Binary Hypothesis Testing

ECE 313

Probability with Engineering Applications

Lecture 17

Professor Ravi K. Iyer

Dept. of Electrical and Computer Engineering

University of Illinois at Urbana Champaign

Today's Topics

- Binary Hypothesis Testing

- Decision Rules:
  - Maximum Likelihood (ML)
  - Maximum a Posteriori Probability (MAP)

- Examples

Hypothesis Testing

- In many practical problems we need to make decisions about populations on the basis of limited information contained in a sample.
- For instance a system administrator may have to decide whether to upgrade the capacity of installation or not. In this case, the choice is binary in nature (e.g. an upgrade either takes place or not)
- In order to arrive at a decision we often make some assumptions or guess about the nature of the underlying population. Such an assertion which may or may not be valid is called a **statistical hypothesis**.
- Procedures that enable us to decide whether to reject or accept hypotheses, based on the available information are called **statistical tests**.

Example 1

- Assume that your office does not have any windows and you want to guess if it is raining outside or not based on the number of people in the office that carry umbrella today.
- Assume that you have two office-mates.

- From the previous observations during 100 non-rainy days in your office, you know that 80% of the time, no one brought an umbrella to the office; on 15 days, only one of your office-mates brought an umbrella, and in 5 days, both your office-mates had an umbrella.

- Assume that you also observed the number of umbrellas in the office in 100 rainy days and found that in 1% of the time, no one has an umbrella, 9% of the time, only one person has an umbrella, and 90% of the time both bring umbrellas to the office.

Example 1 (Cont'd)

- Let *N* be the number of people who bring an umbrella to the office on a given day. *N* takes one of the values {0, 1, 2}.
- Based on your observations, you can write the following conditional probabilities:

P(N = 0 | Not Rainy) = 0.8	P(N = 0 | Rainy) = 0.01
P(N = 1 | Not Rainy) = 0.15	P(N = 1 | Rainy) = 0.09
P(N = 2 | Not Rainy) = 0.05	P(N = 2 | Rainy) = 0.9

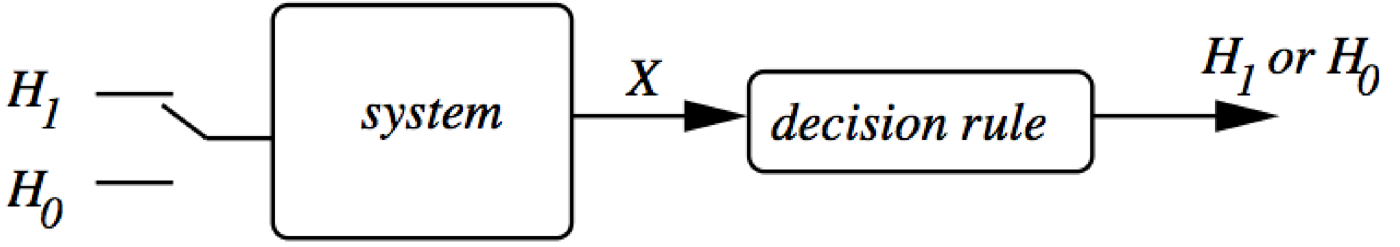
- We can show these conditional probabilities in a matrix:

	N = 0	**N = 1**	**N = 2**
Not Rainy	0.8	0.15	0.05
Rainy	0.01	0.09	0.9

- On a given day, if you see that no one is having an umbrella, what would be your guess?

Binary Hypothesis Testing

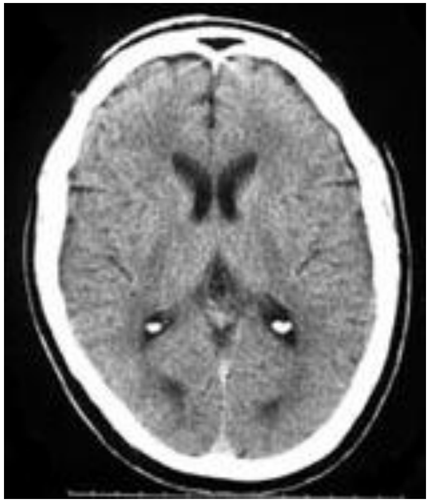
- The basic framework for binary hypothesis testing:



- It is assumed that either hypothesis H_1 is true or hypothesis H_0 is true, as indicated by the position of the switch at the left end.
- Based on which hypothesis is true, a system generates an observation *X.*
- The observation is fed into a decision rule, which then declares either H_1 or H_0
- The system is assumed to be random, so the decision rule can sometimes declare the true hypothesis, or it can make an error

