ECE 313

Probability with Engineering Applications

Decision-making under uncertainty II

Professor Dilip V. Sarwate
Department of Electrical and Computer Engineering

Hypothesis testing model

- One of M mutually exclusive hypotheses $H_0, H_1, \ldots, H_{M-1}$ is true
- **X** is a random variable whose value we can observe, and use, to decide which of the hypotheses is true
- If H_i happens to be the true hypothesis, then the pmf of **X** is $P_i(u)$
- To avoid trivialities, we assume that $M \geq 2$

The decision rule

- We observe the value of **X** and announce our decision as to which hypothesis we believe to be true
- This decision may or may not coincide with reality — our decision may be H_i when in fact H_j is the true hypothesis
- The decision rule (which we are free to choose as we wish) assigns a hypothesis (the decision!) to each possible value of **X**

Specifying the decision rule

- We can specify the decision rule as a table that lists all the values of **X** and the corresponding decisions
- After observing the value of **X**, we merely look up our decision in the table and announce it!
- The process is completely mechanical (or computerized?) — observe **X** and look up the decision from the table

General remarks on decision rules

- There are M choices of hypothesis for the decision for each of the N values of **X**
- Thus, there are M^N different decision rules that we could use
- If M > N, some hypotheses will never be chosen for any value of **X**
- If M < N, several values of **X** will result in the same decision

Simple example of a decision rule

- Consider a binary hypothesis test in which the observation **X** takes on only the values 1 and 2
- Our decision rule is as follows

  If **X** = 1, decide H_0 is the true hypothesis

  If **X** = 2, decide H_1 is the true hypothesis
- Remember that the decision might not be correct (or it might be correct)

Deterministic vs randomized rules

- The decision rule that we have described is called a deterministic decision rule
- A randomized decision rule is one in which, after observing **X**, we choose the decision "randomly" with specified probabilities
- If **X** = 1, toss a coin with P(Heads) = 0.7 and decide H_0 if Heads, H_1 if Tails
- If **X** = 2, toss a coin with P(Heads) = 0.25 and decide H_0 if Heads, H_1 if Tails

Deterministic ⊂ randomized

- If **X** = 1, toss a coin with P(Heads) = 0.7 and decide H_0 if Heads, H_1 if Tails
- If **X** = 2, toss a coin with P(Heads) = 0.25 and decide H_0 if Heads, H_1 if Tails
- Our deterministic decision rule (H_0 if **X** = 1 and H_1 if **X** = 2) is trivially a randomized rule
- We merely toss a two-headed coin if **X** = 1 and a two-tailed coin if **X** = 2 !!

A sigh of relief…

- Randomized decision rules are more "powerful" than deterministic rules, e.g. when M > N …
- Randomized decision rules provide more elegant solutions than deterministic rules
- But, they add more complications to life
- We will not discuss randomized rules in detail any further — just be aware (beware?) of their existence

L is the likelihood matrix

- The pmf of **X** depends on which one of M hypotheses happens to be true
- Let **X** take on values $u_1, u_2, \ldots u_N$
- We write the pmfs in an likelihood array (or likelihood matrix) of M rows (one for each hypothesis) and N columns (one for each of the possible values of **X**)
- We denote this M×N matrix by L (for likelihood)

Properties of the likelihood matrix

- L is an array of M rows and N columns
- Each row corresponds to a hypothesis
- Each column corresponds to one of the values taken on by **X**
- The entry in the i-th row and j-th column is $P_i(u_j)$, the probability that **X** = u_j when H_i is the true hypothesis
- The sum of the entries in each row is 1

Example of a likelihood matrix

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- The values of **X** are shown in orange above the top row
- Each row is a pmf; some values are 0

Why not call it a probability matrix?

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- For each of the possible observed values of **X**, the entries in that column are not a probability assignment for the H_i

The decision rule and L

- We can specify the decision rule via the likelihood matrix L as follows
- For each value of **X**, shade the entry on the row corresponding to the decision for this value
- There is only one shaded entry in each column
- It is allowable for a particular hypothesis to not be the choice in any column

How to choose the decision rule?

- We now know how to specify a decision rule via the likelihood matrix L
- We also know how to calculate error probabilities for any given decision rule — the sum of the unshaded entries on each row of L is the error probability when that hypothesis is the true hypothesis
- But how do we choose the decision rule?
- How do we choose rationally?

