Quantifying Uncertainty

By
S . Komal Kaur
Assistant Professor, CSE Dept.,
VCE

Outline

- Acting Under Uncertainty
- Basic Probability Notation
- Inference using Full Joint Distribution
- Independence
- Bayes Rule and its use

Uncertainty

- Agents may need to handle uncertainty, whether due to partial observability, nondeterminism, or a combination of the two.

- An agent may never know for certain what state it’s in or where it will end up after a sequence of actions.

Uncertainty

- Let action At = leave for air port t minutes before flight departure Will At get me there on time?
- Problems:

1. partial observability (road state, other drivers' plan's, etc.)
2. noisy sensors (traffic reports)
3. Uncertainty in action outcomes (flat tire, out of fuel, etc.)
4. Immense complexity of modelling and predicting traffic.

Uncertainty

- Hence a purely logical approach either

1. risks falsehood: “A25 will get me there on time”, or
2. Leads to conclusions that are too weak for decision making: “A25 will get me there on time if there’s no accident on the bridge and it doesn’t rain and my tires remain intact etc etc.”

- (A1440 might reasonably be said to get me there on time but I’d have to stay overnight in the airport...)

Uncertainty

- Suppose I believe the following:
  - P(A25 gets me there on time|...)=0.04
  - P(A90 gets me there on time |...)=0.70
  - P(A120 gets me there on time |...)=0.95
  - P(A1440 gets me there on time |...)=0.9999
- Which action should I choose?
- That depends on my preferences for missing the flight vs. sleeping at the airport, etc.
- Utility theory is used to represent and infer preferences

Summarizing Uncertainty

- Let's consider an example of uncertain reasoning: diagnosing a dental patient's toothache.
- Let us try to write rules for dental diagnosis using propositional logic, so that we can see how the logical approach breaks down.
- Consider the following simple rule:

$$Toothache \Rightarrow Cavity\ .$$

- The problem is that this rule is wrong. Not all patients with toothaches have cavities;

Summarizing Uncertainty

- Some of them have gum disease, an abscess, or one of several other problems:

  $$\text{Toothache} \Rightarrow \text{Cavity} \vee \text{Gum Problem} \vee \text{Abscess} \ldots$$

- Unfortunately, in order to make the rule true, we have to add an almost unlimited list of possible problems. We could try turning the rule into a causal rule:

  $$\text{Cavity} \Rightarrow \text{Toothache}$$

- But this rule is not right either; not all cavities cause pain.

Summarizing Uncertainty

- The agents knowledge can at best provide only a degree of belief in the relevant sentences.
- Our main tool dealing with degrees of belief is probability theory.

Uncertainty and rational decisions

- To make such choices like airport, an agent must first have preferences between the different possible outcomes of the various plans.
- An outcome is a completely specified state, including such factors as whether the agent arrives on time and the length of the wait at the airport. We use utility theory to represent and reason with preferences.

Uncertainty and rational decisions

- Utility theory says that every state has a degree of usefulness, or utility, to an agent and that the agent will prefer states with higher utility.

Decision theory=utility theory + probability theory

- The fundamental idea of decision theory is that an agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action. This is called the principle of maximum expected utility (MEU)

DT-Agent

function DT-AGENT(*percept*) **returns** an *action*
  persistent: *belief_state*, probabilistic beliefs about the current state of the world
          action, the agent's action

  update *belief_state* based on *action* and *percept*
  calculate outcome probabilities for actions,
    given action descriptions and current *belief_state*
  select *action* with highest expected utility
    given probabilities of outcomes and utility information
  return *action*

Figure 13.1 A decision-theoretic agent that selects rational actions.

Basic Probability Notation

We begin with a set Ω—the sample space

- e.g., 6 possible rolls of a die.
- Ω can be infinite

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ such that:

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

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An event A is any subset of Ω:

$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$

e.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Basic Probability Notation

- Conditional probabilities are defined in terms of unconditional probabilities as follows: for any propositions a and b, we have

$$P(a|b)=P(a \wedge b)/P(b)..........(1)$$

- The definition of conditional probability, Equation (1), can be written in a different form called the product rule:

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

The language of propositions in probability assertions

- Variables in probability theory are called random variables and their names begin with an uppercase letter.
- Thus, in the dice example, Total and Die1 are random variables.
- Every random variable has a domain—the set of possible values it can take on. The domain of Total for two dice is the set {2,...,12} and the domain of Die1 is {1,...,6}.