Example 2

- Suppose the data is from a computer aided tomography (CAT) scan system, and the hypotheses are:
  - ***H_1***: A tumor is present
  - ***H_0***: No tumor is present



- We model the observed data by a discrete random variable *X.* Suppose:
  - If hypothesis H_1 is true, then X has pmf p_1
  - If hypothesis H_0 is true, then X has pmf p_0
- pmf's for the two hypotheses are shown by a ***likelihood matrix***:

		$X=0$	$X=1$	$X=2$	$X=3$
P (X = k | H1) =>	H_1	0.0	0.1	0.3	0.6
P (X = k | H0) =>	H_0	0.4	0.3	0.2	0.1

Example 2 (Cont'd)

- A decision rule specifies for each possible observation (each possible values of *X*), which hypothesis is declared.

- Conventionally we display a decision rule on the likelihood matrix, by underlying one entry in each column, to specify which hypothesis is to be declared for each possible value of *X.*

- An example decision rule: H_1 is declared whenever $X \geq 1$ *:*

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

$\leftarrow$ underlines indicate the decision rule used for this example.

- For example if X = 2 is observed, H1 is declared.

Binary Hypothesis Testing

- In every binary hypothesis testing problem, there are two possibilities for which the hypothesis is true, and two possibilities for which hypothesis is declared:

  H_0 = Negative or null hypothesis,

  H_1 = Positive or alternative hypothesis

- So there are four possible outcomes:
  1. **Hypothesis H_0 is true and H_0 is declared. => True Negative**
  2. **Hypothesis H_1 is true and H_1 is declared. => True Positive**
  3. **Hypothesis H_0 is true and H_1 is declared. => False Positive**
  4. **Hypothesis H_1 is true and H_0 is declared. => False Negative**

Probability of False Alarm and Miss

- By convention, two conditional probabilities are defined:

$$\boxed{p_{false-alarm} = P(declare\ H_1\ true \mid H_0)}$$ => **Type I Error**

$$\boxed{p_{miss} = P(declare\ H_0\ true \mid H_1)}$$ => **Type II Error**

- $p_{false-alarm}$ is the sum of entries in the H_0 row of the likelihood matrix that are not underlined:

$$p_{false-alarm} = 0.3 + 0.2 + 0.1 = 0.6 \quad \Leftarrow$$

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

- p_{miss} is the sum of entries in the H_1 row of the likelihood matrix that are not underlined:

$$p_{miss} = 0.0 \quad \Leftarrow$$

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

Decision Rule

- If we modify the sample decision rule in the previous example to declare H_1 when $X = 1$:

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

- Then the $p_{false\text{–}alarm}$ will decrease to: 0.3
- The p_{miss} will increase to: $0.0+0.3+0.6=0.9$

- So what decision rule should be used?
  – Trade-off between the two types of error probabilities
  – Evaluate the pair of conditional error probabilities $(p_{false\text{–}alarm}, p_{miss})$ for multiple decision rules and then make a final selection

Maximum Likelihood (ML) Decision Rule

- Maximum likelihood (ML) decision rule declares the hypothesis that ***maximizes the probability (likelihood) of the observation***

$$P(X=k \mid H_0) \quad and \quad P(X=k \mid H_1)$$

- ML decision rule is based on ***likelihood matrix***, the matrix of conditional probabilities of $P(X=k \mid H_i)$
- The sum of elements in each row of the likelihood matrix is one.

- Specified by underlying the larger entry in each column of likelihood matrix. If the entries in a column of the likelihood matrix are identical, then either can be underlined.
- For example 2:

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

$$p_{false-alarm} = 0.2 + 0.1 = 0.3$$

- $p_{miss} = 0.0 + 0.1 = 0.1$

Maximum a posteriori probability (MAP) Decision Rule

- Maximum a posteriori probability (MAP) decision rule declares the hypothesis which ***maximizes the posteriori probabilities***

- ***Posteriori probabilities*** are conditional probabilities that an observer would assign to the two hypotheses after making an observation *k:* $P(H_0 \mid X = k) \quad and \quad P(H_1 \mid X = k)$

- So given an observation $X = k$, the MAP decision rule chooses the hypothesis with the larger posteriori conditional probability
- By Bayes' formula, we have:

$$P(H_0 \mid X = k) = \frac{P(H_0, X = k)}{P(X = k)} = \frac{P(H_0, X = k)}{P(H_0, X = k) + P(H_1, X = k)}$$

$$P(H_1 \mid X = k) = \frac{P(H_1, X = k)}{P(X = k)} = \frac{P(H_1, X = k)}{P(H_0, X = k) + P(H_1, X = k)}$$

Maximum a posteriori probability (MAP) Decision Rule (Cont'd)

- MAP decision rule requires the computation of the ***joint probabilities***: $P(H_0, X = k) \;\; and \;\; P(H_1, X = k)$

- Such probabilities cannot be deduced from the likelihood matrix alone, we need to also assume some values for $P(H_0) \; and \; P(H_1)$
- Let the assumed value of $P(H_i)$ be denoted by π_i, so:

$$\pi_0 = P(H_0) \;\; and \;\; \pi_1 = P(H_1)$$

- $\pi_0 \; and \; \pi_1$ are called the ***prior probabilities***, because they are the probabilities assumed prior to the observation is made.

