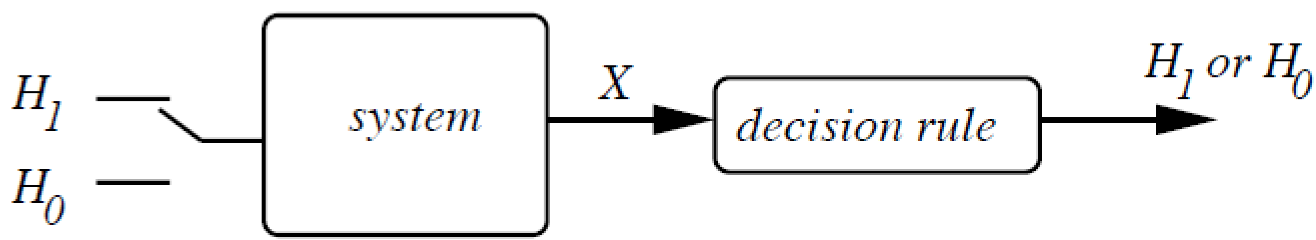
Probability with Engineering Applications
ECE 313 – Section C – Lecture 15

Lav R. Varshney

2 October 2017

Binary Decision-Making

- Last time we were concerned with general inference problems
- Today, let us restrict ourselves specifically to binary decision making
- There are two possible states of the world, H_0 and H_1 (e.g. disease absent, disease present)



Likelihood functions

- We model the observed data by a discrete random variable X
- If hypothesis H_1 is true, then X has the conditional pmf p_1 and if hypothesis H_0 is true then X has pmf p_0
- These are called likelihood functions

Decision rule

- A decision rule specifies, for each possible observation, which hypothesis is declared
- A decision making rule ϕ is a $\{H_0, H_1\}$-valued function of a measurement X, i.e. $\phi(X) \in \{H_0, H_1\}$.
- Equivalently, if S_0, S_1 where $S_1 = S_0^c$ is a binary partition of the measurement space $X \in \Omega$, then

$$\phi(X) = \begin{cases} H_1, & x \in S_1 \\ H_0, & x \in S_0 \end{cases}$$

Likelihood matrix and decision rule

- Likelihood matrix

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0.0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

- Decision rule

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0.0	$\underline{0.1}$	$\underline{0.3}$	$\underline{0.6}$
H_0	$\underline{0.4}$	0.3	0.2	0.1

← underlines indicate the decision rule used for this example.

Outcomes of a decision

true state	decision $\phi(X)$	
H_0	H_0	correct
H_1	H_1	correct
H_0	H_1	false alarm
H_1	H_0	missed detection

False alarms and misses

- We define probabilities of false alarm and missed detection as the following conditional pmfs:

$$p_f = P(\phi(X) = H_1 | H_0)$$
$$p_m = P(\phi(X) = H_0 | H_1)$$

- Note that p_f is the sum of the entries in the H_0 row of the likelihood matrix not underlined
- Note that p_m is the sum of the entries in the H_1 row of the likelihood matrix not underlined

Best decision rules

- The design problem is to determine the best decision rule ϕ, or equivalently the best underlined set S_0

- What criteria make sense?

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0.0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

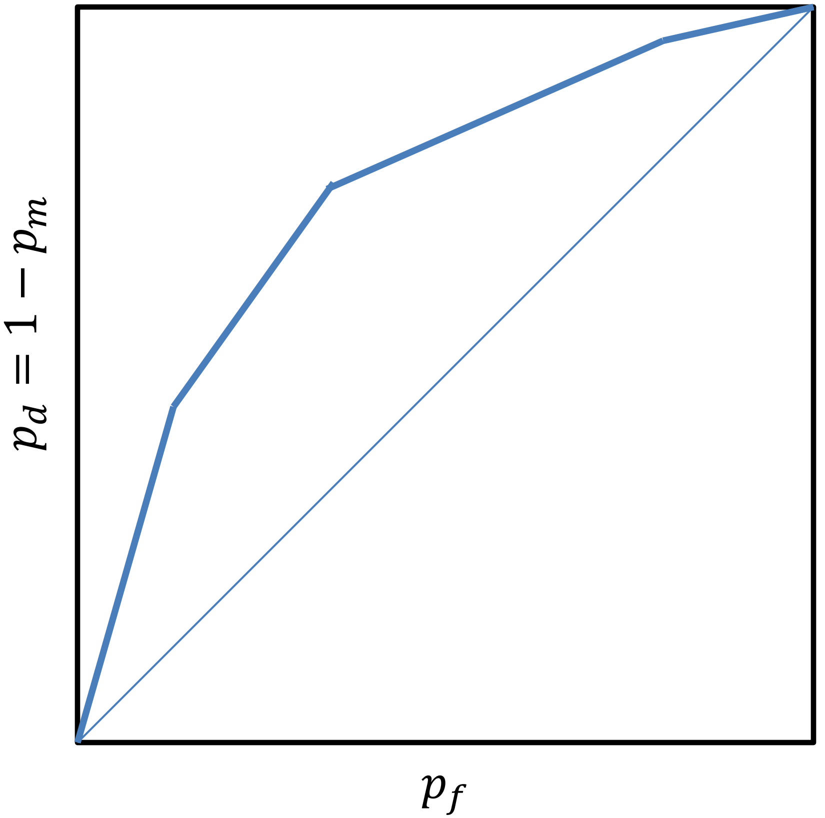
Neyman-Pearson

- Neyman and Pearson suggested that a good decision rule would be one that minimizes missed detection probability p_m subject to upper bound α on false alarm probability p_f

