# LU30-Review of probability – Basic Probability Notation & The Axioms of Probability

| LU Objectives | |
|---|---|
| | To recall basic probability concepts |
| LU Outcomes | CO Number: 3 |
| | Apply probability concepts for reasoning |

## Recap

- In the last topic, “Uncertainty, the agent cannot assume what the environment is and act according to those assumed data
- Hence, the agent should consider multiple possibilities based on **DEGREE OF BELIEF** in the relevant data or sentences
- Here the degree of belief lies between 0 and 1 for all the sentences
- i.e. The identification of truth of the given sentence
- Let us see these in DETAIL

# Probability

- Probabilities are used to compute the truth of given statement, written as numbers between 0 and 1, that describes how likely an event is to occur.
- 0 indicates impossibility and 1 indicates certainly.
  - 1. Tossing a coin 2. Tolling a dice
- Probability based reasoning
  - understanding from knowledge
  - how much of uncertainty present in that event.



Probability (Head) = 50% & Probability (Tail) = 50% as Coin has only 2 sides
Rolling Dice: Probability(1) = 1/6 = 16%

# Probability

- *Probability provides a way of* **summarizing** *the uncertainty, that comes from our laziness and ignorance.*
- Toothache problem - an 80% chance , a probability of 0.8 that the patient has a cavity if he or she has a toothache.
- The 80% summarizes those cases, but both toothache and cavity are unconnected.
- The missing 20% summarizes, all other possible causes of toothache, that we are too lazy or ignorant to confirm or deny.

- **Laziness**: Failure to eliminate expressions, quantifiers etc.
- **Ignorance**: It is a lack of relevant fact, initial conditions etc.

- In a Toothache problem, we might not know for sure what affects a particular patient but we believe that 80% chance that is the probability of 0.8 ….

**NEXT**

# *Probability*

- *Probability provides a way of* **summarizing** *the uncertainty, that comes from our laziness and ignorance.*
- Toothache problem - an 80% chance , a probability of 0.8 that the patient has a cavity if he or she has a toothache.
- The 80% summarizes those cases, but both toothache and cavity are unconnected.
- The missing 20% summarizes, all other possible causes of toothache, that we are too lazy or ignorant to confirm or deny.

- By the previous experience, 80% summarizes those …
- Also, the missing 20% …

- Probabilities between 0 and **1** correspond to intermediate degrees of belief in the truth of the sentence.
- The sentence itself is in *fact* either true or false.
- It is important to note that a degree of belief is different from a degree of truth.
- A probability of 0.8 does not mean "80% true" but rather an 80% degree of belief-that is, a fairly strong expectation.
- Thus, probability theory makes the same ontological commitment as logic-namely, that facts either do or do not hold in the world.
- Degree of truth, as opposed to degree of belief, is the subject of **fuzzy logic**

- In probability theory, a sentence such as
- "The probability that the patient has a cavity is 0.8“,
- is about the agent's beliefs, not directly about the world.
- These percepts create the **evidence,** which are based on probability statements.
- All probability statements must indicate the evidence with respect to that probability is being assessed.
- If an agent receives new percepts, its probability assessments are updated to reflect the new evidence.

# Random Variable

- Referring to a "part" of the world, whose "status" is initially unknown
- *We* will use lowercase for the names of values
  - $P(a) = 1 - P(\neg a)$
- Tossing coin : $P(h) = 1 - P(\neg h)$ : $(0.5 = 1 - 0.5)$
- Rolling dice : $P(n) = 1 – P(\neg n)$ : $(0.16 = 1 - 0.84)$

# Types of random variables

- Boolean random variables
  - Cavity domain (true,false), if Cavity = true then cavity, or
  - if Cavity = false then ¬cavity
- Discrete random variables – countable domain
  - *Weather* might be *(sunny, rainy, cloudy, snow)*
- Continuous random variables – finite set real numbers with equal intervals e.g. interval(0.1)

- Discrete random variables: can have discrete random values for a particular domain
- Eg. Weather domain
- Weather is having 4 values
- If weather is cloudy, then weather = cloudy

**NEXT**

# Types of random variables

- Boolean random variables
  - Cavity domain (true,false), if Cavity = true then cavity, or
  - if Cavity = false then ¬cavity
- Discrete random variables – countable domain
  - *Weather* might be *(sunny, rainy, cloudy, snow)*
- Continuous random variables – finite set real numbers with equal intervals e.g. interval(0.1)

- Continuous random variables: Here we are having finite set of real numbers with equal intervals
- We can take any interval based on our application or domain
- For ex: if x is a random variable, we can write $x <= 2.5$

# Atomic events

- The concept of an **atomic event** is useful in understanding the foundations of probability theory.
- It is a complete specification of the state of the world about which the agent is uncertain.
- It can be an assignment of particular values, to all the variables of which the world is composed

# Atomic events...

- Atomic events have some important properties
- They are mutually exclusive -at most one can actually be the case.
- The set of all possible atomic events is exhaustive – at least one must be the case.
- Any particular atomic event entails the truth or falsehood of every proposition, whether simple or complex
- Any proposition is logically equivalent to the disjunction of all atomic events that required the truth of proposition.

