

Uncertainty

(Chapter 13, Russell & Norvig)

Introduction to Artificial Intelligence

CS 151

Lecture 1

Administration

- PA 1 will be handed out later this week.
- Programming is in MATLAB - the TA will give a MATLAB tutorial on Thursday.

This Lecture

- Probability
- Syntax
- Semantics
- Inference rules

Reading

Chapter 13

Uncertainty

- Let action A_t = leave for airport t minutes before flight from O'hare
- Will A_t get me there on time?

Problems:

1. Partial observability (road state, other drivers' plans, etc.)
2. Noisy sensors (traffic reports)
3. Uncertainty in action outcomes (flat tire, etc.)
4. Immense complexity of modelling and predicting traffic

Hence a purely logical approach either

- 1) risks falsehood: “ A_{135} will get me there on time,” or
- 2) leads to conclusions that are too weak for decision making:
“ A_{135} will get me there on time if there's no accident on I-57
and it doesn't rain and my tires remain intact, etc., etc.”

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport ...)

Methods for handling uncertainty

Default or nonmonotonic logic:

Assume my car does not have a flat tire

Assume A_{125} works unless contradicted by evidence

Issues: What assumptions are reasonable? How to handle contradiction?

Rules with fudge factors:

$A_{135} \pm 0.3$ get there on time

Sprinkler ± 0.99 WetGrass

WetGrass ± 0.7 Rain

Issues: Problems with combination, e.g., Sprinkler causes Rain??

Probability

Given the available evidence,

A_{135} will get me there on time with probability 0.04

(**Fuzzy logic** handles *degree of truth* NOT uncertainty

e.g., WetGrass is true to degree 0.2)

Fuzzy Logic in Real World



Fuzzy Logic Rice Cooker

Made exclusively for Williams-Sonoma, this cooker operates on microchip technology and turns out perfectly cooked rice every time. Programmed settings include sushi rice, brown rice, regular rice, soft grains, slow cooking (for stews and soups), steaming (for vegetables and fish) and quick cooking. The nonstick bowl sits in a cupped heating pad, which ensures even cooking. A timer lets you preset finish times up to 24 hours in advance; the cooker keeps rice warm for 12 hours. Steaming tray, rice paddle, measuring cup and booklet with steaming chart and recipes included. 585W, 5-cup cap. uncooked (10-cup cooked); 13" x 10 $\frac{1}{4}$ " x 9" high. #07-2195378 **\$199.00** ©

Probability

Probabilistic assertions summarize effects of

Ignorance: lack of relevant facts, initial conditions, etc.

Laziness: failure to enumerate exceptions, qualifications, etc.

Subjective or Bayesian probability:

Probabilities relate propositions to one's own state of knowledge

e.g., $P(A_{135} \mid \text{no reported accidents}) = 0.06$

These are **NOT** assertions about the world, but represent belief about the whether the assertion is true.

Probabilities of propositions change with new evidence:

e.g., $P(A_{135} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

(Analogous to logical entailment status; I.e., does $KB \models \alpha$)

Making decisions under uncertainty

- Suppose I believe the following:
 - $P(A_{135} \text{ gets me there on time} \mid \dots) = 0.04$
 - $P(A_{180} \text{ gets me there on time} \mid \dots) = 0.70$
 - $P(A_{240} \text{ gets me there on time} \mid \dots) = 0.95$
 - $P(A_{1440} \text{ gets me there on time} \mid \dots) = 0.9999$
- Which action to choose?
- Depends on my preferences for missing flight vs. airport cuisine, etc.
- Utility theory is used to represent and infer preferences
- Decision theory = utility theory + probability theory

Unconditional Probability

- Let A be a proposition, $P(A)$ denotes the unconditional probability that A is true.
- Example: if Male denotes the proposition that a particular person is male, then $P(\text{Male})=0.5$ mean that without any other information, the probability of that person being male is 0.5 (a 50% chance).
- Alternatively, if a population is sampled, then 50% of the people will be male.
- Of course, with additional information (e.g. that the person is a CS151 student), the “posterior probability” will likely be different.

