r_N \,]^{\mathrm{T}}$$

- Conditional probability distribution function (p.d.f.)

$$p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0) \qquad\qquad p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)$$

- Decisions:

  $\mathcal{D}_0$ (decide for $\mathcal{H}_0$),

  $\mathcal{D}_1$ (decide for $\mathcal{H}_1$)

- Hypotheses and decisions:

  $\mathcal{H}_0$, $\mathcal{D}_0$

  $\mathcal{H}_0$, $\mathcal{D}_1$ (False Alarm or type I error)

  $\mathcal{H}_1$, $\mathcal{D}_0$ (Missed Detection or type II error)

  $\mathcal{H}_1$, $\mathcal{D}_1$

Bayes Criterion: Risk

- Given the "cost" for each pair (hypothesis, decision), make decisions to minimize the average cost

- Definition 1.1 $C_{ij} =$ cost in choosing $\mathcal{H}_i$ (i.e., producing the decision $\mathcal{D}_i$) given $\mathcal{H}_j$ is true, with $i, j = 0, 1$. This will also be indicated as $C = C(\mathcal{D}, \mathcal{H})$

- Definition 1.2 a-priori probabilities

$$\mathbb{P}\{\mathcal{H}_0\} = P_0, \ \mathbb{P}\{\mathcal{H}_1\} = P_1$$

Bayes Criterion: Risk

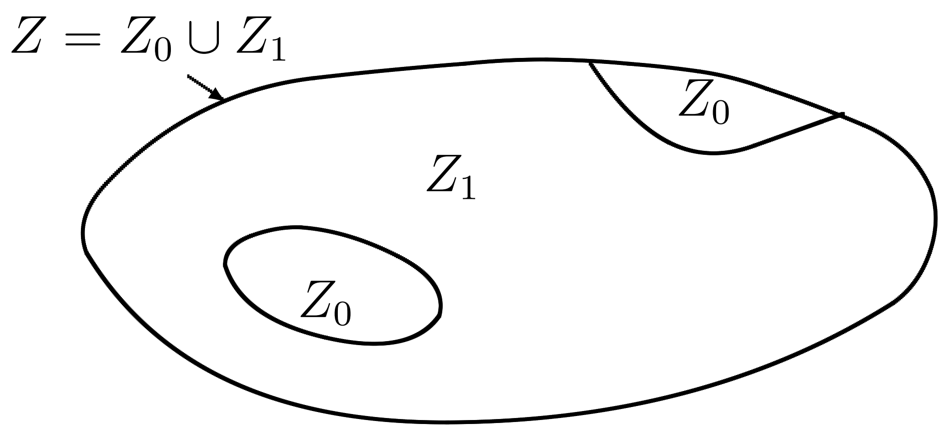
- Definition 1.3 $\text{Average Cost} \triangleq \text{Risk} = \mathcal{R} = \mathbb{E}\{C\}$

$$
\begin{aligned}
\mathcal{R} &= \sum_{\mathcal{D},\mathcal{H}} C(\mathcal{D},\mathcal{H})\mathbb{P}\{\mathcal{D},\mathcal{H}\} \\
&= \sum_{\mathcal{D},\mathcal{H}} C(\mathcal{D},\mathcal{H})\mathbb{P}\{\mathcal{D}|\mathcal{H}\}\mathbb{P}\{\mathcal{H}\} \\
&= C_{00}P_{0|0}P_0 + C_{01}P_{0|1}P_1 + C_{10}P_{1|0}P_0 + C_{11}P_{1|1}P_1
\end{aligned}
$$

where $P_{i|j} = \mathbb{P}\{\mathcal{D} = \mathcal{D}_i | \mathcal{H} = \mathcal{H}_j\}$

Bayes Criterion: Decision Region

- Since we want a decision in any case, the observation space Z must be partitioned in two regions.



- The two regions are:

$$Z_0 = \{\boldsymbol{r} \in Z : \text{decide} \mathcal{H}_0\}$$
$$Z_1 = \{\boldsymbol{r} \in Z : \text{decide for } \mathcal{H}_1\}$$

- Therefore we can express the probabilities as

$$\mathbb{P}\{\mathcal{D} = \mathcal{D}_i | \mathcal{H} = \mathcal{H}_j\} = P_{i|j} = \mathbb{P}\{\boldsymbol{R} \in Z_i | \mathcal{H} = \mathcal{H}_j\} = \int_{Z_i} p_{\boldsymbol{R}|\mathcal{H}_j}\{\boldsymbol{r}|\mathcal{H}_j\} d\boldsymbol{r}$$