- **Mutually Exclusive**: Atmost one can actually be the case
- For eg: Cavity and toothache are unconnected then the statements
  - Cavity and Toothache
  - Cavity and Not Toothache
- the above two can not be the case. We can use one at a time

# Atomic events…

- Atomic events have some important properties
- They are mutually exclusive -at most one can actually be the case.
- The set of all possible atomic events is exhaustive – at least one must be the case.
- Any particular atomic event entails the truth or falsehood of every proposition, whether simple or complex
- Any proposition is logically equivalent to the disjunction of all atomic events that required the truth of proposition.

- Point 4: Eg: Cavity and not Toothache represents truth of cavity and Falsehood of Toothache

- Then,

**How a proposition can be assessed in the absence of other information ?**

# Atomic events…

- Atomic events have some important properties
- They are mutually exclusive -at most one can actually be the case.
- The set of all possible atomic events is exhaustive – at least one must be the case.
- Any particular atomic event entails the truth or falsehood of every proposition, whether simple or complex
- Any proposition is logically equivalent to the disjunction of all atomic events that required the truth of proposition.

- Point 4: Eg: Cavity and not Toothache represents truth of cavity and Falsehood of Toothache

- Then,

**How a proposition can be assessed in the absence of other information ? Prior Probability**

# Prior Probability

- The **unconditional** or **prior probability** associated with a proposition *a*, is the degree of belief according to the absence of any other information;
- it is written as *P(a)*.
- For example, if the prior probability that one have a cavity is 0.1, then we would write
  - *P(Cavity = true)* = *0.1* or *P(cavity)* = 0.1 .
- It is important to remember that *P(a)* can be used only when there is no other information.

"The probability that the patient has a cavity, given that she is a teenager with no toothache, is 0.1"

$$P(cavity \mid \neg toothache \wedge teen) = 0.1 .$$

# Prior Probability…

- we will use an expression **P(** *Weather),* which denotes a ***vector*** of values, for the probabilities of each individual state of the weather.
  - *P( Weather = sunny) = 0.7*
  - *P( Weather = rain) = 0.2*
  - *P( Weather = cloudy) = 0.08*
  - *P( Weather = snow) = 0.02 .*
- we may simply write
- *P( Weather) = (0.7, 0.2, 0.08, 0.02) .*
- This statement defines a prior **probability distribution** for the random variable *Weather.*

# Conditional Probability

- The **conditional** or **posterior** probabilities notation is *P(alb),*
- where *a* and *b* are any proposition.
- This is read as "the probability of *a,* given that *all* we know is *b."*
- For example, *P(cavity| toothache) = 0.8*
- if a patient is observed to have a toothache and no other information is yet available, then the probability of the patient's having a cavity will be 0.8.

- i.e. Cavity is present only because of toothache and there is no other information is available

# Conditional probabilities…

- Conditional probabilities can be defined in terms of unconditional probabilities.
- The equation is

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

- whenever $P(b) > 0$.
- This equation can also be written as
- $P(a \wedge b) = P(a/ b)\ P(b)$ which is called the **product rule.**

# Conditional or Posterior Probability

- Most of the time, however, we have some information called **evidence, that has already been revealed**
- **P(a|b)** – probability of **a** given that all we know is **b**.

Ex. P(Cavity|Toothache) = 0.8

The probability of a person having Cavity given that he has toothache is 0.8

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

$$P(a \wedge b) = P(a|b) * P(b)$$

The product rule says that for a and b to be true, we need b to be true, and we also need a to be true given b.

$$P(b \wedge a) = P(b|a) * P(a)$$

# Basic Axioms of Probability

- All probabilities are between 0 and 1. For any proposition a,

$$0 \leq P(a) \leq 1$$

- Necessarily true (i.e., valid) propositions have probability I, and necessarily false (i.e., unsatisfiable) propositions have probability 0.

