EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 15 - Outline

1. **Decision theory**
   a. **Binary hypothesis testing**
   b. **Maximum Likelihood decisions for discrete RVs**
   c. **Maximum Likelihood decisions for continuous RVs**
   d. **Bayesian hypothesis testing**

Motivating Example. Radar/Sonar Detection

- A radar/sonar system transmits electromagnetic/ ultrasonic pulses in the direction of a target, which may (or may not) be "out there".
- The receiver listens for echoes
  - Echoes may result from the target or from other extraneous effects (noise, large birds, atmospheric inhomogeneities, large metallic flying objects, …)
- Based on observed echoes, the radar system decides whether a target is present or not

Simple Concepts

- System decides if a target is present or not
- The decision may be correct or incorrect
- If the decision is that a target is present
  - This decision may be ***wrong*** if there is no target present
    - **FALSE ALARM**
- If decision is that no target is present
  - This decision may be ***wrong*** if there is a target present
    - **MISSED DETECTION**

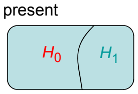
8.1 Binary Test of Hypothesis

Decide Between Two Hypotheses H_0 and H_1,

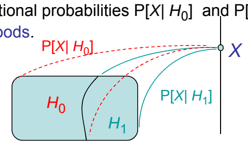
e.g., Radar: H_0 : a target is absent, and H_1: a target is present

One and only one of these hypotheses is true,

i.e., H_0 and H_1 are mutually exclusive exhaustive…

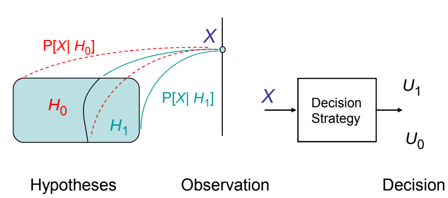


Decision is based on an observation, a random variable X (discrete or continuous):
X is statistically dependent on which hypothesis is true.
The conditional probabilities P[X| H_0] and P[X| H_1] are known & are called the likelihoods.



Decision

Decision maker selects one of the hypotheses.
The event that it selects H_0 is called U_0
The event that it selects H_1 is called U_1

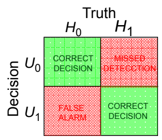


– Decision rule maps X into decision $U \in \{U_0, U_1\}$
– U is derived random variable!

Decision Errors

1) System selects H_1 when H_0 is true
 (target is present when it is actually absent): **FALSE ALARM**

2) System selects H_0 when H_1 is true
 (target is absent when it is actually present): **MISSED DETECTION**



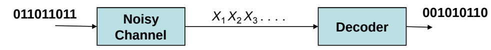
Average Probability of Error: Use Total Probability Theorem

$P[\text{Error}] = P[\text{Error}|H_0]\, P[H_0] + P[\text{Error}|H_1]\, P[H_1]$

Equal **a priori** probabilities on H_0, H_1 → $P[\text{Error}]$ $\frac{1}{2}$ $(P[U_1|H_0] + P[U_0|H_1])$

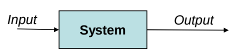
Goal: Develop a decision strategy that reduces these errors

Example 1. Binary Channels



- Modem transmits bits over a noisy channel
  - Bits are 0 (hypothesis H_0) or 1 (hypothesis H_1)
- Output of channel: random variable X generated for each bit, independently, conditioned on value of input bit
- Probability law of X depends on input to channel (0 or 1)
- Decision problem: decode each bit given observation X

Example 2. Significance Testing



- Consider a system (e.g., a production process) with a random output (e.g., defective items)
- Want to characterize the output using a probability model
- Suppose we postulate (or estimate from past measurements) the output to follow some probability model (e.g., defectives follow Bernoulli w.p. p_0=0.01)
- Introduce a change in the system. Decide:

  H_0 (null hypothesis): $p=p_0$ **or** H1: $p \neq p_0$
- Significance Level $\triangleq$ False alarm probability $P[U_1|H_0]$

8.2 Maximum-Likelihood Decision for Discrete RVs

Observation RV X takes on values $x_1, x_2, \ldots x_N$

Write the PMFs for each of the two hypotheses in an array, called the likelihood array, or likelihood matrix:

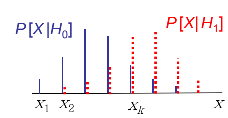
	$X = x_1$	$X = x_2$	..	$X = x_k$	$X = x_N$
$P[X\|H_0]$					
$P[X\|H_1]$					

It is an array of 2 rows and N columns, one row per hypothesis and one column per measurement value

Maximum-Likelihood Decision Rule:

If the observation is $X = x_k$, then

Chose H_1 if $P[X = x_k |H_1] > P[X = x_k |H_0]$
Chose H_0 if $P[X = x_k |H_1] < P[X = x_k |H_0]$
(if tie, select either)



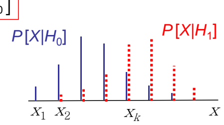
Rationale for the ML Decision Rule

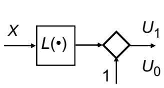
- Select the hypothesis that maximizes the probability of obtaining the observed value of X
- It inherently assumes that the hypotheses are *a priori* equally likely

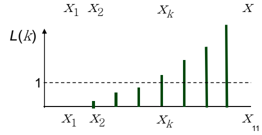
An Equivalent Implementation of ML Decision Rule

Likelihood Ratio $L(k) = \dfrac{P[X = x_k | H_1]}{P[X = x_k | H_0]}$

Chose H_1 if $L(k) > 1$
Chose H_0 if $L(k) < 1$





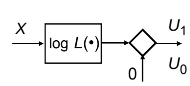


Yet Another Equivalent Implementation:

Log-Likelihood Ratio

$$\log L(k) = \log P[X = x_k | H_1] - \log P[X = x_k | H_0]$$

Chose H_1 if $\log L(k) > 0$
Chose H_0 if $\log L(k) < 0$



Example: Two Coins

Hypotheses: H_0 : coin is biased with probability of heads = 0.2
H_1 : coin is fair (probability of heads = 0.5)

Observation: X = number of heads in 3 flips

Likelihoods: $P[X|H_0]$ ~ Binom(3, 0.2)
$P[X|H_1]$ ~ Binom(3, 0.5)

Likelihood matrix

	X = 0	X = 1	X = 2	X = 3
$P[X\|H_0]$	0.512	0.384	0.096	0.008
$P[X\|H_1]$	0.125	0.375	0.375	0.125

ML Decision Rule: Find and shade the largest entry in each column:
If X = 0 or 1 select H_0
If X = 2 or 3 select H_1

Error Probabilities:
$P[\text{error}|H_0] = P[X{=}2|H_0] + P[X{=}3|H_0] = 0.096 + 0.008 = 0.104$
$P[\text{error}|H_1] = P[X{=}0|H_1] + P[X{=}1|H_1] = 0.125 + 0.375 = 0.5$

Example: Radar

Hypotheses: H_0 : Target is absent
H_1 : Target is present

Observation: The radar system tra[...]its [...]eceives X independent echoes. Each transm[...] [...]ependent Bernoulli trial w[...] a Binomial RV $(n,\ p)$

Likelihoods: $P[X|H_0]$ ~ Binom[...]
$P[X|H_1]$ ~ Bi[...] [...]mple, p_0 [...]05, p_1 = 0.8

Decision makes sense:
Few hits => H0
Many Hits => H1

Likelihood matrix

	X = 0	X = 1	X = 2	X = 3	X = 4	X = 5
$P[X\|H_0]$	0.774	0.204	0.021	0.001	2.97E-05	3.13E-07
$P[X\|H_1]$	0.00032	0.0064	0.0512	0.2048	0.4096	0.32768

ML Decision Rule: Find and shade the largest entry in each column!
If X < 2, select H_0
If X > 1, select H_1

Error Probabilities: $P[\text{error}|H_0] = P[X>1|H_0] = 0.022$
$P[\text{error}|H_1] = P[X<2|H_1] = 0.00672$

General Expression for the ML Rule for a Binomial Test of Hypothesis Problem:

$$L(k) = \frac{P_1(k)}{P_0(k)} = \frac{\binom{n}{k} p_1^k (1-p_1)^{n-k}}{\binom{n}{k} p_0^k (1-p_0)^{n-k}}$$

$$= \frac{p_1^k (1-p_1)^{n-k}}{p_0^k (1-p_0)^{n-k}}$$

$$\log(L(k)) = k \log(\frac{p_1}{p_0}) + (n-k)\log(\frac{1-p_1}{1-p_0})$$

$$k \log(\frac{p_1(1-p_0)}{p_0(1-p_1)}) + n \log(\frac{1-p_1}{1-p_0}) > 0$$

$$k > n \frac{\log(\frac{1-p_0}{1-p_1})}{\log(\frac{p_1(1-p_0)}{p_0(1-p_1)})} \quad \textbf{Declare } H_1$$

Threshold!

8.3 Maximum-Likelihood Decision for Continuous RVs

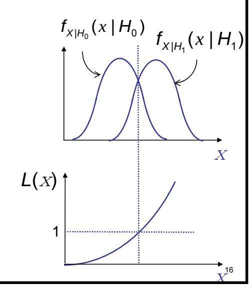
If the observation X is a continuous RV, we write down the likelihood ratio using densities

$$L(x) = \frac{f_{X|H_1}(x|H_1)}{f_{X|H_0}(x|H_0)}$$

Maximum likelihood decision rule:

For each x, select H_1 if $L(x) > 1$,
select H_0 if $L(x) < 1$

If $L(x)$ = 1, select either



Example: Constant Signal in Gaussian Noise

Hypotheses: H_0 : Signal absent, $X = w$, w ~ N(0,1)
H_1 : Signal present, $X = v + w$, v = known constant (signal)

Likelihoods: $f_{X|H_0}$ ~ N(0,1)

$f_{X|H_1}$ ~ N(v,1)

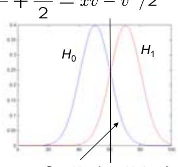
Log-Likelihood Ratio:

$$\log(L(x)) \equiv \log(\frac{f_{X|H_1}(x|H_1)}{f_{X|H_0}(x|H_0)})$$

$$= -\frac{(x-v)^2}{2} + \frac{x^2}{2} = xv - v^2/2$$

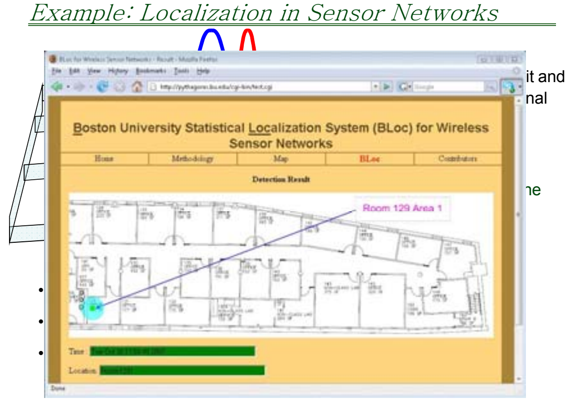
ML Decision Rule: Select H_0 if $X < v/2$
Select H_1 if $X > v/2$

Error Probabilities:
$P[\text{error}|H_0] = P[X > v/2|H_0] = Q(v/2)$
$P[\text{error}|H_1] = P[X < v/2|H_1] = Q(v/2)$



Say H_1 when H_0 true!
False alarm…

Example: Localization in Sensor Networks



[...]it and
[...]nal
[...]ne

8.4 Bayesian Hypothesis Testing

MAP Decision Rule (discrete observation)

For each $X = x$, select the hypothesis that has the maximum *a posteriori* probability (MAP):

$$P\left[H_i \middle| X = x\right] = \frac{P\left[H_i, X = x\right]}{P_X(x)} \quad \text{Bayes rule}$$

If $P[H_1|X=x] > P[H_0|X=x]$, select H_1; else select H_0 (either if a tie)

Since $P_X(x)$ is independent of H_i, this is equivalent to selecting the hypothesis with the largest $P[H_i, X=x]$.

If $P[H_1,X=x] > P[H_0,X=x]$, select H_1; else select H_0 (either if a tie)

Unlike the ML decision rule, the MAP decision rule depends on the prior probabilities $P[H_0]$ and $P[H_1]$.

Average probability of making a correct decision is

$P[\text{error}] = \sum_i P[\text{error} \mid H_i]\, P[H_i]$

Joint Probability Matrix

	$X = x_1$	$X = x_2$	..	$X = x_k$		$X = x_N$
$P[X,H_0]$						
$P[X,H_1]$						

- Sum of entries in column k gives the marginal (unconditional) probability that x_k is observed
- Sum of entries in row i gives prior probabilities $P[H_i]$
- Sum of all entries = 1

The joint probability matrix is obtained from the likelihood matrix by multiplying entries of the i-th row by $P[H_i]$

If all a priori probabilities are equal, I can make the same decision as before by picking the largest entry in every column. If they are not the same, MAP and ML decisions would be different.

Example: Two Coins

Let $P[H_0] = 0.2$, $P[H_1] = 0.8$

Likelihood matrix & ML decision rule

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
H_0	0.512	0.384	0.096	0.008
H_1	0.125	0.375	0.375	0.125

$$P[\text{Error}] = P\left[U = 0 \middle| H_1\right] P[H_1] + P\left[U = 1 \middle| H_0\right] P[H_0]$$

$$\text{ML } P[\text{Error}] = P\left[U_0 \middle| H_1\right] P[H_1] + P\left[U_1 \middle| H_0\right] P[H_0]$$

$$= (0.125 + 0.375) \times 0.8 + (0.096 + 0.008) \times 0.2 = 0.4208$$

Joint Probability matrix & MAP decision rule

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
H_0	0.1024	0.0768	0.0192	0.0016
H_1	0.1	0.3	0.3	0.1

$$\text{MAP } P[\text{Error}] = P[U_0, H_1] + P[U_1, H_0]$$

$$= 0.1 + (0.0768 + 0.0192 + 0.0016) = 0.1976$$

MAP decision rule is an optimal rule that minimizes the average probability of error!

Proof:

For any decision rule $U(x)$,

$$P(Error|X = x) = P(U(x) = 0|H_1, X = x)P(H_1|X = x) + P(U(x) = 1|H_0, X = x)P(H_0|X = x)$$

$$P(Error) = \sum_{x \in S_X} P(Error|X = x) P_X(x)$$

$$= \sum_{x \in S_X} \Big(P(U(x) = 0|H_1, X = x)P(H_1, X = x) + P(U(x) = 1|H_0, X = x)P(H_0, X = x) \Big)$$

For each $X = x$, MAP decision rule selects smallest possible value for that term in the sum.

MAP decision rule can be written in terms of the likelihood ratio $L(x)$

If $L(x) > T$, select H_1

If $L(x) < T$, select H_0 (ties say either)

$$P(H_1, X = x) = P(X = x|H_1)P(H_1)$$

$$P(H_0, X = x) = P(X = x|H_0)P(H_0)$$

MAP: Choose H_1 if $P(H_1, X = x) > P(H_0, X = x)$

$$\Rightarrow P(X = x|H_1)P(H_1) > P(X = x|H_0)P(H_0)$$

$$\Rightarrow \frac{P(X = x|H_1)}{P(X = x|H_0)} > \frac{P(H_0)}{P(H_1)}$$

$$\Rightarrow \text{if } L(x) > T \text{ then choose } H_1$$

MAP and ML are the same type of Decision Rules

Both are threshold rules based on comparing the likelihood ratio $L(x)$ to a threshold T. The only difference is the choice of threshold

$T = 1$ for ML

$T = P[H_0]/P[H_1]$ for MAP