Decision rule specified via L

	0	1	2	3
H_0	**0.2**	0.3	0.2	0.3
H_1	0.12	0.24	**0.64**	0.0
H_2	0.16	**0.1**	0.1	**0.64**
H_3	0.0	0.3	0.3	0.4

- There is only one shaded entry in each column
- Not every hypothesis must be chosen

Is the decision incorrect?

	0	1	2	3
H_0	**0.2**	0.3	0.2	0.3
H_1	0.12	0.24	**0.64**	0.0
H_2	0.16	0.1	0.1	**0.64**
H_3	0.0	**0.3**	0.3	0.4

- Suppose that H_0 is the true hypothesis
- What is the probability that the decision is wrong?

Suppose H_0 is the true hypothesis

	0	1	2	3
H_0	**0.2**	0.3	0.2	0.3
H_1	0.12	0.24	**0.64**	0.0
H_2	0.16	0.1	0.1	**0.64**
H_3	0.0	**0.3**	0.3	0.4

- If **X** has value 0, the decision is H_0
- If **X** takes on any value other than 0, the decision is not H_0

Probability of error when H_0 is true

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- P{error when H_0 is true} = P{**X** ≠ 0}
- P{**X** ≠ 0} = sum of unshaded entries on the H_0 row of the likelihood matrix

Probability of error when H_i is true

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- More generally, if H_i is the true hypothesis, P{error} = sum of unshaded entries on the H_i row of the likelihood matrix

Probability of error when H_i is true

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- P{error} = 0.8, 0.36, 0.36, and 0.7 according as H_0, H_1, H_2, H_3 are the true hypotheses

Maximum-likelihood decision rule I

- We know one of the hypotheses is true
- We have observed the value of **X**, say u_3
- Under which hypothesis would the event already observed, viz. {**X** = u_3}, have had the largest probability?
- The hypothesis which maximizes this probability is the maximum-likelihood decision when {**X** = u_3} is observed
- Repeat for other values of **X**

Maximum-likelihood decision rule II

- The hypothesis for which the probability of the observation is the maximum is called the maximum-likelihood decision when that observation is made
- Statisticians insist that calling these things probabilities is a grievous sin — they are likelihoods, not probabilities!
- For each observation, the maximum-likelihood decision rule maximizes the likelihood of the observation

Finding the ML decision rule

- Under which hypothesis would the already observed event {**X** = u_3} have had the largest probability (or likelihood)?
- Which of the probabilities (likelihoods) $P_i(u_3)$ is the largest?
- Remember that $P_i(u_j)$ is the pmf of **X** when H_i is the true hypothesis
- Operationally, the ML rule says: shade the largest entry in each column of L

This is a maximum-likelihood rule!

	0	1	2	3
H_0	**0.2**	0.3	0.2	0.3
H_1	0.12	0.24	**0.64**	0.0
H_2	0.16	0.1	0.1	**0.64**
H_3	0.0	**0.3**	0.3	0.4

- Notice that the largest entry in each column is shaded
- But what about the column for **X** = 1 ??

This is an ML rule too!

	0	1	2	3
H_0	**0.2**	**0.3**	0.2	0.3
H_1	0.12	0.24	**0.64**	0.0
H_2	0.16	0.1	0.1	**0.64**
H_3	0.0	**0.3**	0.3	0.4

- Which 0.3 entry should we choose?
- It depends on which error probability we want to make smaller

The radar system problem again

- **X** is the number of echoes detected by the radar receiver
- H_0 : **X** ~ Binom(n, p_0)

  H_1 : **X** ~ Binom(n, p_1) where $p_0 \ll p_1$
- The likelihood matrix has 2 rows and n+1 columns
- k-th column entries are $P_0(k)$ and $P_1(k)$

  $\binom{n}{k} \bullet p_0^k (1 - p_0)^{n-k}$ and $\binom{n}{k} \bullet p_1^k (1 - p_1)^{n-k}$

Which is larger?

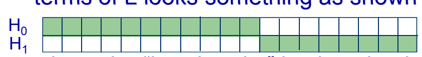
- Which is larger?