Bayes Criterion

- Since $Z = Z_0 \cup Z_1$ we can write

$$\int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_j}(\boldsymbol{r}|\mathcal{H}_j)d\boldsymbol{r} = 1 - \int_{Z_1} p_{\boldsymbol{R}|\mathcal{H}_j}(\boldsymbol{r}|\mathcal{H}_j)d\boldsymbol{r}$$

$$\Rightarrow P_{0|j} = 1 - P_{1|j}$$

Bayes Criterion

- This allows us to write the risk in terms of $P_{0|0}$ and $P_{0|1}$ as follows

$$\mathcal{R} = C_{00}P_{0|0}P_0 + C_{10}\underbrace{P_{1|0}}_{1-P_{0|0}}P_0 + C_{01}P_{0|1}P_1 + C_{11}\underbrace{P_{1|1}}_{1-P_{0|1}}P_1$$

$$= \overbrace{C_{10}P_0 + C_{11}P_1}^{A} + P_0(C_{00} - C_{10})\int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)d\boldsymbol{r}$$

$$+ P_1(C_{01} - C_{11})\int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)d\boldsymbol{r}$$

$$= A + \int_{Z_0} \left[P_1(C_{01} - C_{11})p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) - P_0(C_{10} - C_{00})p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)\right] d\boldsymbol{r}$$

Bayes Criterion: the likelihood ratio

- Our goal is to find the best partition for decision regions that minimize the risk.

- Note: In minimizing risk, we can ignore A since it is independent on the Z partition.

Bayes Criterion

- Thus, the best test (i.e., the best partition) is obtained by assigning those $\boldsymbol{r}$ to Z_0 , for which a negative integrand in the previous equation.
  Hence, the rule is: put in Z_0 all points for which

$$P_1(C_{01} - C_{11})p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) - P_0(C_{10} - C_{00})p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0) < 0$$

  i.e., $P_1(C_{01} - C_{11})p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) < P_0(C_{10} - C_{00})p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)$

  or, equivalently,

$$P_1(C_{01} - C_{11})p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} P_0(C_{10} - C_{00})p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)$$

Bayes Criterion: the likelihood ratio

- It is reasonable to assume that $C_{ij} > C_{jj}$ for $i \neq j$ (i.e., wrong decisions cost more!), which implies that $C_{ij} - C_{jj} > 0$
- We can therefore write the test based on the Bayes criterion as follows

$$\frac{p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)}{p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

- Define the likelihood ratio

$$\Lambda(\boldsymbol{r}) \triangleq \frac{p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1)}{p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)}$$

and the threshold

$$\eta \triangleq \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

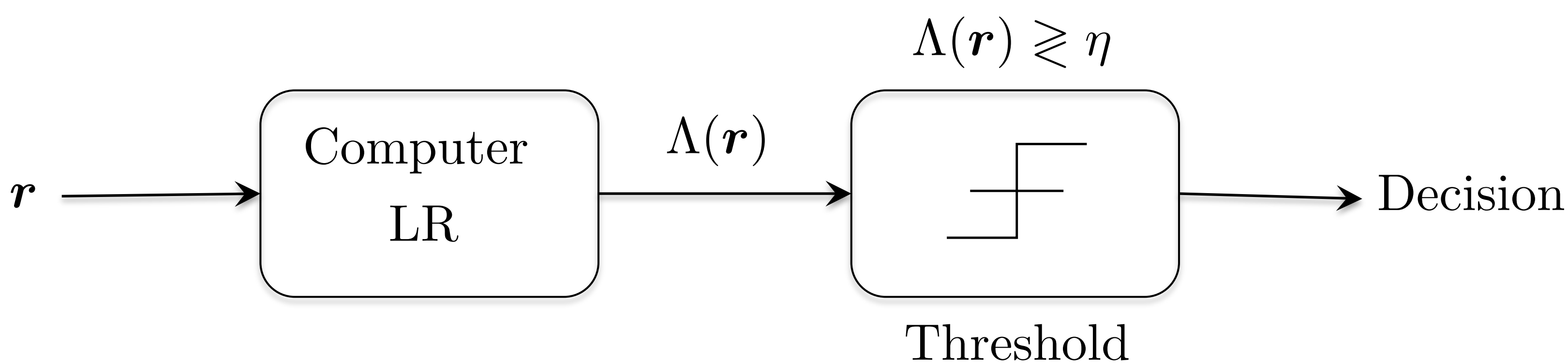
Bayes Criterion: the likelihood ratio

- The decision based on the Bayes criterion can be written as a likelihood ratio test (LRT):

$$\Lambda(\boldsymbol{r}) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \eta$$

or the equivalent log-likelihood ratio test (LLRT)

$$\log \Lambda(\boldsymbol{r}) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \log \eta$$



