EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

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

BOSTON UNIVERSITY

CISE

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 16 - Outline

1. **Decision theory**
   a. **Bayesian hypothesis testing**
   b. **Generalizations: Multiple hypotheses and vector observations**

Minimum Bayes Risk Decision Rule

- General risk function $R = C(U,H)$
  - $C_{10} \equiv C(U_1|H_0)$ is cost of a false alarm
  - $C_{01} \equiv C(U_0|H_1)$ is cost of missed detection
  - Usually assume $C_{00} = C_{11} = 0$ (not essential)
- Given decision rule, U is derived RV and $R = C(U,H)$ is also derived RV!
- Bayesian Hypothesis Test:
  - Select decision rule $U(x)$ so as to minimize the Expected cost $E[C(U(X),H)]$

Decision \ Truth	H_0	H_1
U_0	C_{00}	C_{01}
U_1	C_{10}	C_{11}

Optimal decision rule:

–Choose H_1 if $C_{01}P[H_1|X=x] > C_{10}P[H_0|X=x]$
–Choose H_0 otherwise!

Minimum Bayes Risk (Proof)

$$E[R] = \sum_{x\in S_X} \sum_{H\in\{H_0,H_1\}} C(U(x),H)P(H|X=x)P_X(x)$$
$$= \sum_{x\in S_X} P_X(x)\Big\{\sum_{H\in\{H_0,H_1\}} C(U(x),H)P(H|X=x)\Big\}$$
$$= \sum_{x\in S_X} E[R|X=x]P_X(x)$$

Optimal decision rule to minimize Bayes' risk: for each x, select U(x) to minimize term in brackets!

If $U(x) = 0 \Rightarrow E[R|X=x] = C_{01}P(H_1|X=x)$
If $U(x) = 1 \Rightarrow E[R|X=x] = C_{10}P(H_0|X=x)$

Optimal decision rule:

–Choose H_1 if $C_{01}P[H_1|X=x] > C_{10}P[H_0|X=x]$
–Choose H_0 otherwise!

Optimal decision rule can also be written in terms of likelihood ratio!

$$P(H_1|X=x) = \frac{P(X=x|H_1)P(H_1)}{P_X(x)}$$
$$P(H_0|X=x) = \frac{P(X=x|H_0)P(H_0)}{P_X(x)}$$
$$C_{01}P(H_1|X=x) > C_{10}P(H_0|X=x)$$
$$\Rightarrow C_{01}P(X=x|H_1)P(H_1)) > C_{10}P(X=x|H_0)P(H_0)$$
$$\Rightarrow \frac{P(X=x|H_1)}{P(X=x|H_0)} > \frac{C_{10}P(H_0)}{C_{01}P(H_1)}$$
$$\Rightarrow L(x) > T_1, \text{ choose } H_1$$

MAP, ML, and Bayesian decision rules are similar:

Compare likelihood ratio to a threshold

If greater than threshold, select H_1; else select H_0

We are only arguing about the threshold!

If $C_{01} = C_{10}$, and $P[H_0] = P[H_1]$, $T_1 = 1$,
i.e., Bayes decision rule is identical to ML rule

Example

You are about to step off a long path, known to be safe, to take a shortcut through woods

H_0: Path through woods is safe; $P[H_0] = 0.9$
H_1: Path through woods has land mines $P[H_1] = 0.1$

U_0: Decide that the shortcut is safe and take it
U_1: Decide that the shortcut is unsafe and avoid it

Cost of false alarm C_{10}: take the long path and lose time
Cost of missed detection C_{01} : arm + leg

MAP and Bayesian Decision for Continuous RV observation

When observation X is continuous, the theory is identical!