$$P(true) = 1 \qquad P(false) = 0\,.$$

- The probability of a disjunction is given by

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)\,.$$

# Theory – Test point of view

## Introduction

- Like logical assertions, probabilistic assertions are about possible worlds.
- Logical assertions say which possible worlds are strictly ruled out (all those in which the assertion is false), probabilistic assertions talk about how probable the various worlds are.
- In probability theory, the set of all possible worlds is called the sample space.
- The possible worlds are mutually exclusive and exhaustive
- $\Omega$ is used to refer to the sample space
- $\omega$ refers to elements of the space (possible worlds)

# Axioms of Probability

- The basic axioms of probability theory say that every possible world has a probability between 0 and 1
- Total probability of the set of possible worlds is 1

i.e., $0 \le P(\omega) \le 1$ for every $\omega$ and $\sum_{\omega \in \Omega} P(\omega) = 1$ .

For example, if we assume that each die (when throwing 2 dice) is fair and the rolls don't interfere with each other, then each of the possible worlds is (1,1), (1,2), . . ., (6,6)

**(Each has probability 1/36)**

# Propositions

- The sets of possible worlds are described by **propositions** in a formal language.
- For each proposition, the corresponding set contains just those possible worlds in which the proposition holds.
- **Boolean Random Variables**: Cavity=True
- **Discrete Random Variables**: (sunny, rainy, cloudy, snow). (Weather=snow is the proposition)
- **Continuous Random Variables**: Weight = 45.05Kg (Weight < 45 is proposition)
- The probability associated with a proposition is the sum of the probabilities of the worlds in which it holds:

$$\text{For any proposition } \phi,\ P(\phi) = \sum_{\omega \in \phi} P(\omega)\ .$$

# Propositions

- For example, when rolling fair dice,

  $$P(\text{Total} = 11) = P((5, 6)) + P((6, 5)) = 1/36 + 1/36 = 1/18.$$

- Probability theory does not require complete knowledge of the probabilities of each possible world.

# Atomic Events

- Atomic event is a complete specification of the state of the world about which the agent is uncertain.

  For Ex. **Cavity** and **Toothache** are two Boolean variables, then four distinct atomic events are possible (**Cavity=False ^ Toothache=True** is one example for the atomic event)

# Atomic Events - Properties

- **They are mutually exclusive**: **cavity ^ toothache** and **cavity ^ not toothache** not possible
- **The set of all possible atomic events is exhaustive**: At least one must be the case. The disjunction of all atomic events is logically equivalent to true.
- **Any particular atomic event entails the truth or falsehood of every proposition, whether simple or complex**. For example the atomic event **cavity ^ not toothache** entails the truth of cavity and falsehood of **cavity→toothache**
- **The proposition is logically equivalent to the disjunction of all atomic events that entail the truth of proposition**. For example, the proposition cavity is equivalent to the disjunction of the atomic events: **cavity^toothache** and **cavity^not** toothache

# Prior Probability

Unconditional probability associated with a proposition a is the degree of belief accorded to it in the absence of another information

Ex.

$$P(cavity=True) = 0.1 \text{ or } P(cavity) = 0.1$$

# Conditional or Posterior Probability

- Most of the time, however, we have some information called **evidence, that has already been revealed**
- **P(a|b)** – probability of **a** given that all we know is **b**.

Ex. P(Cavity|Toothache) = 0.8

The probability of a person having Cavity given that he has toothache is 0.8

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

$$P(a \wedge b) = P(a|b) * P(b)$$

The product rule says that for a and b to be true, we need b to be true, and we also need a to be true given b.

$$P(b \wedge a) = P(b|a) * P(a)$$

# Language of Propositions

- Variables in probability theory are called **random variables**
- For example, in dice problem, Total and Die1 are random variables
- Every random variable has a domain which is the set of possible values it can take on.
- A Boolean random variable has the domain {true, false}
- A=true is written as a, while A=false is written as $\neg$a.
- Domains can be sets of arbitrary tokens; For example the domain of Age can be {juvenile, teen, adult} and the domain of Weather might be {sunny, rain, cloudy, snow}.
- Variables can have infinite domains. Either discrete (like the integers) or continuous (like the reals).
- For any variable with an ordered domain, inequalities are also allowed, such as NumberOfAtomsInUniverse $\geq 10^{70}$.