- The joint probabilities $P(H_i, X = k)$ are then determined by the conditional probabilities $P(X = k \mid H_i)$ listed in the likelihood matrix and the prior probabilities π_i :

$$P(H_i, X = k) = P(X = k \mid H_i)\pi_i$$

Maximum a posteriori probability (MAP) Decision Rule (Cont'd)

- MAP decision rule is based on ***joint probability matrix****, the* matrix of joint probabilities of $P(H_i, X = k)$.

- Each row H_i in the joint probability matrix is π_i times the corresponding row of the likelihood matrix. The sum of entries in row H_i is π_i , and the some of all the entries in the matrix is one.

$$\text{Likelihood Matrix} \xrightarrow{\pi_0=0.8 \;\; , \;\; \pi_1=0.2} \text{Joint Probability Matrix}$$

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

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

Maximum a posteriori probability (MAP) Decision Rule (Cont'd)

- Remember : $P(H_i \mid X=k) = \dfrac{P(H_0, X=k)}{P(X=k)}, = \dfrac{P(H_i, X=k)}{P(H_0, X=k) + P(H_1, X=k)}$
- So the posterior probability $P(H_i \mid X=k)$ can be computed by dividing the value in the row H_i , column $X=k$, divided by the sum of entries in the column $X=k$.
- Since the denominators for both $P(H_0 \mid X=k)$ *and* $P(H_1 \mid X=k)$ are the same (both equal to $P(X=k)$), we can only compare the nominator (joint probabilities of $P(H_0, X=k)$)
- MAP decision rule is specified by underlying the larger entry in each column of the joint probability matrix. If the entries in a column of the likelihood matrix are identical, then either can be underlined.
- For the example 2:

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

- $p_{false-alarm} = 0.08/0.8 = 0.1$

  $p_{miss} = (0.00+0.02+0.06)/0.2 = 0.4$

Remember Example 1

- On a given day, if you see that no one is having an umbrella, what would be your guess?
- Using ML Rule **=> Rains when N = 2**
  - Likelihood Matrix:

	N = 0	**N = 1**	**N = 2**
Not Rainy	<u>0.8</u>	<u>0.15</u>	0.05
Rainy	0.01	0.09	<u>0.9</u>

- Using MAP Rule **=> Rains when N = 1**
  - Assume that you know the prior probabilities of raining and not raining:
    - P(Not Rainy) = 0.6, P(Rainy) = 0.2 ,
  - Joint Probability Matrix:

	N = 0	**N = 1**	**N = 2**
Not Rainy	<u>0.480</u>	0.090	<u>0.030</u>
Rainy	0.002	<u>0.018</u>	0.018

Average Error Probability

- Remember the conditional probabilities of:

$$\boxed{p_{false-alarm} = P(declare\ H_1\ true \mid H_0)}$$

$$\boxed{p_{miss} = P(declare\ H_0\ true \mid H_1)}$$

- The average error probability is defined as:

$$\boxed{p_e = \pi_0 p_{false-alarm} + \pi_1 p_{miss}}$$

- $p_{false-alarm}$ is the sum of entries in the H_0 row of the likelihood matrix, (*joint probability matrix*) that are not underlined (divided by π_0)
- p_{miss} is the sum of entries in the H_0 row of the likelihood matrix, *(joint probability matrix*) that are not underlined (divided by π_1)

- p_e is the sum of all the entries in the *joint probability* matrix that are not underlined.
- Among all decision rules, MAP is the one that minimizes p_e => Optimal

Likelihood Ratio Test (LRT)

- Another way to write ML and MAP decision rules is using LRT.
- Define the likelihood ratio $\Lambda(k)$ for each possible observation *k* as the ratio of the two conditional probabilities:

$$\Lambda(k) = \frac{p_1(k)}{p_0(k)} = \frac{P(X = k \mid H_1)}{P(X = k \mid H_0)}$$

- A decision rule can be expressed as an LRT with threshold τ *:*

$$\Lambda(X) \begin{cases} > \tau & \textit{declare } H_1 \textit{ is true.} \\ < \tau & \textit{declare } H_0 \textit{ is true.} \end{cases}$$

- If the threshold τ is increased, then there are fewer observations that lead to deciding H_1 is true.
- As τ increases, $p_{false-alarm}$ decreases, and p_{miss} increases.