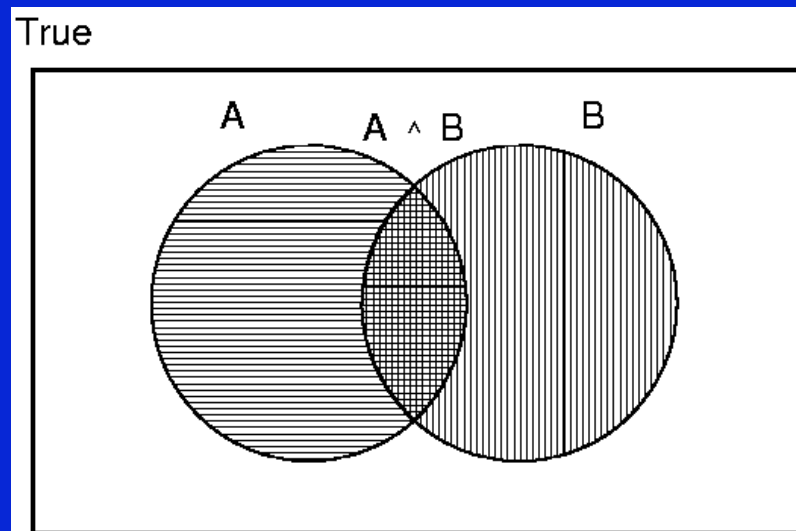
Axioms of probability

For any propositions A, B

1. $0 \leq P(A) \leq 1$

2. $P(\text{True}) = 1$ and $P(\text{False}) = 0$

3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

Syntax

Similar to PROPOSITIONAL LOGIC: possible worlds defined by assignment of values to random variables.

Note: To make things confusing variables have first letter upper-case, and symbol values are lower-case

Propositional or Boolean random variables

e.g., Cavity (do I have a cavity?)

Include propositional logic expressions

e.g., $\neg \text{Burglary} \vee \text{Earthquake}$

Multivalued random variables

e.g., Weather is one of <sunny,rain,cloudy,snow>

Values must be exhaustive and mutually exclusive

A proposition is constructed by assignment of a value:

e.g., Weather = sunny; also Cavity = true for clarity

Priors, Distribution, Joint

Prior or **unconditional probabilities** of propositions

e.g., $P(\text{Cavity}) = P(\text{Cavity}=\text{TRUE}) = 0.1$

$P(\text{Weather}=\text{sunny}) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives probabilities of all possible values of the random variable.

Weather is one of $\langle \text{sunny}, \text{rain}, \text{cloudy}, \text{snow} \rangle$

$P(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$

(normalized, i.e., sums to 1)

Joint probability distribution

Joint probability distribution for a set of variables gives values for each possible assignment to all the variables

$P(\text{Toothache}, \text{Cavity})$ is a 2 by 2 matrix.

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

NOTE: Elements in table sum to 1 → 3 independent numbers.

The importance of independence

- $P(\text{Weather}, \text{Cavity})$ is a 4 by 2 matrix of values:

Weather =	sunny	rain	cloudy	snow
Cavity=True	.072	.01	.008	.01
Cavity=False	.648	.08	.072	.09

- But these are independent (usually):
 - Recall that if X and Y are independent then:
 - $P(X=x, Y=y) = P(X=x) * P(Y=y)$

Weather is one of <sunny,rain,cloudy,snow>

$P(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ and $P(\text{Cavity}) = .1$

The importance of independence

If we compute the *marginals* (sums of rows or columns), we see:

Weather =	sunny	rain	cloudy	snow	marginals
Cavity=True	.072	.01	.008	.01	.1
Cavity=False	.648	.08	.072	.09	.9
marginals	.72	.09	.08	.1	

Each entry is just the product of the marginals (*because* they're independent)!