  $\binom{n}{k} \bullet p_0^k (1 - p_0)^{n-k}$ or $\binom{n}{k} \bullet p_1^k (1 - p_1)^{n-k}$?
- The answer depends on the value of k
- Take the logarithm of the ratio $P_1(k)/P_0(k)$
- If $k \bullet \ln(p_1/p_0) + (n-k) \bullet \ln[(1-p_1)/(1-p_0)] > 0$, then $P_1(k) > P_0(k) \Rightarrow$ the ML rule says H_1
- If $k > \dfrac{n \bullet \ln[(1-p_0)/(1-p_1)]}{\ln(p_1/p_0) + \ln[(1-p_0)/(1-p_1)]}$, say H_1

ML decision in the radar problem

- H_0 : X ~ Binom(n, p_0)

  H_1 : X ~ Binom(n, p_1) where $p_0 < p_1$
- The maximum-likelihood decision rule in terms of L looks something as shown



  where the "break-point" is given by the maximum-likelihood threshold θ_{ML}
- $\theta_{ML} = \dfrac{n \bullet \ln[(1-p_0)/(1-p_1)]}{\ln(p_1/p_0) + \ln[(1-p_0)/(1-p_1)]}$

ML decision rule in words

- H_0 : **X** ~ Binom(n, p_0)

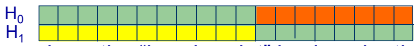
  H_1 : **X** ~ Binom(n, p_1) where $p_0 < p_1$
- The maximum-likelihood decision rule is as follows: If the observed value of **X** exceeds the maximum-likelihood threshold

  $\theta_{ML} = \dfrac{n \bullet \ln[(1-p_0)/(1-p_1)]}{\ln(p_1/p_0) + \ln[(1-p_0)/(1-p_1)]}$

  then announce the decision as H_1 (the target is present). Else announce H_0

Error probabilities for ML decision

- $H_0 : X \sim \text{Binom}(n, p_0)$
  $H_1 : X \sim \text{Binom}(n, p_1)$ where $p_0 < p_1$
- The maximum likelihood decision rule in terms of L looks something as shown



where the “break-point” is given by the maximum-likelihood threshold θ_{ML}
- False alarm probability = sum of ■ (orange)
- False dismissal probability = sum of ■ (yellow)

Back to general case

- We know how to calculate the various error probabilities associated with an arbitrarily-chosen decision rule
- What is the average error probability of the decision rule?
- What do you mean, average?
- There is no such thing as an average error probability
- A hypothesis is either true, or it isn’t true. Where does probability get into it?

What if we repeat the test often?

- If we repeat the hypothesis test over and over; on some trials, H_0 will be true while on other trials, H_1 will be true, and on still other trials, H_2 will be true, and so on
- On some trials, the chances of making an error will be the error probability when H_0 is true, on other trials, we will incur the error probability when H_1 is true, and on still other trials, we will incur the error probability when H_2 is true, and so on

Does probability exist?

- Over a very large number N of hypothesis tests, hypothesis H_i will be true roughly $N \bullet P(H_i)$ times
- Non-Bayesian statistician: Stop right there! You cannot talk of the probability that a hypothesis is true; it makes no sense!
- Either a hypothesis is true, or it is not true (in which case some other hypothesis is true, but there is nothing probabilistic here)
- End of discussion

We pretend that probability exists

- Over a very large number N of hypothesis tests, hypothesis H_i will be true (say) N_i times, and on these N_i tests we will incur error probability $p_i(E)$
- The average of these error probabilities is

$$(1/N)\sum p_i(E)\bullet N_i$$

- Over a very large number N of hypothesis tests, we can pretend that N_i/N is $P(H_i)$, the probability that the hypothesis H_i is true

Conditional probability exists, too

- Over a very large number N of hypothesis tests, we can pretend that N_i/N is $P(H_i)$, the probability that the hypothesis H_i is true
- We can think of $p_i(E)$ as the conditional probability of an error given that the hypothesis H_i is true
- $p_i(E) = P(E|H_i)$
- $(1/N)\sum p_i(E)\bullet N_i = \sum P(E|H_i)\bullet P(H_i) = P(E)$
- Holy theorem of total probability, Batman!

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

The average error probability

- The average probability of making an erroneous decision is

$$P(E) = \sum P(E|H_i) \cdot P(H_i)$$

- Here $P(E|H_i)$, the error probability when hypothesis H_i is true is just the sum of the unshaded entries on the i-th row of the likelihood matrix L
- The i-th row of L is just the conditional pmf of **X** given that H_i is true (i.e. occurred), that is, it is $p_{X|H_i}(u|H_i)$

Better to be right than to be wrong!