# Probablity Distribution

- Probabilities of all the possible values of a random variable can be written as:
  - P(Weather =sunny) = 0.6
  - P(Weather =rain) = 0.1
  - P(Weather =cloudy) = 0.29
  - P(Weather =snow) = 0.01
- This can be abbreviated as
- P(Weather)=(0.6, 0.1, 0.29, 0.01)
- P defines a probability distribution for the random variable Weather
- For continuous variables, it is not possible to write out the entire distribution as a vector, because there are infinitely many values.
- Instead, we can define the probability that a random variable takes on some value x as a parameterized function of x.
- For example, the sentence

  $$P(NoonTemp = x) = Uniform_{[18C,26C]}(x)$$

  This is called **probability density function.**

# Joint probability distribution

- The expression **P(Weather, cavity)** is a four element vector probabilities for the conjunction of each weather type with **Cavity=true**
- It can be represented as 4 * 2 table of probabilities
- If the world consists of just the variables Cavity, Toothache and Weather, then the **full joint distribution** is given by:
- P(Cavity, Toothache, Weather)
- Represented as 2 * 2 * 4 table with 16 entries

## Probability axioms and their reasonableness

- The familiar relationship between the probability of a proposition and the probability of its negation:

$$P(\neg a) = 1 - P(a)$$

- The probability of a disjunction is called the inclusion–exclusion principle:

$$P(a \lor b) = P(a) + P(b) - P(a \land b)$$

# Example for Fully Joint Distributions

| | $toothache$ | | $\neg toothache$ | |
|---|---|---|---|---|
| | $catch$ | $\neg catch$ | $catch$ | $\neg catch$ |
| $cavity$ | 0.108 | 0.012 | 0.072 | 0.008 |
| $\neg cavity$ | 0.016 | 0.064 | 0.144 | 0.576 |

**Figure 13.3** A full joint distribution for the $Toothache$, $Cavity$, $Catch$ world.

- Let us take a domain consisting of just the three Boolean variables Toothache, Cavity, and Catch (the dentist's nasty steel probe catches in my tooth).
- The probabilities in the joint distribution sum to 1

# Inference using Fully Joint Distributions

- To calculate the probability of any proposition, simply identify those possible worlds in which the proposition is true and add up their probabilities.
- For example, there are six possible worlds in which cavity ∨ toothache holds:

$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$ .

- Adding the entries in the first row gives the unconditional or marginal probability of cavity:

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2 .$$

This process is called **marginalization or summing out.**

# Marginalization and Conditioning

- The general marginalization rule for any sets of variables Y and Z is: $$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$
- A variant of this rule involves conditional probabilities instead of joint probabilities, using the product rule: $$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z}) P(\mathbf{z}) \, .$$
- This rule is called **conditioning.**

# Conditional Probability

- For example, we can compute the probability of a cavity, given evidence of a toothache, as follows:

$$\begin{aligned} P(cavity \mid toothache) &= \frac{P(cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6 \end{aligned}$$

- We can compute the probability that there is no cavity, given a toothache

$$\begin{aligned} P(\neg cavity \mid toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4 \end{aligned}$$

- The two values sum to 1.0

## Conditional Probability

- In the above two calculations the term $1/P(toothache)$ remains constant, no matter which value of Cavity we calculate. This is known as the normalization constant which is denoted as $\alpha$.

- We can write the two preceding equations in one:

$$P(Cavity \mid toothache) = \alpha\, P(Cavity, toothache)$$
$$= \alpha[P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$$
$$= \alpha\, [(0.108, 0.016) + (0.012, 0.064)]$$
$$= \alpha\, (0.12, 0.08) = (0.6, 0.4)$$

- We can calculate $P(Cavity \mid toothache)$ even if we don't know the value of $P(toothache)$!

# Inference Procedure

- We begin with the query involving a single variable, X (Cavity in the example).
- Let E be the list of evidence variables (just Toothache in the example), let e be the list of observed values for them, and let Y be the remaining unobserved variables (just Catch in the example).
- The query P(X | e) is as follows: $\mathbf{P}(X \mid \mathbf{e}) = \alpha \, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$
- Here the variables X, E, and Y constitute the complete set of variables for the domain, so P(X, e, y) is simply a subset of probabilities from the full joint distribution.
- It requires an input table of **size O(2n) and takes O(2n) time** to process the table where n is the number of boolean random variables.

## Independence

- Let us expand the full joint distribution by adding a fourth variable, Weather. So the table has $2 \times 2 \times 2 \times 4 = 32$ entries.
- Here P(toothache, catch, cavity, cloudy) = P(cloudy | toothache, catch, cavity)P(toothache, catch, cavity) .
- Since weather does not influence dental variables we can say that: P(cloudy | toothache, catch, cavity) = P(cloudy)
- So P(toothache, catch, cavity, cloudy) = P(cloudy)P(toothache, catch, cavity)
- Thus the 32-element table for four variables can be constructed from one 8-element table and one 4-element table.
- This is called **independence (marginal independence and absolute independence)**

## Independence

- Independence between variables X and Y can be written as follows:

  $P(X \mid Y) = P(X)$ or

  $P(Y \mid X) = P(Y)$ or

  $P(X, Y) = P(X)P(Y)$

# Bayes Rule

- We can write the product rule in two forms:
  $P(a \wedge b) = P(a \mid b)P(b)$ and $P(a \wedge b) = P(b \mid a)P(a)$ .
- Equating the two right-hand sides we get,

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

- This equation is known as **Bayes' rule (also Bayes' law or Bayes' theorem).**
- The more general case of Bayes' rule for multivalued variables is:

$$\mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

- A more general version conditionalized on some background evidence e is :

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})}$$

## Applying Bayes Rule

- Bayes Rule allows us to compute the single term P(b | a) in terms of three terms: P(a | b), P(b), and P(a).
- Often, we perceive as evidence the effect of some unknown cause and we would like to determine that cause.

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}.$$

- The conditional probability P(effect|cause) quantifies the relationship in the **causal direction**, whereas P(cause|effect) describes the **diagnostic direction**.
- In medical diagnosis problem, we often have conditional probabilities on causal relationships (that is, the doctor knows P(symptoms|disease)) and want to derive a diagnosis, P(disease|symptoms).

# Applying Bayes Rule - Example

- For example, a doctor knows that the disease meningitis causes the patient to have a stiff neck, say, 70% of the time.
- The doctor also knows some unconditional facts: the prior probability that a patient has meningitis is 1/50,000, and the prior probability that any patient has a stiff neck is 1%.
- Let s be the proposition that the patient has a stiff neck and m be the proposition that the patient has meningitis.

$$
\begin{aligned}
P(s \mid m) &= 0.7 \\
P(m) &= 1/50000 \\
P(s) &= 0.01 \\
P(m \mid s) &= \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014\,.
\end{aligned}
$$

## Using Bayes' rule: Combining evidence

- What happens when we have two or more pieces of evidence?
- For example, what can a dentist conclude if her nasty steel probe catches in the aching tooth of a patient?

  $$P(Cavity \mid toothache \wedge catch) = \alpha\,\langle 0.108, 0.016 \rangle \approx \langle 0.871, 0.129 \rangle$$

- But such an approach does not scale up to larger numbers of variables.
- By using Bayes' rule to reformulate the problem:

  $$P(Cavity \mid toothache \wedge catch) = \alpha\, P(toothache \wedge catch \mid Cavity)\, P(Cavity)\,.$$

## Using Bayes' rule: Combining evidence

- Toothache and catch variables are independent, however, given the presence or the absence of a cavity.
- Each is directly caused by the cavity, but neither has a direct effect on the other
- That is toothache depends on the state of the nerves in the tooth, whereas the probe's accuracy depends on the dentist's skill, to which the toothache is irrelevant
- So $P(toothache \wedge catch \mid Cavity)$

  $= P(toothache \mid Cavity)P(catch \mid Cavity)$
- So the probability of a cavity is

  $P(Cavity \mid toothache \wedge catch) = \alpha\ P(toothache \mid Cavity)\ P(catch \mid Cavity)\ P(Cavity)$

## Conditional independence

- Conditional independence of two variables X and Y , given a third variable Z, is $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$
- A Joint distribution can also be wriiten as

  $P(Toothache, Catch \mid Cavity)$

  $= P(Toothache \mid Cavity)P(Catch \mid Cavity)$
- For fully joint distribution we can derive a decomposition as follows:

  $P(Toothache, Catch, Cavity)$

  $= P(Toothache, Catch \mid Cavity)P(Cavity)$ (product rule)

  $= P(Toothache \mid Cavity)P(Catch \mid Cavity)P(Cavity)$
- **The decomposition of large probabilistic domains into weakly connected subsets through conditional independence** is one of the most important developments in the recent history of AI.

## Conditional independence

- The full joint distribution can be written as

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$

- Such a probability distribution is called a **naive Bayes model or Bayesian Classifier.**

# Assessment Questions

| | | |
|---|---|---|
| What is random variable? | 2 | R |
| What are the three kinds of random variables? | 2 | R |
| Write the areas of applications of probability | 2 | R |
| Define product rule. | 2 | R |
| Explain the uses of axioms of probability? | 4 | U |
| Why the axioms of probability are reasonable? | 2 | U |
| A domain consists of just three Boolean variables “Toothache, cavity and catch”. Draw the full joint distribution 2×2×2 table. | 6 | U |
| Define the concept of independence. | 2 | R |
| State the Baye's theorem. How is it useful for decision making under uncertainity? | 8 | U |
| State Bayes rule. | 2 | R |