Binary HT performance

- Notation, mainly from Radar terminology:

$$\begin{aligned} \mathcal{H}_0 &: \quad \text{target is absent} \\ \mathcal{H}_1 &: \quad \text{target is present} \end{aligned}$$

- Definition: Probability of Missed Detection

$$P_{\mathrm{M}} = \mathbb{P}\{\mathcal{D}_0 | \mathcal{H}_1\} = P_{0|1}$$

- Definition: Probability of Detection

$$P_{\mathrm{D}} = \mathbb{P}\{\mathcal{D}_1 | \mathcal{H}_1\} = P_{1|1} = 1 - P_{\mathrm{M}}$$

also called the **power of the test** in the statistical literature

Binary HT performance

- Definition: Probability of False Alarm

$$P_{\mathrm{F}} = \mathbb{P}\{\mathcal{D}_1 | \mathcal{H}_0\} = P_{1|0}$$

- Relations:

$$P_{\mathrm{M}} = \int_{Z_0} p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) d\boldsymbol{r}$$

$$P_{\mathrm{F}} = \int_{Z_1} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0) d\boldsymbol{r}$$

MAP Criterion

- Equal Cost: If the cost of correct decisions is K_c and the cost of wrong decisions K_w with $K_w > K_c$, i.e.,

$$C_{i,j} = \begin{cases} K_w & i \neq j \\ K_c & i = j \end{cases}$$

- Then the test becomes

$$P_1 \, p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} P_0 \, p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)$$

- Dividing by $p_{\boldsymbol{R}}(\boldsymbol{r})$ the test can be written in terms of the **a-posteriori** probabilities $\mathbb{P}\{\mathcal{H}_i|\boldsymbol{r}\}$ as

$$\mathbb{P}\{\mathcal{H}_1|\boldsymbol{r}\} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \mathbb{P}\{\mathcal{H}_0|\boldsymbol{r}\}$$

MAP Criterion

- Hence, the decision is for the $\mathcal{H}_i$ which maximizes the a-posteriori probability $\mathbb{P}\{\mathcal{H}_i|\boldsymbol{r}\}$. This is called maximum a-posteriori probability (MAP) criterion.

ML Criterion

- If hypotheses are equally likely, i.e., $P_0 = P_1$, in addition to equal cost, the test can also be written as the comparison of the likelihoods

$$p_{\boldsymbol{R}|\mathcal{H}_1}(\boldsymbol{r}|\mathcal{H}_1) \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} p_{\boldsymbol{R}|\mathcal{H}_0}(\boldsymbol{r}|\mathcal{H}_0)$$

- In other terms the decision is for the $\mathcal{H}_i$ which maximizes the likelihood $p(\boldsymbol{r}|\mathcal{H}_i)$. This is the **ML criterion for hypothesis testing**.

ML Criterion

- Since the two errors cost the same, we can also talk about an **error** event as $\mathcal{D}_0|\mathcal{H}_1$ or $\mathcal{D}_1|\mathcal{H}_0$ and we can define the error probability, P_{e}, where

$$\begin{aligned} P_{\mathrm{e}} &= \mathbb{P}\{\mathcal{D}_1, \mathcal{H}_0\} + \mathbb{P}\{\mathcal{D}_0, \mathcal{H}_1\} \\ &= \mathbb{P}\{\mathcal{D}_1|\mathcal{H}_0\}\mathbb{P}\{\mathcal{H}_0\} + \mathbb{P}\{\mathcal{D}_0|\mathcal{H}_1\}\mathbb{P}\{\mathcal{H}_1\} \\ &= P_{\mathrm{F}} P_0 + P_{\mathrm{M}} P_1 \end{aligned}$$

- Note that when $K_c = 0$,

$$\begin{aligned} \mathcal{R} &= C_{10}\mathbb{P}\{\mathcal{D}_1, \mathcal{H}_0\} + C_{01}\mathbb{P}\{\mathcal{D}_0, \mathcal{H}_1\} \\ &= K_w P_{\mathrm{e}} \end{aligned}$$

- Thus, Bayes Criterion is equivalent to minimizing the error probability for equally likely hypothesis with equal cost and $K_c = 0$