- The average probability of making a correct decision is

$$P(C) = \sum P(C|H_i) \cdot P(H_i)$$

- Here $P(C|H_i)$, the probability of a correct decision when hypothesis H_i is true is just the sum of the shaded entries on the i-th row of the likelihood matrix L
- We now convert the likelihood matrix into the joint probability matrix

Likelihoods to joint probabilities

- The i-th row of the likelihood matrix L is just the conditional pmf of **X** given that H_i is true (i.e. occurred)
- If we multiply the i-th row by $P(H_i)$, we get the probabilities that **X** has the various values and that H_i is true
- Carrying out this operation on all the rows of L gives a matrix called the joint probability matrix J

Properties of joint prob. matrix J

- For $0 \le i \le M-1$, if we multiply the i-th row of L by $P(H_i)$, we get a matrix called the joint probability matrix J
- J = KL where K is a diagonal matrix
- $K = \text{diag}[P(H_0), P(H_1), \ldots, P(H_{M-1})]$
- J is a giant Venn diagram whose (i,j)-th entry is the joint probability $P[\{\mathbf{X} = u_j\} \cap H_i]$
- The sum of all the entries in J is 1

More joint prob. matrix properties

- For $0 \le i \le M-1$, if we multiply the i-th row of L by $P(H_i)$, we get a matrix called the joint probability matrix J
- The sum of all the entries in J is 1
- Since the row sums in L were 1, the row sums in J are just the $P(H_i)$
- The column sums in J give us the unconditional pmf of **X**

Example: Likelihood matrix

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- Suppose that $P(H_0) = 0.4 = P(H_3)$ and $P(H_1) = 0.1 = P(H_2)$

Example: Joint probability matrix

	0	1	2	3
H_0	0.08	0.12	0.08	0.12
H_1	0.012	0.024	0.064	0.00
H_2	0.016	0.01	0.01	0.064
H_3	0.00	0.12	0.12	0.16

- $P(H_0) = 0.4 = P(H_3)$
- $P(H_1) = 0.1 = P(H_2)$

Conditional probabilities of C

	0	1	2	3
H_0	0.2	0.3	0.2	0.3
H_1	0.12	0.24	0.64	0.0
H_2	0.16	0.1	0.1	0.64
H_3	0.0	0.3	0.3	0.4

- $P(C|H_0) = 0.2$, $P(C|H_1) = 0.64$, $P(C|H_2) = 0.64$, and $P(C|H_3) = 0.3$
- $P(C) = (0.2+0.3)\times0.4 + 0.64\times2\times0.1 = 0.328$

P(Correct Decision) from J

	0	1	2	3
H_0	0.08	0.12	0.08	0.12
H_1	0.012	0.024	0.064	0.00
H_2	0.016	0.01	0.01	0.064
H_3	0.00	0.12	0.12	0.16

- $P(C)$ = sum of shaded entries
  = 0.08 + 0.064 + 0.064 + 0.12
  = 0.328

Correct decision probability and J

- For an arbitrary decision rule specified by shading entries in L, the (average) probability of a correct decision is the sum of the corresponding shaded entries in J
- The (average) probability of error is the sum of all the unshaded entries in J
- L and J provide a simple procedure for computing various error probabilities
- L gives conditional error probabilities; J gives average error probabilities

Maximize P(C) by clever choice

- P(C), the average probability of a correct decision, is the sum of the shaded entries in J
- There is one shaded entry in each column
- Which entries are to be shaded is up to us!
- We can maximize P(C) (and thus minimize P(E)) by shading the largest entry in each column of J
- The corresponding decision rule is called a minimum-error-probability decision rule

P(C) for Minimum-Error-Prob. rule

	0	1	2	3
H_0	0.08	0.12	0.08	0.12
H_1	0.012	0.024	0.064	0.00
H_2	0.016	0.01	0.01	0.064
H_3	0.00	0.12	0.12	0.16

- $P(C)$ = sum of shaded entries
  = 0.08 + 0.12 + 0.12 + 0.16
  = 0.48 > 0.328

Summary

- We studied how to set up decision rules in some rational manner
- The ML decision rule chooses the largest entry in each column of the likelihood matrix L
- The joint probability matrix J is obtained by multiplying each row of L by the probability of the hypothesis
- The minimum-error-probability decision rule chooses the largest entry in each column of J

ECE 313 - Lecture 16 Slide 48 of 48

ECE 313

Probability with Engineering Applications

Decision-making under uncertainty II

Professor Dilip V. Sarwate
Department of Electrical and
Computer Engineering