The language of propositions in probability assertions

- “The probability that the patient has a cavity, given that she is a teenager with no toothache, is 0.1” as follows:

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

- Sometimes we will want to talk about the probabilities of all the possible values of a random variable. We could write:
- $P(Weather = sunny) = 0.6$
- $P(Weather = rain) = 0.1$
- $P(Weather = cloudy) = 0.29$
- $P(Weather = snow) = 0.01$,

  but as an abbreviation we will allow

  $P(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$,

Joint Probability

- For example, P(Weather, Cavity) denotes the probabilities of all combinations of the values of Weather and Cavity.
- This is a 4×2 table of probabilities called the joint probability distribution of Weather and Cavity.
- For example, the product rules for all possible values of Weather and Cavity can be written as a single equation:

$$P(Weather, Cavity) = P(Weather \mid Cavity) P(Cavity)$$

Contd..

- Instead of as these 4×2=8equations (using abbreviations W and C):
  - $P(W = sunny \wedge C = true) = P(W = sunny | C = true) P(C = true)$
  - $P(W = rain \wedge C = true) = P(W = rain | C = true) P(C = true)$
  - $P(W = cloudy \wedge C = true) = P(W = cloudy | C = true) P(C = true)$
  - $P(W = snow \wedge C = true) = P(W = snow | C = true) P(C = true)$
  - $P(W = sunny \wedge C = false) = P(W = sunny | C = false) P(C = false)$
  - $P(W = rain \wedge C = false) = P(W = rain | C = false) P(C = false)$
  - $P(W = cloudy \wedge C = false) = P(W = cloudy | C = false) P(C = false)$
  - $P(W = snow \wedge C = false) = P(W = snow | C = false) P(C = false)$.

As a degenerate case, P(sunny,cavity) has no variables and thus is a one-element vector that is the probability of a sunny day with a cavity, which could also be written as P(sunny,cavity) or $P(sunny \wedge cavity)$.

Probability axioms and their reasonableness

- we can derive the familiar relationship between the probability of a proposition and the probability of its negation:

$$\begin{aligned} P(\neg a) &= \textstyle\sum_{\omega \in \neg a} P(\omega) \\ &= \textstyle\sum_{\omega \in \neg a} P(\omega) + \sum_{\omega \in a} P(\omega) - \sum_{\omega \in a} P(\omega) \\ &= \textstyle\sum_{\omega \in \Omega} P(\omega) - \sum_{\omega \in a} P(\omega) \\ &= 1 - P(a) \end{aligned}$$

- We can also derive the well-known formula for the probability of a disjunction, sometimes called the inclusion–exclusion principle:

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

Inference using Full Joint Distribution

- Probabilistic inference—that is, the computation of posterior probabilities for query propositions given observed evidence.
- We begin with a simple example: a domain consisting of just the three Boolean variables Toothache, Cavity, and Catch (the dentist's nasty steel probe catches in my tooth).
- The full joint distribution is a 2×2×2 table as shown in Figure

Inference using Full Joint Distribution

Notice that the probabilities in the joint distribution sum to 1, as required by the axioms of probability.

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

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

Marginal Probability

- For example, there are six possible worlds in which (cavity $\vee$ toothache) holds:
- P(cavity $\vee$ toothache)=0 .108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28 .
- One particularly common task is to extract the distribution over some subset of variables or a single variable.
- For example, 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—because we sum up the probabilities for each possible value of the other variables, thereby taking them out of the equation.

Marginal Probability

- We can write the following general marginalization rule for any sets of variables Y and Z:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z}) \ ,$$

where $\sum_{\mathbf{z} \in \mathbf{Z}}$ means to sum over all the possible combinations of values of the set of variables $\mathbf{Z}$. We sometimes abbreviate this as $\sum_{\mathbf{z}}$, leaving $\mathbf{Z}$ implicit. We just used the rule as

$$\mathbf{P}(Cavity) = \sum_{\mathbf{z} \in \{Catch, Toothache\}} \mathbf{P}(Cavity, \mathbf{z}) \ .$$

Conditional Probability

- 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. Marginalization and conditioning turn out to be useful rules for all kinds of derivations involving probability expressions.

Conditional Probability

- We can compute the probability of a cavity, given evidence of a toothache, as follows:

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

Conditional Probability

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

We can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

The denominator $1/P(toothache)$ can be viewed as a normalization constant α:

$$\begin{aligned}
\mathbf{P}(Cavity|toothache) &= \alpha\,\mathbf{P}(Cavity, toothache) \\
&= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
&= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] \\
&= \alpha\,\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle
\end{aligned}$$

Independence

- Let us expand the full joint distribution by adding a fourth variable, Weather.
- The full joint distribution then becomes P(Toothache, Catch, Cavity, Weather), which has 2 × 2 × 2 × 4 = 32entries.
- For example, how are P(toothache, catch, cavity, cloudy) and P(toothache, catch, cavity) related? We can use the product rule:

$$P(toothache, catch, cavity, cloudy)$$
$$= P(cloudy \mid toothache, catch, cavity) P(toothache, catch, cavity) .$$

Independence

- The following assertion seems reasonable:

  P(cloudy |toothache, catch, cavity)=P(cloudy) .