So, instead of using 7 numbers, we can represent these by just recording 4 numbers, 3 from Weather and 1 from Cavity:

Weather is one of <sunny,rain,cloudy,snow>

$P(\text{Weather}) = \langle 0.72, 0.1, 0.08, (1 - \text{Sum of the others}) \rangle$

$P(\text{Cavity}) = \langle .1, (1-.1) \rangle$

This can be much more efficient (more later...)

Conditional Probabilities

- **Conditional** or posterior probabilities

e.g., $P(\text{Cavity} \mid \text{Toothache}) = 0.8$

What is the probability of having a cavity given that the patient has a toothache? (as another example: $P(\text{Male} \mid \text{CS151Student}) = ??$)

- If we know more, e.g., dental probe catches, then we have

$$P(\text{Cavity} \mid \text{Toothache}, \text{Catch}) = .9 \quad (\text{say})$$

- If we know even more, e.g., Cavity is also given, then we have

$$P(\text{Cavity} \mid \text{Toothache}, \text{Cavity}) = 1$$

Note: the less specific belief remains valid after more evidence arrives, but is not always useful.

- New evidence may be irrelevant, allowing simplification, e.g.,

$$P(\text{Cavity} \mid \text{Toothache}, \text{PadresWin}) = P(\text{Cavity} \mid \text{Toothache}) = 0.8$$

(again, this is because Cavities and Baseball are *independent*)

Conditional Probabilities

- **Conditional** or posterior probabilities

e.g., $P(\text{Cavity} \mid \text{Toothache}) = 0.8$

What is the probability of having a cavity given that the patient has a toothache?

- Definition of **conditional probability**:

$$P(A|B) = P(A, B) / P(B) \quad \text{if } P(B) \neq 0$$

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

$$P(\text{Cavity}|\text{Toothache}) = .04/ (.04+.01) = 0.8$$

Conditional probability cont.

Definition of conditional probability:

$$P(A | B) = P(A, B) / P(B) \quad \text{if } P(B) \neq 0$$

Product rule gives an alternative formulation:

$$P(A, B) = P(A | B)P(B) = P(B|A)P(A)$$

Example:

$$P(\text{Male}) = 0.5$$

$$P(\text{CS151Student}) = 0.0037$$

$$P(\text{Male} | \text{CS151Student}) = 0.9$$

What is the probability of being a male CS151 student?

$$P(\text{Male}, \text{CS151Student})$$

$$= P(\text{Male} | \text{CS151Student})P(\text{CS151Student})$$

$$= 0.0037 * 0.9 = 0.0033$$

Conditional probability cont.

A general version holds for whole distributions, e.g.,

$$\mathbf{P}(\text{Weather}, \text{Cavity}) = \mathbf{P}(\text{Weather} | \text{Cavity}) \mathbf{P}(\text{Cavity})$$

(View as a 4 X 2 set of equations, not matrix mult.-

The book calls this a *pointwise product*)

The **Chain rule** is derived by successive application of product rule:

$$\begin{aligned} \mathbf{P}(X_1, \dots, X_n) &= \mathbf{P}(X_1, \dots, X_{n-1}) \mathbf{P}(X_n | X_1, \dots, X_{n-1}) \\ &= \mathbf{P}(X_1, \dots, X_{n-2}) \mathbf{P}(X_{n-1} | X_1, \dots, X_{n-2}) \mathbf{P}(X_n | X_1, \dots, X_{n-1}) \\ &= \dots \\ &= \prod_{i=1}^n \mathbf{P}(X_i | X_1, \dots, X_{i-1}) \end{aligned}$$

Bayes Rule

From product rule $P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$, we can obtain Bayes' rule

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

Why is this useful???