– Only difference: Likelihood Ratio is defined in terms of densities

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

– The rest is the same: ML, MAP, Bayesian optimal decision rules are: Select H_1 if $L(x) > T$, otherwise select H_0

– Reduce likelihood ratio comparison to a condition on the variable X

$$L(X) > T \Leftrightarrow X \in B_T$$

Prob. Detection = $P[X \in B_T|H_1]$

Prob. False Alarm = $P[X \in B_T|H_0]$

Critique of the Bayesian approach

– Hard to assign prior values $P[H_0]$, $P[H_1]$

– Costs are pretty arbitrary guesses, also

– "*The answer can be made to come out to whatever you want it to be!*"

– "*Your personal prejudices and beliefs get injected into the problem*"

Philosophical response of the Bayesians…

– Non-Bayesians are using Bayesian statistics but refusing to admit the fact ($C_{01} = C_{10}$, $P[H_0]$, $P[H_1]$)

– Non-Bayesians are using their (unstated) belief that the hypotheses are equally likely when using ML

Both are right

– Bayesian approach is very valuable *if used carefully and thoughtfully*

– Bayesian approach can be abused very easily, and used to provide support for any theory that one wants to make up!

Engineering view

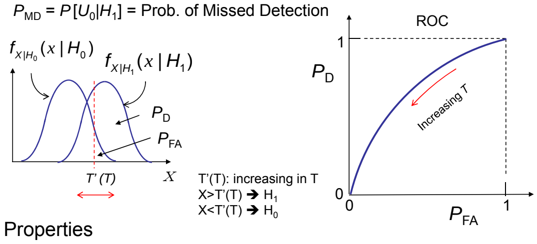
– Ignore philosophy. We have one degree of freedom to choose: selection of the threshold. Choice of any threshold can be justified by subjective rationalization in Bayesian terms!

Receiver Operating Characteristic (ROC)

The ROC is a plot of the probability of detection $P_D = (1 - P_{MD})$ versus the probability of false alarm P_{FA} for varying threshold T. It shows the tradeoff between these probabilities

$P_{FA} = P[U_1|H_0]$ = Prob. of False Alarm

$P_{MD} = P[U_0|H_1]$ = Prob. of Missed Detection



Properties

- Starts at (0,0) and ends at (1,1)
- Concave (consider randomized decision rules)

Example

Likelihood matrix

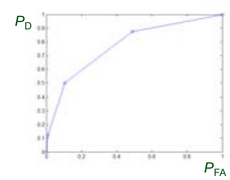
	$X = 0$	$X = 1$	$X = 2$	$X = 3$
H_0	0.512	0.384	0.096	0.008
H_1	0.125	0.375	0.375	0.125
$L(x)$	0.244	0.976	3.906	15.625

```
LR = [0.512 0.384 0.096 0.008;
0.125 0.375 0.375 0.125];
ROC =zeros(2,size(LR,2)+1);
ROC(:,1) = [1;1];
E = cumsum(LR')';
ROC(:,2:end) = [1-E(1,:); 1-
E(2,:)];
plot(ROC(1,:)',ROC(2,:)','o');
hold on;
plot(ROC(1,:)',ROC(2,:)');
```

- ROC curve: Discrete
  - 5 possible decision rules
  - Can get values between corners if use randomized decision rules (i.e., flip a coin to choose T from discrete set…)
    - Beyond our interest in course

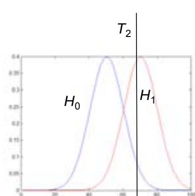
$$P(U=1|H_0) = \sum_{x:U(x)=1} P(X=x|H_0);$$

$$P(U=0|H_1) = \sum_{x:U(x)=0} P(X=x|H_1);$$



Example: Constant Signal Detection in Gaussian Noise

- Signal level: v, known constant
  - Noise: $w \sim$ N(0,1)
- Hypotheses:
  - H_1: Signal is present: $X = v + w$
  - H_0: No signal is present: $X = w$
- Likelihoods
  - $f_{X|H_0} \sim$ N(0,1)
  - $f_{X|H_1} \sim$ N(v,1)
- Objective: Compute ROC



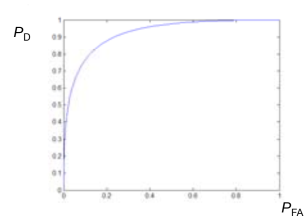
$$\log(L(x)) \equiv \log\left(\frac{f_{X|H_1}(x|H_1)}{f_{X|H_0}(x|H_0)}\right)$$

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

Declare H_1 if $xv - v^2/2 > \log(T_1) \Leftrightarrow x > \frac{\log(T_1) + v^2/2}{v} = T_2$

- For a given threshold T_2, can find P_{FA}, P_D
  - Choose threshold = choose tradeoff
  - Different decision rules = different thresholds

$$P_{FA} = 1 - \Phi(T_2), P_D = 1 - \Phi(T_2 - v)$$




```
X = linspace(-4,4)';
Pfa = 1 - normcdf(X,0,1);
Pd = 1 - normcdf(X,2,1);
plot(Pfa,Pd)
```

Example: Computer Manufacturing

- 2 factories: 1 has probability of failure per part = 0.001, other has prob. failure 0.005
- Get huge batch of parts, don't know which factory made them
- H_0: factory 0. H_1: factory 1
- Test parts until first failed part is found. Measure X = number of parts tested until first failure found

• Likelihoods
  - $P[X|H_0]$ ~ geometric, parameter (0.001)
  - $P[X|H_1]$ ~ geometric, parameter (0.005)

• Likelihood Ratio, $L(n) = \frac{0.005(1-0.005)^{n-1}}{0.001(1-0.001)^{n-1}} = 5\left(\frac{1-0.005}{1-0.001}\right)^{n-1}$

$$\log(L(n)) = (n-1)\log(0.995/0.999) + \log(5)$$

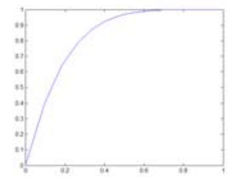
• ML rule: Say H_1 if $n < 1 + \log(5)/[\log(0.999/0.995)] = n_0$
  - Say H_0 otherwise

• If $P[H_0]$, $P[H_1]$ are different, MAP rule:

  Say H_1 if $n < 1 + \log(5P[H_1]/P[H_0]) / \log(0.999/0.995) = n_0$
  - H_0 otherwise

$$\begin{aligned} P_{FA} &= P(U_1|H_0) \\ &= P[n < n_0|H_0] \\ &= geocdf(n_0, 0.001) \\ P_D &= 1 - P_{MD} \\ &= 1 - P[U_0|H_1] \\ &= 1 - P[n > n_0|H_1] \\ &= geocdf(n_0, 0.005) \end{aligned}$$

P_D



P_{FA}

```
X = linspace(0,10000)';
Pf = geocdf(X,0.001);
Pd = geocdf(X,0.005);
plot(Pf,Pd)
```

- Bayesians: pick threshold
  - Use Costs, prior probabilities
- Engineers: Pick threshold
  - Look at ROC, explore tradeoffs

8.5 Generalizations: *N*-ary Test of Hypotheses & Vector Observations

N-ary Test of Hypothesis

- Hypotheses: $H_0, \ldots, H_{N-1}$
- Likelihoods: $P[X = x|H_i]$ (X is discrete)
  $f_{X|H_i}(x|H_i)$ (X is continuous)
- Decisions $U \in \{U_0, \ldots, U_{N-1}\}$

Decision Rules

MAP: For each $X = x$, select $U(x) = U_i$ whenever

$$i \in \arg\max_{j\in\{0,\ldots,N-1\}} P(X = x|H_j)P(H_j) \text{ discrete}$$
$$i \in \arg\max_{j\in\{0,\ldots,N-1\}} f_{X|H_j}(x|H_j)P(H_j) \text{ continuous}$$

MAP minimizes average Prob. of error

ML: For each $X = x$, select $U(x) = U_i$ whenever

$$i \in \arg\max_{j\in\{0,\ldots,N-1\}} P(X = x|H_j) \text{ discrete}$$
$$i \in \arg\max_{j\in\{0,\ldots,N-1\}} f_{X|H_j}(x|H_j) \text{ continuous}$$

Vector Observations

Observation: A pair of RVs, i.e., a random vector **X** = (X,Y)

Likelihoods: $P_{X,Y|H_i}(x,y|H_i)$ for X, Y discrete
$f_{X,Y|H_i}(x,y|H_i)$ for X, Y jointly continuous

Binary Test of Hypotheses, Vector Observation

Likelihood ratio $L(x,y) = \frac{P(X=x,Y=y|H_1)}{P(X=x,Y=y|H_0)}$ discrete
$= \frac{f_{X,Y|H_1}(x,y|H_1)}{f_{X,Y|H_0}(x,y|H_0)}$ continuous

ML rule: $L(x,y) > 1$ choose H_1; else, choose H_0

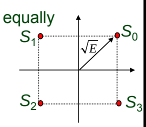
MAP: $L(x,y) > \frac{P(H_0)}{P(H_1)}$ choose H_1; else, choose H_0

Bayes' Risk: $L(x,y) > \frac{C_{10}P(H_0)}{C_{01}P(H_1)}$ choose H_1; else, choose H_0

Example: Quaternary Phase Shift Keying

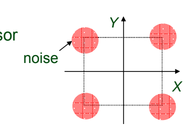
A transmitter sends one of 4 equally likely symbols corresponding to hypotheses H_0, H_1, H_2, H_3

Hypotheses n = 0, 1, 2, 3 correspond to four phases with equally spaced angles:



$$\begin{aligned} S_n &= \sqrt{E}\exp\{jn\pi/2 + j\pi/4\} \\ X_n &= \sqrt{E}\cos(n\pi/2 + \pi/4) \\ Y_n &= \sqrt{E}\sin(n\pi/2 + \pi/4) \end{aligned}$$

Observation: The real & imaginary parts of the phasor



$$\begin{aligned} X &= X_n + W \\ Y &= Y_n + V \end{aligned}$$

W and V ~ N(0,1) are additive independent Gaussian noise introduced by the receiver

ML Detector

Likelihoods: jointly Gaussian RVs

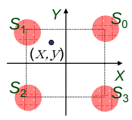
$$f_{X,Y}(x,y|H_n) \sim N\left(\begin{pmatrix} \sqrt{E}\cos(n\pi/2+\pi/4) \\ \sqrt{E}\sin(n\pi/2+\pi/4) \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

$$n \in \arg \max_{k\in\{0,\ldots,3\}} \log(f_{X,Y|H_k}(x,y|H_k))$$

$$n \in \arg \max_{k\in\{0,\ldots,3\}} -\log 2\pi - \frac{\begin{bmatrix} x-x_k & y-y_k \end{bmatrix} \begin{bmatrix} x-x_k \\ y-y_k \end{bmatrix}}{2}$$

ML decision rule: Given (x,y), select n for which $\ln f_{X,Y}(x,y|H_n)$ is maximum.

$\Rightarrow$Select hypothesis in same quadrant of the observation (x,y),



Probability of Error: Use symmetry, independence of X,Y

$$P(Error) = 1 - P(X>0, Y>0|H_0) = 1 - \Phi(\sqrt{E/2})^2$$

8.6 Significance Testing as a Test of Hypotheses

Does a model describe the observations of an experiment?

One hypothesis H_0: model is correct

Alternative H_1: model is not correct

Significance level: $\alpha = P[s \in R \mid H_0]$

= Probability of rejecting H_0 when it is true

= False alarm probability

α is specified, Test is based on the decision rule

Example

- H_0: X is N(0,1)
- Test R: $|X| > 3$
- Significance level: $P[X \in R|H_0] = 2\,Q(3)$