  From this, we can deduce

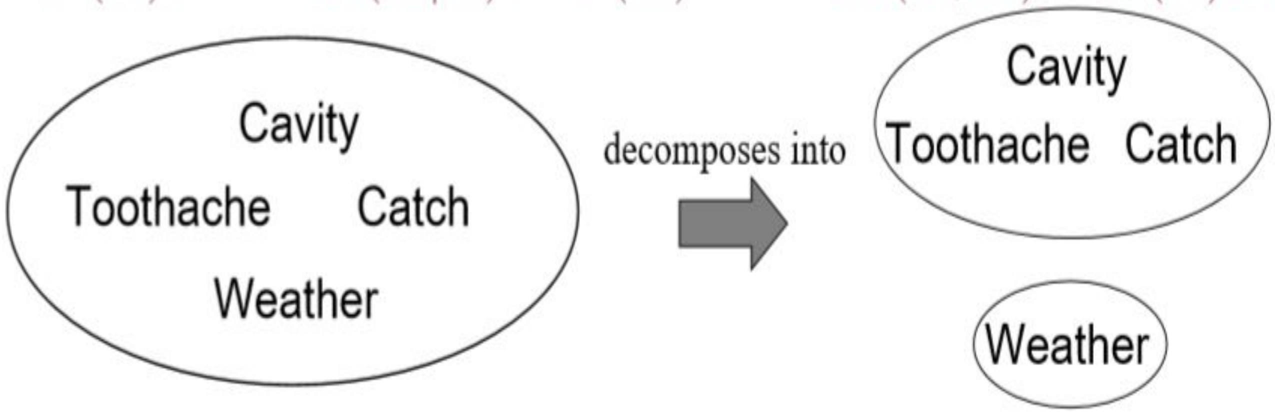
  P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)

- Thus, the 32-element table for four variables can be constructed from one 8-element table and one 4-element table.
- This decomposition is illustrated schematically in Figure. The property is called independence (also marginal independence and absolute independence).

Independence

Definition: A and B are independent iff

$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entries reduced to 12

Bayes' rule and its use

- The product rule: It can actually be written in two forms:

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

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

- Equating the two right-hand sides and dividing by P(a), we get

  $$\mathbf{P(b|a)=P(a|b)P(b) /P(a)}$$

  This equation is known as Bayes' rule

Bayes Rule

- The more general case of Bayes' rule for multivalued variables can be written in the P notation as follows:

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

- Useful for assessing diagnostic probability from causal probability:

$$\mathbf{P(Cause|Effect)=P(Effect|Cause)P(Cause)/P(Effect)}$$

Bayes'Rule

$$
\begin{aligned}
P(m) &= 1/50\,000 \\
P(s) &= 0.01 \\
P(s|m) &= 0.7
\end{aligned}
$$

What is the probabilitiy of meningitis given that I have a stiff neck?

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \times 1/50\,000}{0.01} = 0.0014$$

Using Bayes' rule: Combining evidence

- We can try to reformulate the toothache problem using Bayes' rule:

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

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

we need to know the conditional probabilities of the conjunction toothache $\wedge$ catch for each value of Cavity

Using Bayes' rule: Combining evidence

- Mathematically, this property is written as $P(toothache \wedge catch \mid Cavity)$

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

- This equation expresses the conditional independence of toothache and catch given Cavity, to obtain the probability of a cavity:

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

$\alpha P(toothache \mid Cavity)P(catch \mid Cavity)P(Cavity)$.

Conditional independence

- The general definition of conditional independence of two variables X and Y , given a third variable Z, is

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

- For Example,

P(Toothache, Catch |Cavity)

= P(Toothache |Cavity)P(Catch |Cavity)

Bayes'Rule and conditional independence

$$\mathbf{P}(Cavity|toothache \wedge catch)$$
$$= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\,\mathbf{P}(Cavity)$$
$$= \alpha\,\mathbf{P}(toothache|Cavity)\,\mathbf{P}(catch|Cavity)\,\mathbf{P}(Cavity)$$

This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\,\Pi_i \mathbf{P}(Effect_i|Cause)$$

Thank You