For assessing diagnostic probability from causal probability:

e.g. Cavities as the cause of toothaches

Sleeping late as the cause for missing class

Meningitis as the cause of a stiff neck

$$P(Cause | Effect) = \frac{P(Effect | Cause)P(Cause)}{P(Effect)}$$

Bayes Rule: Example

$$P(Cause | Effect) = \frac{P(Effect | Cause)P(Cause)}{P(Effect)}$$

Let M be meningitis, S be stiff neck

$$P(M) = 0.0001$$

$$P(S) = 0.1$$

$$P(S | M) = 0.8$$

$$P(M | S) = \frac{P(S | M)P(M)}{P(S)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

A generalized Bayes Rule

- More general version conditionalized on some background evidence E

$$P(A | B, E) = \frac{P(B | A, E)P(A | E)}{P(B | E)}$$

Full joint distributions

A **complete probability model** specifies every entry in the joint distribution for all the variables $\mathbf{X} = X_1, \dots, X_n$
i.e., a probability for each possible world $X_1 = x_1, \dots, X_n = x_n$

E.g., suppose Toothache and Cavity are the random variables:

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

Possible worlds are mutually exclusive $\nexists P(w_1 \wedge w_2) = 0$

Possible worlds are exhaustive $\nexists w_1 \vee \dots \vee w_n$ is True

hence $\sum_i P(w_i) = 1$

Using the full joint distribution

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

What is the unconditional probability of having a Cavity?

$$\begin{aligned} P(\text{Cavity}) &= P(\text{Cavity} \wedge \text{Toothache}) + P(\text{Cavity} \wedge \sim \text{Toothache}) \\ &= 0.04 + 0.06 = 0.1 \end{aligned}$$

What is the probability of having either a cavity or a Toothache?

$$\begin{aligned} P(\text{Cavity} \vee \text{Toothache}) &= P(\text{Cavity}, \text{Toothache}) + P(\text{Cavity}, \sim \text{Toothache}) + P(\sim \text{Cavity}, \text{Toothache}) \\ &= 0.04 + 0.06 + 0.01 = 0.11 \end{aligned}$$

Using the full joint distribution

	Toothache=true	Toothache = false
Cavity=true	0.04	0.06
Cavity=false	0.01	0.89

What is the probability of having a cavity given that you already have a toothache?

$$P(\text{Cavity} \mid \text{Toothache}) = \frac{P(\text{Cavity} \wedge \text{Toothache})}{P(\text{Toothache})} = \frac{0.04}{0.04 + 0.01} = 0.8$$

Normalization

Suppose we wish to compute a posterior distribution over random variable A given $B=b$, and suppose A has possible values $a_1 \dots a_m$

We can apply Bayes' rule for each value of A :

$$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$$

...

$$P(A=a_m | B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$$

Adding these up, and noting that $\sum_i P(A=a_i | B=b) = 1$:

$$P(B=b) = \sum_i P(B=b|A=a_i)P(A=a_i)$$

This is the normalization factor denoted $\alpha = 1/P(B=b)$:

$$P(A | B=b) = \alpha P(B=b | A)P(A)$$

Typically compute an unnormalized distribution, normalize at end

e.g., suppose $P(B=b | A)P(A) = \langle 0.4, 0.2, 0.2 \rangle$

then $P(A|B=b) = \alpha \langle 0.4, 0.2, 0.2 \rangle$

$$= \langle 0.4, 0.2, 0.2 \rangle / (0.4+0.2+0.2) = \langle 0.5, 0.25, 0.25 \rangle$$

Marginalization

Given a conditional distribution $P(X | Z)$, we can create the unconditional distribution $P(X)$ by marginalization:

$$P(X) = \sum_z P(X | Z=z) P(Z=z) = \sum_z P(X, Z=z)$$

In general, given a joint distribution over a set of variables, the distribution over any subset (called a **marginal distribution** for historical reasons) can be calculated by summing out the other variables.

Conditioning

Introducing a variable as an extra condition:

$$P(X|Y) = \sum_z P(X | Y, Z=z) P(Z=z | Y)$$

Why is this useful??