- In statistics, α is called the *size* of the statistical test, and $\beta = 1 - p_m$ is called the *power* of the test

Neyman-Pearson

- One can explore the tradeoff between α and β using the receiver operating characteristic (ROC)



Maximum likelihood (ML)

- The ML decision rule declares the hypothesis which maximizes the probability (or likelihood) of the observation
- Operationally, the ML decision rule can be stated as follows: Underline the larger entry in each column of the likelihood matrix (if entries in a column of the likelihood matrix are identical, either can be underlined)

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
H_1	0.0	0.1	$\underline{0.3}$	$\underline{0.6}$
H_0	$\underline{0.4}$	$\underline{0.3}$	0.2	0.1

$\leftarrow$ underlines indicate the ML decision rule

Likelihood ratio test

- The ML rule can be rewritten in a form called a *likelihood ratio test* (LRT) as follows
- Define the likelihood ratio $\Lambda(k)$ for each possible observation k as the ratio of the two conditional probabilities:

$$\Lambda(k) = \frac{p_1(k)}{p_0(k)}$$

- The ML rule is equivalent to deciding H_1 if $\Lambda(X) > 1$ and deciding H_0 if $\Lambda(X) < 1$

Likelihood ratio test

- Can be rewritten more compactly as:

$$\Lambda(X) \begin{cases} > 1 & \phi(X) = H_1 \\ < 1 & \phi(X) = H_0 \end{cases}$$

- More general decision rules are also likelihood ratio tests, with general threshold τ in place of the specific choice of 1 here
- Note that varying τ traces out the ROC

Prior probabilities

- Often we may have prior beliefs about which hypothesis will arise, e.g. a disease may be known to be rare
- These probabilities π_0 and π_1 are called *prior probabilities*, since they are the probabilities assumed prior to when the observation X is made

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 56, NO. 10, OCTOBER 2008 4553

Quantization of Prior Probabilities for Hypothesis Testing

Kush R. Varshney, *Graduate Student Member, IEEE*, and Lav R. Varshney, *Graduate Student Member, IEEE*

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 60, NO. 9, SEPTEMBER 2012 4537

Quantization of Prior Probabilities for Collaborative Distributed Hypothesis Testing

Joong Bum Rhim, *Student Member, IEEE*, Lav R. Varshney, *Member, IEEE*, and
Vivek K Goyal, *Senior Member, IEEE*

Bayes rule

- Use Bayes rule to combine priors and likelihoods and determine *posterior probabilities* (after making measurement)

$$P(H = H_i | X = x) = \frac{\pi_i p_i(k)}{\pi_0 p_0(k) + \pi_1 p_1(k)}$$

Bayes rule

- Together the conditional probabilities in the likelihood matrix and the prior probabilities determine the joint probabilities $P(H_i, X = k) = \pi_i p_i(k)$ (the numerator in Bayes)
- The joint probability matrix is the matrix of these, in the same layout as the likelihood matrix

Bayes rule

$\pi_0 = 0.8$ and $\pi_1 = 0.2$. Then the joint probability matrix is given by

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0.00	0.02	0.06	0.12
H_0	0.32	0.24	0.16	0.08.

- Note that row for H_i of the joint probability matrix is π_i times corresponding row of likelihood matrix

Maximum a posteriori (MAP) rule

- We can design a decision rule to minimize error probability: $p_e = \pi_0 p_f + \pi_1 p_m$
- It can be proven that the rule that maximizes the posterior probabilities does this
- MAP rule: underline the larger entry in each column of the joint probability matrix

	$X=0$	$X=1$	$X=2$	$X=3$
H_1	0.00	0.02	0.06	$\underline{0.12}$
H_0	$\underline{0.32}$	$\underline{0.24}$	$\underline{0.16}$	0.08

$\leftarrow$ underlines indicate the MAP decision rule

Maximum a posteriori (MAP) rule

- MAP rule declares hypothesis H_1 if $\pi_1 p_1(k) > \pi_0 p_0(k)$
- Equivalently if $\Lambda(k) > \pi_0/\pi_1$, where Λ is the likelihood ratio
- This is the LRT with threshold π_0/π_1

- Note that MAP reduces to ML when priors are equal