Likelihood Ratio Test (LRT) (Cont'd)

- If the observation is *X = k:*
- The ML rule declares hypothesis H_1 is true, if $p_1(k) > p_0(k)$ and otherwise it declares H_0 is true.
- So the ML rule can be specified using an LRT with $\tau = 1$:

$$\Lambda(X)\begin{cases} >1 & declare\ H_1\ is\ true. \\ <1 & declare\ H_0\ is\ true. \end{cases}$$

- The MAP rule declares hypothesis H_1 is true, if $\pi_1 p_1(k) > \pi_0 p_0(k)$ and otherwise it declares H_0 is true.
- So the MAP rule can be specified using an LRT with $\tau = \dfrac{\pi_0}{\pi_1}$:

$$\Lambda(X)\begin{cases} >\dfrac{\pi_0}{\pi_1} & declare\ H_1\ is\ true. \\ <\dfrac{\pi_0}{\pi_1} & declare\ H_0\ is\ true. \end{cases}$$

- *For uniform priors $\pi_0 = \pi_1$, MAP and ML decision rules are the same*

Example 3

- Suppose you have a coin and you know that either:
  - H1: the coin is biased, showing heads on each flip with probability 2/3; or
  - H0: the coin is fair, showing heads and tails with probability 1/2

- Suppose you flip the coin five times. Let *X* be the number of times heads shows.

- Describe the ML and MAP decision rules using LRT.
- Find $p_{false-alarm}, p_{miss}$, and p_e for both decision rules.
- Use the prior probabilities of $(\pi_0, \pi_1) = (0.2, 0.8)$.

Example 3 (Cont'd)

- *X* (number of times that head shows up) has a binomial distribution with *n = 5*, and:
  - $p = 2/3$ for H_1 (Coin is biased)
  - $p = 1/2$ for H_0 (Coin is fair)

- Remember for a binomial distribution: $P(X = k) = \binom{5}{k} p^k (1-p)^{5-k}$
- So we have:

$$P(X = k \mid H_1) = \binom{5}{k}\left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{5-k} \qquad P(X = k \mid H_0) = \binom{5}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{5-k}$$

- The rows of the likelihood matrix consist of the pmf of *X:*

	X = 0	*X = 1*	*X = 2*	*X = 3*	*X = 4*	*X = 5*
H_1	$\left(\frac{1}{3}\right)^5$	$5\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^4$	$10\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3$	$10\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2$	$5\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)$	$\left(\frac{2}{3}\right)^5$
H_0	$\left(\frac{1}{2}\right)^5$	$5\left(\frac{1}{2}\right)^5$	$10\left(\frac{1}{2}\right)^5$	$10\left(\frac{1}{2}\right)^5$	$5\left(\frac{1}{2}\right)^5$	$\left(\frac{1}{2}\right)^5$

Example 3 (Cont'd)

- In computing the likelihood ratio, the binomial coefficients cancel, so:

$$\Lambda(k) = \frac{\left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{5-k}}{\left(\frac{1}{2}\right)^5} = 2^k \left(\frac{2}{3}\right)^5 \approx \frac{2^k}{7.6}.$$

- The ML decision rule is:
  - Declare H_1 whenever $\Lambda(X) \geq 1$, or equivalently $X \geq 3$.

- The MAP decision rule is:
  - Declare H_1 whenever $\Lambda(X) \geq \frac{0.2}{0.8} = 0.25$, or equivalently $X \geq 1$.

Example 3 (Cont'd)

- For the ML rule:

$$p_{false-alarm} = 10\left(\frac{1}{2}\right)^5 + 5\left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right) = 0.5$$

$$p_{miss} = \left(\frac{1}{3}\right)^5 + 5\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^4 + 5\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3 = \frac{51}{243} \approx 0.201$$

$$p_e = (0.2)\,p_{false-alarm} + (0.8)\,p_{miss} \approx 0.26$$

- For the MAP rule:

$$p_{false-alarm} = 1 - \left(\frac{1}{2}\right)^5 \approx 0.97$$

$$p_{miss} = \left(\frac{1}{3}\right)^5 = \frac{1}{243} \approx 0.041$$

$$p_e = (0.2)\,p_{false-alarm} + (0.8)\,p_{miss} \approx 0.227$$

- As expected the probability of error (p_e) for the MAP rule is smaller than p_e for the ML rule.