Intuition: it is often easier to assess each specific circumstance, e.g.,

$P(\text{RunOver} | \text{Cross})$

$$\begin{aligned} &= P(\text{RunOver} | \text{Cross}, \text{Light}=\text{green})P(\text{Light}=\text{green} | \text{Cross}) \\ &\quad + P(\text{RunOver} | \text{Cross}, \text{Light}=\text{yellow})P(\text{Light}=\text{yellow} | \text{Cross}) \\ &\quad + P(\text{RunOver} | \text{Cross}, \text{Light}=\text{red})P(\text{Light}=\text{red} | \text{Cross}) \end{aligned}$$

Absolute Independence

- Two random variables A and B are (absolutely) **independent** iff
 $P(A, B) = P(A)P(B)$

- Using product rule for A & B independent, we can show:
 $P(A, B) = P(A | B)P(B) = P(A)P(B)$
Therefore $P(A | B) = P(A)$

- If n Boolean variables are independent, the full joint is:

$$P(X_1, \dots, X_n) = \prod_i P(X_i)$$

Full joint is generally specified by $2^n - 1$ numbers, but when independent only n numbers are needed.

- Absolute independence is a very strong requirement, seldom met!!

Conditional Independence

- Some evidence may be irrelevant, allowing simplification, e.g.,

$$P(\text{Cavity} \mid \text{Toothache}, \text{PadresWin}) = P(\text{Cavity} \mid \text{Toothache}) = 0.8$$

- This property is known as **Conditional Independence** and can be expressed as:

$$P(X \mid Y, Z) = P(X \mid Z)$$

which says that X and Y independent given Z.

- If I have a cavity, the probability that the dentist's probe catches in it doesn't depend on whether I have a toothache:

- $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$

i.e., Catch is conditionally independent of Toothache *given* Cavity

This doesn't say anything about $P(\text{Catch} \mid \text{Toothache})$

The same independence holds if I haven't got a cavity:

- $P(\text{Catch} \mid \text{Toothache}, \sim\text{Cavity}) = P(\text{Catch} \mid \sim\text{Cavity})$

Conditional independence contd.

Equivalent statements to

$$P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity}) \quad (*)$$

1.a $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$

$$\begin{aligned} P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) &= P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \\ &= P(\text{Catch} \mid \text{Cavity}) P(\text{Toothache} \mid \text{Cavity}) / P(\text{Catch} \mid \text{Cavity}) \quad (\text{from } *) \\ &= P(\text{Toothache} \mid \text{Cavity}) \end{aligned}$$

1.b $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$

$$\begin{aligned} P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) &= P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad (\text{product rule}) \\ &= P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) \quad (\text{from 1a}) \end{aligned}$$

Using Conditional Independence

Full joint distribution can now be written as

$$\begin{aligned} &P(\text{Toothache}, \text{Catch}, \text{Cavity}) \\ &= P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) P(\text{Cavity}) \\ &= P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity}) P(\text{Cavity}) \quad (\text{from 1.b}) \end{aligned}$$

Specified by: $2 + 2 + 1 = 5$ independent numbers

Compared to 7 for general joint

or 3 for unconditionally independent.

Belief (Bayes) networks

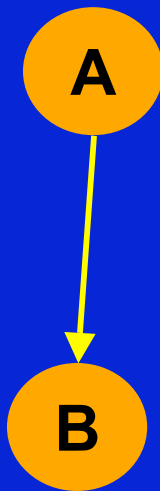
A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions.

Syntax:

1. a set of nodes, one per random variable
2. links mean parent “directly influences” child
3. a directed, acyclic graph
4. a conditional distribution (a table) for each node given its parents: $P(X_i \mid \text{Parents}(X_i))$

In the simplest case, conditional distribution represented as a conditional probability table (CPT)

A two node net & Conditional probability table



- Node A is independent of Node B, so it is described by an unconditional probability $P(A)$
- $P(\neg A)$ is given by $1 - P(A)$
- Node B is conditionally dependent on A. It is described by four numbers, $P(B | A)$, $P(B | \neg A)$, $P(\neg B | A)$ and $P(\neg B | \neg A)$.
- This can be expressed as 2 by 2 conditional probability table (CPT).
- But $P(\neg B | A) = 1 - P(B | A)$ and $P(\neg B | \neg A) = 1 - P(B | \neg A)$.
- Therefore, only two independent numbers in CPT

Example

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:

